Editorial

Discipline-Centered Post-Secondary Science Education Research: Distinctive Targets, Challenges and Opportunities

Brian P. Coppola[1] and Joseph S. Krajcik[2]

[1]*Department of Chemistry, University of Michigan, Ann Arbor, Michigan*
[2]*College of Education, College of Natural Science, Michigan State University, East Lansing, Michigan*

Received 17 June 2014; Accepted 17 June 2014

This is the second *JRST* Special Issue on *Discipline-Centered Post-Secondary Science Education Research*. The response to our focus on the distinctive role of the discipline in shaping science education research at the post-secondary level (Coppola & Krajcik, 2013) has been overwhelmingly positive. In this issue, our selection of papers raises questions on how meaningful learning outcomes at the college and university levels are influenced by the rich background and prior knowledge of post-secondary students, on how gaining understanding of relevant subject matter intersects with the ability to use it productively, and on how to bridge the transition from working in school settings to the real world needs of the baccalaureate class as they become professionals and practitioners.

Understanding Meaningful Learning

Promoting deep versus surface learning (Marton & Saljo, 1976) is an unquestioned axiom of education. Deep learning (synthesizing and integrating new knowledge in the context of prior knowledge, permitting critical problem-solving in new and unfamiliar situations) is deemed to be better than surface learning (memorization of unlinked facts with no integration into existing contexts, providing recall of information and procedural heuristics over understanding).

In his critique of situated cognition, Bereiter (1997) provides the fictionalized account of two students doing equally well when they take Algebra I, thus having been in the same situations, who nonetheless diverge when they take Algebra II. Flora, it is supposed, has used meaningful learning strategies (Ausubel, 1963) in Algebra I, while her classmate Dora has used rote and recall strategies. And while they both did well in Algebra I, presumably because the assessments could not sort them out, Dora fails Algebra II because she has not learned in a way that allows her to transfer and apply the facts and heuristics she only recalls. Dora has learned how to recognize and reproduce what is needed to do Algebra I problems, and how to take Algebra I tests, but she has not learned Algebra.

Correspondence to: Brian P. Coppola, E-mail: bcoppola@umich.edu
DOI 10.1002/tea.21165
Published online 12 July 2014 in Wiley Online Library (wileyonlinelibrary.com).

Bereiter's point in no way takes instructors off the hook for making good choices in creating thoughtful learning environments which facilitate meaningful learning. The full intent of an excellent teacher about a learning environment is never self-evident, nor can it presuppose that its users, the students, will automatically follow the prescription. Without stating the aphorism explicitly, Bereiter rightly reminds instructors and researchers alike: *you can lead a horse to water*. . . The choices made by Flora and Dora are as important as the teacher's. The perceptions and decisions made by the students in a learning environment are as critical to understanding that environment, and its educational success, as are the perceptions and decisions made by instructors. Seminal work by Ramsden and Entwistle (1981) laid the groundwork for understanding the intimate, reciprocal relationship between the design of a university-level learning environment ("teaching") and decisions made by university-level learners, as well as its connection to fundamental positive learning characteristics (Ning & Downing, 2012).

In his 1991 Editorial in *this Journal*, "Mantras, False Dichotomies, and Science Education Research," Jim Wandersee took exception to the then-common way of expressing Flora's strengths: she was learning process, not content. Wandersee explicitly and justifiably characterized "content versus process" as a canard, a false dichotomy, and implicitly suggested the usefulness of the relationships shown in Figure 1 (Ege, Coppola, & Lawton, 1997). He identified a third type of student, we'll call her Cora, with his whimsical label of "intellectual amnesiac": someone who knows how to think, but who has nothing to think about. More seriously, this category is an excellent reminder that understanding a new topic (content and process) does not take place in isolation, but embedded in the existing sets of facts, relationships, and strategies that already exist in the mind of the learner. Not having all the facts about something does not preclude Cora from creating relationships and conclusions based on what she does know. In fact, some might say Cora is building hypotheses, critical components of which are recognizing that your factual knowledge is limited and that your propositions depend on an array of assumptions.

Wandersee's formulation rightly characterizes Flora, in the "expert learning" quadrant, as a person who can take advantage of a well-structured learning environment. She builds her new knowledge by integrating what she is learning with her prior knowledge, including relevant content, the situational conditions for its use, existing processes for making meaning, and the ability to identify and evaluate useful analogies from outside the immediate domain of

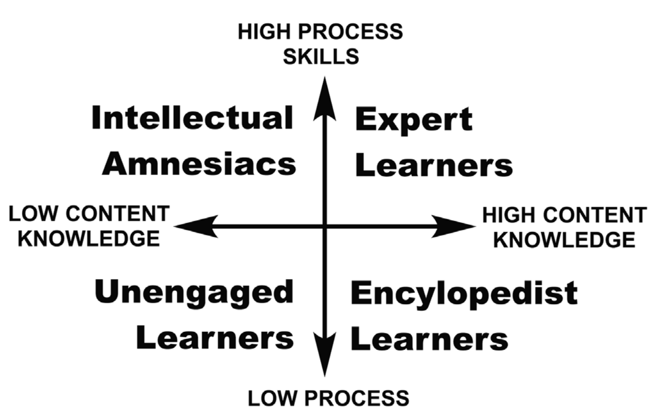

Figure 1. Resolving the content/process dichotomy.

interest. In the dichotomous "content versus process" competition, Flora can be seen as sacrificing her learning when she accumulates factual information, which is unfair because having factual information is a critical component to learning. However, factual information needs to be accumulated in an environment that promotes connection, coherence, and integration in order to be meaningful.

Although Dora has made a poor choice when she only memorizes and recalls facts and heuristic operations, as does a teacher whose learning environment only promotes these goals, she may have a foundation for learning based on relevant (and not random) associations. As an "encyclopedist," she has made a connection between a topic (the encyclopedia topic) and some relevant information (the encyclopedia entry). As in her Algebra I class, where, presumably, the assessments were built on rote and recall, she can do as well as Flora only if a critical component of the learning environment is flawed. Make no mistake: Dora is in trouble.

Let's be clear. There is no point, at any level of education, where the legitimate need to have factual information can be an excuse for poor instruction. That is, teaching and testing around a rote, recognition, and recall scheme because one version or another of this argument is used: students cannot possibly understand anything unless they first know this list of facts (Momsen, Long, Wyse, & Ebert-May, 2010). Unless the environment supports the integration of those facts into a larger framework in which students can build upon their prior learning and experiences, that list of facts will not really be known in any meaningful way, at all. For all her apparent success in Algebra I, Dora is behind the eight ball when she starts Algebra II, and she has low odds of recovering from what was ultimately a bad experience, and for persisting in an area that might need her to understand Algebra (Seymour & Hewitt, 1997).

The underlying context for Figure 1 is that a learner, and particularly one at the post-secondary level, comes into a new learning situation with years of experience to drawn from, representing an idiosyncratic mélange of facts, relationships, and processing skills, and covering a range of accuracy. This context is the one in which the new learning will occur. Acknowledging this view of learning is a point made by recent reports guiding the direction of K-12 science education (National Research Council, 2007, 2012). In addition, the idea that imperfect prior knowledge and experience can potentially derail, or at least have adverse effects upon, new learning, is also going to be amplified for post-secondary students. An important change in thinking, from "fixing misconceptions" to "building upon knowledge in pieces," is truly reflective of a student-centered perspective, acknowledging that the learning cannot be separated from its situation (diSessa, 1993; Özdemir & Clark, 2007).

In Figure 2, we want to return concretely to this point about how important a different balance in the strengths and weaknesses suggested by Figure 1 can be. Imagine that in moving from Bereiter's fictitious Algebra I to Algebra II class (1997), Flora, Dora and, now, Cora, need to extend from using simple integers to non-integers and negative numbers. Flora has the relevant recall from Algebra I, and she has integrated her knowledge of manipulating non-integers and proceeds with her learning in Algebra II. Dora's strategy for handling problems with integers, as it turned out, involved her profound ability to recall multiplication tables and their patterns, but this does not help her deal with the non-integers and negative numbers (regardless of how well she understands them as topics).

Cora, who did not always do well in upper level math, provides an intriguing set of answers: five out of six of her responses are correct. While it might be tempting to encourage her as a natural Algebra talent, comparable to Flora, who simply got a little sloppy on one of the easy examples, it turns out that Cora has a huge conceptual problem. She is exceptionally poor at multiplication and division and quite good at addition and subtraction. If you have not noticed it by now, five of these examples are from the subset of those where the operations of

Algebra I	Flora	Dora	Cora
2 x Y = 4	Y = 2	Y = 2	Y = 2
2 x Z = 6	Z = 3	Z = 3	Z = 4
Algebra II			
1.1 x A = 12.1	A = 11	?	A = 11
3.5 x B = 4.9	B = 1.4	?	B = 1.4
-1 x C = -0.5	C = 0.5	?	C = 0.5
0.75 x D = -2.25	D = -3	?	D = -3

Figure 2. Flora, Dora, and Cora do Algebra.

addition and multiplication produce the same numerical answer (Hoffmann & Coppola, 1996). Teachers who view Cora as doing quite well, as pupil who just needs to take a little more care with her work, are not using a truly student-centered approach when analyzing the situation. And researchers who might study these three students cannot use the ability to produce correct answers as a surrogate for conceptual understanding, which is an important and often-neglected lesson.

A "meaningful versus rote learning" debate (Mayer, 2002) is no less a false dichotomy than "content versus process." Cora has taken a commonsense (Friedman, Forbus, & Sherin, 2011) approach, constructing a quite consistent performance based on an analogy (Gentner & Smith, 2013; Taylor, Friedman, Forbus, Goldwater, & Gentner, 2011); that is, she uses, quite consistently and deployed correctly, the operation of addition for that of multiplication. Friedman et al.'s (2011) subjects, similar to the college graduates nearly 25 years ago, in "A Private Universe" (Schneps & Sadler, 1988), also use their everyday experiences quite consistently when constructing their explanations about the seasons. What they and Cora lack is the relevant information, and the conditions of its use, not the intrinsic ability to use it. Thus we adapt Wandersee's procedure for resolving false dichotomies into Figure 3, and remind ourselves that

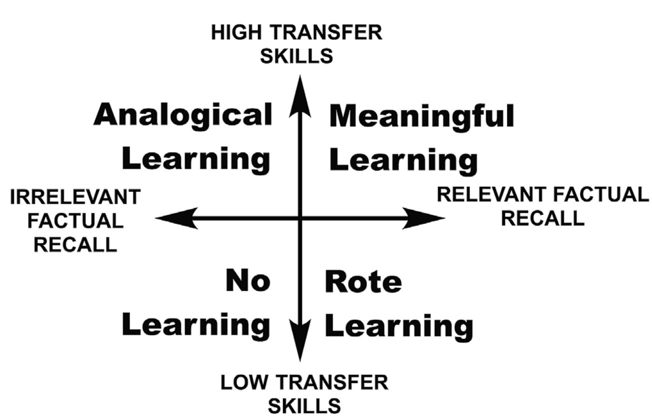

Figure 3. Resolving the rote/meaningful dichotomy.

integrating relevant factual information with how to use it is critical to Flora's *Meaningful Learning* (Linn & Elyon, 2006), and that Cora's *Analogical Learning* quadrant (operating well with incomplete or incorrect information) can be just as problematic as Dora's *Rote Learning* quadrant (incorrect or incomplete operation on the relevant content).

At least four important factors mediate the relationships shown in Figure 3.

(1) *Nothing is learned in isolation*. If this diagram maps the ways in which one can enter into the learning space, then it also sits in the rich context of everything else the student has learned as well as how that student has learned.

(2) *No mechanism for learning is implied*. The diagram does not propose a mechanism or progression, but characterizes how different balances between the attributes can lead to quite different performances. Because of Flora's rich prior context, she enters into her learning of Algebra II with a keen and productive interaction of her transfer skills and her recall and use of relevant factual information. Dora and Cora need wildly different advice in order to think about how to move them towards meaningful learning, and it must rely on an accurate diagnosis of the approaches they are using.

(3) *Rote Learning is not the same as Memorization*. Despite the Western stereotype, there is a great deal of evidence that "memorization" is neither monolithic, universally non-productive, nor a surrogate for "rote learning." Twenty years of cross-cultural studies with students in China, Japan, and India have shown that students can hold a strong orientation toward memorization as an effective pathway to understanding (Chan & Rao, 2009; Marton, Dall'Alba, & Tse, 1996; Wong & Wen, 2001). There is a difference between memorization in which some relevant associations are made and memorization in which the facts are a more-or-less random collection.

(4) *Students are autonomous*. The roles of the student as a learner and an autonomous individual, along with the decisions that he or she makes, are critical to remember. The most perfectly designed learning environment is still a trough from which the student needs to decide to drink. Novak and Cañas (2008), in summarizing Ausubel's (1963) definition of meaningful learning over rote learning, requires (p. 3):
   1. The material to be learned must be conceptually clear and presented with language and examples relatable to the learner's prior knowledge.
   2. The learner must possess relevant prior knowledge.
   3. The learner must choose to learn meaningfully.

The consequences of these ideas for post-secondary science education research are that more adult learners bring more prior knowledge and learning experience to the table, and they have probably not learned everything they know through the same starting point or by the same pathway. Their answers cannot be categorized simply as right, rote, or wrong. Ausubel's third point is critical, and on which Bereiter's commentary about situated cognition pivots: the learner's agency to integrate ideas, to make connections to prior knowledge and skills, to learn meaningfully when the opportunity exists.

Pursuing meaningful (deep, not surface) learning extends to researchers, too! Simply evaluating the accuracy of Cora's answers (Figure 2) is not enough to understand what she knows. Cora is not correct, at all, just because she is able to produce mostly correct answers. Cora's coworkers, answering interview questions about the origins of the seasons, are not wholly incorrect, either. Their prior knowledge is not only the factual information they recall, which might be incorrect or incomplete, but also the processes they bring to the information they use. As researchers, we must go far beyond just evaluating Cora's answers in order to understand what she knows, including understanding her prior knowledge, her learned thinking processes, and details about the environments in which she learned them. She might be reflecting what she does or does not understand, but she might also be reflecting exactly what the expectations and practices were in her prior experiences.

Like all learning environments, post-secondary learning environments are complex, involving many critical resources that exist outside of the classroom itself, and which might have been generated by the current students, by past students, and by the institution. There is no common core in higher education, so the educational specifics of a class (goals, methods, implementation, assessment) are usually highly idiosyncratic, often to the individual instructor, and so examining student learning and achievement cannot be dissociated from these, nor easily tagged to only the classroom environment. The detailed interaction between students and the components of their learning environment must be understood, from the standpoint of the intent of the designed and implemented (by the instructor), from how it was received and utilized (by the learner), and from how aligned these two things are with the assessments that are used.

The Special Issue

We received 32 submissions from our call for papers, which is twice the response from last year's solicitation. The peer-review process resulted in nine manuscripts that were sent out for revisions, and ultimately yielded the five articles comprising this issue. We have two articles in the area of undergraduate chemistry (general and organic), one in graduate chemistry, one in evolutionary biology, and one that is multidisciplinary. Teaching for students from underrepresented population is featured in several articles, as is a direct comparison of education in an online versus face-to-face environment.

Dr. Vicente Talanquer, Professor of Chemistry and Biochemistry at the University of Arizona, a post-secondary science education researcher, frequent contributor to, and reviewer for, *this Journal*, and an award-winning classroom educator, has authored the closing essay for the Special Issue. He provides thoughtful and provocative challenges to the discipline-based research community to advance the field.

Article1: Culturing Reality: How Organic Chemistry Graduate Students Develop Into Practitioners

In the first article, Bhattacharyya and Bodner examine the highly discipline-centered transition of first-year organic chemistry graduate students, through the pedagogical orientation of their highly skill-based course in organic synthesis, by comparing the evolution of their problem-solving abilities in organic synthesis to the strategies used by third-year graduate students. Using in-depth, longitudinal interviews around a performance-based task, the authors explore highly contextualized ideas and artifacts using ethnomethodology, which is based on gathering data through recording the normal daily experiences of its subjects.

Similar to Feldon et al.'s account on the effect of graduate teaching on research (2011), the task used by the third-year students in this study is authentic: developing an independent research proposal. The appreciation by the learners that their education has involved authentic, or "real world," activities is concluded as a key feature in the observed growth in their epistemological understanding. These authors also triangulate their interviews with the students with additional data derived from understanding the learning environment in which this development takes place. Implications for improving the transition at the undergraduate level towards a more legitimate participation in the discipline include the explicit connection with the primary literature in the design of undergraduate research and textbooks, which is aligned with earlier work on examinations (Coppola, Ege, & Lawton, 1997).

Article 2: Replicating Peer-Led Team Learning in Cyberspace: Research, Opportunities, and Challenges

Varma-Nelson and her coworkers have studied the replication of the face-to-face Peer-Led Team Learning (PLTL) program in a distance, but synchronous, cyber-environment (cyber-PLTL; cPLTL). Using a strong experimental design, where a group of the same peer leaders facilitate instruction under both conditions, a collection of different data sources is used. The authors demonstrate that adding detailed discourse analyses on student work from a set of 24 comparable PLTL and cPLTL group sessions (12 each, with three samples from each of four instructors), using features from a deep learning model, reveals interesting differences in what students do in each of these conditions.

This study benefited from digging past reporting the simple effect: that the less finely grained measures of academic performance, such as aggregated grades, pointed to comparable outcomes under each of the conditions. By exploring the details of what students were actually doing during their sessions, the authors present a compelling hypothesis, based on their evidence, that students in the cyber-groups may demonstrate a higher degree of constructivist orientation in their learning than in the face-to-face groups. Students in the cyber-group, perhaps mediated by the more formal infrastructural demands of their setting were more focused on the problem-solving process of their work compared with students in the more informal, easily digressed face-to-face setting, where mutual agreement about "getting the right answer" tended to shut down any additional conversation.

Article 3: Ethnically Diverse Students Knowledge Structures in First-Semester Organic Chemistry

Understanding the origins of why certain groups remain underrepresented in the STEM fields is an important problem. Lopez and his coworkers have continued to explore the organic chemistry setting, which is a key gateway course. These researchers have used student-generated concept maps as a source of evidence about the knowledge structures constructed during learning, in comparing the understanding and achievement of relatively large group of 90 students from diverse ethnic backgrounds while taking this course. In order to assess the disciplinary validity of the propositions contained in the maps, disciplinary experts were needed. A second, holistic analysis of the maps was also performed to provide complementary data.

Although prior academic performance and ethnicity are difficult to disentangle, the researchers have used their multiple methods of analysis to point strongly to the role of prior achievement as the mediator in differences they observe. As the development of expertise in the discipline follows from conceptual understanding combined with socialization, understanding the things that differentiate students on the first day of class is critical. The issues raised by these researchers are not at all settled, by their own account, but their study directs them, and others, to explore more deeply the question of who is walking into the classroom by what their experiences have been, and not so much by other demographic information. As such, their work points to the importance of learning more about our students than the indices of demographic information.

Article 4: Deconstructing Evolution Education: The Relationship Between Micro- and Macroevolution

In their paper, Novick and her coworkers look at a specific and generally troublesome feature of learning evolutionary biology, namely, the relationship between microevolutionary concepts, such as natural selection, with corresponding macroevolutionary ones, such as Tree of Life (ToL) thinking. In a sample of 124 students assessed for their prior knowledge of natural selection, half of the subjects received self-paced but highly directed and explicit instructional materials related to ToL thinking (the representations, their interpretation, and use), while the other half engaged a comparable level of effort on general science reasoning activities.

The findings in this study, derived from testing the integrated understanding of micro- and macroevolution held by these students, counteracts a dogmatic belief that understanding natural selection automatically transfers to an understanding of ToL thinking. As in Bhattacharyya and Bodner (Article #1), the success of the instructional environment is attributed, in some part, to its connection with authentic, or "real work," (Coppola, in press) representation of science.

Article 5: College Chemistry Students Understanding of Potential Energy and Atomic–Molecular Interactions

Energy has emerged as an explicit, crosscutting concept in K-12 science education (Chen, Eisenkraft, Fortus, & Krajcik, 2014; National Research Council, 2012). In their study, Becker and Cooper have looked at the existing understanding of potential energy, in the context of chemistry, with first- and second-year undergraduate chemistry students, as a way to understand their prior knowledge about this area, from their precollege education, and how it may or may not have been integrated into their university education. Using a set of written, open-ended surveys with 333 students from three courses, in addition to semi-structured interviews with 18 students from these classes combined with four other upper division students, the researchers probed students' understanding of what potential energy means at the atomic–molecular level.

These authors found that while students' explanations fell into three more or less useful and comprehensible categories, their understanding of how those categories (capacity for work, stored energy, and stability) related to notions of potential energy were incomplete, incorrect, and/or incoherent. These intuition-based explanations, inevitably derived from the course contexts in which they were learned, functioned operationally in these classes, but nonetheless broke down when examined in detail. The question posed by the K-12 Framework for Science Education (National Research Council, 2012) is whether reduction of the historically diverse way in which the disciplines treat energy to a common, foundational core, can improve both the depth of understanding held by learners within the disciplinary units as well as across the multi-disciplinary spectrum. These findings suggest that this is likely to be a significant challenge.

Closing Commentary: DBER and STEM Education Reform: Are We Up to the Challenge?

In his essay, Talanquer lightly characterizes the community of post-secondary science education research as being constrained by its origins as well as the history and traditions of precollege science education research. Understanding how post-secondary disciplinary expertise might affect the design and scope of research was one of the topics we also speculated about in our previous editorial (Coppola & Krajcik, 2013).

In describing a set of challenges, Talenquer artfully challenges the community itself. At the same time, he sees some of the underlying strengths of the papers included in this Special Issue in how they begin to model the way to break down some of the walls that he sees surrounding the DBER community.

Talanquer's challenges are:

(1) *Collaboration* with coworkers in science education, psychology, educational psychology, and the learning sciences in general, is important because all of the underlying theories derive from these areas which continue to evolve there. It is unnecessary for discipline-centered post-secondary to reinvent wheels, and better focus on working in the most informed way possible, including the interaction with our campus coworkers

who need our disciplinary expertise to help translate and apply their core ideas in our settings. Three of our articles (#2, #3, #4) illustrate the value of these collaborations.

(2) *Conceptual Integration* is education's Holy Grail, or perhaps its Unified Field Theory, to draw a different comparison. The underlying presumption is that the world was a perfectly well integrated and holistic place before the continuous fracturing and disintegration that has taken place as our ability to produce knowledge has so far outstripped any individual's ability to know what is known (Coppola & Daniels, 1998). Article #5 supposes that an underlying conceptual re-integration across the area of energy can benefit the student learning in the disciplines as well as their ability to cross between the disciplines, and it provides a sense of this as a great challenge.

(3) *Development of Expertise* must be of continuously emergent importance as a learner moves from the precollege to graduate settings. How and when to represent more and more authentic work, connected to the state of the art held by the stewards of the disciplines, is a powerful concern for research at the post-secondary level. Article #1 provides leadership in examining the post-secondary to graduate level transition.

(4) *Diversity*, increasing the persistence, success, and representation of historically underrepresented populations, continues to be a huge challenge at ever level of the educational enterprise. Research on multiple fronts is absolutely needed in order to break stereotype, dispel myth, and identify the key pressure points for effecting change. One of the articles (#2) looks at the question of student performance in the context of the given intervention, while two of them (#3, #4) include the explicit component of how prior knowledge and experience may influence achievement.

(5) *Translation of Research Results into Practice* is a complex problem. We suspect that as you move into the post-secondary level, the needs and expectations from the disciplines, their traditions and their dispositions, will influence greatly the design and implementation of ideas. Myths about magic bullets (Feldon, 2010) abound: "teacher proof" materials, discipline-neutral implementation, discounting and ignoring the autonomous decisions of students, the diversity and complexity of the learning environment outside of the classroom. Two of these articles (#2, #4) describe activities that are highly informed and anchored to the post-secondary context, while a third (#5) alludes to a new program that derives from conceptual integration.

Talenquer's final challenge is for researchers to move their collaborations into areas where the understanding of the subject matter is not the only focus, which is highly represented, but to deepen the meaningfulness of that understanding beyond the surface features of learning, which is far less represented. He also argues for increasing the breath of study, to how we achieve, and can more effectively achieve, the scientific and intellectual dispositions that ought to emerge in a truly educated person in our postsecondary education system.

Evolving Our Own Meaningful Understanding of Teaching, Learning, and Assessment

In this section, we are going to follow Professor Talanquer's lead and suggest four target areas where research in post-secondary science education might evolve to the benefit of both its constituency (tertiary level instructors) and its community (education research, in general). We are going to use the development of these two Special Issues as our inspiration. These four themes recurred through the process of vetting the first round of submissions for the two Special Issues, mirroring common areas were we, and reviewers, provide feedback to help improve manuscripts generally.

(1) Do not only report the effect, examine what produces the effect (what, why, how, under what conditions. . .).

Evaluation versus Research. Based on our collective experience as editors, one of the first filters used by reviewers is the question: is this an evaluation report or is it a research study? This question comes up more frequently in response to studies carried out by the emergent group of investigators from the post-secondary, discipline-based community than it does for others. One hypothesis for this is that the ubiquitous offices for institutional research and evaluation have dominated the collection of data for academic accountability and regulatory compliance in college and university settings. We cannot be the only ones to encounter an administrator who reminds us "these are just data, and you can do what you want with data; why do the standards of research have a bearing on this?"

Generating results by short-circuiting research standards is a cottage industry for so-called Institutional Evaluation. A recent article, for example, was titled "The Counterfactual Self-Estimation of Program Participants: Impact Assessment Without Control Groups or Pretests," a method in which "program participants are capable of estimating the hypothetical state they would be in had they not participated" (Meuller, Gaus, & Rech, 2014). Administering a single retrospective survey, after a mathematics class, that asks students to self-assess the learning gains they have made in mathematics because of that class, is constrained by many theoretical and methodological limitations (Finney, 1981; Poggio, Miller, & Glasnapp, 1987). Reporting this out as evidence for an increase in quantitative reasoning skills, which is not problematic for Institutional Evaluation, obviously does not come near the standards for pushing the field forward with respect to what can account for increased learning, nor do these reports met the standards for publishable work in *JRST* (or many other places, we imagine).

Currently, concept inventories are proliferating (Libarkin, 2008). These short, multiple-choice, standardized tests exist at the border between evaluation and research. The earliest and most commonly used inventories, and nearly all those to come later, are built upon the older misconceptions literature which, as indicated above, has yielded to new ideas about meaningful learning. How has the psychometric validity been affected by this? In addition, while these tests are usually focused, they still cover much ground, often with only 1–2 items to reveal student understanding on a given topic. Another lingering concern is the degree to which students, through the content bias that the examinations might have on instructors, are being more narrowly prepared for these specific situations. Smith and Tanner (2010) provide a thoughtful and balanced review of the benefits and limitations of these exams.

The connection between selecting a correct answer and the application of conceptual understanding (one of three goals from the transformational teaching model) is not automatic. Regardless of how well selecting the correct answer might be derived from conceptual understanding, correct answers can also derive from other sources that have nothing to do with conceptual understanding at all (remember Cora and Dora?).

If an instructional changes, it is worth investigating. The two most common sources of evidence are at least suspect in their superficial nature: (a) fewer students fail (the DFW rate, students who earn grades of D, F, and who have withdrawn from the course), and (b) a short, standardized multiple choice exam is used as a pre- and post-test (these are the concept inventories), and a gain score is reported (Smith and Tanner, 2010). The observations of higher grades and improved standardized exam scores are unquestionably true, but without an aligned chain of evidence gathered through a triangulated research design, claims about why the outcomes are different cannot exceed correlation. Cora: again. The field needs to learn more about why these changes have been successful.

Advocates for "scientific (evidence-based) teaching" (Handelsman et al., 2004; Handelsman, Miller, & Pfund, 2007) ought to not compromise on the existing standards of evidence for the complex, social science of education research. Every claim of effective instructional intervention ought to include evidence for and alignment between the intentionality in the pedagogical design, observation and open coding of the implementation, interview and/or observation of student work, analysis of artifacts, independent performance-based assessment, and serious consideration of all alternative hypotheses and whether there is any evidence for falsification.

The observed outcomes from the various "active learning" classrooms are real (Freeman et al., 2014), but, scientifically, their enthusiastic advocates need to follow good scientific practices and separate the observation from its attribution. Conceptual understanding is not observed. What is observed is typically two things: fewer students get failing grades, and students show gain scores on short, standardized, multiple-choice exams. Without evidence to the contrary, multiple hypotheses for these outcomes are potentially operating (beyond the accomplishment of meaningful learning). To date, for example, we are unaware of serious research that has started with the hypothesis that students in these settings are being targeted for test-training, resulting in robust heuristics that allow them to recognize, select or generate correct answers more efficiently. Is this the same as meaningful learning with conceptual understanding? We do not know; it needs to be examined.

Without a doubt, the active classroom observations have changed the way a large fraction of instructors think about teaching and assessment, which may be the most positive result. Perhaps simply having fewer students fail is a desired outcome, because it might increase the fraction of students who stay in the science pipeline, and give those who leave a lingering positive impression of science. But are these active classrooms also developing more Floras, and not simply improving the testing skills of Dora and Cora? There are many unanswered questions for which research can provide evidence.

(2) Shift the focus from understand how well students do on science exams to how well they are learning science.

As described previously (Coppola & Krajcik, 2013), discipline-centered, post-secondary science education is more likely to be carried out by practicing scientists who carry deep and complex disciplinary dispositions as a integral part of their understanding. College and university science instruction is not bounded by highly defined standards of content, depth, scope, or sequence. As a result, even within a disciplinary department, the sense of what constitutes a legitimate understanding of a given topic can vary from one sub-discipline to another. While an organic chemist might rely on a broad descriptive application of electrostatic attraction and repulsion (electrophile/nucleophile; HOMO/LUMO) in describing chemical reactivity, perhaps without ever invoking the term, a physical chemist might well be interested in the exact solutions to equations that describe the properties of a hydrogen atom with Coulomb, overlap, and exchange integrals that include terms for proton–proton repulsion, electron–electron repulsion, and proton–electron attraction.

Beyond the introductory chemistry level, in which topics are as usually as constrained and pre-defined as at the precollege level, there is simply no universal notion of how crosscutting ideas such as free energy, acid-base chemistry, and Molecular Orbital Theory, which appear in every class, are used by authentic practitioners. These differences are even further exaggerated as you move from one scientific area to another, which is not a problem, but simply the way the traditions of each discipline have evolved to handle their different needs.

Consequently, post-secondary science education research that seeks to understand work in the discipline needs to start, at least, with the prevailing dispositions and disciplinary cultures that exist in these areas. And the view of what constitutes the university-level education in any one of the scientific disciplines is idiosyncratic: there are no examples of sustained, long-term curricula or standard curriculum materials in higher education, probably because disciplinary practitioners are likely to draw from their first-hand scientific experience on a highly *ad hoc* basis.

This chaotic picture of instructional design and implementation is far less likely to be experienced at the precollege level. It is easier to carry out research when the topical items, and how they are taught, have been agreed upon by the field and specified by a set of standards. At least for now, post-secondary science education is still in the hands of disciplinary experts who can draw deeply from their scientific knowledge and their personal experience with scientific practices, both of which vary greatly as one moves from area to area and from university to university, if not course by course. In fact, arguably, what drives "the prevailing understanding of chemistry... physics... biochemistry..." at the post-secondary level is not really captured well at all by textbooks, which are a distance echo from where the disciplinary practitioners draw their understanding, but might well be better captured by meta-analyses of dissertations and journal publications.

(3) Explore the breadth of learning that a college-educated person acquires, as well as the breadth of the learning environment in which it is acquired.

Implicitly, Slavich and Zimbardo (2012) contend that a college-educated person in a given area has (a) mastered a conceptual understanding in that area, (b) improved their ability to learn, and (c) accrued positive, learning-related attitudes, values, and beliefs. Strong, discipline-centered research evidence, as opposed to evaluation, that can support all three of these outcomes would be a welcome addition to our understanding about higher learning. Learning how to document added values, such as scientific skepticism, complex reasoning, creativity, personal leadership, or civic contributions, would provide powerful antidotes to popular public criticism about whether undergraduates are actually learning anything (Arum & Roska, 2011):

> "With regard to the quality of research, we tend to evaluate faculty the way the Michelin guide evaluates restaurants," Lee Shulman, former president of the Carnegie Foundation for the Advancement of Teaching, recently noted. "We ask, 'How high is the quality of this cuisine relative to the genre of food? How excellent is it?' With regard to teaching, the evaluation is done more in the style of the Board of Health. The question is, 'Is it safe to eat here?"' Our research suggests that for many students currently enrolled in higher education, the answer is: not particularly.

Along these same lines, the traditions of instructional development and education research at the precollege level focus heavily on the classroom and the student–teacher interaction. At the post-secondary level, however, the classroom is a fraction, and sometimes a small fraction, of the learning environment available to students. Resources and how to use them, generated by instructors, peers, the institution, and by the students themselves, creates a broad array of options and potentially millions of different combinations that a student might try. In-depth research on the diverse character of the university-level learning environment (*how we teach for meaningful learning*) is needed, as is the highly aligned and deeply detailed results of how resources are used successfully (*how students learn meaningfully*), as well as how meaningful learning is assessed. Because no system is perfect, or likely to be close to it, research is going to reveal weaknesses as well as strengths, inadequacies in addition to successes, and as many places where improvement is needed as places where effective instructional, learning, and/or assessment practices are being implemented.

(4) Excellent reporting and excellent research are synergistic.

An important criterion used by reviewers is whether the paper "meets the standards for publication." The clarity of the written report of the research is certainly seen as reflecting the clarity of thinking that has gone into the research itself. Anecdotally, some excellent authors report that thinking about what is needed for the paper helps guide them in the earliest design stages of a research project. Although, we are not reviewing what makes an excellent paper, we have seen the following easily avoided aspects recur in our submissions, and in what the reviewers comment upon the most.

Failing to connect the dots: The narrative of our science education research papers is not driven by giving a blow-by-blow retelling of the diary of a project. Unless it is critical to the findings, reviewing the circuitous route from an original idea to the final results is not purposeful. These papers are highly stylized and constructed narratives, driven by the basic coherence and alignment features of an argument. Although we might select a number of important ideas, here, the one that is overridingly problematic is a paper that does not provide enough information to review it. The reviewer is a keenly interested member of the readership who is driven by one question: do I believe this? And in order to believe it, the author needs to clearly lead the reader along a path: Is the context for this study compelling, that is, is answering the research question (a) something that the community cares about, (b) a contribution to our understanding, and (c) conceivably answerable in the scope of a finite study? Are the design, methods, and data appropriate and defensible (can they answer the question)? Is there sufficient detail about every decision that is key to understanding the study, such that the experiment can be reproduced if need be?

Internal bias: In 1834, Justus Liebig wrote "the most beautiful theories are destroyed by these damned experiments. . ." (Berzelius, 1982) which is a lovely reminder that we are all enamored, at some level, by our favorite hypotheses to the exclusion of others. By far, good papers are returned for revisions for one reason more than any other: over-attribution of the evidence in making a claim, failing to clearly separate the observations from their interpretation.

External bias: In a post Bayh-Dole world (Coppola, 2001), where intellectual property ownership of discoveries remains with the faculty investigator, much of the motivation for research has moved into Pasteur's Quadrant (Stokes, 1997), where the commercialization of the applied results from research accompanies the pursuit of knowledge. The same has been true for education research. Prior to this time, the idea of partnering basic education research with a commercial interest was not common; today, funding might well be tied with identifying a publisher or some other form of distribution for the product being studied. Not every paper that studies an educational system is automatically trying to sell that system, so it cannot be assumed. On the other hand, we encounter examples of papers whose *raison d'être* appears to be driven by getting a high-profile publication, where the advertising department for a given product is ready to move on an ad campaign that begins with the same four words (all the time): "The data show that. . ." This type of success clearly contributes to one's ability to compete effectively for continued funding, so the stakes are high. It does not matter that commercialization has crept into basic research because we are simply at that point in our scientific cultural history. What does matter is that researchers and authors are conscientious about not compromising their research practices, including the implication of external bias.

Acknowledgments

As guest editors for these two Special Issues, and as a regular associate editor and editor for *JRST*, we wanted to open up the dialog on Discipline-Centered Post-Secondary Science Education Research. We are grateful for the positive response from the community to last year's issue, and for the high quality work represented, in both issues, in their ten papers and two closing commentaries by Professors Susan Singer and Vicente Talanquer. Collectively, we think these articles clearly point to the promise that this research has to offer.

References

Arum, R., & Roska, J. (2011). Are undergraduates actually learning anything? Chronicle of Higher Education, January 18, 2011. http://chronicle.com/article/Are-Undergraduates-Actually/125979/

Ausubel, D. P. (1963). The psychology of meaningful verbal learning. New York: Grune and Stratton.

Bereiter, C. (1997). Situated cognition and how to overcome it. In D. Kirshner & J. A. Whitson (Eds.), Situated cognition: Social, semiotic, and psychological perspectives (pp. 281–300). Hillsdale, NJ: Erlbaum.

Berzelius, J. J. (1982). Berzelius und Liebig—Ihre Briefe von 1831-1845 ; Göttingen, Germany: Jürgen Cromm (p. 94): Die schönsten Theorien werden durch die verdammten Versuche über den Haufen geworfen, es ist gar keine Freude mehr Chemiker zu sein.

Chan, C. K. K., & Rao, N. (Eds.). (2009). Revisting the Chinese Learner: Changing Contexts, Changing Education CERC Studies in Comparative Education 25. Hong Kong: Comparative Education Research Centre.

Chen, R. F., Eisenkraft, A., Fortus, D., & Krajcik, J. (Eds.). (2014). Teaching and learning of energy in K-12 education. New York, NY: Springer.

Coppola, B. P. (2001). The Technology Transfer Dilemma: Preserving morally responsible education in a utilitarian entrepreneurial academic culture. HYLE: International Journal for Philosophy of Chemistry, 7, 155–167.

Coppola, B. P. (in press). Do real work, not homework. In J. Garcia-Martinez & E. Serrano-Torregrosa (Eds.), Chemistry education: Best practices, innovative strategies and new technologies. New York: Wiley.

Coppola, B. P., & Daniels, D. S. (1998). Mea Culpa: Formal Education and the Dis-Integrated World. Science and Education, 7, 31–48.

Coppola, B. P., Ege, S. N., & Lawton, R. G. (1997). The University of Michigan Undergraduate Chemistry Curriculum. 2. Instructional strategies and assessment. Journal of Chemical Education, 74, 84–94.

Coppola, B. P., & Krajcik, J. S. (2013). Discipline-centered post-secondary science education research: Understanding university level science learning. Journal of Research in Science Teaching, 50(6), 627–638.

diSessa, A. A. (1993). Toward an epistemology of physics. Cognition and Instruction, 10, 105–225.

Ege, S. N., Coppola, B. P., & Lawton, R. G. (1997). The University of Michigan Undergraduate Chemistry Curriculum. 1. Philosophy, curriculum, and the nature of change. Journal of Chemical Education, 74, 74–83.

Feldon, D. F. (2010). Why magic bullets don't work. Change, 42(2), 15–21.

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., . . . Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. Science, 333, 1037–1039.

Finney, H. C. (1981). Improving the reliability of retrospective survey measures. Evaluation Review, 5, 207–229.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8410–8415.

Friedman, S. E., Forbus, K. D., & Sherin, B. (2011). Constructing and revising commonsense science explanations: A metareasoning approach. Proceedings of the AAAI Fall Symposium on Advances in Cognitive Systems.

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. In D. Reisberg (Ed.), The oxford handbook of cognitive psychology (pp. 668–681). New York, NY: Oxford University Press.

Handelsman, J., Ebert-May, D., Beichner, R. J., Bruns, P., Chang, A., DeHaan, R., . . . Wood, W. B. (2004). Scientific teaching. Science, 304, 521–522.

Handelsman, J., Miller, S., & Pfund, C. (2007). Scientific teaching. New York, NY: W.H. Freeman and Company.

Hoffmann, R., & Coppola, B. P. (1996). Some heretical thoughts on what our students are telling us. Journal of College Science Teaching, 25, 390–394.

Libarkin, J. C. (2008). Concept Inventories in Higher Education Science. Prepared for the National Research Council Promising Practices in Undergraduate STEM Education Workshop 2 (Washington, DC, Oct. 13–14, 2008).

Linn, M. C. Elyon, B.-S. in Handbook of Educational Psychology, P. A. Alexander, P. H. Winne, Eds. (Erlbaum, Mahwah, NJ, ed. 2, 2006), pp. 511–544.

Marton, F., Dall'Alba, G., & Tse, L. K. (1996). Memorizing and understanding: The keys to paradox? In D. A. Watkins & J. B. Biggs (Eds.), The Chinese learner: Cultural, psychological and contextual influences (pp. 69–84). Hong Kong: CERC and ACER.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning—1: Outcome and process. British Journal of Educational Psychology, 46, 4–11.

Mayer, R. E. (2002). Rote versus meaningful learning. Theory Into Practice, 41(4), 226–232.

Meuller, C. E., Gaus, H., & Rech, J. (2014). The counterfactual self-estimation of program participants: Impact assessment without control groups or pretests. American Journal of Evaluation, 35(1), 8–25.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Sciences Education, 9(4), 435–440.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K-8. Committee on Science Learning, Kindergarten Through Eighth Grade. In R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.), Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council. (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Ning, H. K., & Downing, K. (2012). Influence of student learning experience on academic performance: The mediator effects of self-regulation and motivation. British Educational Research Journal, 38, 2.

Novak, J. D., & Cañas, A. J., (2008). The theory underlying concept maps and how to construct and use them, technical report IHMC CmapTools 2006-01 Rev 01-2008, Florida Institute for Human and Machine Cognition, available at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConcept-Maps.pdf

Özdemir, G., & Clark, D. B. (2007). An overview of conceptual change theories. Eurasia Journal of Mathematics. Science & Technology Education, 3(4), 351–361.

Poggio, J. P., Miller, M. D., & Glasnapp, D. R. (1987). The adequacy of retrospective judgments in establishing instructional validity. Educational and Psychological Measurement, 47, 783–793.

Ramsden, P., & Entwistle, N. J. (1981). Effects of Academic Departments on Students' Approaches to Studying. British Journal of Educational Psychology, 51, 368–383.

Schneps, M. H., & Sadler, P. M. (1988). A private universe. Pyramid films (http://www.learner.org/teacherslab/pup/).

Seymour, E., & Hewitt, N. C. (1997). Talking about leaving: Why undergraduates leave the sciences. Boulder, CO: Westview Press.

Slavich, G. M., & Zimbardo, P. G. (2012). Transformational teaching: Theoretical underpinnings, basic principles, and core methods. Educational Psychology Review 24(4), 569–608.

Smith, J., & Tanner, K. (2010). The problem of revealing how students think: Concept inventories and beyond. CBE Life Sciences Education, 9(1), 1–5.

Stokes, D. (1997). Pasteur's Quadrant. Washington, DC: The Brookings Institution.

Taylor, J. L. M., Friedman, S. E., Forbus, K. D., Goldwater, M., & Gentner, D. (2011). Modeling structural priming in sentence production via analogical processes. Proceedings of the 33rd Annual Conference of the Cognitive Science Society (CogSci). Boston, MA.

Wong, K.-C., & Wen, Q. (2001). The impact of university education on conceptions of learning: A Chinese study. International Education Journal, 2(5), WCCES Commission 6 Special 2001 Congress Issue (http://iej.cjb.net).