

Assessment of student learning: how the UNC system can become a national leader

Stephen R. Porter

Department of Leadership, Policy and Adult and Higher Education
North Carolina State University
srporter@ncsu.edu

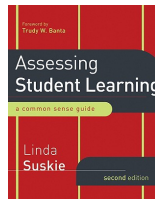
November 12, 2013

Should the UNC System be conducting assessment or research?

Upcraft and Schuh have noted that assessment differs from research in its purpose...

Traditional empirical research is conducted to test theories, while assessment is ... a distinct type of research whose purpose is to inform and improve one's own practice rather than make broad

generalizations. **Ewell has called this a craft-based rather than scientific approach.**



Who would Peter Ewell choose?



So what happens when ...

I make a mistake?

So what happens when ...

I make a mistake?

You make a mistake?

So what happens when ...

I make a mistake?

You make a mistake?

If earning a college degree is truly a life-altering experience, we should treat it as such when when doing assessment

Conspiracy theory

Institutions like the current approach to assessment

- Shoddy research does not show how the institution is really performing
- Can easily be ignored

Assessment practitioners like it as well

- Anyone can do it; no need for extensive training in research methods
- Craft-based approach means every technique is valuable; can't be criticized

How to become a national leader

Do something completely different:

Adopt a “scientific” approach to measuring student learning

What does this mean?

- Ignore anyone who says they do assessment
- Collect the right data
- Adopt the appropriate research design
- Understand that student motivation affects almost everything

You face three major challenges

- 1 Adopting a valid measure of student learning

You face three major challenges

- 1 Adopting a valid measure of student learning
- 2 Making sure each institution's outcome is representative of the institution
 - Students must be sampled randomly
 - Sampled students must participate in the assessment
 - Students must treat the assessment as a high-stakes exam

You face three major challenges

- 1 Adopting a valid measure of student learning
- 2 Making sure each institution's outcome is representative of the institution
 - Students must be sampled randomly
 - Sampled students must participate in the assessment
 - Students must treat the assessment as a high-stakes exam
- 3 Taking into account institutional differences when comparing institutions

Measuring student learning

Two main approaches

- Surveys that ask students how much they have learned (e.g., NSSE)
- Exam-based instruments that measure current level of ability

Exam-based instruments are the only credible approach

Measuring student learning

Four main players in this area

- Collegiate Assessment of Academic Proficiency (CAAP)
- Collegiate Learning Assessment (CLA)
- Critical Thinking Assessment Test (CAT)
- Proficiency Profile (PP)

All measure critical thinking; some also measure other areas, such as writing skills

Measuring student learning: content validity

CLA and CAAP have strong content validity, in terms of measuring critical thinking

- PP less so, because their test items do not force students to critically evaluate a long, complex argument
- CAT - unknown

Measuring student learning: construct validity

All four measures are highly correlated with one another

Growth over time

- Only CLA and CAAP (via the Wabash Study) have measured growth over time
- Both indicate about .40 SD increase after four years of college
- Studies show a similar effect size for gain in standardized test scores from grades 9 to 12

Measuring student learning: criterion validity

All four measures are correlated with GPA and SAT scores

Not much relationship with survey data, such as the NSSE (this is to be expected)

Major problem with the CAT

Unlike the other measures, faculty at each institution score the CAT

- Creates huge reliability issues
 - Are faculty within an institution, and across institutions, scoring the same way?
- K-12 tells us what happens when measures are used to judge institutions
 - Can we really expect faculty to score their own students fairly?
 - Easy to game value-added: low scores at entry, high scores at exit

Recommendations

- 1 Use either the CLA or the CAAP

Representing institutions: sampling

Sampling students will be a challenge

- Unfortunately, the CLA and CAAP allow institutions to determine who will be sampled
- This is like allowing principals to determine which of their students will undergo standardized testing

Course-based sampling

- Sounds like a good idea - easy administration with captive audience
- Guarantees that students who skip class will not be measured

Simple random sample

- Single best way to ensure representativeness, but ...

Representing institutions: sampling

We know that survey response rates vary dramatically across institutions in predictable ways

- Predictors of nonresponse are likely correlated with student learning
- One worry is that Institution A and Institution B both randomly sample, but participants differ in ways correlated with learning

So there are two issues here:

- 1 Ensuring that institutions draw representative samples of students
- 2 Motivating students to participate in the assessments

Representing institutions: student non-participation

You can do one of two things:

Nothing

- Just accept the fact that student participants may differ across institutions
- A value-added approach will help here
 - Idea is that by measuring change over time, you are controlling for differences between institutions
 - Some of these differences will be why students choose to participate in assessments

Representing institutions: student non-participation

Or take steps to enhance the participation rate

- Highly compensate students for their time
 - Feds do this with their national surveys; highly effective
- Require all students to take the assessment (e.g., can't register for courses until fill it out)
 - Solves the nonresponse problem at the expense of the motivation problem

Representing institutions: student motivation

Even if every student in the sample agrees to participate, they still must perform a cognitively difficult task that does not matter to them

- Contrast this test-taking situation to the SAT, GRE and LSAT
- This is probably the single biggest hurdle facing efforts to measure postsecondary learning

Research indicates some possible ways to turn these into “high-stakes” exams

- Compensate students for performing well
- Tie performance to something on campus
 - Course credit
 - Academic honors
- Tie performance to something off campus
 - Include score on transcript that is sent to graduate schools and employers

Representing institutions: student motivation

Innovative experimentation in this area is one way to establish national leadership

But over the long-term, institutions should adopt similar approaches for comparability

Recommendations

- 1 Use either the CLA or the CAAP
- 2 Use random sampling outside of institutional control
- 3 Highly compensate sampled students for their participation
- 4 Link performance to an extrinsic reward of some kind

Taking into account institutional differences

Thought experiment: suppose we had

- A perfectly valid measure of critical thinking
- A representative sample of students at each institution
- And students performed at their best

Would we want to compare institutions on this metric?

Taking into account institutional differences

Thought experiment: suppose we had

- A perfectly valid measure of critical thinking
- A representative sample of students at each institution
- And students performed at their best

Would we want to compare institutions on this metric?

Obviously not; some schools will perform better than others due to higher admission standards

- Half the variation in NSSE benchmarks is due to student and institutional characteristics

Taking into account institutional differences

Some researchers have used statistical models to take into account institutional differences

- Highly sensitive to which variables are used as controls
- Not credible among many researchers
- Opaque and difficult to explain to a broader audience

Stronger approaches use measurements at entry and exit

Taking into account institutional differences

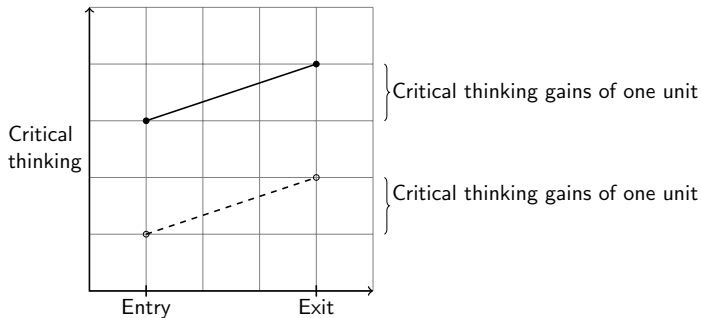
Main issue is when to measure

- Some advocate for measuring first-years and seniors at same time, and adjusting for differences between the two groups
- Same issue as before; depends on variable selection and is not credible

Only credible approach is a true longitudinal design

- By using the exact same group of students over time, worry less about differences between students and institutions
- Transparent way of describing growth at School A versus School B

Comparing two students or two institutions

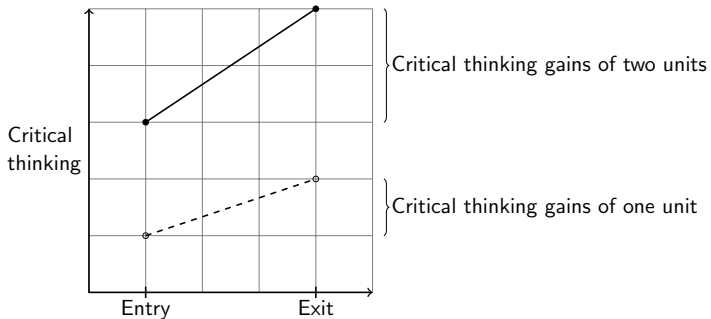


Taking into account institutional differences

But there is one problem

- If we don't want to worry about student nonresponse, motivation, or differences between institutions, then the growth for participants must be the same for non-participants
- This is a strong assumption
- But much more plausible than cross-sectional statistical adjustment of institutions or contemporaneous cohorts of students

Hopefully this isn't happening



Recommendations

- 1 Use either the CLA or the CAAP
- 2 Use random sampling outside of institutional control
- 3 Highly compensate sampled students for their participation
- 4 Link performance to an extrinsic reward of some kind
- 5 Adopt a true longitudinal design for institutional comparisons and benchmarking

Questions?