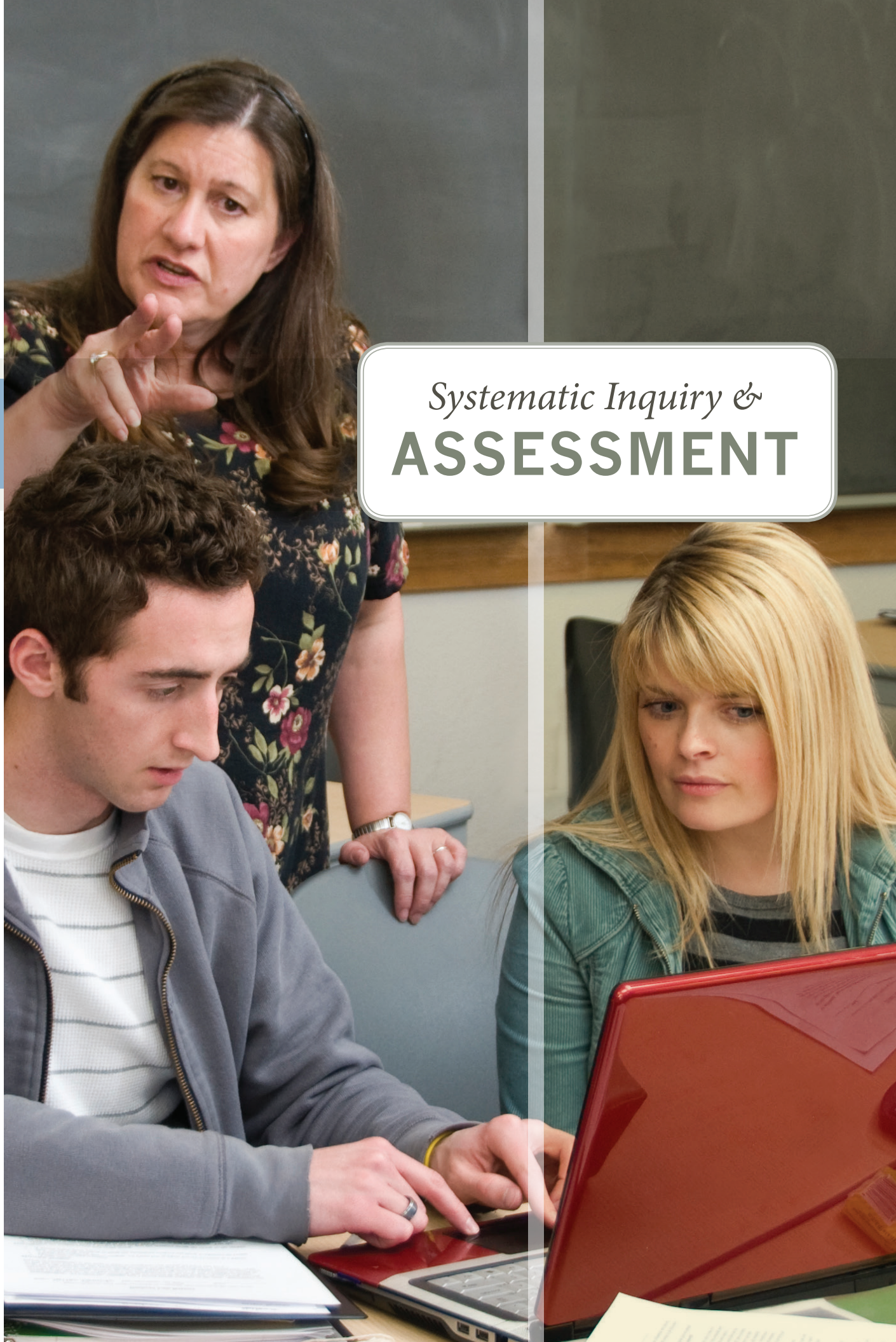
BYU IDAHO

FACULTY CONVERSATION SERIES **BOOK THREE**



Systematic Inquiry & ASSESSMENT

As educators, being professional

Success in education is challenging to define, quantify, and measure —

failure, on the other hand, it easy to identify and relatively simple to quantify. Educators can point to test scores, persistence rates, graduation rates or average course grades as evidence that their students are not failing. They have a much more difficult time producing evidence that their students are learning in significant and permanent ways.

This leads to a common and problematic situation in educational institutions. Very often, because it is difficult to get a firm handle on success, a lack of conspicuous failure is accepted as a working definition for educational success. Yet lack of failure is not the same as success and "good enough" is not good enough in professional educational settings.

All educational institutions struggle with defining and measuring success in such a complex and profoundly human endeavor as learning and teaching. There are, of course, many important aspects of most human endeavors which cannot be usefully quantified or measured.

means assessing our work

Even though the outcomes of education are often qualitative, they can still be defined and very often measured. It is true that qualitative things should not be measured quantitatively. Some things must be evaluated more in terms of discernment and observation than numerically, but they can still be assessed.

Yet because of the difficulty involved, some educators have simply given up on the struggle, arguing that since they cannot really quantify the most important aspects of learning and teaching, they simply have to trust that good people, involved in a good process, will lead to good outcomes. And after all, there is plenty of evidence that we are not failing.

For these people, attempts to rigorously define expected learning outcomes, or attempts to methodologically measure and compare the effectiveness of the process or the resulting learning, can be interpreted as a lack of trust in the goodness or the professionalism of the people involved.

Just the opposite it true, however. It is because we are professionals that we need to clearly articulate what we envision as the purposes of our instruction. It is because we are professionals that we need to measure, assess, and compare our effectiveness where possible with others engaged in the same effort.

We do this because we care about real success — the learning of our students.

In this book, we will look at the reasons that systematic inquiry and assessment are critical for our success at BYU–Idaho. We will discuss the process of inquiry and assessment as a driver for ongoing development and improvement and we will look at different types of systematic inquiry and assessment.

REASONS AND PROCESSES OF SYSTEMATIC INQUIRY AND ASSESSMENT | Ongoing systematic inquiry has not always been emphasized in educational contexts, least of all in higher education. Because the tradition or culture of systematic inquiry is new to many, we review the underlying reasons both for defining outcomes and for measuring our progress against those outcomes. Then we look briefly at the history of educational assessment that has not always emphasized or welcomed data-driven processes, preferring to rely on tradition and anecdote instead. Lastly, we review a general process of inquiry and assessment that provides feedback on our effectiveness at various levels.

LEARNING FROM EXISTING SCHOLARSHIP | While we have much to learn about our effectiveness through reflection and data analysis of our work, much work has already been done to study what is effective versus what is only marginally effective in promoting student learning. Much of this research has been compiled into books that can introduce us to the scholarship of learning and teaching. A book list is introduced as a starting point for faculty members interested in understanding best practices.

CLASSROOM RESEARCH | Classroom research is a type of research conducted by professional educators on their own practice. They apply the rigor and methodology of scholarly research to explore ways to improve student learning in their own classrooms. Classroom Research explores ways to apply the literature of best practices to our own instructional experiences.

PROGRAM LEVEL ASSESSMENT | While inquiry and assessment of our work is important on the course level, it is possibly even more important on the program level. After all, students come to pursue a given curriculum for their major, not just to take a disjointed selection of independent courses. Program level assessment helps us integrate the many courses in our programs into a coherent student experience and aids us in assuring that the learning outcomes are met for the students we serve.

reasons for and processes of

SYSTEMATIC

INQUIRY & ASSESSMENT

THE NEED TO DEFINE SUCCESS (OUTCOMES)

This is meant to be a working document, so feel free to write your own thoughts, questions and ideas within the sidebars, or consider the questions provided.

If I am walking with my son in the woods, and we stop and sit on a log for a talk, I do not need to have thought out ahead of time what the talk will be about. In fact, I might destroy the moment as well as the learning and teaching opportunity if I try to over-structure the interaction. In such a setting, mostly what I do is listen. I listen to him and to the Spirit to know what to say and how to teach him.

If I am part of a formal, organized attempt to educate large numbers of people, however, I cannot avoid defining expected learning outcomes. I only have so much time with the students who only have so much energy for my class. Besides that, the discussion is no longer just between me and the student, but between me, the student, the teachers of subsequent classes the student might take, the standards of the discipline, and the future employer or graduate school. I am constrained to develop very clear ideas about what the students need to achieve during their time in the courses I teach.

There is simply no way to build a successful, large scale educational endeavor based on the walking-through-the-woods-and-sitting-on-a-log model of teaching. I cannot say, “Oh, the students will have the chance to be with me and some good materials for some 120 hours over the course of the semester. I do not know where the discussion might lead or what they might learn exactly, but I am sure that it will be a fabulous education.”

While such a situation might indeed provide a powerful learning experience for a few people, some of the time, it is simply not a workable solution for educating large numbers of diverse people, over an extended period of time, with multiple teachers or classes and the need for specific, consistent program level results.

Even as a dad in the woods with my boy, I have pre-determined things that I would like to see happen in his life. I have given previous thought to what lessons he might need. I still listen to him and I still seek the input of the Spirit, but that does not mean that I do not have ideas about the kinds of lessons I think he ought to learn. As professional instructors, there is simply no way to avoid having to define the purposes and goals of our educational endeavors.

DEFINING

THE GOALS OF OUR WORK

is necessary for success

Even if I say, "I just want my students to discover what they need and want to learn," I have determined the kind of learning I hope to achieve with my teaching. Defining outcomes is unavoidable.

But that is not a bad thing. President Henry B. Eyring, in the September 2004 *Ensign*, said:

> *The place to begin is with our aim, our vision of what we seek in the lives of our young people. … we must raise our sights.*

So in our attempt to give our students a quality educational experience, we have to begin by determining what a quality experience is, what it looks like, and how we know. We then need to measure our progress against those outcomes.

TO PONDER:

"Would you tell me, please, which way I ought to go from here?"

"That depends a good deal on where you want to get to," said the Cat.

■ *Alice's Adventures in Wonderland*, by Lewis Carroll, Chapter 6.

Seeing bleeding in an historical context in not very troubling.

What would be the reaction, however, if we saw a modern doctor using these traditional techniques?

THE NEED TO GATHER DATA

For years, doctors bled patients based on a metaphorical understanding of human health derived from Hippocrates' theory of the four humors. An imbalance in the humors or an overabundance of the sanguine indicated the need for bleeding. This practice that originated with the ancient Greeks persisted into the late nineteenth century.

Two thousand years of practice, anecdotal evidence, and the personal experience of physicians established the "effectiveness" of the practice. In general medical practice, bloodletting was the principal form of treatment for over a hundred different illnesses and ailments.

Only when medicine as a field rejected the anecdotal and the metaphorical as a basis for its underlying theory did the practice of bloodletting fall into disuse. Instead, the theory of medicine came to be based in statistical empiricism applied to clinical observations. Statistics replaced tradition, and metaphor was ousted by research. Medicine is better as a result.

What then do we make of the following comments by Grant Wiggins, a leading thinker and commentator on education in the last decades?

> *I firmly believe that education is currently where medicine was in the late 18th century – not yet but almost a science of best practice; we're still in the village-barber, old-wives-tale phase of teaching.*

> *Can't we be more clear and less loosey-goosey about just what is and isn't negotiable in instruction, given the stated goals and what they logically demand of the use of class time and the learners' minds?*

For centuries, much education was based on metaphorical understandings of learning and anecdotal evidence gleaned from years of personal experience. Only recently has the need to change the basis of educational decision-making met with the research methods and data which would allow us to do so.

We now have almost a century's worth of statistical information about what works powerfully and what works only marginally. Additionally, the last two decades have given us far greater insight into the workings of the brain than ever before.

It is time that we as educators embrace — at least in part — a data-based understanding of our profession. It is time that we participate in reviewing and producing such data. In short, it is time that our profession become rooted not only in metaphor or personal experience, but in systematic inquiry into the real basis of learning and its implications for teaching.

Education will be better as a result.

It is true that education is not a science if science means a technology where one has merely to turn the crank to achieve results. Neither is it an art, however, if by art one means merely the personal expression of the educator's individual creativity. Rather, it is a craft which is to be learned and mastered. It is a craft which should be guided and directed by what is known about best practices.

Again, Grant Wiggins is insightful:

> *In few professions are novices allowed to free-lance. No doctor or electrician can blithely invent basic technique or simply decide not to use by-the-book solutions to diagnoses or problems. In fact, in medical education . . . few doctors would deviate from prescribed responses to common ailments unless those prescribed approaches failed to work. Why should teaching be any different?*

His final question hangs heavy. Why indeed should teaching be any different? Shouldn't we have a good grasp on what works well most of the time before we begin inventing or improvising?

Yet knowing what works well requires us to collect data about our relative success as measured against our purposes and expected outcomes. It requires practicing our profession in the context of ongoing systematic inquiry into our effectiveness.

TO PONDER:

Where performance is measured, performance improves. Where performance is measured and reported, the rate of improvement accelerates.

■ Thomas S. Monson

THE PROCESS OF SYSTEMATIC INQUIRY

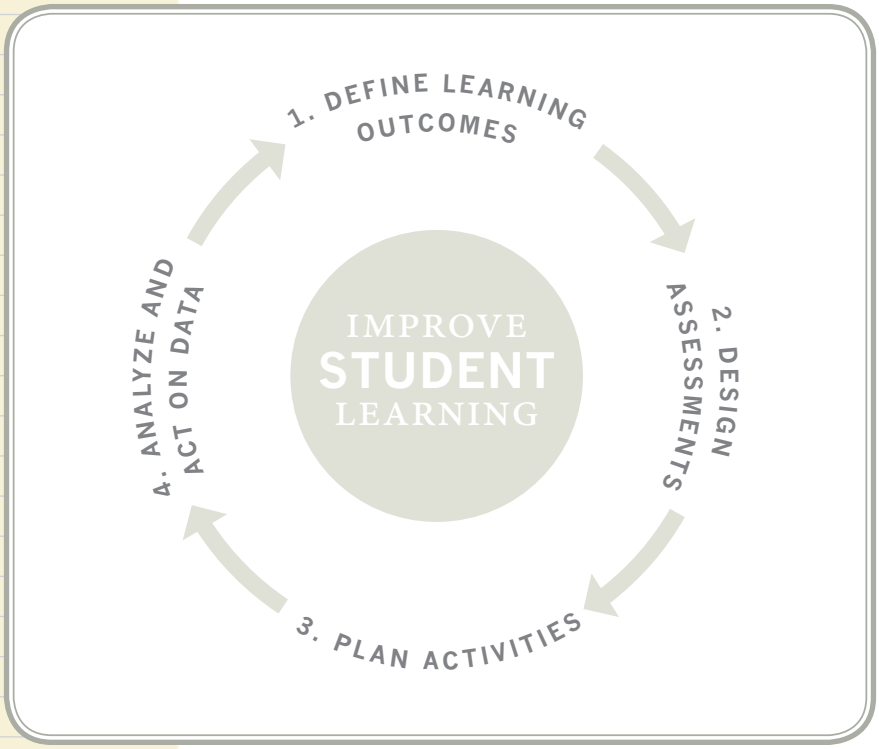
TO PONDER:

Prove all things; hold fast to that which is good.

■ 1 Thessalonians 5:21

We can collect data, observations, and evidence that tell us whether what we are currently doing is meeting our desired outcomes or not. If we keep our goals firmly in mind, and assess against them regularly, we will know how we are doing and have a good chance of improving over time.

If we slip at all, either in defining our outcomes or in measuring against them we are flying blind to who knows where. While there will always be room for inspiration, individuality and flexibility in our work, we cannot ignore the value of data in making important decisions about how to best accomplish our educational goals. The cycle for ensuring quality at every level of the university therefore looks something like the graphic below.



This process forces us to reconsider not only what we do, but to reassess the very structures of the system that defines so much of what we do. As much as we need to assess student performance, we need to use student performance to assess ourselves and our programs.

we assess our students' work, then

USE THEIR WORK TO ASSESS

OUR INSTRUCTION

THREE TYPES OF SYSTEMATIC INQUIRY

There are basically three ways in which we can engage in systematic inquiry and assessment.

LEARNING FROM EXISTING SCHOLARSHIP | The first way in which we engage in systematic inquiry is to recognize that there is scholarship in the field of instruction, instructional design and human learning to which we can compare and against which we can measure our practice. Becoming acquainted with best practices as defined in the literature is one aspect of what is commonly called the Scholarship of Learning and Teaching (SOLT).

Former BYU–Idaho President Bednar, in an August 2001 faculty meeting, clarified:

> *We should be excellent scholars; and our scholarship should be focused on the processes of learning and teaching.*

TO PONDER:

The concept of staff development is changing from what was once conceived as an externally organized, formal program, toward any process in which a teacher (working individually or in a group) systematically attempts to understand herself, her students, the school context, and/or new practices in order to improve her teaching. The process may be initiated and conducted by the teacher(s) with or without the help of an external facilitator. Within this newer conception, teacher inquiry or action research may be though of as professional staff development.

■ Virginia Richardson, "*Teacher Inquiry as Professional Staff Development*," NSSE No. 93, Pt. 1, 186

YOUR THOUGHTS:

What kind of systematic inquiry and assessment do you currently do? What would be the first step in improving your systematic inquiry and assessment?

CLASSROOM RESEARCH | The second way in which we can engage in systematic inquiry is by embracing the idea that as professionals we need to consistently collect data about our own performance and its effects on student learning. We elicit this data from students and peers. We engage in Classroom Research and modify our teaching activities according to our findings.

Classroom Research is research that we undertake in our own courses to determine how what we do effects the student learning. It is a research that uses formal and rigorous scholarly methodologies to answer questions about our classroom effectiveness. It tends towards the practical rather than the theoretical and guides improvement efforts on the ground.

Such research is at the basis of the vision articulated by John Hattie, which asserts that "the most powerful way of thinking about a teacher's role is for teachers to see themselves as evaluators of their effects on students." Classroom research gives us data to help evaluate the effect we have. Understanding our effect might mean brushing up on some basic tools in descriptive statistics which are introduced here.

PROGRAM LEVEL ASSESSMENT | Systematic inquiry also requires us to ask about effectiveness at the program level. We need to answer the question, "Do our graduates know what they need to know and can they do what they need to do?" That question is not sufficiently answered by looking at their GPA or their transcript, but it can be answered with the right research design and data collection.

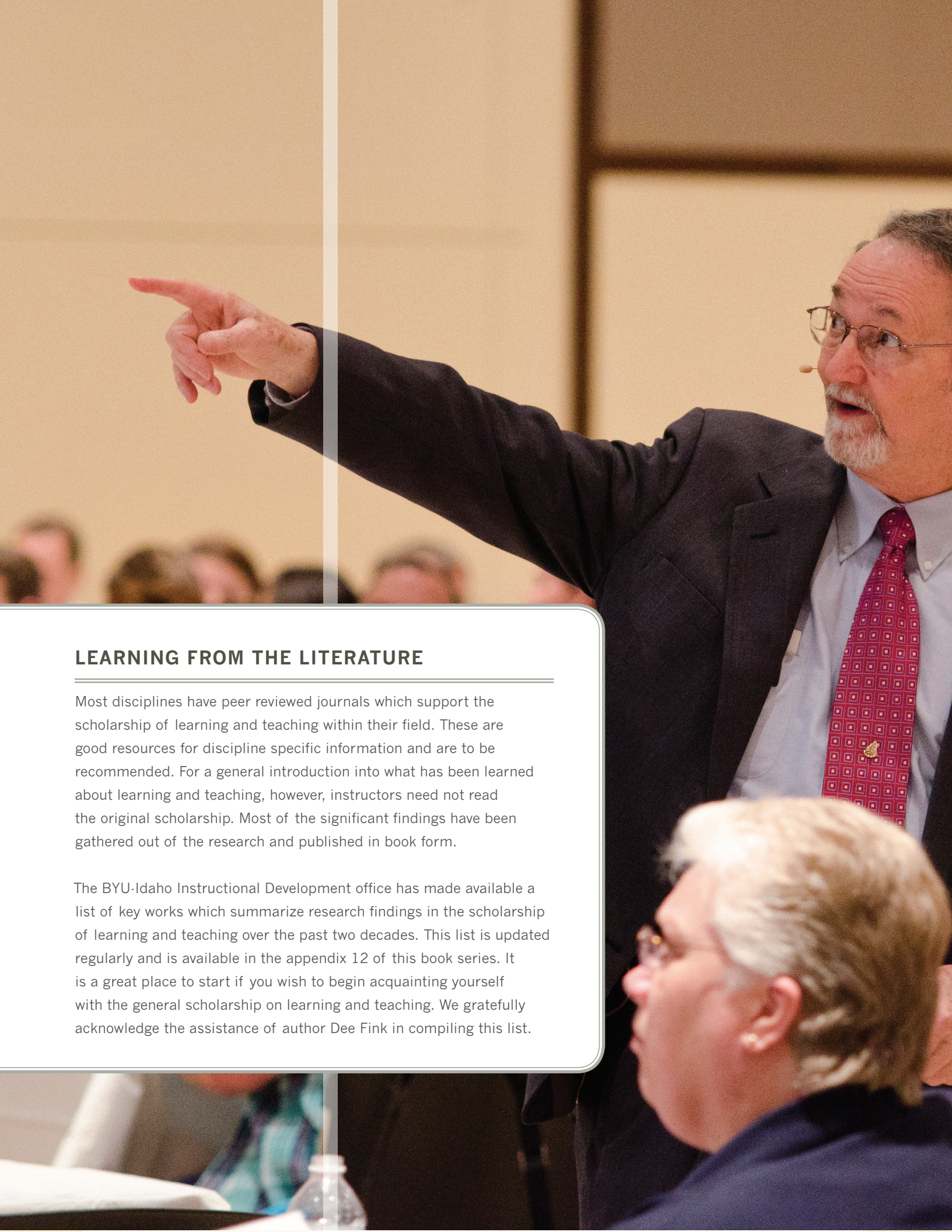
We need systems in place that gather data about our course outcomes, our programs, the integration of our curriculum, the development of students from matriculation to graduation and the cumulative learning attributed to our instructional efforts.

The remainder of this book will take a closer look at these three types of systematic inquiry and assessment.

learning from existing

SCHOLARSHIP

AS SYSTEMATIC INQUIRY



LEARNING FROM THE LITERATURE

Most disciplines have peer reviewed journals which support the scholarship of learning and teaching within their field. These are good resources for discipline specific information and are to be recommended. For a general introduction into what has been learned about learning and teaching, however, instructors need not read the original scholarship. Most of the significant findings have been gathered out of the research and published in book form.

The BYU-Idaho Instructional Development office has made available a list of key works which summarize research findings in the scholarship of learning and teaching over the past two decades. This list is updated regularly and is available in the appendix 12 of this book series. It is a great place to start if you wish to begin acquainting yourself with the general scholarship on learning and teaching. We gratefully acknowledge the assistance of author Dee Fink in compiling this list.

collecting data through

CLASSROOM RESEARCH

AS SYSTEMATIC INQUIRY

TO PONDER:

Acquiescence is not a characteristic of an action researcher. He is resourceful, committed, tenacious, and above all, curious. He will not be satisfied with a given system if he sees elements of the system as unsatisfactory. He will seek to change it. In doing so, he refuses to be a servant, but becomes an acting agent. He rises above the role of a skilled technician and becomes an educator.

- Jean McNiff, *Action Research, Principles and Practice*, McNiff, 1988, 50

CLASSROOM RESEARCH

Every discipline, course, and instructor is different, so sometimes other people's scholarship of learning and teaching does not provide the answers we need in our own teaching. In those cases, we can engage in our own scholarship of learning and teaching by doing what is called classroom research. Such research can help us apply and adapt research on best practices to our own unique situations.

Classroom research applies the methods of traditional scholarly research to the questions of instruction. It begins with a literature review and the formal articulation of a research question. This is followed by systematic data collection and interpretation. Based on the results of the data collection and interpretation, a plan of action is formulated, implemented and evaluated. The research cycle can then begin again even as the results are shared with others.

The purpose here is to get data on how we are doing and how the design of our course is being received, make changes based on the data, and measure for improvement. Student scores can give us some insight when doing this type of research, but there are many other ways as well to take the pulse and measure the health of our instructional efforts. Following are some ways that faculty commonly use to determine the success of their teaching in the context of action research.

PEER REVIEW | Peers in our discipline are uniquely positioned to review our syllabi, comment on our lesson plans, and help us think through our intended outcomes. We need to welcome and use this rich source of data and insight into our instructional efforts.

PRE- AND POST-TESTING | If we start with an assumed baseline of zero, any learning that the students demonstrate can be attributed (some-times falsely) to our instructional efforts. Giving a pre-test allows us to set a better baseline to compare against. Using the same test for pre and post class measurements, allows us to compare apples to apples in terms of what students gained from the course.

KNOWLEDGE SURVEYS | An alternative to traditional pre and post tests are knowledge surveys, where students simply indicate their confidence level around answering certain questions rather than actually answering them. Students complete knowledge surveys at the beginning, middle and end of a course, generating data about their perceived ability to answer question relevant to the course content and outcomes. Research comparing responses from knowledge surveys, examination results, and final course grades suggests that knowledge surveys provide meaningful measures of learning gains.

YOUR THOUGHTS:

What aspects of your own teaching would you like to better understand? What kind of action research might help you pursue that understanding?

seek resources
TO HELP YOU CONDUCT
CLASSROOM RESEARCH

USING NATIONAL, NORMED TESTS | Many disciplines have national, normed concept inventories or other standardized tests for students in their field. These tests are useful both because the questions are vetted for statistical validity and reliability, and because a large body of student performance data already exists for them.

USING MID-SEMESTER SURVEYS | Many faculty find that the simplest way to get assessment data on student learning is to simply ask the students. Students are not qualified to tell you everything you need to know about your course, but they are uniquely qualified to tell you if they are learning and give insights into why or why not. There is little or no research indicating that end of semester course evaluations aid in the formative process of improving courses or instruction. There is significant research, however, indicating that moving the feedback mechanism to the middle of the semester, when there is still time to act on the insight generated, can significantly improve student learning.

STUDENT FEEDBACK TEAMS | Student feedback teams are a process for involving students in improving the course in an ongoing way. A small group of selected students meets weekly outside of class to discuss the course experience and articulate feedback for the instructor. The instructor meets with the team on alternate weeks to discuss the student insight although they are in no way bound to act on every student suggestion. Research into this process has documented that students and faculty working together in this way arrive at about the same ideas for improvement as did a professional developer working with the faculty member.

STRUCTURED FOCUS GROUPS | It is generally accepted that properly structured group processes produce better results than individuals working alone. Structured focus groups allow us to derive feedback data about student learning not simply by aggregating individual data, but by working with the course participants *as a group* to generate valid and valuable data about student learning.

START/STOP/CONTINUE FEEDBACK | Surveys can take a lot of time, and you would not want to do them too often, but a quick method for eliciting student feedback is just to hand them a 3 x 5 card and ask them to write three things:

- *What should we start doing that would help your learning?*
- *What would help your learning if we stopped doing it?*
- *What should we continue that is helpful to your learning?*

OFFICE OF INSTRUCTIONAL DEVELOPMENT | Our office has many tools and resources to help faculty collect data on their instruction, including resources on conducting and interpreting formal action research using direct measures of student learning.

SCOTS | Specially trained student consultants can give various kinds of feedback on a course from a student perspective, or collect data from participants in the course. These Student Consults on Teaching, or SCOTs also help in doing student focus groups, videotaping classroom procedures or any of a number of feedback and data collection options.

Sample comments by BYU–Idaho faculty members:

"The SCOT program allowed me to gather data about my teaching from multiple perspectives. I was able to work with my consultant in designing questions and activities to give me specific feedback and insights into my teaching. The SCOT program helped me rethink my teaching and better apply the learning model."

"I loved having myself filmed. It's an eye-opener! This program is wonderful. It's a great hands-on tool rather than some vague teaching idea."

STATISTICS FOR CLASSROOM RESEARCH

YOUR THOUGHTS:

How might getting statistical data about our work change our experience as teachers?

Original educational research often requires skill in inferential statistics, where the findings relative to a small population of students are viewed as representative of students generally. However, for most educators who engage in classroom research simple descriptive statistics are sufficient. Descriptive statistics are used to accurately describe a data set and changes or correlations within that data set, without trying to claim that the data is representative.

As a teacher begins to think about the role of data in his or her teaching and efforts to improve, a few tools from descriptive statistics become very useful. Among these are the calculation of effect size, normalized gain, the T-test, ANOVA or analysis of variance, T-test, and the Chi squared tests.

Both effect size and normalized gain are ways to interpret data from a pre- and post- test.

EFFECT SIZE | The effect size takes the difference between the post- and the pre-test, and divides that number by the standard deviation. This simply means that the average gain on a test (post test score minus pre-test score) is adjusted to account for whether the spread in scores was wide or narrow (standard deviation.)

$$\text{Effect Size} = \frac{\text{Average (post-test)} - \text{Average (pre-test)}}{\text{Spread (average standard deviation of pre-test)}}$$

NORMALIZED GAIN | Normalized gain starts with the average gain on the assessment (difference between the post-test and the pre-test scores), but then divides by the average possible increase available. In other words, the denominator in this calculation is the total possible on the exam minus what students already knew coming into the exam.

$$\text{Normalized Gain} = \frac{\text{Average (post-test)} - \text{Average (pre-test)}}{100\% - \text{average pre-test expressed as \%}}$$

T-TEST | This test is used to compare average values between two groups. It might be used, for example, to explore success rates in a class between men and women, or between RM's and non-RMs, or between any other two groups.

ANOVA | The analysis of variance is simply a way to look at differences between variables in a single set of data or across multiple data sets. It is a generalization of the T-test to more than two groups.
A one-way ANOVA, for example might look at participation scores between three or more sections of a course and allow the teacher to quantify and compare them. A multivariate ANOVA is an analysis of variance that looks at more than one variable. For example, a two-way (or two variable) ANOVA might look at the role that rest and exercise have in student performance. Performing a multivariate analysis allows the teacher to look at correlations for each variable independently (rest and exercise) or in combination with one another.

MULTIPLE REGRESSION | This test allows, for example, a teacher to look at the scores for each unit test and determine if they are correlated with the scores on the final examination. Using this technique, a teacher might find that one or more of the unit tests account for a large portion of the variance in the final while a third unit test correlates not at all with scores on the final.

CHI-SQUARED GOODNESS OF FIT | This test is appropriate when a teacher wants to compare his or her data set to a hypothesized standard distribution. For example, an instructor may wish to compare student scores to national averages on a standardized test. This test allows them to compare not only the average score, but the spread of the scores to the national spread and average.

CHI-SQUARED TEST FOR INDEPENDENCE | This is used to explore relationship between two categorical variables. A teacher may wish to know, for example, if students who have completed the pre-requisite course actually do better in the subsequent course.

TO PONDER:

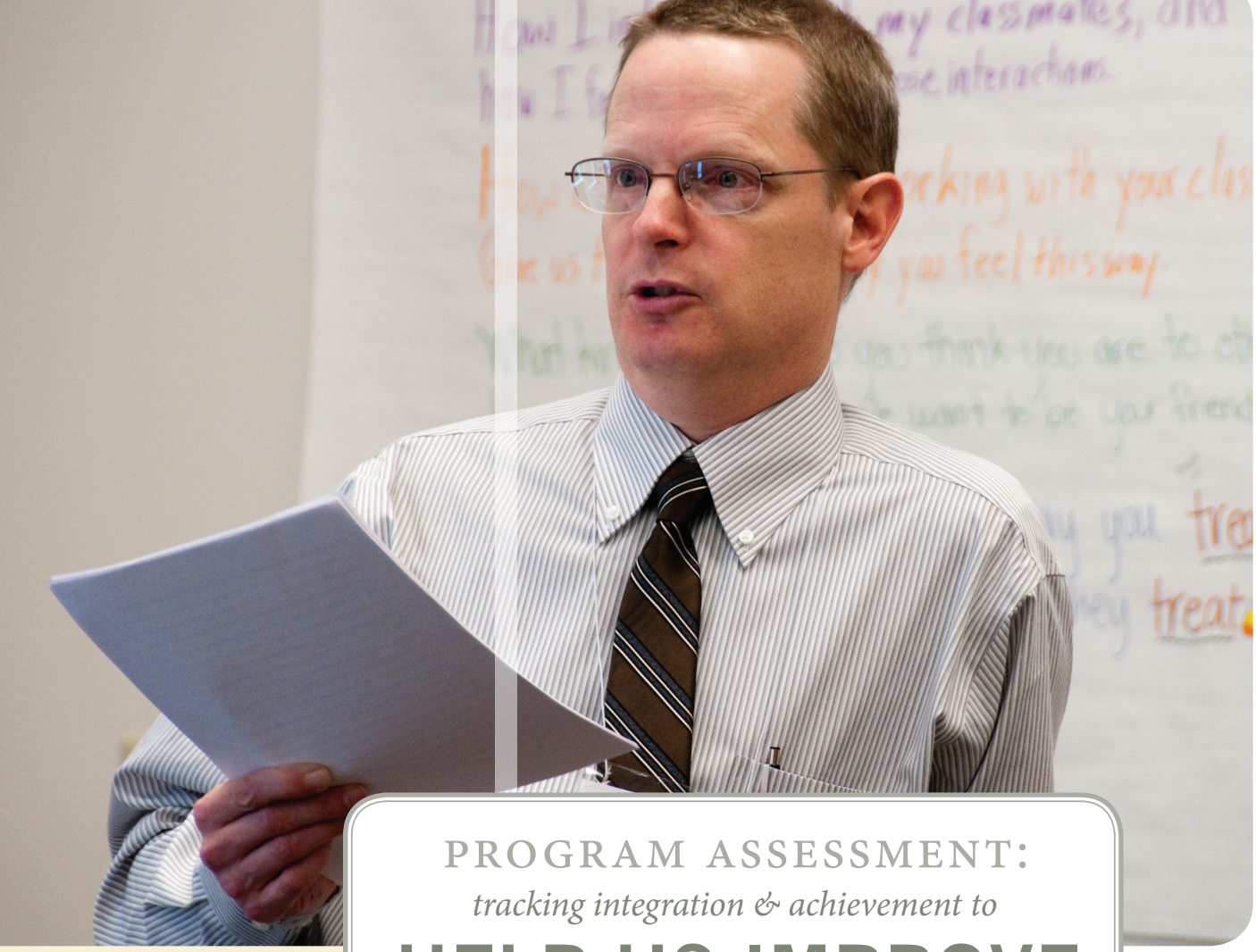
Cognitive psychology tells us that the unaided human mind is vulnerable to many fallacies and illusions because of its reliance on its memory for vivid anecdotes rather than systematic statistics.

■ Steven Pinker.
2011 *Observer* Interview

program level

ASSESSMENT

FOR SYSTEMATIC INQUIRY



PROGRAM ASSESSMENT:
tracking integration & achievement to
HELP US IMPROVE

DESIGNING PROGRAM LEVEL ASSESSMENT

Program level assessment helps ensure two things, first that all the parts of a program add up to the education we want for our students, and second, that we are continually trying to improve that program based on feedback.

All too often, students experience their education not as an integrated and aligned experience preparing them to know, do, and become, but as a loosely-coupled series of courses taken to fulfill graduation requirements. Many universities – even teaching institutions – define their mission not in terms of providing an integrated and aligned educational experience for their students to know, do and become, but in terms of acting as knowledge reservoirs for the disciplines and the sub-specialties.

By aligning what we do with what we hope to achieve for our students, and by consistently measuring our progress, program level assessment improves these situations, helping students experience their education as coherent progress toward specific learning outcomes, and helping universities define themselves in terms of producing high levels of student learning as well as conserving and expanding disciplinary knowledge.

Program level assessment follows the same basic pattern of systematic inquiry presented earlier in this book, with a few modifications to accommodate the added complexity of dealing with the multiple courses within a program.

The steps for an assessment process at the program level are as follows:

1. ***Define program level outcomes***
   a. *Define more specific sub-outcomes as needed*
   b. *Map those outcomes to the courses and experiences of the program using a process called curriculum mapping*

2. ***Design Assessments***
   a. *Determine from the curriculum map appropriate data capture sources which will be used to assess each outcome*
   b. *Where appropriate, create rubrics to assess the data in terms of the program outcomes*

3. ***Plan activities***
   a. *Establish an assessment schedule which articulates when data will be collected and the cycle of outcomes assessed*

4. ***Analyze and act on data***
   a. *Establish a procedure for reviewing and interpreting assessment data and determining changes to be made in the program*
   b. *Revise program curriculum and / or instruction based on findings*

Programs wishing to initiate or review their program level assessment process have various institutional resources at their disposal to help in the process. The University's Outcomes and Assessment website, can be found at: *www.byui.edu/outcomes-and-assessment*

TO PONDER:

The quality of an assessment plan does not depend on its complexity or comprehensiveness. Quality is, rather, a matter of the plan's ability to provide the faculty who can act on it evidence regarding who is and is not learning.

Good assessment is valid and reliable but above all, good assessment is useful.

■ BYU–Idaho Outcomes and Assessment Website

STEP 1a: DEFINING OUTCOMES

Program level outcomes are high-level outcomes. They indicate the knowledge, attitudes, values and skills that students will take with them upon graduation and will retain several years after graduation. A typical program will have between six and eight program level outcomes. Some programs like nursing or education that have external accreditors and examinations may well have more outcomes than this that they need to track.

Articulating program level outcomes is a collaborative effort that should include all of the full-time instructors in a program and may well include adjunct instructors, students and advisory bodies as well. For help with the specific process of defining program level outcomes, see the outcomes portion of Book Four in this series.

As seen in the example on the facing page, it may also prove useful to define sub-outcomes to the key program outcomes.

CONTINUED RESEARCH:

Among the most referenced works on outcomes and assessment currently at BYU–Idaho are:

- *Assessment Clear and Simple: A Practical Guide for Institutions, Departments, and General Education* by Barbara E. Walvoord
- *Feedback for Learning* by Grant Wiggins

STEP 1b: CURRICULUM MAPPING

A curriculum map is simply a matrix with the various program level outcomes listed on the X axis and the courses and required experiences of the program listed on the Y axis. The purpose is to establish and locate those aspects of the curriculum that either introduce, reinforce or assess the outcomes of the program. These points are identified in the matrix with an I, R, a, or A respectively. The 'A' indicates a data point that has been selected for use in program level assessment processes.

When used in designing program level assessment, the purpose of the map is not so much to align all the course outcomes with the program outcomes, as it is to identify the student experiences and assignments that lead to the program level outcomes. Using the curriculum map helps programs move from an input model of assessment (faculty with terminal degree, number of credit hours required, average GPA of students accepted, etc.) to an output model of assessment. For example, student writing is assessed not by reporting the presence of a required writing course in the discipline, but by collecting actual student writing assignments from identified courses and subjecting them to analysis using a rubric developed for that purpose.

I INTRODUCED
R REINFORCED
A ASSESSED

PHYSICS DEPARTMENT OUTCOMES

PROGRAM OUTCOMES:

1) Understand fundamental concepts and principles of physics
2) Apply knowledge and understanding of physics
3) Communicate and share knowledge and experiences in physics
4) Work well in teams and/or groups
5) Be prepared for internships, jobs in industry, and graduate school
6) Need and use of scientific and spiritual reasoning

Other Skills

DETAIL OUTCOMES:	121	123	150	220	250	270	291	314	328	332	333	336	385	388B	408	412	433	473	488	Elect.	FCI	CSEM	MBE	EBAPS	Lawson	ETS	GRE	Thesis
Classical mechanics	I		R	R		R	R	R		R		R	R				R	R		R	A					A	A	
Electromagnetism				I	R	R	R				R	R	R			R				R		A				A	A	
Quantum Mechanics						I		R				R	R			R	R	R		R						A	A	
Composition of Matter				R	R	R	R			R	R	R	R			R	R	R		R			A			A	A	
Conservation Laws	I	R		R		R	R			R	R	R	R			R	R	R		R						A	A	
Experimentation	I	R	I		R						R	R		A	A				R									
Ask questions	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R	R	R				A				A
Define problems	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R	R	R								A
Predict outcomes	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R	R	R								A
Evaluate solutions	I	R	R	R	R	R	R			R	R	R	R	A	A	R	R	R	R	R	A	A		A	A		A	A
Construct models	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R	R	R	A				A		A	A
Written			I	R	R	R		R	R	R	R	R	R	A	A	R	R	R	R	R								A
Oral presentation	I		R		R	R		R	R		R	R	R	A	A	R	R	R	R									A
Poster		I							I R					A	A													A
Conference attendance									I					A	A													
Networking									I					A	A													
In-class groups	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R		R								
Out-of-class groups	I	R	R	R	R	R	R	R		R	R	R	R	A	A	R	R	R		R								
Group projects			I		R			R				R	R	A	A													
Mentoring									I					A	A				R									
Identify areas of interest									I					A	A				R	R								A
Identify opportunities									I			R	R	A	A				R	R								
Required preparation									I			R	R	A	A					R	A	A				A	A	
Required skill set									I			R	R	A	A					R								A
Conduct research									R					A	A	R			R									A
Professionalism	I	R	R	R	R				R					A					A									A
Understand scientific reasoning	I	R	R	R	R	R	R	R	R	R	R	R	R	A	A	R	R	R	R	R	A	A		A	A		A	A
Understand spiritual reasoning								I						A	A													
Recognize limitations								I						A	A									A				
Know when & how to apply both								I						A	A													
Computer programming	I		R		R					R	R	R	R A	A	A	R												
Data analysis	I	R	R	R	R	R	R			R	R	R	R	A	A	R	R	R	R	R								A
Mathematical	I	R	R	R	R	R	R			R	R	R	R	A	A	R	R	R	R	R							A	A
Electronics			I		I						R	R		A	A													

STEP 2a: CHOOSING DATA CAPTURE POINTS

While it is certainly possible and desirable to use third party assessments or data generated specifically for program assessment, there is something both elegant and practical about using data that is already being generated in the program's courses. The trick is to find assessment data that already speaks to the program outcomes, or assessment artifacts that could be subjected to a program-level rubric. For example, it might be appropriate to take the mean and median of the final exam from a class that correlates closely with a program outcome. Writing samples collected in the form of internship reports, however, would need to be evaluated according to a writing rubric in order to render data useful for the writing outcome in the program assessment. Simply taking the scores from those papers would not be appropriate because they have likely been graded on criteria other than just writing.

TO PONDER:

Most institutions have routinized data collection, but they have little experience in reviewing and making sense of data. It is far easier to sign up for a survey offered by an outside entity or to have an associate dean interview exiting students than to orchestrate a series of complex conversations with different groups on campus about what the findings from these data mean and what actions might follow.

- Blaich and Wise, *From Gathering to Using Assessment Results: Lessons from the Wabash Study.* National Institute for Learning Outcomes Assessment. Jan 2011

STEP 2b: DESIGNING AND USING RUBRICS

CRITERIA FOR A GOOD RUBRIC:

- *Valid – characterizes skilled performance*
- *Authentic – approximates real-world situations*
- *Generalizable – applies to other similar tasks, not just this one*
- *Appropriate – describes a reasonable range of examined ability*
- *Understandable – defines criteria clearly enough for students to understand and use*
- *Useful – explains weaknesses and suggests how to improve future performance*

Once we have defined the learning outcomes, mapped the curriculum, and identified artifacts that could be used to assess the outcomes, we need to determine how to analyze and interpret the artifacts to produce useful and consistent data. An important tool for this step is the rubric.

A rubric is a set of criteria used for judging performance quality (e.g., from poor to exceptional). Rubrics are an effective method for scoring student work, allowing the chance to clarify expectations and providing a standard for feedback. They are particularly useful in program level assessment as they help standardize both the criteria and the scoring of acceptable levels of performance.

Rubrics provide a framework for identifying and judging the quality or performance of virtually any behavior that is observable—from balancing a check book to balancing a chemical equation, from explaining the causes of a labor dispute to selecting good spud seeds, from wiring an electrical circuit to writing a persuasive essay. Artifacts that do not already provide data in numerical terms will need to be interpreted using a rubric.

To develop a rubric:

1. *Make a list of what excellent performance looks like (consider the process, result, or both)*
2. *Gather, recall, or imagine examples with a broad range of quality*
3. *Sort examples as good, unacceptable, or borderline and subdivide large categories*
4. *Determine what is unique about each category*
5. *Organize statements of defining characteristics into the rubric; associate points to each descriptor of each criteria*
6. *Work with colleagues to review the rubric for overlooked or overemphasized criteria*
7. *Evaluate the rubric in terms of its ability to help focus on important criteria and its ability to standardize assessment of student work according to the outcome statement*

RUBRIC EXAMPLE

	DISTINGUISHED	PROFICIENT	DEVELOPING	UNSATISFACTORY
Creativity & Communication *40 Points*	31-40 POINTS: *The logo is creative and engages the imagination. Symbols clearly connect to the product or exhibit. Symbols are combined in a clever and interesting way.*	21-30 POINTS: *The logo design works, but is expected. Symbols connect to the exhibit or product but also may be confused for other products or services.*	11-20 POINTS: *Lack of research and exploration are evident in the logo. Communication is not clear. Symbols are ambiguous. Symbols are not combined well and*	0-10 POINTS: *Logo is obvious or derivative. Shows little evidence of original thought or effort.*
Principles Of Design *40 Points*	31-40 POINTS: *Logo is well designed. The individual elements work well together. The principles of design are applied correctly. Line quality is*	21-30 POINTS: *Logo is overall good but has some design flaws and could be improved.*	11-20 POINTS: *Logo has serious design flaws in value, line quality, image structure, or any of the design principles.*	0-10 POINTS: *Logo shows little evidence of the application of design principles.*
Craftsmanship & Following Directions *20 Points*	16-20 POINTS: *Overall good craftsmanship. Correct size (10 inches wide x 8 inches tall), good mounting (no spray mount where it shouldn't be), clean edges, good placement of logo on board.*	11-15 POINTS: *Minor problems with craftsmanship and/or following directions.*	6-10 POINTS: *Major problems with craftsmanship or following directions.*	0-5 POINTS: *Lack of craftsmanship is sufficient to seriously detract from the logo itself. Differs significantly from instructions.*

The graphic above shows one way of organizing a rubric. Once the general lay-out of the rubric has been decided, the cells within the rubric are filled in with the characteristics of each performance outcome at each level. Points may be assigned to each cell within the matrix to help standardize the use of the rubric across courses and between multiple assessors.

The three criteria shown in this example are only for the purposes of demonstration. An actual rubric will have as many criteria as are necessary and practical in assessing a given outcome.

TO PONDER:

When used correctly, [rubrics] serve the purposes of learning as well as of evaluation and accountability. Like portfolios, exhibitions, and other authentic approaches to assessment, rubrics blur the distinction between instruction and assessment.

- Heidi Goodrich Andrade "*Using Rubrics to Promote Thinking and Learning*" in What Do We Mean By Results. Feb. 2000 Vol. 57, No 5, pp 13-18

STEP 3a: DESIGNING AN ASSESSMENT CYCLE

Most programs will not assess every outcome every year. Rather, in most cases, a program should establish a three year cycle where a portion of the outcomes are assessed each year. A typical schedule would see a program collect data throughout the academic year and then analyze it in department meetings prior to the beginning of the new academic year.

In most cases, it would be appropriate for BYU–Idaho programs to conduct their assessment before the seven-week summer break. This aligns the program assessment process with the academic calendar and gives sufficient time before the beginning of a new academic year to work changes into the curriculum and instructional plans and to generate reports needed for the curriculum council, program assessment or accreditation.

THE END PURPOSE
of assessment needs to be
ACTION

If a program chooses to establish a cycle different from the one described above, variables like the following should be considered:

- *The availability of faculty members to conduct the assessment*
- *The relationship of the assessment to the beginning of a new semester where changes can be implemented based on the assessment findings*
- *The relationship of the assessment to the budget cycle in case the report indicates the need for budget resources in a certain area*

It is important to design an assessment cycle that balances the need for regular data and feedback with the extra work and effort involved in collecting, assessing and interpreting the data. In many cases, because programs simply do not have the time and resources to dedicate to an extensive assessment process, less is more.

STEP 4a: REVIEWING & INTERPRETING DATA

While data collection is important, it is only the first step in establishing a process of continual review and improvement. Reviewing and interpreting the data is critical. At least once a year, all the full-time faculty that teach in a program, and possibly the adjunct faculty as well, should meet together to review the data that has been collected and to discuss what it means for the program.

- *Can strengths be identified based on the data?*
- *Does the data give insight into areas that could be improved?*
- *Is the focus of the improvement efforts to be curricular or instructional?*

This meeting is also a good time to review the curriculum map in light of the findings from the assessment process.

- *Are there alignment issues between the courses that need to be addressed?*
- *Does the content covered in courses need to be adjusted?*
- *Are there ways that individual instructors could better coordinate to improve student outcomes?*

This meeting is more than data review and interpretation, however. It should result in a concrete and detailed plan of changes, either curricular or instructional, that will be implemented in the program to improve on the achievement of the program level outcomes. These changes are documented in the yearly program report. The next time that the same set of outcomes come up for assessment, this plan acts as a baseline for review.

- *Did the changes implemented have an effect on the outcomes of the program? Why or why not?*
- *Do the changes implemented account for the change in the data or are other variables at play that need to be addressed?*

STEP 4b: REVISING THE PROGRAM

Based on the most recent experience with the assessment plan, are there changes that need to be made to the outcomes, the data collection, the rubrics, the assessment cycle or other aspects of the program level assessment plan?

TO PONDER:

Assessment leaders should avoid doing presentations in which the data and conclusions are simply handed out to faculty. If faculty do not participate in making sense of and interpreting assessment evidence, they are much more likely to focus solely on finding fault with the conclusions than on considering ways that the evidence might be related to their teaching.

[T]alk about some patterns that [you] see in assessment data and then ask, "What do you think this means?" The goal in these conversations is not to accept just anything that people say in interpreting the data but to engage in a "Yes, that sounds reasonable but how is that consistent with what students say on these questions?" In many ways, good discussion about assessment data resembles a good seminar discussion about a book. People cite the text, in this case the data, and then dig in, push back, consider their own experience, and try to find broad themes.

■ Banta and Blaich, "Closing the Assessment Loop" Change Magazine. Jan · Feb 2011

BYU
IDAHO

SUMMARY

Avoiding educational failure is not the same thing as promoting educational success. To truly pursue success as professional educators, we need to clearly define the learning outcomes we hope to achieve from our efforts and we need to systematically gather data to help us understand if we are attaining our goals.

Systematic inquiry and assessment in a university setting is usually pursued in one of three ways. First, faculty need to become familiar with the research that has already been done on university learning and teaching and which has been used to define best practices in the field. Even the most veteran teacher should not neglect research based on thousands, even millions of data points only on the basis of their own experience. Next, faculty members can engage in classroom research, applying their skills as scholars to understanding and improving their own teaching. Lastly, inquiry and assessment at the program level offers the chance for our programs to review curricular alignment and track success in meeting program level objectives.

PRODUCED BY: THE ACADEMIC OFFICE — BYU–IDAHO