2013

Lessons Learned from Collaborations in Chemistry Assessment across Universities: Challenges in Transfer and Scale

Pamela L. Paek
Center for Assessment

Thomas Holme
Iowa State University, taholme@iastate.edu

Lessons Learned from Collaborations in Chemistry Assessment across Universities: Challenges in Transfer and Scale

Abstract

This chapter reviews a collaborative effort to cross-pollinate and share work around chemistry assessments across several universities. The goal was to find ways to synthesize separate projects and capitalize on applying developed instruments and assessments beyond a single university, and in new situations, to increase scale and check for generalizability. Discussion of the successes and challenges of scale and transfer of the collaboration is detailed in this chapter.

Disciplines

Curriculum and Social Inquiry | Educational Assessment, Evaluation, and Research | Higher Education | Other Chemistry | Science and Mathematics Education

Comments

Publication Date (Web): September 26, 2013 | doi: 10.1021/bk-2013-1145.ch011

Chapter 11

Lessons Learned from Collaborations in Chemistry Assessment across Universities: Challenges in Transfer and Scale

Pamela L. Paek[1] and Thomas A. Holme[*,2]

[1]Center for Assessment, Austin, Texas 78733
[2]Department of Chemistry, Iowa State University Ames, Iowa 50011
[*]E-mail: taholme@iastate.edu

This chapter reviews a collaborative effort to cross-pollinate and share work around chemistry assessments across several universities. The goal was to find ways to synthesize separate projects and capitalize on applying developed instruments and assessments beyond a single university, and in new situations, to increase scale and check for generalizability. Discussion of the successes and challenges of scale and transfer of the collaboration is detailed in this chapter.

By definition, collaboration is "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem (*1*)." In theory, collaborations appear to be an easy way to combine the power of multiple minds in the joint effort of developing a product or set of products that could not be done by a single individual (*1*, *2*). However, in practice, there are logistical issues (*3*), different mindsets and habits of mind (*4*), as well as unspoken end goals or motivations (*5*) that can impact the effect collaborations can have. Additionally, underestimating these issues moderates the amount of transfer and scale that is possible in multi-site collaborations (*6*, *7*).

This chapter discusses the synthesis of the multiple projects across multiple universities to address a larger issue in undergraduate chemistry. In this collaboration, the goal was to develop a system of chemistry assessments that could be used collectively to inform instruction for undergraduate chemistry instructors. The thought was that the combined use and comparison of different

In Trajectories of Chemistry Education Innovation and Reform; Holme, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2013.

measures would provide insight into how the instruments better determined similar or different types of knowledge and understanding than a single project usually would entail. This collaborative effort synthesized work together by having separate projects interact. It was not just a summary of different discrete projects, rather, a variety of combinations that would allow projects to interact in various ways, within and across different universities, chemistry courses, and combinations of uses.

The outcomes of the overall collaborative effort, both in terms of the successes and challenges, will be examined through five main themes: (1) previous partnerships prior to this large-group collaboration, (2) similarities and differences in project goals, (3) university support for conducting and scaling up chemistry education research, (4) the use of original and modified instrument measures for comparison across years, classrooms, and universities and (5) how instructor beliefs and values affect how they use different information for assessing student learning. The first two ideas are grouped under the theme of collaboration, while the latter two themes focus on scale and transfer. The third idea about university support, deals with both collaborations as well as scale and transfer.

Collaborations: Previous Partnerships Prior to This Large-Group Effort

This collaborative effort consisted of eight principal investigators from eight different universities. Each of these PIs is a distinguished faculty member specializing in chemistry education, all with a focus on synthesizing assessment ideas and instruments within undergraduate chemistry. They are listed below in alphabetical order with their department and current university affiliation:

- Stacey Lowery Bretz, Department of Chemistry & Biochemistry, Miami University
- Melanie Cooper, Department of Chemistry, Michigan State University
- Thomas Holme, Department of Chemistry, Iowa State University
- Jennifer Lewis, Department of Chemistry, University of South Florida
- Norbert Pienta, Department of Chemistry, University of Georgia
- Angelica Stacy, Department of Chemistry, University of California, Berkeley
- Ronald Stevens. Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles
- Marcy Towns, Department of Chemistry, Purdue University

As proof of concept to show collaborations yielded successful results, all eight principal investigators (PIs) on the larger project worked together in some capacity – mainly through pairwise partnerships – previous to the award of this collaborative grant through the National Science Foundation's Course, Curriculum, and Laboratory Improvement (CCLI) grants: Collaborative Research: A Model for Data-Driven Reform in Chemistry Education (Award ID

DUE-0817409). One of the reasons NSF was interested in supporting this larger collaborative effort was because of the previous successes of the PIs working together in pairs or triads, showing some initial synergies for advancing their work in chemistry education. Another was that each of the principal investigators was in their own right leaders in chemistry assessment, and as a group they might contribute to a new wave of how these assessments and corresponding information could be used to improve instructional practices of undergraduate chemistry instructors.

What is interesting to note is that none of the PIs were from schools of education, rather, they were housed in departments of chemistry, or in one case, a different, but closely related, science department. So, the fact that they were all based in their strong content knowledge of chemistry, yet interested in employing educational practices like development and measurement of content-based and non-cognitive assessments to their content area, was indeed a major undertaking and leap forward to advance how assessments could inform undergraduate chemistry instruction. In short, chemistry faculty members specializing in chemistry assessment is new area of study that currently only has a small number of researchers invested in the work. This group wanted to help support chemistry instructors by better understanding what undergraduate chemistry students understood and had misconceptions about from a content perspective. From a non-content cognitive perspective, the group was also interested in developing and using measures of students' self-efficacy and metacognitive skills, which they postulated could be used to target students more at risk because of a lack of confidence or lacking of study skills to effectively succeed, primarily in a first-semester, gateway, chemistry course.

Collaborations: Similarities and Differences in Project Goals

Given the similar outcomes and intent of previous individual and pair-wise partnerships, the collaborative effort of eight principal investigators along with similar research interests for measuring student cognitive and non-cognitive factors, the project envisioned collaborations to expand both the depth of the assessment research and the scale of the application of assessment instruments. Thus, the overarching interest of the collaborators was ultimately to both improve chemistry instruction and develop students' metacognitive skills. From the outset, all partners in this project seemed to have the same overall goal, where each PI would be able to contribute his/her part. An outside evaluator with specialization in developing and evaluating educational assessments, Pamela Paek, was charged with analyzing the findings of this collaborative effort. This chapter demonstrates this evaluation, along with recommendations for future collaborations.

Collaboration Proposal and Plans

The collaboration was initially proposed as a three-year study, to provide time to conduct synergistic activities beyond single universities across multiple years in chemistry assessment. This format would allow PIs to replicate findings over multiple years, potentially begin some longitudinal studies, and allow for adaptations and modifications as needed to extend or refine current instruments for further study and analysis. However, the funding level for the project was notably less than what was initially proposed and the timeframe was reduced to 18-months. As such, the project's change of scope focused more on the initial activities that could be done across universities, but not necessarily gather data beyond a single year or be able to move too far with scale and transfer, given the timeframe.

As in any research project, an 18-month research study involving students and faculty would mean potentially only 2 semesters worth of data, assuming each researcher would be able to have research instruments ready to administer, human subjects approval, and other logistics in place for other instructors on board to implement, as well as graduate students to support the work. While this 18-month funding window in principle allowed a basis for initiating collaborations across universities, the logistics for doing so required substantially more time to successfully carry off within each institution, and was compounded when trying to scale beyond a single university and researcher.

Since all collaborators had been previously involved in using and developing different assessment instruments, the goal of the first six months was to use existing instruments to gather baseline data, provide comparative data for students and instruments at the institutions involved in the project, and address logistical issues that may arise from the collaboration. The challenge here was not about continuing research on work that was already in place, but around the collaborative efforts to cross-pollinate assessment instruments and conduct comparative studies between institutions, and across cohorts within institutions. This challenge proved to be the most difficult, outside of the issues of faculty members within institutions not utilizing the instruments for the purpose of informing instruction. So, scale became problematic not only within universities beyond the main researcher onsite, but also across universities.

The researchers were realistic in that the limited timeframe would inhibit scope, as they would be able to establish baseline data, but not be able to conduct longitudinal studies. There would also be limited opportunities for repeating studies to confirm initial observations related to content changes for assessments that were still being developed and refined. Additionally, the ability to discover several synergies between assessment instruments was limited due to how quickly each researcher was able to pool resources to utilize collaborators' instruments and integrate those with their own instruments at each university site.

While the group requested and was approved for a no-cost extension of this grant, the time that was afforded was merely to provide setup for the activities they wanted to engage in. The original timeframe was extended to allow for more time on instrument development, plans for cross-site studies, and initial analyses of these results. Even with the more limited funds acquired via this grant, with the extended time, the PIs were able to produce a myriad of presentations and publications that highlighted the different partnerships in this collaboration, one of which was a joint publication of all eight PIs (*8*).

Refocused Collaborative Efforts in a Compressed Timeframe

A large focus of the collaboration was on development and uses of affective and metacognition measures, as seen in the use of the full Metacognitive Awareness Inventory (MCAI) (*9*, *10*), a modified version of the MCAI, a modified version of the Attitude toward the Subject of Chemistry Inventory (ASCI) (*11*, *12*), and CHEMX (*13*, *14*). The goals of using these measures were to see how performance on these assessments related to students' performance on chemistry assessments. Data collection included gathering pre-test performance to compare to later chemistry assessment performance as well as gathering post-test performance to evaluate change on the non-cognitive measures and how this change was related to understanding of chemistry content. The goal was to analyze the relationship of these data points and to see how instructors may be able to use the pre-test measures as ways to intervene and better support students' efforts and approaches to learning.

A second area of focus was on developing ways to better assess students' approaches to learning, as evidenced by the use of an instrument measuring key concepts (*15–17*) that include the misconceptions students may have on these complex topics (*18*, *19*), reasoning concepts using the Test of Logical Thinking (TOLT) and Group Assessment of Logical Thinking (GALT) (*20*), as well as the use of IMMEX (*21*, *22*). The goal of using these measures was to find a way to better understand students' problem solving strategies and key areas of misconceptions, to help instructors have the data they need to support students and target instruction more effectively. Figure 1 provides a timeline of the various implementation of instruments over the various schools in the project. Schools appearing above the arrows indicate the location where the instrument was developed or adapted, and those below the arrows are the additional schools the used that instrument and thereby provided cross-validation data available to this project.

While there were similarities and interests in the different assessments developed or under development, there was no whole group effort to use one set of assessments across all eight universities. Rather, the collaboration appeared to be a continuation, with some expansions, of previous partnerships that extended into new and refined instruments, or extending the pairs into groups of three or four. The one common thread for several of these small groupings was the collaborator who focused on the measurement effort, as her role was to analyze each new assessment to validate the overall construct. This included, for example, investigating factor loadings, and providing advice on possible refinements, using statistics, for future versions of each assessment. This activity, however, did not extend to each collaborative implementation of assessment carried out within the project.

Even with the compressed timeframe and complexities of conducting a multi-PI research endeavor, a significant amount of collaborative research was conducted, with results published of these collaborative efforts in twelve publications, including one joint publication of all principal investigators and the evaluator of this collaborative (*8*). The whole group publication described the intent of the collaboration and the goals for working across universities, instruments, and contexts, to show how these different measures could work synergistically to move the chemistry assessment field forward and assist in the way chemistry instructors could improve their teaching, and thus increase the success of their students' master of core content in their classes. Funding from the grant was often used to support graduate students within the individual PI research groups, so seven of the publications from the project include graduate student co-authors.

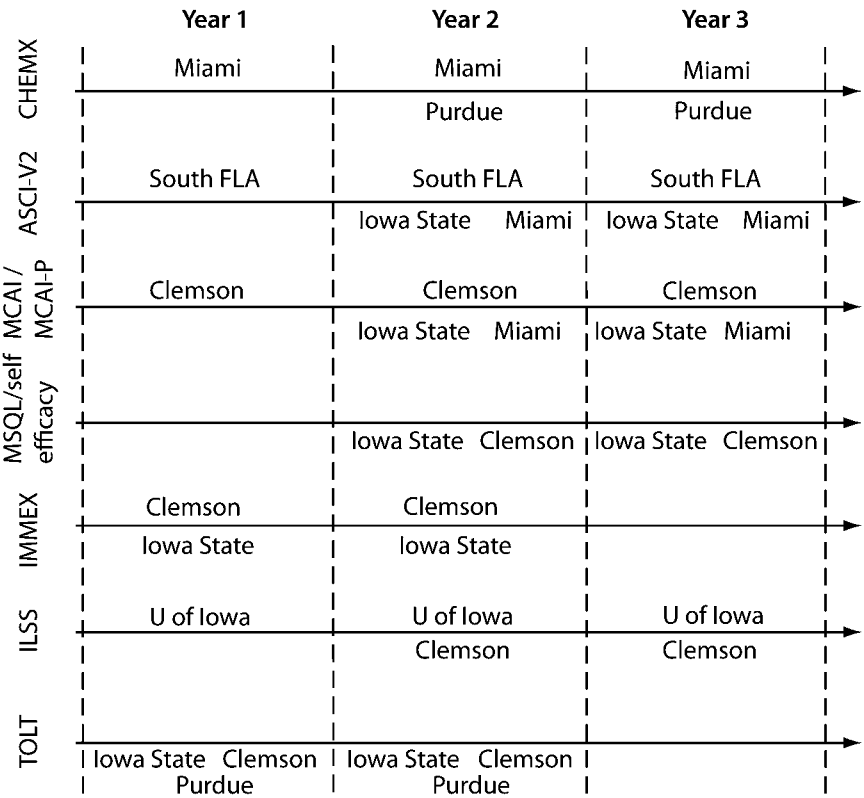

Figure 1. Timelines for implementation of assessment instruments at collaborating universities. Instruments are listed down the left side. A university listed above the arrow is the instrument developer and those listed below the arrow are instrument users.

It is an understatement to say collaborations take considerable energy and effort to keep everyone focused on the same sets of goals and outcomes. Two meetings of all PIs on the project were held in order to enhance communication and re-establish group priorities during the project. Nonetheless, with more time and funding to support the diversity of expertise and opinion, this collaboration could have been even more successful in terms of the increased transfer and scale across sites of the multiple instruments that have been developed. While quite a bit of progress was made, interviews with each of the PIs indicated their desire for more years of data to study. As such, both increased time and funding were needed to harness the overall strength of this group of researchers, to make the impact they wanted to make collectively, which they only began to do with this initial grant.

University Support for Conducting Research

Each collaborator was very thoughtful and inclusive in the way they included their graduate students in their research. The number of publications that included graduate students as coauthors testifies to the fact that these PIs knew how to mentor and support their students. In many cases, the students were given autonomy to lead parts of their research projects, as well as publish not only as secondary authors, but first authors of the research. If left to do their own research without a need to transfer and scale their work to other instructors within their department, each PI demonstrated high success in their independent endeavors, as well as through their students. It was only when the collaborators had to rely on faculty peers or college or university administrative efforts that their research agenda was compromised: by lack of buy-in, support, and investment from other faculty members, and administration to truly use the developed measures as they were intended.

When collaborators would ask other chemistry instructors to administer their assessments and use the data to inform their practice, the results from these peers were always lower than the partners in the collaborations. These findings may be a result of faculty peers not understanding the benefits as clearly as the PIs for using the information. These cooperating faculty members may also not have shared the same level of interest for improving their instruction. So from the outset, different motivations for using the assessments lead to different implementation and use of the information. In fact, two of the PIs specifically analyzed instructor's perceptions of what content they prioritized, and the types of information they used to support those beliefs (*23*). This is discussed further in the section below related to transfer and scale.

The other challenges that assessment studies often face arise from a need for student level data beyond what a PI may automatically have for students in their course. For instance, data such as previous science performance, GPA, demographic information, or other data that could be used to adjust for prior performance, or be used to demonstrate potential differences by subgroups is often unavailable or quite difficult to obtain. For courses taught by faculty members outside of the collaboration, PIs were either unable, or significantly delayed (more than two years at one university) to gain access to such data, which was critical for their analyses to demonstrate similarities and differences of classes, and to even make headway for generalizing results. Without this information, it was impossible to begin to study how well assessments scaled or transferred when used by other faculty members. This type of hold-up obviously affects the timeliness of research for individuals or groups to be published, but also demonstrates the lack of support mechanisms at universities to provide discipline-based education research (DBER) faculty members the data they need to conduct their research. Given that all universities participating in the current project have top tier research ratings, one would think that administrative efforts to facilitate research would be more universal for DBER faculty members, allowing them to contribute to the university's reputation as strong research institutions. This comment is not to say that the collaborators were directly blocked from access to additional data. Rather, this observation reflects an overall sense of institutional apathy towards being more proactive to support their faculty researchers. It is the lack of action and/or attention of university administration that is the issue here. With less tenacious individuals, the amount of publications and research that would have resulted would likely have been slim or even none. The role of institutional barriers to successful scale-up of assessment research represents a key finding of this analysis of large scale, cross-institution collaborations, even though the finding is a result within what may broadly be considered research university environments. The point to be made is that until individual university administrations are more proactive and supportive of the data they can provide faculty research, collaborations across universities will be further hindered in their ability to transfer and scale research.

Transfer and Scale: The Use of Original and Modified Instrument Measures for Comparison Across Years, Classrooms, and Universities

McDonald and colleagues define scale as "the practice of introducing proven interventions into new settings with the goal of producing similarly positive effects in larger, more diverse populations (*7*)." Part of scale includes modification and transfer, where initial research provides information to potentially improve and refine initial measures and hypotheses, and then replicate the results within similar settings or scale to other settings.

One premise of the members of the collaboration was that chemistry instructors are more likely to adopt assessment instruments that require little time to use and analyze. As such, early on in the project, a shortened version of a previously published instrument (*24*) was proposed and validated (*11*). This new ASCI instrument as well as a modified version of MCA-I were then tested across three universities, to see what findings would hold or differ across sites and demographics.

Similarly, a preliminary study of CHEMX across these three universities was conducted, to analyze how well results generalized within and across sites (*14*). While there were the hopes for the overall project to carry out work of this nature across more instruments, including more years within and across universities, in the timeframe available, only these smaller studies were achievable. However, these studies showed promise for how this information could be used to inform instructors about cognitive factors, both content and non-content related. These findings would then need to be incorporated by instructors in how they would use such data to change their practice, which is a point of further study for this team of researchers. This collaboration was only on the verge of exploring

In Trajectories of Chemistry Education Innovation and Reform; Holme, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2013.

generalizability and practical issues around implementation to study transfer and scale. This broader implementation step is where the next issue of transfer and scale comes into play.

Transfer and Scale: How Faculty Beliefs and Values Impact How Different Information Is Used for Assessing Student Learning

Within any academic discipline there are two main approaches for dealing with content coverage within a course: treating a broad range of topics rather lightly or addressing fewer topics in greater depth. A significant amount of material can be covered at a relatively superficial level while more integrated in-depth exploration of fewer topics may mean deeper understanding of a smaller range of content. Choices made in this regard have a large impact on assessment choices that accompany instruction. This project was implemented largely within introductory chemistry courses. Because these courses include a large and growing list of topics, with high expectations of mastery, coverage that balances breadth and depth is a constant challenge for instructors of these courses. As Cooper states, "general chemistry... covers too much material, thereby sacrificing depth for breadth (*25*)."

What compounds the problem of depth versus breadth is instructors' understanding of how students learn. A relative lack of familiarity often leads to the failure to use instructional strategies that would engender more student motivation and interest as well as sound pedagogical techniques for ensuring mastery of content. While not universally true, chemistry faculty members at research universities may treat the teaching of undergraduate courses as a less attractive part of their academic responsibilities. Therefore, it is unwise to assume that instructors in the large-lecture introductory courses are particularly interested in the ways students learn. It would also be a notable assumption that many of them have a profound understanding of educational assessment. Operationally, it appears more than likely that they are interested in instruments that survey a wide host of concepts (hence, wider breadth) rather than depth. Such assessment is in line with faculty beliefs that the purpose of these introductory courses is generally to provide an overview of the subject. In fact, an article focusing on the questions to ask instructors about assessment, not just as a compliance task, but actually making meaning of what assessments can do (*26*), demonstrates one form of professional development that could help research faculty make more meaning out of assessment efforts in introductory courses. Ideas such as this help frame the reasons for why assessment development work is important. In addition, research that investigates the reasons instructors enumerate as to why change may not be happening becomes vital because of different beliefs and values of what teaching is, and what is important to learn (*23*).

The role of assessment within higher education is not wholly ignored, but the willingness of many instructors to commit limited time resources to enhancing their measures of student learning is apparent. This situation may have a particularly large effect in science courses that occupy a service role in the curriculum of a majority of the enrolled students. Additionally, faculty members who lack strong foundations or interest in instructional methods may inadvertently eliminate potential chemistry majors because they did not engage all students to be successful in learning the content of these general chemistry courses. The research that the collaborators in this project conducted on misconceptions, reasoning, conceptual frameworks, and problem solving all demonstrated that if this information was not somehow attended to, students would generally lack a true understanding of complex ideas. Without such a depth of knowledge these students can be expected to be less successful at carrying the content information into future chemistry courses (if they continued). They would also miss out on opportunities to improve the way they could reason and problem solve, because they were not given more opportunities to improve upon those higher-order thinking skills. In short, without using assessment information of the type provided by this collaboration, other faculty continue to miss opportunities to help students learn the content more deeply, improve upon their general approaches to learning—including self-efficacy and metacognition—and overall, inadvertently contribute to the attrition of science majors. Thus, the challenges with obtaining buy-in from fellow faculty members that the collaborators faced in this project become a particularly important observation. *Even with high quality, publishable results from DBER efforts, the transfer of these ideas to other instructors represents a central challenge in the cause of using sound evidence of student learning in the reform of teaching and learning.* Not only does this impact the teaching efforts of instructors, but in the longer term serves to limit the pool of new science talent because students are less engaged and instructors are less focused on improving the ways their students can learn and grow. It will be important that future grants include time and money to support faculty buy-in and professional development to ensure more success in transfer and scale of reform efforts.

Summary

Ultimately, the yet unachieved goal of this collaborative was to point to a new era within chemistry assessment. Not only can there be measures to improve the ways different assessments can unpack students' misconceptions, knowledge, and interest, they can also inform instruction, to assist more students to achieve success in chemistry. And further, better understanding the expectations of faculty members informs the fledgling chemistry assessment research in what these instructors see as critical to assess and teach. These studies of the academic environment provide more insight into why certain measures or data would not be used or disregarded to inform their practice. As such, headway is being made, but more time and research is needed to forge ahead with a new era of new chemistry assessment use.

For change to happen within and across each educational level, the structures and support processes must be revised to accommodate and successfully implement (as well as sustain) large-scale change across all levels. This collaboration suggests that for scale-up, there is a need for both buy-in and support from instructors as well as administration to successfully implement these reform efforts. Additionally, it is important to understand the contexts in which faculty members operate and to better understand their own beliefs about student learning and the types of data they value for measuring student learning. Without a better understanding of sources of resistence to the use of assessment instruments, scale-up will prove challenging, especially if results vary due to faculty beliefs and values of how and what students should learn in these undergraduate chemistry courses. This point must not be taken lightly – as this issue is systemic in education, from K-12 to higher education.

For the assessment and evaluation community, there are certainly significant issues related to scaling and transferring research to more practical venues. Because of the successful and perseverant nature of each of these collaborators on this project, they were able to champion their own work, move forward with their own research agendas, and still individually be quite prolific in their publications. The progress of STEM education reform cannot become solely dependent on the resilience and persistence of individuals for research to be successful. Rather, the development of infrastructure that would enable participants to be successful in collaborative efforts becomes apparent, and a must-needed investment. Several different levels of institutional involvement are important, in particular: (1) university support from both peers and administration; (2) grant funding that provides adequate resources and opportunities for collaborators in various locations to work together onsite and over longer time scales; (3) resources to hire a person in charge of coordinating separate endeavors as their sole responsibility; and (4) an understanding from funding agencies about how long it takes to develop, cultivate, and support collaborations. There is little that can be done in a short timeframe such as 18 months, and for scale and transfer to happen, more time, money, and logistical support is needed to provide any opportunities to replicate results over multiple years across sites. Collaborators on projects of this type need to spend more time to cross-pollinate and learn from each others' efforts and results, and continue to build and refine how each participant's efforts contributes to the overall project. As shown in many individual projects, this work can be done if there is one PI and one project as the focus. There has been relatively little work, however, that attempts to understand collaborative efforts that arise from multiple individual PI projects. This is definitely a missed opportunity to synthesize individual work into something larger that has such potential for positive change to undergraduate science teaching and learning.

This chapter identified some characteristics of collaborative projects that funding agencies can use to discern probable success and to learn what it would take to help those project do well. To summarize, the need for grants that span a longer period, similar to those in scale-up studies would be most appropriate for any collaborative effort, as collaboration can be viewed as a version of scale —with slight modifications or development, rather than full scale-up and efficacy trials. If collaborations—those where PIs combine their research agendas into a larger set of studies, not just working jointly on a project within or across departments and/or universities—could be classified in their own category, that would also prove helpful to address the complexity of what it takes to truly cross-pollinate multiple research projects.

References

1. Hafernik, J. J.; Messerschmitt, D. S.; Vandrick, S. Collaborative research: Why and how? *Educ. Res.* **1997**, *26*, 31–35.
2. Wasser, J. D.; Bresler, L. Working in the interpretive zone: Conceptualizing collaboration in qualitative research teams. *Educ. Res.* **1996**, *25*, 5–15.
3. Cummings, J. N.; Kiesler, S. Collaborative research across disciplinary and organizational boundaries. *Social Stud. Sci.* **2005**, *35* (5), 703–722.
4. Steinuehler, C.; Duncan, S. Scientific habits of mind in virtual worlds. *J. Sci. Educ Technol.* **2008**, *17*, 530–543.
5. Kotecha, P.; Walwyn, D.; Pinto, C. *Deepening Research Capacity and Collaboration Across Universities in SADC*; Southern African Regional Universities Association: Johannesburg, 2011.
6. Coburn, C. E. Rethinking scale: Moving beyond numbers to deep and lasting change. *Educ. Res.* **2003**, *32*, 3–12.
7. McDonald, S.; Keesler, V. A.; Kauffman, N. J; Schneider, B. Scaling-up exemplary intervention. *Educ. Res.* **2006**, *35*, 15–24.
8. Holme, T.; Bretz, S. L.; Cooper, M. M.; Lewis, J. E.; Paek, P. L.; Pienta, N.; Stacy, A.; Stevens, R. H.; Towns, M. H. Enhancing the Role of Assessment in Curriculum Reform in Chemistry. *Chem. Educ. Res. Pract.* **2010**, *11*, 92–97.
9. Sandí-Ureña, G. S. *Design and Validation of a Multimethod Assessment of Metacognition and Study of the Effectiveness of Metacognitive Interventions*. Ph.D. Dissertation, Clemson University, Clemson, SC, 2008.
10. Cooper, M. M.; Sandi-Ureña, G. S.; Stevens, R. H. Reliable multi-method assessment of metacognition use in chemistry problem solving. *Chem. Educ. Res. Pract.* **2008**, *9*, 18–24.
11. Xu, X.; Lewis, J. E. Refinement of a chemistry attitude measure for college students. *J. Chem. Educ.* **2011**, *88*, 561–568.
12. Brandriet, A. R.; Xu, X.; Bretz, S. L.; Lewis, J. E. Diagnosing changes in attitude in first-year college chemistry students with a shortened version of Bauer's semantic differential. *Chem. Educ. Res. Pract.* **2011**, *12*, 271–278.
13. Grove, N.; Bretz, S. L. CHEMX: Assessing students' cognitive expectations for learning chemistry. *J. Chem. Educ.* **2007**, *84* (9), 1524–1529.
14. Xu, X.; Lewis, J. E. *Investigating Reliability and Validity of CHEMX Scores from Three Institutions*; 2009, unpublished manuscript.
15. Murphy, K.; Holme, T.; Zenisky, A.; Caruthers, H.; Knaus, K. Building the ACS exams anchoring concept map for undergraduate chemistry. *J. Chem. Educ.* **2012**, *89*, 715–720.
16. Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring concepts content map I: General chemistry. *J. Chem. Educ.* **2012**, *89*, 721–723.
17. Holme, T.; Murphy, K. Assessing Conceptual versus Algorithmic Knowledge: Are We Engendering New Myths in Chemical Education? In *Investigating Classroom Myths through Research on Teaching and Learning*; Bunce, D. M., Ed.; ACS Symposium Series 1074; American Chemical Society: Washington, DC, 2011; pp 195−206.

In Trajectories of Chemistry Education Innovation and Reform; Holme, T., et al.;
ACS Symposium Series; American Chemical Society: Washington, DC, 2013.

18. Villafañe, S. M.; Loertscher, J.; Minderhout, V.; Lewis, J. E. Uncovering students' incorrect ideas about foundational concepts for biochemistry. *Chem. Educ. Res. Pract.* **2011**, *12*, 210–218.
19. McClary, L. M.; Bretz, S. L. Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength. *Int. J. Sci. Educ.* **2010**, *34* (5), 2317–2341.
20. Jiang, B.; Xu, X.; Garcia, C. A.; Lewis, J. E. Comparing two tests of formal reasoning in a college chemistry context. *J. Chem. Educ.* **2010**, *87*, 1430–1437.
21. Beal, C. R.; Stevens, R. H. Improving students' problem solving in a web-based chemistry simulation with embedded metacognitive messages. *Technol., Instruct., Cognit. Learn.* **2011**, *8*, 255–271.
22. Cooper, M.; Stevens, R.; Holme, T. Assessing Problem-Solving Strategies in Chemistry Using the IMMEX system. In *Proceedings of the National STEM Assessment Conference*, National Science Foundation: Washington, DC, 2007; pp 118−129.
23. Bruck, L. B.; Bretz, S. L.; Towns, M. H. Faculty perspectives of undergraduate chemistry laboratory: Goals and obstacles to success. *J. Chem. Educ.* **2010**, *87*, 1416–1424.
24. Bauer, C. F. Attitude towards chemistry: A semantic differential instrument for assessing curriculum impacts. *J. Chem. Educ.* **2008**, *85*, 1440–1445.
25. Cooper, M. M. The case for reform of undergraduate general chemistry curriculum. *J. Chem. Educ.* **2010**, *87*, 231–232.
26. Bretz, S. L. Navigating the landscape of assessment. *J. Chem. Educ.* **2012**, *89*, 689–691.