Expertise in University Teaching & the Implications for Teaching Effectiveness, Evaluation & Training

Carl Edwin Wieman

Universities face the challenge of how to teach students more complex think-ing and problem-solving skills than were widely needed in the past, and how to teach these to a much larger and more diverse student body. Research advanc-es in learning and teaching over the past few decades provide a way to meet these challenges. These advances have established expertise in university teach-ing: a set of skills and knowledge that consistently achieve better learning out-comes than the traditional and still predominant teaching methods practiced by most faculty. Widespread recognition and adoption of these expert practices will profoundly change the nature of university teaching and have a large ben-eficial impact on higher education.

University teaching is in the early stages of a historic transition, changing from an individual folk art to a field with established ex-pertise, much as medicine did 150 years ago. What is bringing about this transition and what can we expect of it? To answer, I start with the na-ture of expertise and how it applies to the context of academic disciplines. In particular, I discuss how such expertise defines disciplines and how re-search and other scholarly work plays an essential role in establishing dis-ciplinary expertise. Then I show how recent research has established exper-tise in university teaching: a set of instructional practices that achieve better student outcomes than traditional teaching methods. These advances also illustrate the essential role that disciplinary expertise has in effective uni-versity teaching and provide perhaps the best justification for the research university as an educational institution. However, while disciplinary exper-tise is a necessary part of good university teaching, it is far from sufficient: there are many other elements of teaching expertise. I conclude by arguing that the widespread recognition of expertise in university teaching will im-prove both the effectiveness and efficiency of teaching by making it a more

collective and coherent endeavor with better-defined standards for evaluation and training.

There is a general process by which expertise is established in any human endeavor; this applies to both academic disciplines and university teaching. In many areas of human activity, including music, sports, and medicine, the concept of "expertise" is well known. In these areas, there are individuals who can consistently achieve measurably better results than most people. Much of the research and discussion on expertise has focused on what it is about uniquely high-performing individuals that sets them apart. But what is the nature of expertise more generally? What are the requirements for associating expertise with an area of activity? And how does this concept of expertise apply to academic disciplines and university teaching?

There is a large literature on expertise, both what it is and how it is acquired. I will use the definition given by cognitive psychologist Anders Ericsson, slightly paraphrased: *expertise is a specific set of skills and knowledge that are not widely shared and can be seen to consistently produce measurably better results when applied to relevant tasks*.[1] Thus, for an activity to involve expertise, there must be readily identifiable tasks, and there must be measurable outcomes. The research shows that a person's level of expertise or, equivalently, "competence level" steadily increases with the amount of time spent in appropriate learning activities. For mature disciplines, reaching the highest levels (becoming an "expert") requires thousands of hours of practice.[2] When I refer to an "expert" here, I mean a recognized successful practitioner in the discipline; for example, the equivalent of a typical university faculty member.

From the studies of expertise across multiple fields, including my own research looking at different academic disciplines, I argue that, in the context of academic disciplines, expertise is primarily defined in terms of a set of decisions. It is *applying the skills and knowledge of the discipline to make decisions with limited information in relevant novel contexts*. The quality of those limited-information decisions – be they which scholarly question or problem to pursue, which information is relevant and which irrelevant, choosing methods of analyses, how to structure an argument, choosing standards of evidence, or justification of conclusions – all rely on the standards of the discipline. An activity can only exist as a recognized discipline if there are consensus standards that are used to evaluate the quality of scholarly work (such as the quality of the decisions embodied in that work) and, correspondingly, the quality of scholars in a field (for example, in academic hiring and promotion decisions). A requirement for the establishment of such standards is a foundation of "research"/ scholarly work that has demonstrated that, among the possible alternative

decisions that a person might make, there are particular choices and processes for making such decisions that consistently achieve better results.

In some activities, particularly sports, there are clear quantitative measures of overall performance, and so the "research" proceeds rapidly, establishing which practices and training methods lead to improvements in outcomes. In a new video game, for example, the establishment of expertise in game performance happens very rapidly. In academic disciplines, the outcomes, and the connections between performance elements (like decisions) and outcomes, are more complex. Then the research process proceeds more slowly, as extensive research is needed to establish what factors do and do not impact outcomes, and over what range of contexts and performers.

To establish levels of competence and guide improvement, it is also essential to resolve expertise in a field into the set of subskills or practices required in the ultimate performance. For example, rather than simply having standards as to what constitutes well-played violin music, there are accepted standards as to what is good fingering technique, bowing technique, and so on that the "research" by music teachers has shown are important for achieving the ultimate goal of good music. Thus, there are standards that guide the learner in practicing and mastering that subskill, even while they are doing other things wrong and good music is not being produced. In academics, such standards for subskills would apply to the outcome of the decisions listed above, such as choice of question or sources of evidence. Making such decisions in an expert way involves both having the relevant knowledge and having the reasoning skills to guide when and how that knowledge is used. In total, these standards for subskills, encompassing appropriate knowledge and its use to make decisions, largely define expertise in a discipline. With sufficient practice, some of these decisions become automatic, carried out with little conscious thought, thereby increasing the speed of the process.

The role of research in establishing expertise is illustrated by the field of medicine. In the 1400s, the definition of what it meant to be a good doctor was quite arbitrary and varied according to individual idiosyncrasies. Anyone and everyone could believe, and announce to the world, that they were a good doctor, even though different doctors employed a wide variety of practices. A similar situation exists today with regard to education; almost everyone who has been to school, let alone taught a class, believes that they are an expert, in that their opinion has equal or greater weight as that of anyone else.

Over the subsequent centuries, medical research led to the establishment of knowledge, principles, and methods that produced consistently better results. A practitioner who knew and applied these produced better outcomes (healthier, more long-lived patients) than those who did not, making

it possible to set objective standards for who was a competent doctor. This included standards about the components of expert practice such as washing hands between patients, knowing which diagnostic tests to use, and prescribing the most effective treatments. The transformation of medicine illustrates how fields change as a research base is established, leading to the recognition of expertise in the field. This establishment of research-based medical expertise led to changes in the training and conduct of medicine, with resulting improvements in both outcomes and the rate of further progress. The transition of alchemy into the modern discipline of chemistry is another example illustrating how an academic discipline with expertise develops following the creation of an adequate research base.

Teaching has traditionally not been an area for which well-defined expertise exists; it is more often characterized as an "art" wherein each individual is encouraged to choose their preferred style. While there has been a generally accepted goal – learning – what that means and how it can be measured has been ill-defined and variable. It is striking to read the many recent OECD (Organisation for Economic Co-operation and Development) reports on improving the quality of university teaching and see that none of them actually define teaching quality or how it could be measured. "Good" teachers are often described in terms of personal characteristics like "enthusiasm," "concern with students," and "interest in their subject." Judgments of teaching quality have traditionally depended largely on individual preferences, much like the judgment as to whether a painting is attractive or not, or whether a person is likeable. At the level of the institution or academic department, efforts to "improve teaching" often focus on the curriculum: what topics are covered in what order. Research on learning, however, implies that such curricular choices play at best a secondary role in determining meaningful student learning outcomes, particularly learning to think more like an expert in the discipline. The lack of agreed-upon standards for teaching quality allows everyone to consider themselves to be a good teacher by some standard, and most do.

Research during the past few decades has changed this situation for university teaching, although this change has yet to be widely recognized. These advances in research now make it possible to define expertise in university-level teaching and, correspondingly, define teaching quality in an objective expertise-based manner. The research comes from a combination of studies in cognitive psychology and the science of learning, studies in university science and engineering courses, and, most recently, from brain research. This includes hundreds of laboratory and classroom studies involving controlled comparisons of different teaching methods, primarily, but not exclusively, measuring student learning.

Much of the classroom research is the result of the relatively new field of "discipline-based education research" (DBER), which has developed over the past few decades.[3] This research focuses primarily on undergraduate learning of the science, technology, engineering, and mathematics (STEM) disciplines at research universities, and is carried out by faculty in the respective disciplines (physics, biology, computer science, so on).[4] This is distinct from the educational research that is carried out in schools of education, which is largely confined to the K–12 level.

The standards of DBER have rapidly evolved, and different disciplines are still at different stages of progress in this evolution. Not long ago, such university education "research" consisted of instructors trying some change in their teaching of a course and measuring the impact in some idiosyncratic way, primarily how much the students liked it. Now, quality DBER, which is what I am discussing here, is similar to medical research. It requires controlled comparisons of different ways to teach particular material, and the impacts are measured using validated, often published, and widely used tests that probe learning. Research protocols are similar to those for other human-subjects research and have the same institutional review.

DBER has led to new types of assessments of learning, new teaching methods, and comparisons of learning achieved with different methods of instruction. The research has explored the importance of many different factors for student learning, course completion, and, occasionally, student retention in a major. The teaching methods that have been found to be the most effective are well aligned with cognitive psychology research on learning, sometimes by intention and other times not.[5] This alignment is particularly evident in the research on teaching expert thinking, which has illustrated the need for explicit practice of the mode of thinking to be learned along with guiding feedback.

The assessments of learning in DBER that have been the most sensitive and impactful are "concept inventories." Such inventories are carefully developed to probe the extent to which students can apply relevant disciplinary concepts like an expert in the field to novel situations appropriate to the course content. Their primary use is to measure the effectiveness of the teaching in the class as a whole, rather than the learning of the individual students per se. Such inventories now exist for material covered in a number of standard introductory science and math courses and a few upper-level science courses. These provide researchers with good instructor-independent measures of learning that can be widely used, and hence allow widespread, carefully controlled comparisons of different teaching methods. These assessments are based on the unique disciplinary frameworks for making decisions that experts use, rather

than based on remembering pieces of knowledge or a memorized procedure. As such, learning to do well on these assessments of "expert thinking" is more sensitive to instructional practices than typical exam questions and less sensitive to "teaching to the test." These kinds of assessments have become a uniquely valuable tool for research on the relative effectiveness of different types of university teaching, but for practical reasons, they only measure a subset of the relevant expert thinking. There are other aspects that must be measured in different ways, including things like deciding on choices of possible solutions or designs, recognizing the range of real-world situations in which the discipline can be useful to understand and predict important phenomena, and the learner deciding they can master and enjoy working in the discipline.

Researchers also look at more conventional outcomes, such as failure rates and course and exam grades, but those are more sensitive to the characteristics of the incoming students and the idiosyncrasies of individual instructors, and thus are less reliable measures. Nevertheless, they still have reasonable validity if there are consistent standards and the instructor is careful in the exam construction, because of the degree of standardization of the undergraduate STEM curriculum, textbooks, and instructional goals across universities. Unfortunately, this is not true for many STEM exams that, often unintentionally, primarily test the student's memory of basic terminology, facts, and procedures.

DBER in university STEM courses is a relatively young field and is not widely known. It has primarily been carried out in the United States and funded by the National Science Foundation. It tends to be published in specialized journals (*Physical Review Physics Education Research*, *CBE – Life Sciences Education*, *Chemistry Education*, and *Journal of Engineering Education*, among others), with an occasional article published in *Science* or *Proceedings of the National Academy of Sciences*. There is limited awareness of DBER within the broader university faculty and administration, with the level of knowledge varying significantly by discipline. With a few exceptions, DBER is also little-known outside of North America. Some recent reports and reviews have attempted to synthesize and disseminate the findings of DBER and its implications for improving university teaching.[6]

DBER has established that there are particular principles and practices that consistently achieve better student outcomes than the traditional didactic lecture and high-stakes exam. This has typically been shown through experiments involving controlled comparisons. These effects are sufficiently large that, when one takes incoming student preparation into account by measuring learning gains rather than just outputs, the choice of teaching practices

results in larger differences than any other identified variables associated with the teacher (for instance, rated quality as a lecturer) or the students. The results have been replicated within and across instructors, institutions, courses, and disciplines.[7]

Such results have been shown in all the disciplines in which extensive classroom studies have been carried out, including all science and engineering disciplines at the university level and, to a lesser extent, mathematics. There have been some studies in other types of higher education institutions and a few recent, small studies in the social sciences.

It would be worthwhile to carry out similar controlled comparisons of learning in a broader range of disciplines such as history and classics. There are theoretical reasons to think that the same teaching methods would likely also work well in such fields, if properly adapted. The methods that have been consistently effective reflect fundamental mechanisms for learning from cognitive psychology (see Figure 1), particularly for learning to think like an expert in the discipline, as mapped onto the particular course and student population.[8] The DBER that has produced the biggest gains in learning has involved looking at the decisions that students make in solving problems after receiving traditional instruction and how they differ from those of scientists, and then designing educational activities that involve the students explicitly practicing making such decisions with feedback. Sam Wineburg has identified some key elements of historian expertise, including how historians determine the credibility of historical artifacts and what conclusions they decide they can draw from them, and how their thinking in this regard differs from college students who have taken a history course. It seems like these aspects of historian thinking could be directly incorporated into the corresponding research-based methods developed in STEM, likely with corresponding improvements in learning.

In this discussion, I have been careful to distinguish university teaching from teaching at the K–12 level. In *The Cambridge Handbook of Expertise and Expert Performance*, psychologist James Stigler and education scholar Kevin Miller present an excellent discussion of the challenges faced in establishing and defining K–12 teaching expertise in the United States.[9] As they have discussed, there are a number of confounding variables outside the control of the K–12 teacher, most notably the local context, that make K–12 teaching harder to characterize and harder to study. It is useful to contrast the K–12 context they describe with teaching in research universities where most DBER has been carried out. Variables such as classroom behavior, the subject matter mastery of the teacher, the scheduling of teaching and assessment activities, and the extent of variability in the student backgrounds are all major

Figure 1
Principles and Practices of Effective Teaching
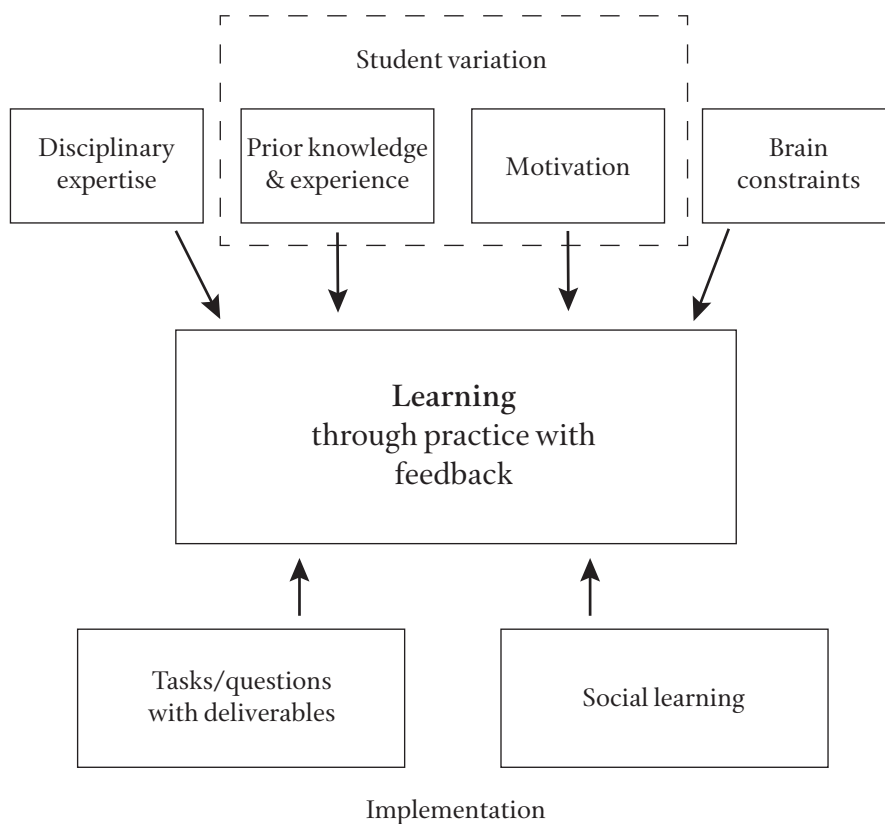
Note : This figure represents the full span of research on the principles and practices involved in learning to make good decisions in a specific disciplinary context. At the center are the essential components of learning. This represents the intense practice of the specific elements of thinking to be learned, ideally the decision-making skills that experts in the subject use in relevant situations, combined with feedback that guides improvement in that thinking. The top row of boxes represents factors that enable and facilitate this learning process. Much of the apparent variation across the student population comes in through the motivation and prior knowledge boxes, both of which depend heavily on the learners' prior experiences. The two boxes in the bottom row represent consistent elements in the implementation of highly effective teaching practices.

issues in K–12, but these are much smaller factors at the university level (even though nearly all university teachers complain about the level and uniformity of the preparation of their students). The U.S. K–12 context is also highly variable across schools, districts, and states, and these differences play a large role in the educational practices and assessment. In contrast, the context of university teaching is far less variable: relative to K–12, there is a high degree of standardization of the curriculum, the textbooks, the student populations and behavior, the instructional settings, the subject mastery of the instructors, and the desired learning outcomes. This makes the classroom research at the university level far simpler and cleaner, and it provides more definitive results than research in K–12 teaching. In the future, greater K–12 standardization through vehicles such as the Common Core State Standards Initiative and Advanced Placement courses might provide more K–12 uniformity. Stigler and Miller do propose three "teaching opportunities" that they believe would be the characteristics of expert teachers, if sufficiently clean research results could be obtained; these overlap with what I present below.

When expertise is first being established in a field, the distinctions as to different levels of competence are relatively crude. One can become an "expert," a top performer, merely by recognizing basic decisions that need to be made and, in those decisions, accounting for the basic factors that have been shown to be most relevant. As university teaching is a new area of expertise, one can achieve relatively high levels of mastery merely by using the basic principles and practices that have demonstrated improved learning. The description of expertise here is limited to this relatively coarse level. As any discipline matures, more complexity and nuance are seen to result in higher quality decisions, and thus more subtle factors become recognized as elements of expertise. This will eventually happen in teaching.

Before I can talk about what constitutes expert teaching, I need to define the intended learning goals that such expert teaching will reliably achieve. Often, the stated goals (or "objectives") of courses are expressed in terms of "understanding" or "appreciating" various topics. From extensive discussions with faculty members as to what they mean by such vague statements, I claim that the goals of the great majority of university STEM courses can be summarized as: teaching students to think about and use the subject like a practitioner in the discipline, consistent with the student's background and level. In practice, this means making relevant decisions and interpretations using the reasoning and knowledge that define expertise in the discipline. Of course, the level of sophistication with which the students might learn to do

that and the complexity and range of the contexts in which they are capable of making such decisions will vary widely according to the course. For the dedicated fourth-year chemistry major, that decision might be how best to synthesize a molecule in an industrial setting, while for a major from another discipline taking their one required chemistry course, it might be deciding not to pour hydrochloric acid down the drain or deciding not to invest in a company that claims it has a process for turning seawater into gold. But "thinking like a chemist" is needed for all these decisions. Thus, I am taking the basic goal of most university courses as having students learn to think more like an expert in their respective discipline.[10]

The most basic principle that every teacher should know about teaching this sort of thinking is that the brain learns the thinking it practices, but little else. To have students learn to recognize relevant features and make relevant decisions more like an expert in the field, they must practice doing exactly this. The longer and more intense the practice, the greater the learning. There is a biological origin to this requirement, as such intense mental practice modifies and strengthens particular neuron connections, and the new thinking capabilities of the learner reside in this "rewired" set of neurons. There is much research on how the brain changes the way it organizes and accesses relevant information as it learns, and on the connection between the functional and structural changes that occur in the brain during extended learning of expertise.[11]

The basic principle that people learn from practice with appropriate feedback is placed at the center of Figure 1. To my knowledge, practice and feedback are part of all research-based instruction that shows significantly better learning outcomes than the traditional lecture. These are also the two most basic elements of "deliberate practice," which has independently been found to be essential for the acquisition of expertise.[12] The first element in this context means having the learners actively and intently practicing the thinking to be learned. One particularly important and often overlooked feature in teaching is that thinking like an expert is primarily about making particular decisions. So, the learning task must involve the learners *actually making relevant decisions*. Too often, instruction only involves the teacher modeling a solution process by telling students the decisions that the expert has made. The differences in learning between a student being told the desired outcome of a decision versus having the student make the decision, even if incorrect, and then reflect upon the outcome of their decisions while supported by instructor guidance are profound.[13] These differences are easy to appreciate if you think about learning to find your way through a strange city. If you go between two locations by simply following the directions for each turn provided by a

person or mapping program, you will be incapable of telling another person how to do it or reproduce it on your own. If you had to form a mental map of the city and explicitly decide on the turns, you will have learned far more. In this case, you practiced making decisions and strengthened neuron connections in the necessary way to learn, and it does not matter if some of those turns were wrong, and you had to revise your route: you still learned better the correct decisions. This also carries over to your learning better how to transfer your knowledge to a new context, such as going between new locations or dealing with road closures. The same principles apply to learning problem-solving decisions in a discipline.

Effective teaching is about first designing learning activities that have the student carrying out tasks that require them to make decisions using the specific reasoning processes, including the associated requisite knowledge, to be learned. The second element is good feedback, which means feedback that is timely, specific, nonthreatening, and actionable.[14] To be able to provide such feedback requires that the instructor monitor the learner's thinking in some way, and then use that information to provide feedback to guide the improvement in that learner's thinking (often labeled as "formative assessment"). Under this broad general principle of practice with feedback, there is a detailed set of factors that have been shown to play an important role in supporting this learning process.[15] These are illustrated in Figure 1. Each of the boxes in the upper row represents a well-studied principle involving established mechanisms of learning. Good instructional design incorporates these principles into the design of the practice tasks and the types of feedback provided. The two boxes in the bottom row represent research on how best to implement these in instructional settings. If and how the instruction incorporates the best practices represented in all of these boxes is a measure of teaching expertise.

Disciplinary expertise. Embedding expertise in the subject into the instructional activities is a fundamental requirement. This expertise includes recognizing what decisions need to be made in relevant contexts, along with the tools, reasoning, and knowledge of the discipline to make good decisions.[16] In this regard, good instructional tasks should directly reflect the standards that define expertise in the discipline discussed above, as mapped onto the context of the specific course being taught. This involves many different decisions, but an example of the most general and basic is, when confronted with an authentic problem/question and context, deciding what the key features and information are, and what information is irrelevant to solving the problem. Artificially constrained "textbook type" problems remove practice in this critical decision skill.

Motivation. Serious learning is inherently hard work that involves prolonged strenuous mental effort. The motivation to engage in that effort plays a large part in the learning outcomes. Motivation is obviously enhanced by making a subject interesting and relevant to the learner, which often means framing the material in terms of a meaningful (to the learner!) context and problem that can be solved.

A less obvious element in motivation is having a "growth mindset," the learners' belief that they can master the subject and a sense of how to attain that mastery, a belief that can be powerfully affected by both prior experiences and teacher behaviors.[17] Too often teachers fail to recognize the impact of the various messages they convey through what they say or how they grade. For example, an exam that measures all of what students should have learned and only that, compared with the more typical exam that focuses on the most challenging material that will provide the best differentiation between students, send very different signals to students. The first shows them all of what they are learning and is motivating, while the second leaves many students, for example those who only get a 50 percent score after intensive study, with a demotivating sense of failure and frustration, even if that is the class average.

Prior knowledge and experience. To be effective, instructional activities must match with and build upon what the student already knows and believes about the subject and how to learn it. Research has shown that it is important for effective instruction to recognize and address even very specific aspects of the learners' thinking about particular topics, such as whether a student believes that heavier objects fall more rapidly than lighter objects when teaching introductory physics.

Both prior knowledge and what does and does not motivate students are highly dependent on their prior experiences. Hence, these are the areas where most of the observed variations in the student populations are apparent. The expert teacher will recognize it is inadequate to ask students what they know or come to conclusions based on the syllabi of prior courses the students have taken. Instead they will measure what the students know and can do, initially and ongoing through the course. They will then optimize learning by adjusting their instruction to match best the characteristics of their student population.

Brain constraints. The next box, constraints of the brain, refers to 1) the limited capacity of the short-term working memory of the brain (five to seven new items, far less than introduced in a typical class session) and its well-studied impacts on learning; and 2) the processes that hinder and help long-term retention of information. The limited capacity of working memory means that anything peripheral to the desired learning that attracts the learner's

attention will reduce the desired learning. This includes new jargon, attractive images, or even amusing stories or jokes. The biggest problem with long-term retention is not in remembering material in the first place; rather, it is correctly retrieving it later after additional material has been learned. That new material interferes with the retrieval process. To avoid this interference, as new material is learned, it needs to be intermingled with the recall and application of old material. This is not the usual practice in STEM courses wherein novice teachers cover the topics in a strict chronological order.

The two boxes at the bottom of Figure 1 represent key elements for the implementation of research-based teaching:

Tasks/questions with deliverables. To ensure that students are practicing the desired thinking, they need to be given tasks or questions that explicitly require that thinking. Explicit deliverables achieve engagement in the task and provide essential information to the teacher for giving effective feedback. For example, in a genetics class, students would consider the blind fish in Mexican caves. They would be asked to consider what they could decide about the number of genes containing the blindness mutation from the distribution of blindness in the offspring of true-breeding lines of fish bred from lines in two different caves. In a large class (two hundred to three hundred students), the instructor would have the students answer using a personal response system (PRS), followed by small-group discussion (that the instructor and TAs monitor) and a second vote. In a smaller class, students would have to write out their prediction with the reasoning, to be turned in for participation credit, possibly in addition to the PRS questions. In a physics class, they would be given a problem to solve for a particular physical situation, such as predicting how much electricity could be produced from a hydroelectric plant: the first step would be to write out which physics concepts are most relevant to solving the problem and why, to be turned in later and minimally graded; the instructor and TAs would circulate and read students' responses during class. In a large class, this could be followed with a PRS question testing them on their choices. In all of these cases, there should be follow-up homework questions, and it should be explicit that there will be quite similar questions on future exams.

Social learning. Interacting with peers during the learning process is a valuable and commonly used facilitator of learning.[18] It supports learning in multiple ways. Students get timely knowledge and feedback from their peers, they learn the standards of discourse and argument of the discipline, and they develop metacognitive skills through their critique of others' reasoning and hearing others question their own. Finally, there are unique cognitive processes that are triggered by social interactions that produce learning. Even anticipating that one will teach a peer about a topic has shown to improve learning

over just studying the topic. And, of course, such group activities provide opportunities for the students to learn collaborative skills. Important elements of teaching expertise are to know how to avoid the potential pitfalls of group work, how to set and monitor norms of behavior, and how to structure the group activities to achieve all of the potential benefits.

The set of factors and practices represented in Figure 1 largely determine learning outcomes at the university level for the disciplines and institution types in which they have been tested. There are many examples where very experienced faculty have changed their teaching practices to incorporate these principles and practices, usually moving from lecture to research-based instruction, and achieved substantial improvements in student learning outcomes. Research is ongoing as to how best to take these factors into account in the design and implementation of the learning process across the full range of disciplines, topics, and students. However, the relevance and benefits can be understood in terms of established general mechanisms of learning, and thus it is likely that they will apply across nearly all higher education settings and academic disciplines.[19]

If a teacher is applying these practices in a discipline in which they have not been studied, the respective disciplinary standards of expertise and associated decisions must provide the foundation of the educational practice tasks that learners carry out, as well as the feedback they receive. This emphasizes the need for every good university teacher to have a high level of disciplinary expertise.

In summary, the experimental study of how learning takes place and how best to facilitate it in university teaching has provided a rich body of evidence establishing the basis of expertise in teaching. Research consistently shows better student outcomes compared with lectures when students are fully engaged in challenging tasks that embody expert thinking and they receive guiding feedback : the principles represented in Figure 1. This success is the basis for my claim that expertise in university teaching exists. An expert teacher will be aware of these principles and use suitable research-tested practices to incorporate all of them into their instruction.

In one respect, it is somewhat surprising that the research results are so consistent.[20] As in every discipline, there are countless ways for a novice to do such complex tasks poorly, even if trying to follow best practices. These research-based teaching practices are regularly being adopted by faculty with little teaching expertise, usually, though certainly not always, to good effect. I believe that a likely reason for this consistency is that research-based teaching is, to a substantial extent, self-correcting. In nearly all forms, it provides opportunities for the instructor to know what the students are thinking and

struggling with – far better opportunities than instructors get when lecturing. When instructors are first adopting these methods in even modestly informed ways, they almost always comment on how much better they now understand student thinking and difficulties compared with when they were teaching by lecturing, and how this new understanding of student thinking is changing their teaching. These new insights allow them to recognize and correct weaknesses in their instruction, thereby improving learning.

Although university teaching expertise can now be defined, it is not widely known and practiced. Again, the situation with university teaching is like medicine in the mid-1800s. Although research had established a basis for science-based medical practice, many "doctors" were unaware of that science. Their practice was based primarily on tradition and individual superstitions with no accepted standards. That changed during the late 1800s and early 1900s. There is reason to hope for a similar transition in university teaching.

The establishment of expertise in teaching has implications for the training, evaluation, and cultural norms for how teaching is carried out. In every discipline, the relevant standards of expertise play a large part in the practice and training in the discipline. Once there are well-defined and generally accepted standards of expertise, these provide standards on which to base both evaluation and training. This includes standards for being certified as competent, either formally as in medical or legal licensure, or informally as in the process of review of scholarly work for publication or judging the qualifications of faculty job applicants. In the case of university teaching, a teacher now can, and should be, evaluated on their level of teaching expertise: how familiar they are with the principles and practices represented in Figure 1 and to what extent they use these in teaching. Training needs to provide them with this expertise.

Evaluation of teaching quality at the university level has long been problematic. Currently, the dominant method is student course/instructor evaluation surveys. There are obvious problems with such evaluations, as well as some particularly compelling recent studies showing substantial gender bias.[21] As I have written elsewhere, the basic requirements for any good evaluation system are:

- *Validity.* Results correlate with the achievement of the desired student outcomes and allow meaningful comparisons of quality across different instructors and departments.
- *Fairness.* Only depends on factors under the instructor's control.
- *Guides Improvement.* Provides clear guidance as to what should be done to improve.[22]

Student course evaluations fail badly at meeting any of these criteria. Most important for this discussion, they have been clearly shown to fail at both reflecting the extent of expert teaching practices being used and reflecting improvements in learning.

However, it is now possible to evaluate teaching based on standards of expertise. One example of this is the Teaching Practices Inventory (TPI) developed by Sarah Gilbert and me (see Appendix I).[23] It is a survey that can be completed quickly (about ten minutes per course) and reflects nearly all the decisions that an instructor makes in designing and teaching a course. It provides a detailed objective characterization of most of the instructional practices used in a course and, correspondingly, the extent of use of research-based effective practices. It is not perfect; it does not show the effectiveness with which these practices are being used. It is analogous to measuring if doctors are washing their hands between patients, but not how well they are washing. We and others have seen that this level of measurement is sufficient to easily distinguish between the different levels of teaching expertise present in a typical sample of university science faculty. The TPI shows a high degree of discrimination across a typical sample of university faculty, with the highest scoring faculty also having very high measures of student learning outcomes. TPI results allow meaningful comparisons to be made across faculty, departments, and institutions.

The use of such expertise-based evaluation of teaching would make it more like the evaluation of research, allowing institutions to include teaching both in their evaluation and incentive systems in a far more meaningful and intentional way than is currently possible. It would also make it straightforward to set clear criteria for the level of teaching competence expected for new faculty hires and for promotion and tenure decisions.

Effective training of teachers, similar to good training in any area of expertise, involves practicing the relevant thinking and actions in authentic contexts, along with feedback to guide improvement. As in academic disciplines, the most important part of training in teaching is to practice the relevant decision processes, recognizing what information is most important to guide those decisions and using it accordingly. This will require training that is both more extensive and more targeted than most existing university teacher training programs.

The list of elements that needs to be covered in training university teachers reflects all aspects of teaching a course and all the principles represented in Figure 1. This may seem overwhelming compared with what is now typical, but it is small compared with the training faculty received to become

experts in their disciplines. I have seen that faculty can reach a respectable level of teaching expertise in something in the range of fifty hours of training; less time than is required to complete most university courses.[24] That is sufficient to allow faculty members to switch from teaching by traditional lecture and exams to research-based methods and achieve good results. Of course, this small amount of time (fifty hours) required to be reasonably competent in teaching, compared with the thousands of hours required for high competence in a mature discipline, is a reflection of the immaturity of the field and the current level of expertise. As the level of teaching expertise increases, the standards of competence and corresponding expectations of training and quality will likely also increase.

I should emphasize that it does not require any additional time to teach using these new research-based methods instead of traditional teaching; it only requires time to learn how. But in my experience, nearly all faculty that successfully adopt these methods find that it makes teaching a far more enjoyable and rewarding activity. Consequently, many of them voluntarily choose to spend more time on teaching than they had previously.[25]

The typical university teacher training program is too unfocused, as it is usually designed to serve faculty from all disciplines at the same time. As with the specificity needed for training of any type of expertise, effective development of teaching expertise will require training programs that focus on the teaching of the particular discipline and student population that the faculty member will encounter. While the principles are general, it is a very large step from them to knowing how to apply them to teaching a specific discipline and level.

One training option is to have an individual "coach," an approach successfully used in many areas of expertise. Such a coach for university teaching would have expertise both in the relevant discipline and in teaching in that discipline, and would be well informed about the student population and the other important contextual constraints. The coach would individually review the trainee's instructional activity designs, observe their implementation in class, and provide feedback to guide improvement. A vital skill is also knowing the way things can fail, and help the trainee anticipate and avoid such failures. The use of such disciplinary teaching coaches has been shown to be an effective model in the Science Education Initiative (SEI; see Appendix II). The SEI provided funding to departments to hire disciplinary experts, typically new Ph.D.s, with a strong interest in teaching, who were then trained in the research on teaching and learning and implementation methods, and on how to work with faculty to support and coach them in transforming their teaching. "Master-apprentice" training involving a novice teacher team-teaching a

course with an experienced expert teacher faculty member captures most of the same elements and has also been shown to be effective.[26]

There is a fundamental change in the social culture of a discipline when it develops widespread recognition of expertise, a change that we can expect in university teaching in the coming years. The establishment of recognized expertise in a discipline enables increased collaboration/collective work and building upon prior work. When a field is recognized as an area of expertise, like physics, chemistry, or history, that means there is a commonly accepted set of standards and principles, along with accompanying common language, for discussion. This commonality makes it both possible and desirable to share ideas and methods and pursue collaborative projects, as well as have disciplinary conferences and journals. In contrast, teaching at the university level is now widely seen as an isolated activity, with faculty in a department almost never coming to view each other's classes and seldom discussing or collaborating on teaching activities or methods. This contrast in culture is directly related to differences in the level of recognized expertise.

It can be understood by considering the hypothetical situation of a physicist whose office is in a building otherwise occupied exclusively by ancient poetry scholars. There would be little value in the physicist going and talking with those faculty to discuss ideas about physics, or to find new ideas for experimental designs (and vice versa, if it were a poetry scholar exiled to the physics building). Assuming no Internet, the physicist would sit at a desk trying to invent everything in isolation. But that same physicist, if located in a building full of physicists, would be engaged in peer discussions about scientific ideas and methods, gaining new information and insights and making far more progress as a result. These physicists would be pursuing their own specific goals, but within a commonly accepted framework of principles, knowledge, and standards: the core of physics expertise that facilitates discussion and sharing for mutual benefit. This framework supports interaction and sharing of ideas while still allowing room for identifiable individual contribution, essential components of every academic discipline.

Teaching is currently seen as a matter of individual taste and style. Each time faculty members teach a new course, they usually design it largely from scratch, at best taking small elements from previous offerings of the course at their institution and nothing from other institutions. This perception of teaching as a solitary activity is encouraged by the institutional policies for how teaching is allocated and evaluated. Each individual course is typically assigned to an individual faculty member who then has full responsibility for all aspects of that course, with very little oversight or expectations as to what will be taught and how.

The recognition of expertise in university teaching will go hand-in-hand with it becoming a more collective enterprise within departments and institutions, much as is the case for scholarly work in the disciplines. I observed this in the UBC Science Education Initiative.[27] There were far more frequent and substantial discussions about teaching among the faculty in a department after a number of the faculty became moderately expert. This socialization of teaching will in turn make teaching more efficient and effective. In scholarly research, by building on past work, an individual can accomplish far more than if they had to invent everything on their own. As practices established through DBER have spread, there have been early examples of this happening for teaching in some disciplines. While many elements of expert teaching are the same across disciplines, it is likely that socialization of teaching will still be largely confined to the existing disciplinary boundaries. That is because of the large role that the disciplinary expertise plays, including student knowledge and beliefs about the discipline, in the design and implementation of educational activities.

T he lecture method that dominates university teaching has remained much the same for hundreds of years. The concept of education through an expert relaying information to a room full of novices predated the printing press, but to a large extent remains the norm today. The treatment of teaching as an individual art form has shaped its practice and evaluation. This is in striking contrast to the nature of the academic disciplines, which have changed and advanced enormously. These medieval methods of teaching are now confronting the challenges posed by the increased complexity of thinking that it is desirable for students to learn, and the greatly increased numbers and diversity of students that need a good university education. The acquisition of basic information is now of limited value, while complex reasoning and decision-making skills that can be broadly applied have high value in many aspects of modern society.

The establishment and recognition of teaching expertise has far-reaching implications. Much as happened in medicine as it moved from its medieval roots to modern, research-based methods, the expertise established by these research advances in teaching provide a standard for the quality of practice, hiring, evaluation, and training. The adoption of such standards will result in immediate and ongoing improvements in educational effectiveness. The establishment of such consistent standards also enables the conduct of teaching in a more collective way, using and building on previous work. This promises to improve both the effectiveness and efficiency of instruction. While higher education is facing many challenges, the rise of teaching expertise offers a

path to a dramatic improvement in how it pursues its educational mission. This would be a historic change, and while such changes never come easily, it would provide broad societal benefits. As well as enhancing the educational value provided by universities, it would more clearly demonstrate their unique educational contribution.

Many examples of teaching activities that incorporate these principles in various disciplines are given in Appendix III, accessible at http://www.amacad.org/daedalus/teachingexpertise.

ABOUT THE AUTHOR

Carl Edwin Wieman, a Fellow of the American Academy since 1998, is Professor of Physics and Professor of Education at Stanford University. He is the author of *Improving How Universities Teach Science: Lessons from the Science Education Initiative* (2017) and has recently published in such journals as *Journal of Educational Psychology*, *Physical Review Physics Education Research,* and *Biochemistry and Molecular Biology Education*.

ENDNOTES

1 Anders Ericsson and Robert Poole, *Peak: Secrets from the New Science of Expertise* (New York: Eamon Dolan/Houghton Mifflin Harcourt, 2016); and K. Anders Ericsson, Ralf Th. Krampe, and Clemens Tesch-Römer, "The Role of Deliberate Practice in the Acquisition of Expert Performance," *Psychological Review* 100 (3) (1993): 363–406.

2 Ibid.

3 Susan R. Singer, Natalie R. Nielsen, and Heidi A. Schweingruber, eds., *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (Washington, D.C.: National Academies Press, 2012).

4 In what follows, I use the label "university" to refer to research universities: those large institutions with substantial numbers of undergraduate and graduate degrees, conventional academic departments, substantial programs of scholarly work, and so on.

5 Daniel L. Schwartz, Jessica M. Tsang, and Kristen P. Blair, *The ABCs of How We Learn* (New York: W. W. Norton & Company, 2016).

6 Singer et al., *Discipline-Based Education Research*; President's Council of Advisors on Science and Technology, *Engage to Excel: Producing One Million Additional College Graduates With Degrees in Science, Technology, Engineering, and Mathematics* (Washington, D.C.: Executive Office of the President, 2012); and Scott Freeman, Sarah L. Eddy, Miles McDonough, et al., "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proceedings of the National Academy of Sciences* 111 (23) (2014): 8410–8415.

7 Freeman et al., "Active Learning Increases Student Performance in Science, Engineering, and Mathematics."

8 See also Schwartz et al., *The ABCs of How We Learn*.

9 K. Anders Ericsson, Neil Charness, Robert R. Hoffman, and Paul J. Feltovich, eds., *The Cambridge Handbook of Expertise and Expert Performance*, 2nd ed. (Cambridge: Cambridge University Press, 2018).

10 A notable exception is the typical service course for nonmath majors taught by mathematics faculty.

11 Ericsson et al., *The Cambridge Handbook of Expertise and Expert Performance*.

12 Ericsson and Poole, *Peak*; and Ericsson et al., "The Role of Deliberate Practice in the Acquisition of Expert Performance."

13 Ibid.; and Singer et al., *Discipline-Based Education Research*.

14 Schwartz et al., *The ABCs of How We Learn*.

15 Here we are considering the brain of the typical university student, neglecting any "clinical" anomalies present in special cases.

16 There are many calls for university students to learn "critical thinking." As this is usually defined, it is equivalent to making better decisions in realistic situations. But a closer examination of what this means to any particular advocate of teaching critical thinking is usually that the students should learn to use the skills and knowledge of their discipline in making decisions of the sort valued by their discipline, with the assumption that this represents a generic skill that all students should have. There is an extensive body of research indicating that there is no such generic skill: any authentic decisions will necessarily involve discipline-specific knowledge and reasoning, and hence any measure of "critical thinking," including those currently used with claims they are generic, such as the Collegiate Learning Assessment, are in fact not generic. If the context and nature of the decisions involved changed, so would a student's performance.

17 Schwartz et al., *The ABCs of How We Learn*.

18 Ibid.

19 This is different from the all-too-common example of a novice teacher applying some technique without understanding the principles on which it is based or the benefits it might provide, and thereby achieving little apparent benefit. An example (a real one) is introducing the use of clicker questions and peer discussion in a political science course with little understanding of suitable questions or goals of

discussion, and then judging the effectiveness of this teaching method according to the changes (or not) observed in the quality of the writing of the students' term papers.

20 Freeman et al., "Active Learning Increases Student Performance in Science, Engineering, and Mathematics."

21 Carl Wieman, "A Better Way to Evaluate Undergraduate Teaching," *Change: The Magazine of Higher Learning* 47 (1) (2015): 6–15; Lillian MacNell, Adam Driscoll, and Andrea N. Hunt, "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching," *Innovative Higher Education* 40 (4) (2015): 291–303; and Amy L. Graves, Estuko Hoshino-Browne, and Kristine P. H. Lui, "Swimming against the Tide: Gender Bias in the Physics Classroom," *Journal of Women and Minorities in Science and Engineering* 23 (1) (2017).

22 Wieman, "A Better Way to Evaluate Undergraduate Teaching."

23 Originally developed for characterizing teaching in sciences, with some very small wording changes, it is now being used on at least a limited basis for all academic disciplines. Carl Wieman and Sarah Gilbert, "The Teaching Practices Inventory: A New Tool for Characterizing College and University Teaching in Mathematics and Science," *CBE – Life Sciences Education* 13 (3) (2014): 552–569; and Carl Wieman Science Education Initiative at the University of British Columbia, "CWSEI Teaching Practices Inventory," October 3, 2014, http://www.cwsei.ubc.ca/Files/CWSEI_TeachingPracticesInventory_Oct2014.pdf.

24 Carl Wieman, *Improving How Universities Teach Science: Lessons from the Science Education Initiative* (Cambridge, Mass.: Harvard University Press, 2017).

25 Ibid.

26 Ibid.

27 Ibid.

Appendix I
CWSEI Teaching Practices Inventory:
For Use in the Natural and Social Sciences

To create the inventory we devised a list of the various types of teaching practices that are commonly mentioned in the literature. We recognize that these practices are not applicable to every course, and any particular course would likely use only a subset of these practices.

We have added places that you can make additions and comments and we welcome your feedback.

It should take only about 10 minutes to fill out this inventory.

Please fill out the inventory for the current or just completed Term, lecture sections only.

Course number:
Section #(s) or Instructor name:
Total number of students in your class or section (approximate):

I. Course information provided to students via hard copy or course webpage

Check all that occurred in your course:

☐ List of topics to be covered
☐ List of topic-specific competencies (skills, expertise, . . .) students should achieve (what students should be able to *do*)
☐ List of competencies that are not topic related (critical thinking, problem solving, …)
☐ Affective goals – changing students' attitudes and beliefs (interest, motivation, relevance, beliefs about their competencies, how to master the material)
☐ Other (please specify)

If you selected other, please specify:

II. Supporting materials provided to students

Check all that occurred in your course:

- ☐ Student wikis or discussion boards with little or no contribution from you
- ☐ Student wikis or discussion boards with significant contribution from you or TA
- ☐ Solutions to homework assignments
- ☐ Worked examples (text, pencast, or other format)
- ☐ Practice or previous year's exams
- ☐ Animations, video clips, or simulations related to course material
- ☐ Lecture notes or course PowerPoint presentations (partial/skeletal or complete)
- ☐ Other instructor selected notes or supporting materials, pencasts, etc.
- ☐ Articles from related academic literature
- ☐ Examples of exemplary papers or projects
- ☐ Grading rubrics for papers or large projects
- ☐ Other (please specify)

If you selected other, please specify:

III. In-class features and activities

A. Various

Give approximate average number:

Average number of times per class: pause to ask for questions:
Average number of times per class: have small group discussions or problem solving:
Average number of times per class: show demonstrations, simulations, or video clips:
Average number of times per class: show demonstrations, simulations, or video where students first record predictions (write down, etc.) and then afterwards explicitly compare observations with predictions:
Average number of discussions *per term* on why material useful and/or interesting from students' perspective:
Comments on above (if any):

Check all that occurred in your course:

- ☐ Students asked to read/view material on upcoming class session
- ☐ Students read/view material on upcoming class session *and* complete assignments or quizzes on it shortly before class or at beginning of class
- ☐ Reflective activity at end of class, e.g. "one-minute paper" or similar (students briefly answering questions, reflecting on lecture and/or their learning, etc.)
- ☐ Student presentations (verbal or poster)

Fraction of typical class period you spend lecturing/talking to whole class (presenting content, deriving mathematical results, presenting a problem solution, . . .):

- ☐ 0 – 20%
- ☐ 20 – 40%
- ☐ 40 – 60%
- ☐ 60 – 80%
- ☐ 80 – 100%

Considering the time spent on the major topics, approximately what fraction was spent on the *process* by which the theory/model/concept was developed, including the experimental methods and results that support specific theories?

- ☐ 0 – 10%
- ☐ 11 – 25%
- ☐ more than 25%

B. Individual Student Responses (ISR)

If a student response method is used to collect responses from all students *in real time in class*, what method is used?

Check all that occurred in your course:

- ☐ Raising hands
- ☐ Raising colored cards
- ☐ Electronic (e.g. "clickers") with student identifier

☐ Electronic anonymous
☐ Written student responses that are collected and reviewed in real time
☐ Other (please specify)

If you selected other, please specify:

Number of ISR questions posed followed by student-student discussion per class:
Number of times ISR used as quiz (counts for marks and no student discussion) per class:

IV. Assignments

Check all that occurred in your course:

☐ Homework/problem sets assigned or suggested but did not contribute to course grade
☐ Homework/problem sets assigned and contributed to course grade at intervals of 2 weeks or less
☐ Paper or project (an assignment taking longer than two weeks and involving some degree of student control in choice of topic or design)
☐ Encouragement and facilitation for students to work collaboratively on their assignments
☐ Explicit group assignments
☐ Other (please specify)

If you selected other, please specify:

V. Feedback and testing; including grading policies

A. Feedback from students to instructor during the term

Check all that occurred in your course:

☐ Midterm course evaluation
☐ Repeated online or paper feedback or via some other collection means such as clickers
☐ Other (please specify)

If you selected other, please specify:

B. Feedback to students

(check all that occurred in your course)

☐ Assignments with feedback from instructor, teaching assistant, or peer before grading or with opportunity to redo work to improve grade
☐ Students see graded assignments
☐ Students see assignment answer key and/or grading rubric
☐ Students see graded midterm exam(s)/quizzes
☐ Students see midterm exam(s)/quizzes answer key(s)
☐ Students explicitly encouraged to meet individually with you
☐ Other (please specify)

If you selected other, please specify:

C. Testing and grading

Number of tests during term that reflect course expectations (e.g. midterm exams, but not final exams):

Approximate fraction of test scores from questions that required students to explain reasoning:

Approximate breakdown of course grade (% in each of the following categories):

Final exam:
Midterm/other exam(s):
Homework assignments:
Paper(s) or project(s):
In-class activities:
In-class quizzes:
Online quizzes:
Participation:
Lab component:
Other:

If you selected other, please specify:

VI. Other

Check all that occurred in your course:

☐ Assessment given at beginning of course to assess background knowledge
☐ Use of instructor-independent pre-post test (e.g. as concept inventory) to measure learning
☐ Use of a consistent measure of learning that is repeated in multiple offerings of the course to compare learning
☐ Use of pre-post survey of student interest and/or perceptions about the subject
☐ Opportunities for students' self-evaluation of learning
☐ Students provided with opportunities to have some control over their learning, such as choice of topics for course, paper, or project, choice of assessment methods, etc.
☐ New teaching methods or materials were tried along with measurements to determine their impact on student learning

VII. Training and guidance of Teaching Assistants

Check all that occurred in your course:

☐ No TAs for course
☐ TAs must satisfy English language skills criteria
☐ TAs receive 1/2 day or more of training in teaching
☐ There are Instructor-TA meetings every two weeks or more frequently where student learning and difficulties and the teaching of upcoming material are discussed
☐ TAs are undergraduates
☐ TAs are graduate students
☐ Other (please specify)

If you selected other, please specify:

VIII. Collaboration or sharing in teaching

☐ Used or adapted materials provided by colleague(s)
☐ Used "Departmental" course materials that all instructors of this course are expected to use

Discussed how to teach the course with colleague(s):

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

Read literature about teaching and learning relevant to this course:

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

Sat in on colleague's class (any class) to get/share ideas for teaching:

☐ 1 Never
☐ 2
☐ 3
☐ 4
☐ 5 Very Frequently

IX. General

Open-ended comments:

Please write any other comments here. If this inventory has not captured an important aspect of your teaching of this course, or you feel you need to explain any of your above answers, please describe it here:

Approximately how long did it take you to fill out this inventory?

We thank you for taking the time to fill out this inventory.

Appendix II
Background of the CWSEI

The Carl Wieman Science Education Initiative (CWSEI) at the University of British Columbia and its smaller partner at the University of Colorado Boulder were large-scale finite-duration experiments (approximately $10 million and $5 million, respectively) in institutional change. They showed that it is possible for large research-intensive university science departments to make major changes in their teaching, and they revealed the processes that help and hinder such change. An extensive discussion of this experiment is given in Carl Wieman, *Improving How Universities Teach Science* (2017).

At the University of British Columbia, the Initiative changed the teaching of about 170 science faculty members and courses, with the fraction of transformed faculty and credit hours reaching 90 percent in some departments. These faculty are finding teaching to be more rewarding, and their students are far more engaged and learning more. Teaching became much more of a collaborative intellectual activity in these departments, with faculty sharing methods and results and seeking out ideas from others. The transformed teaching is characterized by: detailed learning goals for the course that express what students should learn to do in operational terms; in-class active-learning activities such as peer instruction, think-pair-share, and worksheets that have students practicing expert thinking by answering questions in small groups monitored by the instructor and TAs and interspersed with regular instructor feedback and guidance; different forms of assessment aligned with course goals, such as graded homework, more-frequent lower-stakes exams, and two-stage exams that students complete individually and then as a group; reflective exercises such as two-minute papers at the end of a class; and brief preclass preparations such as targeted readings.

Such results were not easy nor shared across all departments. The three most important elements were: supporting department-level change, incentives, and maximizing faculty buy-in.

Supporting department-level change. At universities, each department decides what and how to teach, and so the department is the unit of educational change. The CWSEI used a competitive grant program by which departments competed for up to $1.8 million over six years to transform teaching. Potential grants of this scale produced discussions of undergraduate teaching needs and opportunities that had never happened before. The success of the funded departments was strongly influenced by disciplinary culture and the quality of the departmental leadership and administration, which varied greatly. New

structures and people, such as a teaching initiatives committee with responsibility and resources, were required, as the traditional departmental structures, when left unchanged, were never effective at supporting innovation.

A key component in every successful department were science education specialists (SESs) with deep expertise in the respective discipline combined with expertise in teaching and learning in the discipline. The SESs were hired by the department and worked collaboratively with a sequence of faculty to transform courses and, in the process, the teaching of the faculty. The SESs act as nonthreatening coaches, providing expert guidance and support to faculty members as they try new things in their courses. With SES guidance, a faculty member was likely to implement research-based teaching methods in an effective manner from the beginning, and hence have a positive teaching experience. The SESs also provide expert and time-saving assistance in developing new course materials and assessments. It was usually easy to find good SES candidates with the necessary disciplinary knowledge and interest in education, typically new Ph.D.s, but it was necessary to set up an extensive training program for them in the relevant research and best research-based teaching methods.

Incentives. Incentives need to be provided for both the departments and the individual faculty members to take the time to learn new teaching methods. The formal incentive system is a powerful disincentive to improving teaching. At all universities, the evaluation system does not recognize that research has shown there are fundamental differences in the effectiveness of different teaching methods, and hence the system penalizes any time away from research to learn better methods. The CWSEI showed that it does not cost more money or time to teach using these more effective methods, but it does cost money to bring about change. One incentive is having the dean and department chair clearly convey that better teaching is an important institutional goal, but most other incentives involve money in one form or another, largely to minimize and compensate for the time required to learn.

Maximizing faculty buy-in. Instead of starting with specific courses to transform, it was more effective to start with any willing faculty members and accommodate them according to what courses and process of change work best for them. Some faculty were happy to carry out a total course transformation all at once, but for many others, an incremental approach worked better, from both psychological and logistical perspectives. Even modest changes usually showed positive results. Almost immediately the use of active learning methods gave faculty a better understanding of their students' thinking, and hence how to make their teaching more effective. There are many fears associated with making change. The most effective ways to address these fears were not

by providing data, but rather by having faculty talk to their colleagues who had transformed their teaching and watch the teaching of a good transformed course in their department. For many faculty members, it can take one or two years of hearing about these ideas and discussing them with their colleagues before they decide to change, with no obvious large differences between young and older faculty members.

The CWSEI has published a large body of resources on its website. These include peer-reviewed research papers on various aspects of teaching and learning and extensive guidance for instructors. The following links also feature a variety of guides on details of design and implementation of research-based instruction and videos showing demonstrations.

- For a collection of documents offering detailed advice for departments and faculty members on how to redesign courses, see "Course Transformation Resources," http://www.cwsei.ubc.ca/resources/course_transformation.htm.
- For a collection of short guides for instructors (on assessment, clicker use, student engagement, and so on) that illustrates in concrete terms the pedagogical philosophy (active engagement of students) underlying these initiatives, see "Instructor Guide," http://www.cwsei.ubc.ca/resources/instructor_guidance.htm. The advice is highly practical.
- For a collection of videos that show, among other things, what active learning looks like, see "Science Education Initiative (SEI) Videos," http://www.cwsei.ubc.ca/resources/SEI_video.html.
- For an annotated bibliography of papers on the research behind many aspects of active learning, see "Recommended Papers," http://www.cwsei.ubc.ca/resources/papers.htm.