A Brief Guide to Classroom Inquiry

Version 1.0

1. Evaluating Understanding

2. Developing a Research Focus to Investigate Student Learning in the Classroom: Think SMALL

3. A Model of Classroom Inquiry

4. Scholarship of Teaching for Understanding

5. Making Students' Thinking Visible During the Learning Process

6. Using "Think Alouds" to Evaluate Deep Understanding by Lendol Calder and Sarah Eva Carlson

7. The Scholarship of Teaching and Learning: A Beginner's View by Curtis Bennett

Evaluating Understanding

How do you know if students understand something? It is relatively easy to determine whether students remember something—you ask them a question and then compare their answers to the original information they were supposed to learn. But, evaluating someone's understanding is more complex.

On most matters, a person's understanding is somewhere between two extremes with abject ignorance on one end and well developed understanding on the other.

No Clue---------1----------l----------l----------l----------l----------Developed Understanding

Toward the "no clue" end of the continuum we would say a person's understanding is incomplete, underdeveloped, fragmented, naïve, inchoate, half-baked, incipient, superficial, or trivial. Toward the "developed understanding" end of the continuum we would say a person's understanding is rich, elaborate, profound, thorough, expert, or well developed.

Moreover, misunderstanding and misconceptions can be part of the individual's grasp of a subject anywhere along this continuum. Even when understanding is well developed it may not be "perfect." Partial understanding is the norm in most matters. As scholars note

> *In even the most mature person, understanding is a mixture of insight and misconception, knowledge and ignorance, skill and awkwardness.*
>
> (Grant Wiggins and Jay Mc Tighe in Understanding by Design)

Consequently, we should not ask "if students understand," but "to what extent they understand," or "in what ways do they understand."

It is tempting to view understanding as a kind of static mental entity that students can call forward on demand just the way we expect them to recall information from memory. But understanding is not a static representation of knowledge in the students' mind so much as it is a capacity to do thought provoking things with knowledge one has developed or is developing. Understanding involves being able to "think with" new knowledge and not just "think about" it.

There are several important methodological issues related to evaluating students' understanding (and evaluating any type of learning).

What do you mean by "understanding"? As long as "understanding" is an elusive concept in your class, it will be difficult to evaluate it effectively. In my experience, I was not able to evaluate understanding effectively until I could say things like, "On this assignment, understanding is the ability to . . .

- use concepts X, Y and Z to solve novel problems
- develop causal relationships to explain how X affects Y
- use different theoretical perspectives to predict X, Y, and Z
- use relevant evidence to develop a personal perspective
- use relevant evidence to develop a compelling argument

These statements operationalize understanding, indicating what understanding "looks like," and what students have to do in order to demonstrate their understanding. They also serve as a guiding principle for the design of assignments and class activities.

How do you judge the quality of understanding? Understanding is not easily quantified. What would it mean, for example, to say that a student understands 70% of the novel *Moby Dick*? How do we look at two different performances of understanding and make credible judgments about the differences between them? How do you systematize the evaluation regardless of whether the performance is a written paper, a class discussion, or a group presentation?

You need a scorecard—a rubric. Rubrics are scoring guidelines for evaluating performance. A rubric indicates the criteria used to distinguish among qualitatively different levels of performance. Essentially, a rubric for understanding is a template or model that defines important qualities of understanding and further distinguishes different levels of understanding.

One of the best developed rubrics for evaluating understanding is the work of Grant Wiggins and Jay Mc Tighe in their book, *Understanding By Design*. They identify six ways in which students can demonstrate their understanding—by explaining, interpreting, applying, developing a perspective, empathizing, and through self-knowledge. Their framework distinguishes between qualitatively different "levels" of understanding for each of these facets. See the attached *Understanding By Design* rubric.

Is your evaluation method credible? Two key technical properties of evaluation instruments are validity and reliability. If you are not familiar with these concepts, it is worth learning about them (See references in the appendix). Most instructors are not equipped to do extensive validity and reliability studies. But instructors can develop rubrics and evaluation procedures that have construct validity and inter-rater reliability.

> Construct validity refers to the extent to which your method measures the underlying theoretical construct it purports to measure. For example, if you define understanding as "the ability to apply knowledge to solve new problems," then that's what your method measures and not some other abilities such as reading comprehension or You can improve construct validity by developing and refining a sound model of understanding as it applies to the students and subject matter in your class.
>
> Inter-rater reliability refers to the extent to which evaluators agree on the ratings or evaluation of students. A measure of learning is reliable if two evaluators independently arrive at the same scores for a group of students. For example, suppose you develop a rubric for evaluating student understanding on a written assignment. The rubric is reliable to the extent that different instructors who evaluate the students' work arrive at the same scores for students (i.e., a student who gets a "low score" by one evaluator also gets a low score by the other). You can improve inter-rater agreement by developing and refining the scoring guidelines. In the event that you have multiple instructors to evaluate student performance, it is advisable to hold a norming session to discuss the criteria and how to apply them. You can also check inter-rater reliability by having instructors independently evaluate student performance and then determine the degree of consistency between them.

If you develop instruments and procedures that are valid and reliable, you will be better able to support your claims that you have measured the qualities of learning your study in intended to measure.

Additional Resources about Evaluating Understanding

The first four references are excellent starting points for theory, research and practice related to teaching for and learning with understanding.

Bransford, John D., Brown, Ann L., & Cocking, Rodney R. Editors 1999. *How people learn: Brain, mind, experience and schooling*. Washington, DC: National Academy Press.

This book-length report summarizes important developments in the science of learning. Accessible to a non-specialist audience, the book examines such topics as differences between novices and experts, conditions that improve students' ability to apply knowledge to new circumstances and problems, the design of learning environments, teacher learning, and effective teaching in history, mathematics, and science. This volume provides teachers with a thorough grounding in contemporary theory and research, and highlights important implications for teaching. The entire book is online at http://books.nap.edu/html/howpeople1/

Stone Wiske, Martha. Editor 1998. *Teaching for understanding: Linking research with practice*. San Francisco: Jossey-Bass Publishers.

This book is the product of a six-year collaborative research project by school teachers and researchers at the Harvard Graduate School of Education. Although it focuses on pre-collegiate teaching, it is applicable to university-level teaching as well. According to the TfU model, there are four fundamental elements in teaching for understanding—generative topics that afford possibilities for deep understanding in a subject, goals that explicitly state what students are expected to understand, performances of understanding through which students develop and demonstrate understanding, and ongoing assessment. The book provides interesting examples of these elements from actual classrooms and examples of student performance. This volume should be valuable for any instructor who views better student understanding as a primary goal of the scholarship of teaching.

Wiggins, Grant 1998. *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco: Jossey-Bass Publishers.

This book, a precursor to *Understanding by design* by the same author, challenges common assessment practices and offers a comprehensive approach to the design and practice of assessment intended to improve student performance. The book examines authentic assessment, the nature of feedback, how to use assessment to promote understanding, how to assess understanding, how to design assessments and create assessment systems. It is itself an important contribution to the scholarship of teaching that provides fundamental grounding is how and why to evaluate student learning and performance. I have used the book extensively to develop a more consistent assessment philosophy and also as a handbook to guide in the design of assessment materials.

Wiggins, Grant and McTighe, Jay. 1998. *Understanding by design.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

This book proposes that understanding is revealed to the extent that one can explain, interpret, apply, empathize, and have perspective and self-knowledge. The authors describe a process by which teachers can design experiences and materials to be consistent with these facets of understanding. A key component of the process is a way to assess understanding. Toward this end, they offer a rubric that defines different "levels" of understanding and suggest ways to evaluate different facets of understanding. This is a valuable book for those who want to translate abstract notions of understanding into concrete, observable aspects of student performance. See the Understanding by Design website http://ubd.ascd.org/index.html

Additional References about Assessment and Evaluation

Angelo, T.A. & Cross, K. P., (1993). *Classroom assessment techniques: A handbook for college teachers (second edition)*. San Francisco: Jossey-Bass Publishers.

If there is a "classic" about classroom assessment, this is it. The book is an excellent resource which has a large number of readymade CAT's (Classroom Assessment Techniques).

The following are good resources related to evaluation and assessment

They are written for a wide faculty audience and are not overly technical. They do not focus on classroom level assessment but many of the principles apply to classroom inquiry.

Light, R. J., Singer, J. D., & Willett, J. B., (1990). *By Design: Planning Research on Higher Education*. Cambridge, MA: Harvard University Press.

This is a very accessible book about doing educational research. It is not about classroom inquiry but does have useful information about validity and reliability.

Erwin, T. D., (1991). *Assessing student learning and development: A guide to the principles, goals and methods of determining college outcomes*. San Francisco: Jossey-Bass Publishers.

Palomba, C. A. & Banta, T. W. (1999). *Assessment essentials: Planning, implementing and improving assessment in higher education.* San Francisco: Jossey-Bass Publishers.

Middle States Commission on Higher Education, (2002). Student learning assessment: options and resources. Philadelphia: Middle States Commission on Higher Education. Available online at http://www.msache.org/SLA_Summary.pdf

How to Create Rubrics to Evaluate Student Understanding

The *Understanding by Design* model is an excellent starting point for evaluating understanding. The model has predetermined categories for different levels of understanding and you may be able to modify it to suit your specific needs and subject matter.

Some rubrics may be specific and analytical, while others can involve more holistic judgments. Here are two examples from my own work.

Thinking with vs. Thinking about subject matter. An important feature of understanding is the ability to use newly learned disciplinary concepts to explain other concepts, solve problems, formulate a perspective and so forth. In this way, understanding is the ability to "think with" new knowledge. In contrast, "thinking about" the subject matter involves describing and telling what one knows.

I use this distinction to make holistic judgments about students' understanding based on the extent to which they:

- integrate course concepts vs. mention or describe course concepts
- use course concepts to develop ideas vs. rely on personal opinion or intuitive beliefs
- use course concepts analyze and explain vs. use course concepts descriptively (not explain).

Types of explanations. A more specific rubric is one I have used to judge the quality of students' explanations. In this case, the rubric zeroes in on whether the explanation uses course concepts to make a causal connection between ideas. The rubric distinguishes between causal explanations that make specific connections among ideas vs. generic explanations that hint at the causal relationship but are non-specific vs. non-explanations in which the student just describes course concepts without making a causal connection. For examples see
http://kml.carnegiefoundation.org/gallery/bcerbin/Course_Overview/Reciprocal_Teaching_Class_Prob/RT_Analysis/rt_analysis.html

Developing your own rubric. A different approach is to start by "looking at the data." Put the evidence of student understanding in front of you and start asking questions about how students' responses they are similar to and different from one another. For example, suppose your study investigates how students understand a particularly difficult concept in your discipline. You have devised a task in which students respond to several questions at the beginning and end of class. Develop a rubric to evaluate the end-of-class responses.

1. Identify examples of good and poor understanding on the assignment. Pick a few of the best and worst pieces of work. Make a list of the characteristics shared by the good pieces and a list of characteristics shared by the poor pieces. The goal is to develop criteria that distinguish good and poor understanding, the dimensions that define quality of understanding.

2. Define gradations of quality. Define the characteristics of the best understanding. Define the characteristics of the poorest understanding. These provide anchor points. Fill in the middle levels of quality. You can make as many gradations or levels as you want. The trick is to identify clearly what distinguishes the "Best" understanding from the "Next Best" and then what that is next best and so on.

3. Practice using the rubric and revise as needed. You want to make sure you can distinguish levels of understanding. The only way to do this is to practice using the rubric to make sure it includes all the relevant dimensions of understanding, and to make sure you can use it reliably. You can tell

it works well if you can read a set of student work and say with confidence that all the work in the top category really belongs there—that all the pieces have the same level of quality on the critical dimensions. If not, something is wrong. What if you apply the rubric and then discover that work in the same category actually looks different on the critical dimensions? You may not be clear yourself about what the actual dimensions are (e.g., you could be using some criteria that you have not yet made explicit) or you may be applying the criteria inconsistently (e.g., you may need to define the criteria more carefully so that you can use them without thinking twice about what they mean).

Understanding By Design Rubric

Explanation	Interpretation	Application	Perspective	Empathy	Self-Knowledge
sticated: an unusually ugh, elegant, and tive account (model, y, or explanation); fully orted, verified, and ed; deep and broad; goes beyond the information .	*Profound*: a powerful and illuminating interpretation and analysis of the importance/meaning/ significance; tells a rich and insightful story; provides a rich history or context; sees deeply and incisively any ironies in the different interpretations	*Masterful*: fluent, flexible, and efficient; able to use knowledge and skill and adjust understandings well in novel, diverse, and difficult contexts.	*Insightful:* a penetrating and novel viewpoint; effectively critiques and encompasses other plausible perspectives; takes a long and dispassionate, critical view of the issues involved.	*Mature*: disposed and able to see and feel what others see and feel; unusually open to and willing to seek out the odd, alien, or different.	*Wise*: deeply aware of th boundaries of one's own others' understanding; ab recognize his prejudices a projections; has integrity- and willing to act on wha understands.
pth: an atypical and ling account, going nd what is obvious or was explicitly taught; s subtle connections; supported by argument vidence; novel thinking ayed.	*Revealing*: a nuanced interpretation and analysis of the importance/meaning/ significance; tells an insightful story; provides a telling history of context; sees subtle differences, levels, and ironies in diverse interpretations.	*Skilled*: competent in using knowledge and skill and adapting understandings in a variety of appropriate and demanding contexts.	*Thorough*: a revealing and coordinated critical view; makes own view more plausible by considering the plausibility of other perspectives; makes apt criticisms, discriminations, and qualifications.	*Sensitive*: disposed to see and feel what others see and feel; open to the unfamiliar or different.	*Circumspect*: aware of o ignorance and that of oth aware of one's prejudices knows the strengths and l of one's understanding.
loped: an account that ts some in-depth and nalized ideas; the student king the work her own, beyond the given-there pported theory here, but ficient or inadequate nce and argument.	*Perceptive*: a helpful interpretation or analysis of the importance/meaning/ significance; tells a clear and instructive story; provides a useful history or context; sees different levels of interpretation.	*Able*: able to perform well with knowledge and skill in a few key contexts, with a limited repertoire, flexibility, or adaptability to diverse contexts.	*Considered*: a reasonably critical and comprehensive look at all points of view in the context of one's own; makes clear that there is plausibility to other points of view.	*Aware*: knows of different points of view and somewhat able to empathize with others; has difficulty making sense of odd or alien views.	*Thoughtful*: generally aw of what is and is not understood; aware of one prejudices; knows the strengths and limits of on understanding.
ive: an incomplete nt but with apt and ntful ideas; extends and ens some of what was ed; some "reading een the lines"; account mited support/argument/ or sweeping alizations. There is a y, but one with limited g and evidence.	*Interpreted*: a plausible interpretation or analysis of the importance/meaning/ significance; tells a clear and instructive story; provides a history or context.	*Apprentice*: relies on a limited repertoire of routines; able to perform well in familiar or simple contexts, with perhaps some needed coaching; limited use of personal judgment and responsiveness to specifics of feedback/situation.	*Aware*: knows of different points of view and somewhat able to place own view in perspective, but weakness in considering worth of each perspective or critiquing each perspective, especially one's own; uncritical about tacit assumptions.	*Developing*: has some capacity and self-discipline to "walk in another's shoes," but is still primarily limited to one's own reactions and attitudes; puzzled or put off by different feelings or attitudes.	*Unreflective*: generally unaware of one's specific ignorance; generally unaw of how subjective prejudgments color understandings.
e: a superficial account; descriptive than tical or creative; a entary or sketchy nt of facts/ideas or glib alizations; a black-and- account; less a theory an unexamined hunch or wed idea.	*Literal*: a simplistic or superficial reading; mechanical translation; a decoding with little or now interpretation; no sense of wider importance or significance; a restatement of what was taught or read.	*Novice*: can perform only with coaching or relies on highly scripted, singular "plug-in" (algorithmic and mechanical) skills, procedures, or approaches.	*Uncritical*: unaware of differing points of view; prone to overlook or ignore other perspectives; has difficulty imagining other ways of seeing things; prone to egocentric argument and personal criticisms.	*Egocentric*: has little or no empathy beyond intellectual awareness of others; sees things through own ideas and feelings; ignores or is threatened or puzzled by different feelings, attitudes, or views.	*Innocent*: completely una of the bounds of one's understanding and of the of projection and prejudic opinions and attempts to understand.

From: Wiggins, G. and McTighe, J. (1998). *Understanding by Design*. Association for Supervision and Curriculum Development. Alexandria, VA.

Developing a Research Focus to Investigate Student Learning in the Classroom: Think Small

So, you want to investigate teaching and learning in your own classroom. Where do you begin? A key to successful classroom inquiry is finding an interesting and manageable research focus. If you have never done systematic classroom inquiry, try to resist the urge to attack BIG issues and problems. Given one's normal professional responsibilities, it is very difficult to take on the additional commitment of a large scale research project. Instead, think small, very small. But, small does not mean that the question or issue is trivial or inconsequential. Ultimately, the topic you pursue should have some significance for teaching in your field, and should be of some significance to you as well. Below are three avenues into a classroom inquiry project in a course you teach.

Avenue A: Inquiry can start with questions about student learning and development. Classroom inquiry often starts with "problems of practice" arising from instructors' personal experiences. In the broadest sense problems of practice focus on what, how and why students learn or do not learn what they are taught. This distinguishes them from other types of teacher problems. For example, new teachers spend a large amount of time learning basic teaching skills—how to give a lecture that lasts the entire class period, how to write an examination, how to assign grades, how to get a discussion going and so on. Although these are urgent matters for new teachers, they are not problems of practice as they focus only on the teacher's acquisition of basic classroom skills, and not on how those practices affect student learning.

Problems of Practice (in the Field of Psychology)

. . . the driving force behind inquiry into teaching and learning most often is found in what Hoshmand (1994) describes as "problems of practice." She suggests that problems encountered during practice can form the basis for inquiry, and although her examples are based on clinical practices, they are easily translated into problems that can arise in the practice of teaching of psychology. One common type of problem is encountered "when an intervention based on a particular rationale does not work" (Hoshmand 1994: 184). Almost every teacher has experienced a failed attempt in the classroom, no matter how brilliantly planned. A fairly typical response is to chalk it up to a bad day and move on. But increasingly, psychologists are turning such experiences into opportunities to investigate why the teaching intervention was not effective.

Hoshmand's framework encourages us to interview students, reflect on what was observed in class during the unsuccessful intervention, solve the problem, attempt to correct it, and further reflect on what type of intervention would be more successful and why, leading to another attempt at the now revised intervention and repeating the cycle of reflective inquiry. This process of inquiry among psychologists is often informed by other work on teaching and learning by scholars in the field (e.g., Bransford, Brown, and Cocking 1999). In turn, the result of this type of inquiry may extend, modify, or even contradict prior understandings, leading to important new research questions. Building on and further developing existing understanding is, by its very nature, at the core of scholarly work (Shulman 1999b).

Excerpt from "Disciplinary Styles in the Scholarship of Teaching and Learning: A View from Psychology," by Susan Nummedahl, Janette Benson & Stephen Chew in Huber, M. and Morreale, S. (2002). *Disciplinary styles in the scholarship of teaching and learning: Exploring common ground.* Washington, DC: American Association for Higher Education.

Problems of practice focus directly on student learning and development, such as

- gaps between the instructor's expectations for student learning and their performance
- specific concepts and ideas that are especially difficult for many students
- some students who just don't "get it"
- misconceptions and beliefs that prevent students from understanding the subject matter

In these cases, inquiry begins with a specific question about student learning. Why do so many students have difficulty with concept "X"? Why are students unable to do "Y" even after considerable teaching and practice? Why do students consistently misunderstand "Z"? Why do some students fail this course? What are the dominant misconceptions students have in this class and how can I get them to move beyond their person-on-the-street theories?

In general the classroom inquiry process to investigate these questions would involve several major phases

Design
- devise ways to make student thinking (relative to the problem) visible and subject to observation
- devise tools to evaluate the quality, characteristics, or forms of student thinking

Gather evidence of student thinking
- make systematic observations of student thinking

Make sense of the evidence
- analyze and interpret students' thinking
- Optional: Use the evidence to revise, re-teach and collect additional data

Report the results
- produce an account of the research project
- present the work for peer review

Avenue B: Inquiry can start with an existing teaching and learning episode. Alternatively, instructors might organize classroom inquiry around a single teaching and learning episode. A teaching episode may be thought of as the teaching and learning that takes place in a single class period or that takes place with respect to a single assignment or activity (in or out of class).

A single teaching episode is an appealing level of analysis for several reasons. First, teachers organize and plan lectures and class activities that can take place within single class periods. A single class period is a manageable unit of analysis; it is much easier to monitor students' learning carefully in a single class period than it is during several weeks or months of a class. It is also easier to analyze the structure of a single class period, and to observe the intended connections between teaching and learning than it is to do so in larger segments of a course.

A classroom inquiry project could investigate teaching and learning in a single class period, with respect to a single assignment or class activity. A classroom inquiry project could involve testing and revising an existing class assignment, exercise or teaching episode.

The inquiry process would involve
- devising ways to observe and evaluate student performance during the episode, assignment, or activity
- field testing the episode by collecting data about students' engagement, learning and behavior in class
- analyzing the evidence and proposing possible ways to revise the episode in order to foster better or different learning.
- field testing the revised episode (in a subsequent semester) and collecting data about student performance
- analyzing the evidence and proposing additional ways to improve the episode
- producing an account of the research project

- presentation and peer review of the work

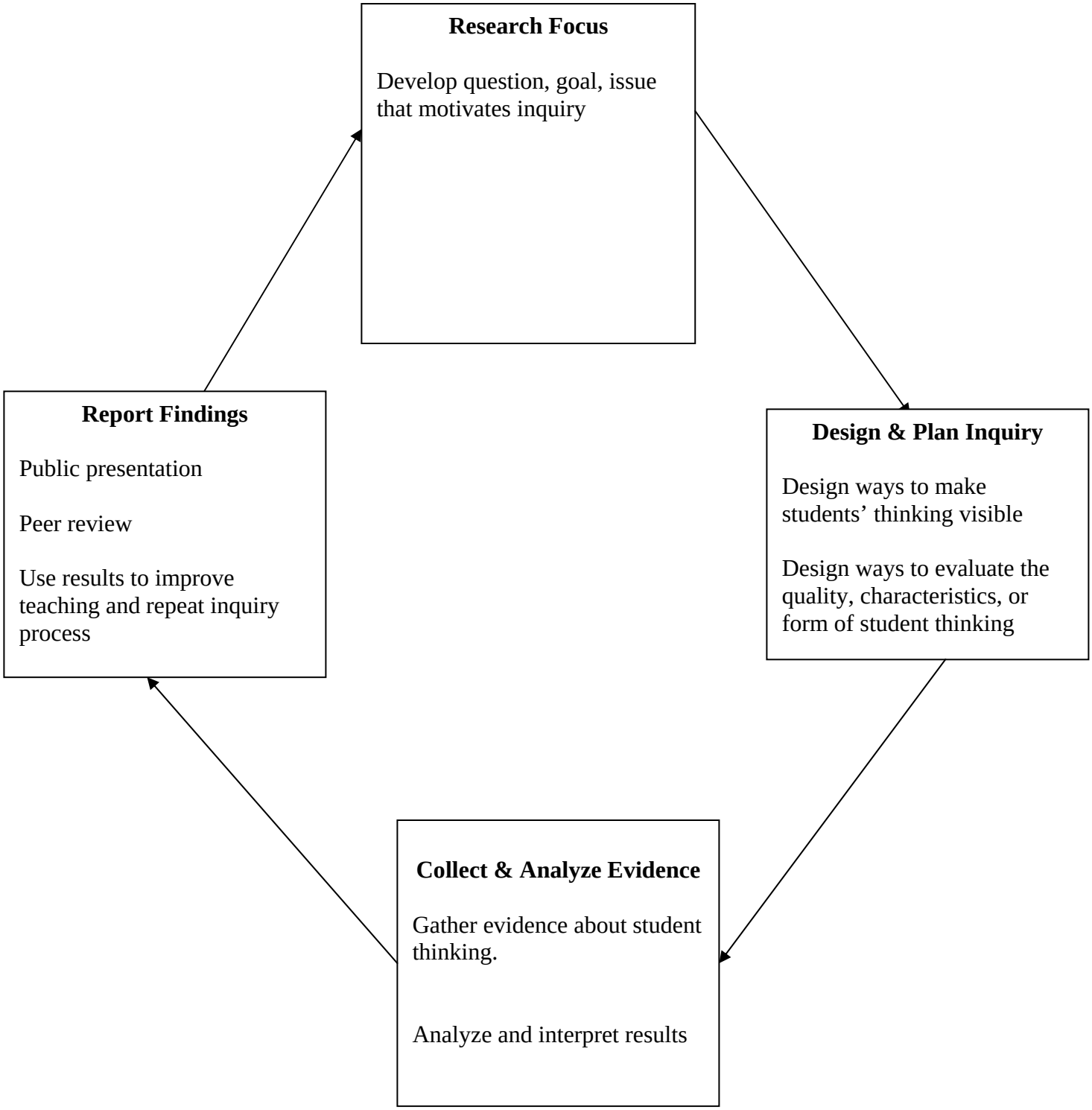
Avenue C: Inquiry can start with a new attempt to address a student learning goal. An instructor's goal for student learning can be the focal point of classroom inquiry. A classroom inquiry project could be built around how to address the goal in one or two specific class periods. The instructor would design, test and revise a new (rather than existing) assignment, exercise or teaching episode to address the student learning goal.

The inquiry process would involve

- designing a new teaching episode, assignment, activity
- devising ways to observe and evaluate student performance during the episode, on the assignment, activity
- field testing the episode, assignment, activity and collecting data about students' engagement, learning and behavior
- analyzing the evidence and proposing possible ways to revise the experience in order to foster better or different learning.
- field testing the revised episode, assignment (in a subsequent semester) and collecting data about student performance
- analyzing the evidence and proposing additional ways to improve the experience
- producing an account of the research project
- presentation and peer review of the work

A Model of Classroom Inquiry

Investigating Student Learning in the Classroom



Scholarship of Teaching for Understanding

You might consider developing a scholarship of teaching project around your efforts to promote deeper understanding.

Below are some sketchy suggestions about possible projects.

- How an idea develops. Track changes in students' understanding of a particular concept or topic during a term. This might take place over a week or the entire term.

- Range of understanding. Do more in-depth analysis of how understanding develops in a few students at different ability levels. Examine differences in the range of understanding in your class.

- Develop a "background knowledge probe" to determine students' knowledge of topics before you teach them. Compare their responses to their understanding at the end of the unit/instruction.

- Misconceptions. Keep track of the kinds of misconceptions students bring to your class and develop in your class.

- Understanding difficult ideas. Focus on how students develop understanding of particularly difficult concepts. Every field has topics and concepts that are especially difficult for students in introductory level classes or other classes.

- Enduring understanding ("month-long learning"). Email students a month (or some other elapsed time) after the class, and ask them to answer questions about key questions and ideas from the course.

- Transfer of learning. Ask instructors in the "next" class to administer a "post-test" to determine students' long-term understanding and ability to use concepts they learned in your class.

- Thinking with the subject matter. Devise a series of "transfer tasks" (i.e., assignments in which students apply newly-learned concepts to novel situations or problems). Examine the extent to which they can use new knowledge.

- Progressive discourse. How does students' understanding develop from class discussions? Devise a way to evaluate the quality of class discussion in terms of how students interact and what kinds of ideas they produce.

- Field test TfU assignments. Try out a new TfU assignment. Evaluate students' learning and use the results to redesign the assignment. Test it again the following term.

- Leaving legacies. Ask students in your class to create materials, analyses, advise, etc. for students taking the class in the future.

Making Students' Thinking Visible During the Learning Process

A hitting coach tells a player, "When you start to swing you drop your back shoulder. As a result you're swinging under the ball. That's a major reason why you are making poor contact and popping it up so much." Hitting a baseball is a complex act, but it is directly observable. The baseball coach can watch all facets of the act—the stance, how the player moves into the pitch, the body mechanics of the player given different types of pitches and so forth.

Thinking is a complex act too. But, intellectual activity is not always overt and teachers may have little direct access to students' thinking. Too often, teachers view only the final product of students' thinking—the completed paper or examination, and not the intermediate phases of thinking that lead up to the product. This is the equivalent of being a baseball coach whose sole source of information about his team is what he reads in the newspaper.

A significant challenge in classroom inquiry is trying to make students' thinking visible and subject to observation and analysis at key times during the learning process. This involves engaging students in activities, exercises, and assignments through which their understanding becomes externalized, such things as

To illustrate, consider the following episode from an elementary school math lesson. At the start of class, the teacher introduces a problem, ½ + 1/3 = ?, for the class to solve. This is the first time they have seen addition of fractions with unlike denominators. After the children have worked out their answers, the teacher asks three children to put their answers on the board.

Child A: ½ + 1/3 = 2/5
Child B: ½ + 1/3 = 2/5
Child C: ½ + 1/3 = about ¾.

Based on their answers about all we can conclude is that all three children are wrong. But, the teacher also asks the children to explain how they arrived at the answer—to talk through their solutions. As they do so, it is evident that each has a different understanding of the mathematical concepts involved.

Child A: ½ + 1/3 = 2/5 ("I added the numerators first and got 2 and then the denominators and got 5. That gave me 2/5.")

Child B: ½ + 1/3 = 2/5 ("First I changed ½ to 2.1 and then I changed 1/3 to 3.1. Then I added 2.1 + 3.1 and got 5.2. Then I changed this back to a fraction—2/5.")

Child C: ½ + 1/3 = about ¾. ("Well, I don't know how to add the fractions. But I thought about how big the two fractions are. I imagined a pizza and what ½ and 1/3 of it would look like. It just seemed like if you added ½ a pizza and 1/3 of a pizza you'd get about ¾ of a pizza. I don't know if it's right, it's just kind of an estimate.")

As students describe how they did the problem, we get a glimpse of how they make sense out of the subject matter and reveals significant differences in their understanding. If we focus only on the endpoint of their thinking, we find out whether they get the correct answer but we have no basis for explaining how and why students learn or do not learn what they were taught.

Strategies for making students' thinking visible. To the extent possible, we want access to students' thinking as they engage the subject matter and tasks in our classes. These opportunities already exist when

students answer questions, participate in discussions, work out problems, write papers, and in general, "show their work."

So, why use "strategies" if students already reveal their thinking in class? The answer to this question depends upon whether or not students produce the kind of evidence you want relevant to your research question. In some cases a naturalistic observation may suffice—you simply observe or record their classroom activity. In other cases, you may need to interrupt or intrude on students' activity to ask specific questions. Or, you may need to design a new task that externalizes their thinking.

Strategies can be as simple as asking students to write a brief response to a question in class or more highly structured performances of understanding (e.g., presentation of projects). The key feature is that they evoke students' understanding and reveal how students make sense of the subject matter relevant to your research question (e.g., engage students in such things as using knowledge to explain, interpret, analyze, compare, make analogies, extrapolate, and find connections among unrelated facts and ideas).

Examples of Strategies to Make Students' Thinking Visible.

The type of evidence you need will determine the type of strategies you use. The list below is not ordered in any particular way.

Background knowledge probe. To determine students' knowledge and understanding of a new topic ask them one or more questions that you hope they will understand by the end of the class period (or unit or eventually). These should focus on the major concepts rather than details. Over time you might develop a set the questions that works particularly well. (See *Classroom Assessment Techniques Second Edition* by Tom Angelo and Pat Cross, p. 119)

Written responses. Ask students to answer a question, solve a problem or identify gaps in their understanding at the end of class. Have students summarize their grasp of the material (i.e., the "minute paper") or identify ideas that do not make sense (i.e., muddiest points). Alternatively, pose a question, problem or exercise that elicits students' understanding of the subject at any time during a class period. These can vary from brief insertions (e.g., the instructor stops and asks students to write a brief answer to a question) to more extensive "performances of understanding" (e.g., a challenging exercise or activity through which students demonstrate and develop their understanding).

Think-Pair-Share. Pose a thought-provoking question in class. Give students a minute or two to write a response, then ask them to compare their responses to classmates sitting next to them. Ask several students who have different views to report their responses to the entire class.

ConcepTests. Harvard physicist, Eric Mazur, uses conceptests during a class period to determine students' understanding of key concepts. He follows a sequence similar to the think-pair-share technique: 1) Instructor poses question, 2) Students think about it and then record individual answers, 3) Students try to convince neighbors of their answer, 4) Students revise their answers, and 5) the instructor tallies answers and explains the correct answer. About one third of class time is used for conceptests. (See *Peer Instruction* published by Prentice Hall for a full explanation.)

Class presentations can be significant performances of understanding. However, not all presentations demonstrate or develop understanding, especially if students' have scripted the entire event. Presenters should respond to questions that probe their grasp of the topic as part of the presentation. Questions might be submitted before, during or after the presentation.

Class discussions can be performances of understanding to the extent that they involve progressive discourse—i.e., students actually develop ideas, not just talk about stuff. This depends upon how the instructor structures the discussion and the topic or question motivating the discussion. Instructors also need a way to monitor students' understanding during the discussion. This could be a checklist that indicates the types of ideas, perspectives, etc. that ought to appear in the discussion.

Online assignments and discussions Students post their ideas to an electronic discussion forum and respond to classmates' work, or submit their work electronically to the instructor prior to class (e.g., students write answers to complex questions and email them to the instructor one day before class).

Personal Response Systems. Some classrooms have Personal Response Systems in which students use keypads on their desks to select their answers to questions. Instructors can stop periodically to ask an understanding question and get immediate feedback from the entire class. You can also do this procedure without the expensive technology by asking for a show of hands.

Drafts of significant papers and projects. Drafts of papers and projects can reveal intermediate forms of thinking and understanding. Of course, whether the drafts reflect qualitative differences depends on how the instructor structures the revision process.

Think aloud protocols. Cognitive psychologists use a research technique called verbal protocol analysis to gain access to individuals' thought processes while they perform a task. The person is instructed to say whatever comes to mind and to keep talking for the duration of the task. This method, sometimes called "Think Alouds," has been used to study complex psychological processes. For example, much of what we know about how people solve complex problems problem solving processes is derived from protocol analysis. The technique produces a verbal record that closely reflects the thought processes that occur during the task performance. Refer to the article by Lendol Calder which describes how he has used think aloud protocols to analyze students' understanding in an introductory history class.

Monitoring the development of understanding throughout the semester Instructors can examine the development of important concepts and skills across the entire semester (and beyond the semester if so desired). Students' understanding of central ideas may evolve throughout the semester, and it is easy to overlook this development as you focus on day-to-day material. Examine development by asking students periodically to address questions or problems related to core concepts (i.e., the most central ideas you want students to understand deeply in the class). The questions could be the same ones asked earlier in the term or they could be a series of questions that are progressively more sophisticated.

To make this a manageable take, decide how many different pieces of work to examine and how many students to sample. For example, one strategy is to select several pieces of work from 2-3 students at different performance levels (high performing, typical performing, low performing).

Assessing student understanding beyond the end of the course. Sometimes instructors want to know whether student understanding endures beyond the class where they first developed it. If you have the time and inclination you can contact a sample of students (email is good, phone calls are better) to do a follow up assessment. Another option is to ask instructors in follow up classes to use an assessment instrument in their classes. This works well if the instrument can be used as a diagnostic pre-test in the follow up course.

In summary, thinking aloud is a natural part of the college classroom. Making students' understanding visible is not difficult, but in order to be useful we need to think about how, when, and why to observe students' in the process of learning and thinking.

Performances of Understanding: A Pedagogical Framework for Making Students' Understanding Visible

Understanding is the ability to use knowledge in thought provoking ways to explain, interpret, analyze, compare, make analogies, and discern relevant connections among unrelated facts and ideas.

Performances of understanding are tasks, activities, assignments through which students demonstrate and develop their understanding of important knowledge and skills. Instructors have their own ideas about what constitutes understanding in their disciplines and what performances of understanding provide the most compelling evidence about students' understanding. Nonetheless, good performances of understanding

- Relate directly to understanding goals
- Develop and apply understanding through practice
- Engage multiple learning styles and forms of expression
- Promote reflective engagement in challenging, approachable tasks
- Demonstrate understanding—means of monitoring, publicizing and learning from students' understanding.

Understanding develops. Sometimes it develops quickly, but development may be slow going and halting if concepts are particularly difficult. Instructors can use a sequence of understanding performances to support students' development and to make their thinking visible. Each performance serves a different function.

1. Messing about—initial inquiry not yet structured by disciplined-based methods or concepts; beginning of unit and draw S's into the domain of a generative topic; open-ended and approachable on multiple levels so S's can engage them no matter what their prior level of understanding; help S's see connections between topic and own interests/previous experiences

2. Guided inquiry (Getting the hang of it)—engages S in using ideas and modes of inquiry central to understanding the identified goals; T provides guidance/support as S applies disciplinary concepts/methods to integrate their growing body of knowledge and to perform complex understanding

3. Culminating performance—demonstrated mastery of designated understanding goals; synthesize other understanding.

For additional information see

Stone-Wiske, Martha. Editor (1998). *Teaching for understanding: Linking research with practice*. San Francisco: Jossey-Bass Publishers.

This book is the product of a six-year collaborative research project by school teachers and researchers at the Harvard Graduate School of Education. Although it focuses on pre-collegiate teaching, it is applicable to university-level teaching as well. According to the TfU model, there are four fundamental elements in teaching for understanding—generative topics that afford possibilities for deep understanding in a subject, goals that explicitly state what students are expected to understand, performances of understanding through which students develop and demonstrate understanding, and ongoing assessment. The book provides interesting examples of these elements from actual classrooms and examples of student performance. This volume should be valuable for any instructor who views better student understanding as a primary goal of the scholarship of teaching.

Blythe, Tina, & Associates (1998). *The teaching for understanding guide*. San Francisco: Jossey-Bass Publishers.

Using "Think Alouds" to Evaluate Deep Understanding
by
Lendol Calder and Sarah-Eva Carlson

"Deep understanding" is what teachers want for students. But how do we know when it has been achieved? And are certain assessments better than others at shedding light on what students really know and understand? The story is told of a deaf English public schools inspector who assessed learning on the basis of facial reconnaissance. "I have not been able to hear anything you have said," he would admit to students, "but I perceive by the intelligent looks on your faces that you have fully mastered the text." Are today's assessments any more valid than this? The following essay describes the authors' experience with invalid student learning assessments and their subsequent employment of an effective technique called the *think aloud method*.

Think alouds are a research tool originally developed by cognitive psychologists for the purpose of studying how people solve problems. The basic idea behind a think aloud is that if a subject can be trained to think out loud while completing a defined task, then the introspections can be recorded and analyzed by researchers to determine what cognitive processes were employed to deal with the problem. In fields such as reading comprehension, mathematics, chemistry, and history, think alouds have been used to identify what constitutes "expert knowledge" as compared to the thinking processes of nonexperts. For first year assessors, think alouds offer a promising method to uncover what conventional assessment methods often miss: hidden levels of student insight and/or misunderstanding.

Experienced teachers know that popular assessment methods conceal as much as they reveal. Papers and exams, for example, offer little help for figuring out why a student has recorded a wrong answer or struggled unsuccessfully with an assignment. Conventional assessments also run into problems of validity. Because they rely on students' ability to articulate themselves in formal language, papers and exams tend to conflate *understanding* with *fluency*. But sometimes, especially with first-year students, the tongue-tied harbor deep understandings even though they perform poorly. The reverse is true, as well; sometimes articulate students are able to say more than they really understand. "The thorniest problem" of assessment, according to Grant Wiggins and Jay McTighe (1998), calls for differentiating between the quality of an insight and the quality of how the insight is expressed.

We first utilized think alouds when assessing a new design for a first-year history course. The new design shifted emphasis away from tidy summaries of historical facts and knowledge toward the central questions, methods, assumptions, skills, and attitudes that characterize history as a discipline. Students completed eight identical assignments in the course, and student learning was measured by comparing the students' first and last papers. The results were disheartening. It was the rare student who showed steady progress from week to week, and few of the final papers were superior to the first ones. On the basis of this evidence, it seemed the new course was a failure.

But different evidence suggested otherwise. In course evaluations and self-reports, students insisted they had learned a great deal, a claim that certainly squared nicely with the intelligent looks on their faces at the end of the term. Puzzled by the conflicting evidence, we turned to think alouds for help.

Our procedure was as follows. From sixty students in the course, twelve were selected to participate in a think aloud study, representing a cross-section of students in terms of gender, grade point average, and major/nonmajors. For their participation, subjects were paid ten dollars an hour. In week one of the course, we sat down with each student in a room equipped with a tape recorder. After training subjects how to think out loud, we presented them with documents concerning the Battle of the Little Bighorn, a subject most knew little about. Then we asked our subjects to think out loud while "making sense" of the

documents. This was essentially the same task they would perform eight times over the length of the course, though in this case their thoughts would not be filtered by the task of composing an essay. With the tape recorder running, subjects read through the documents aloud, verbalizing any and all thoughts that occurred to them. When subjects fell silent, we would prompt them to think out loud or to elaborate on their thoughts as they attempted to make sense of the historical evidence.

Our think aloud sessions lasted anywhere from forty to ninety minutes. After all twelve sessions were completed, the tape recordings were transcribed for analysis. Analysis took the form of coding each discrete verbalization in the transcript according to the type of thinking it exemplified. We were able to identify fifteen different types of thinking processes displayed in the think alouds, from the uncategorizable ("it sure is hot in here") to comprehension monitoring ("I don't understand that part") to the six types of historical thinking we were particularly looking for, such as *sourcing a document* ("I can't trust Whittaker; he wasn't there"), *asking a historical question* ("I wonder what caused this battle?"), or *recognizing limits to knowledge* ("I need to see more evidence than this"). After coding each think aloud independently, we used a common rubric to rate each subject's proficiency on the six thinking skills taught in the course. For this, we used a 5-point Likert scale where "1" indicated the undeveloped ability of an average high school senior and "5" indicated a sophistication comparable to that of a professional historian. We then compared our coded transcripts until reaching consensus on how to rate the students' abilities in the six key areas. To prevent our bias as course designers from influencing the results, we contracted with an outside analyst to help us code the transcripts and rate students' abilities.

At the end of the term the twelve subjects completed a second think aloud. When these sessions had been transcribed and coded and the subjects' abilities rated, we compared the first and second think alouds to determine whether students had made gains in their understanding of what it means to "think historically." The think alouds opened a fascinating window into the thinking patterns of students before and after the course. Overall, the think alouds revealed cognitive enhancements that were not as dramatic as claimed in student self-reports, but much greater than indicated by using comparisons of early and late papers. Other surprises were equally interesting. Under-performing students struggled less with historical thinking than with reading itself. Moreover, in the second set of think alouds, we noted that some of the best insights and meaning making came from students who, in the gradebook, were steady "B" and "C" performers. For them, deep understandings seemed to evaporate when they tried to wrestle their thoughts to paper. This told us that we had work to do if we wanted to distinguish between assessing *understanding* and assessing *students' ability to communicate their understanding*. The real roadblocks to learning historical thinking, we discovered, are poor reading comprehension and prose writing.
On our campus, the potential of think aloud protocols has not been lost on other faculty. For example, library staff are using think alouds to assess how first-year students find information when they go to the library. Information gained from the study will be used to help library staff identify misconceptions and teach against common missteps students make when doing research.

Think alouds are not perfect assessment instruments. The advantage of think alouds is that they give us insight into our students' struggle to formulate problem solving strategies, employ skills, and develop insights. Papers, exams, and ex post facto commentary by students are helpful in their own way. But they make the process of understanding seem more orderly than it is, covering up the confusion, the disorientation, the mimicry of correct responses, and the lucky guesses-all of which are good to know about when assessing teaching and learning.

As the emphasis in first-year pedagogy switches from teaching to learning, from "what am *I* going to do today" to "what are *they* going to do today," the days using simply papers and exams to assess student learning are long gone. Teachers need more procedures capable of opening up the hidden world of

learning. Think alouds can be helpful this way, especially in courses emphasizing the development of cognitive skills.

The Authors

Lendol Calder is chair of the department of history at Augustana College, Rock Island, IL. Sarah-Eva Carlson is an Augustana senior and research assistant for the think aloud project.

Bibliography

On course design and assessment:
Grant Wiggins and Jay McTighe, Understanding by Design (Association for Supervision and Curriculum Development, 1998).

On think alouds-their history, effectiveness, and procedures:
Maarten W. van Someren, Yvonne F. Barnard, and Jacobijn A. C. Sandberg, The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes (Academic Press, 1994).
K. Anders Ericsson and Herbert Alexander Simon, Protocol Analysis: Verbal Reports as Data (MIT Press, 1993).
Debra K. Meyer, "Recognizing and Changing Students' Misconceptions: An Instructional Perspective," College Teaching 41 (Summer 1993): 104-108.

Disciplinary Uses of Think Alouds:
Reading
Catherine Crain-Thoreson, Marcia Z. Lippman, and Deborah McClendon-Magnuson, "Windows on Comprehension: Reading Comprehension Processes as Revealed by Two Think-Aloud Procedures," Journal of Educational Psychology 89 (December 1997): 579-591.
Linda Kucan and Isabel L. Beck, "Thinking Aloud and Reading Comprehension Research: Inquiry, Instruction, and Social Interaction," Review of Educational Research 67 (Fall 1997): 271-299).
Rebecca Swearingen and Diane Allen, Classroom Assessment of Reading Processes (Houghton Mifflin, 1997), pp. 21-25.
Suzanne E. Wade, "Using Think Alouds to Assess Comprehension," Reading Teacher 43 (Mar 1990): 442-451.

Chemistry
Craig W. Bowen, "Think-Aloud Methods in Chemistry Education: Understanding Student Thinking," Journal of Chemical Education 71 (March 1994): 184-190.

Mathematics
Alan H. Schoenfeld, Cognitive science and mathematics education (Lawrence Erlbaum Associates, 1987).

History
Lendol Calder, "Looking for Learning in the History Survey," American Historical Association Perspectives 40 (March 2002): 43-45.
Samuel S. Wineburg, "Probing the Depths of Students' Historical Knowledge," American Historical Association Perspectives 30 (March 1992) : 1-24.
Samuel S. Wineburg, "Historical Problem Solving: A Study of the Cognitive Processes Used in the Evaluation of Documentary and Pictorial Evidence," Journal of Educational Psychology 83 (March 1991): 73-87.

The Scholarship of Teaching and Learning: A Beginner's View

Curtis D. Bennett
Mathematics
Loyola Marymount University
Carnegie Scholar 2000-2001

I received my Ph.D. in 1990. Over the last 11 years I have striven to be a scholarly teacher, by which I mean that I approach my teaching in a scholarly fashion and use my scholarship in mathematics to improve my teaching. However, I only started investigating my teaching with the intention of making my work public about a year ago. Speaking as a beginner, there are many things that I wish I knew before I got started, and there are many thoughts I would like to share with others who are just starting off on the scholarship of teaching and learning (SoTL). So, in the spirit of David Letterman:

My SoTL Top Ten List

1. **Investigate what you are interested in.** As Randy Bass says, "What matters most is for teachers to investigate the problems that matter most to them" (Bass 1999, p. 7). Doing any serious scholarship takes time. Consequently, you have to judge how much time you can devote to any one project. SoTL is no different. Moreover, in many places it is not well rewarded, so you need to see a personal value in what you are doing. If you have a burning question that you want answered, that is wonderful. In this case, don't let anyone else deflect you from trying to answer your question. However …

2. **Before you can decide if something "works" you need to define "works."** The main question that most of us have is a variant on the theme: "Does method y work in teaching subject x?" In fact, my initial research question was: "Do semester long research projects work in teaching students to think more like mathematicians?" While the question sounded really nice, there were a lot of undefined terms hidden inside of it. I had already operationally defined "semester long research projects." On the other hand, I had only a vague idea about what I meant by "think like mathematicians." Worse yet, I didn't really know what I meant by "work." Everyone has his or her own understanding of this term, so finding a common understanding is much harder. As in any research, the first step is to turn the problem you are interested in solving into something that is researchable and that can be communicated to others. For a hard scientist, this comes down to research design; for a mathematician, this means defining your terms; and for someone studying education, it means establishing a framework. In truth, these are all basically the same idea. Thus …

3. **Your academic training can be applied to SoTL.** While the exact methodology you have learned for your scholarly research may not be appropriate to investigating your question in SoTL, the basic tools of your trade probably are. As a mathematician, my skill in writing mathematical proofs probably won't help me discover what my students know. On the other hand, my experience with defining terms extremely carefully has come in exceedingly handy. Moreover, since I am interested in knowing how the students approach mathematical problems, using my knowledge about why mathematicians approach problems in various ways helps me understand and categorize student responses. As another example, Mills Kelly, an historian doing SoTL, discovered that his training in history actually applied to his work on SoTL (Hutchings, 2000 and Kelly, 2000). Scientists might not be able to design a pure research study on different teaching techniques, but they can certainly apply the thought experiments they do. Remember, just because you can't control all variables doesn't mean that you can't do your best to control some. To paraphrase Robert Solow: Just because a perfectly aseptic environment is impossible, doesn't mean you operate in a pig sty. This leads us to …

4. **SoTL is not neat and tidy (or as one Carnegie Scholar says, "Make Peace with Reality").** As much as you might want students to interpret questions in the way you mean them, it won't happen. In my research, I asked students to state questions as part of their weekly assignments. At current count I have about 4 cases out of 40 where the students actually did what they were asked. If I do the study again, I will change the assignment. However, for now I have to deal with what my students actually did. Students are people, and people are unpredictable. You have to allow for the fact that just because you want things to happen, doesn't mean they will. Of course, related to this is a need to look at other methodologies, which brings us to …

5. **Read! Read! Read! But don't forget to think.** In any field, keeping up with the literature is important. When embarking on a new journey into scholarship of teaching and learning, you will probably need to catch up, not only on what people are doing, but also on how to do it. The first recourse for this is to read. But while you are reading, you also need to keep an eye on what you want to do. As in any field, it is easy to spend all of your time reading about what other people have done. Thus one needs to think about which articles must be read today and which articles can wait until later. In addition, you need to decide if the research methodologies suggested are actually the right methodologies for you. These all require spending time thinking and reflecting on what you are doing. To help with this you might consider working with faculty from other fields. If you need to do sociological research, you should at least hook up with a sociologist briefly. One of the greatest time-savers for me has been my collaboration with a faculty member at MSU who does sociological research. She has told me things it would have taken me weeks to learn from a book. Similarly, other faculty can suggest articles for you to read. Of course, none of this is all that different from what we do in our research fields; it is just that we are making more bridges outside of our field. Of course to build those bridges (particularly as a mathematician or hard scientist), you should …

6. **Be prepared to confront your research prejudices.** As a mathematician, I had long believed that sociological research involving numbers was a bunch of bunk. Of course, once I started becoming educated about such research, I discovered that I had serious misconceptions about this and other fields. Consequently, I have been forced to drop many of my prejudices towards other fields. Starting in on such research forces you to open your eyes to a greater reality. Of course, as you discover things, you will find that …

7. **Your research questions will change rapidly.** Once you start investigating topics of teaching and learning, it is amazing how quickly new questions spring up. Often these new questions must be answered before you can even attempt to make any progress on the old. On the flip side, it is also the case that, like any research, you end up with more questions than you start with, which is to say …

8. **True scholarship is a never-ending story.** As scholars we know that an answer to a good question leads to more questions. In fact, doing the scholarship of teaching and learning naturally leads to investigating your practice over and over again. At each stage you reflect on what you have done, work to improve your teaching, and collect information to support your arguments to others. This brings us to making your work public. My suggestion is that …

9. **Your research should speak to you first, your peers second, and the greater academic community third.** Shulman argues that the scholarship of teaching and learning needs to reside in the disciplines (Shulman, 2000). To undertake such research, it should be something you find interesting. If colleagues teaching similar classes are to benefit, your work must speak to your peers. Hence it is important that your work be presented to your peers first and the greater academic community second. Of course, for it to be listened to, you should …

10. **Tell a good story.** Perhaps the most important thing I have learned from Lee Shulman is the importance of telling good stories. The better the story, the more people will want to read what you tell them, and the more they read, the better they will understand what you are trying to say. Anyway, that's my story.

References

Bass, R. (1990), The Scholarship of Teaching: What's the Problem?, *Inventio*, Vol. 1 (1), 1-9.

Hutchings, P. (2000), Introduction, Approaching the Scholarship of Teaching and Learning, *Opening Lines, Approaches to the Scholarship of Teaching and Learning*, Carnegie Foundation, 2000, pp. 1-10.

Kelly, T. M. (2000), For Better or Worse? The marriage of web and classroom, *Opening Lines, Approaches to the Scholarship of Teaching and Learning*, Carnegie Foundation, 2000, pp. 53-62.

Shulman, L. (2000), From Minsk to Pinsk: Why a scholarship of teaching and learning?, *The Journal of Scholarship of Teaching and Learning*, Vol. 1 (1), 48-52