



Higher Education Outcomes Assessments: Measurements or Contests?

**Thomas Van Essen
Educational Testing Service**

**Presentation at
Conference on “Modeling and Measurement of Competencies
in Higher Education”**

**Maritim proArte Hotel Berlin, Germany
February 24, 2011**

Disclaimers, Prejudices and Assumptions

- OECD Expert Group
- AHELO
- Background in test development
- Higher education outcomes assessment is worth doing
– *but only if you can do it well*

Organization

- Why outcomes assessments matter
- Assessments as contests and assessments as measurements
- Things to worry about
 - Two fundamental issues
 - Some technical issues
- Some recommendations

IREG-5

- The role of outcomes assessments in academic rankings
- Skills matter

Outcomes Assessments as the Death of Higher Education

- Faculty will teach to the test at the expense of all the other attributes of higher education

Why Higher Education Will Survive Them

- Over-emphasizes the power of tests
- Under-emphasizes the strength and diversity of universities
- Makes certain assumptions about the nature of the instrument and its role in the higher education ecosystem.
- These are not necessary assumptions

Measurements and Contests

- The measurement view
 - What is being measured and why?
 - The object is information that will inform our diagnosis
 - How can we improve?
- But the best intentioned measurements can become contests

The Contest View

- The only thing that matters is relative position
 - The SAT
- To the extent that participants in the assessment perceive it as a contest they will behave in certain ways
- And if it is a contest we have an obligation to make it fair

The Contestants

- Students
 - Competing against each other for something of value in the labor market
- Departments
 - Competing for funding
- Institutions
 - Competing for resources within national systems
- National Systems
 - Competing for prestige and global resources

Fairness as Validity

- When we say an assessment is unfair we are really saying that it is not valid for a certain group.
- Fairness is not a “nice to have”
 - If you don’t have fairness you don’t have a valid test

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment
 - The domains to be assessed
- Technical and Operational Issues
 - Student motivation
 - Comparability and sampling
 - Translation and scoring
 - Institutional value add
 - Time
 - Cost
 - Transparency

What Is The Purpose of the Assessment?

- Profound and unacknowledged confusion about the audience for assessment data and what they intend to do with.
- We must be very clear about our answer to this question because there are a number of complex interdependencies between the decisions we make about users of the data and the assessment design.

Data Consumers

- Stakeholders interested in improving higher education
 - Governments, education systems, schools, departments within schools
- But what about students?
 - Direct or indirect consumers of data?
 - Students matter because their motivation
 - They may even matter for ethical reasons

Direct vs. Indirect Data

- Direct data
 - An individual student report that says something meaningful about what students know and can do
- Indirect data
 - Information about the department or school in which they are enrolled or intend to enroll

Indirect Data

- Will students within an institution be motivated to provide data that is only of value to the institution or potential students?

Soft vs. Hard Data

- Hard data
 - Tied to something consequential
- Soft data
 - Formative
- These obligate us to very different designs

Student Data Decision Tree

- Do we want to provide student level data?
 - If yes
 - Hard or soft?
 - If hard
 - what is the standard of reliability and technical quality?
 - If soft
 - what is the standard of reliability and technical quality?

We Have to Decide On the Purpose at the Outset

- There will be tremendous political pressure to make a single assessment be all things to all people
- That is impossible

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed
- Technical and Operational Issues
 - Student motivation
 - Comparability and sampling
 - Translation and scoring
 - Institutional value add
 - Time
 - Cost
 - Transparency

The Domain(s) To Be Measured

- Tension between general skills and subject specific skills
 - Critical thinking and problem solving both verbal and quantitative is the main point of higher education
 - But most higher education occurs at the disciplinary level

General Skills

- We need more research
- What does it mean to think like a social scientist or a scientist or a humanist?
 - What are the appropriate ways for making, supporting and evaluating arguments in general fields?
- How can we measure them?

Specific Skills

- How can we avoid the danger of making the assessments reductive and over simple?
 - What are we trying to measure that is different from licensure tests?
- AHELO's approach seems on the right track
 - How does one think like an engineer or an economist

Domain(s) Need to Be Specified Up Front

- The broader the universe of educational systems in which we want to make comparisons the harder this will be.

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation
 - Comparability and sampling
 - Translation and scoring
 - Institutional value add
 - Time
 - Cost
 - Transparency

Student Motivation

- Assume that this is not a high stakes situation
- If soft data is provided we need to see how it motivates performance across comparison groups
- If no student level data is provided
 - Requirement
 - In order to graduate students will be required to take the assessment
 - Incentives
 - Differing values across comparison groups
 - Intrinsic interest of the assessment
 - Can it be made fun or rewarding or painless?
 - Ignore “unmotivated” responses
 - Problematic

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling
 - Translation and scoring
 - Institutional value add
 - Time
 - Cost
 - Transparency

Comparability and Sampling

- Given the diversity of higher education what are the appropriate comparison groups?
 - In the US four year research institutions are very different from two year community colleges
 - We probably need to define 2 or 3 institutional comparison groups
 - Assign each participating institution to a group

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling ✓
 - Translation and scoring
 - Institutional value add
 - Time
 - Cost
 - Transparency

Translation and Scoring

- We know how to create instruments in multiple languages
- Higher order skills measured will require constructed response tasks
 - Need procedures in place avoid institutional and national bias
 - Complex cross scoring
 - Multiple choice anchor items

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling ✓
 - Translation and scoring ✓
 - Institutional value add
 - Time
 - Cost
 - Transparency

Institutional Value Ad

- Admitted students come in with different levels of ability and preparation
- We need a way to put matriculated students on a common scale
 - Some block of multiple choice items designed for this specific purpose

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling ✓
 - Translation and scoring ✓
 - Institutional value add ✓
 - Time
 - Cost
 - Transparency

Testing Time

- Time is money
- But stakeholders who are serious about getting useful outcomes measures must understand that it will take time
 - 1-2 hours per student twice a year

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling ✓
 - Translation and scoring ✓
 - Institutional value add ✓
 - Time ✓
 - Cost
 - Transparency

Cost

- Money is money
- This will be a relatively expensive proposition
 - What is the value of this kind of work?
 - Who should pay for it?
 - National education systems?
 - International organizations?

Things To Worry About In Terms of Fairness

- Fundamental Issues
 - The purpose of the assessment ✓
 - The domains to be assessed ✓
- Technical and Operational Issues
 - Student motivation ✓
 - Comparability and sampling ✓
 - Translation and scoring ✓
 - Institutional value add ✓
 - Time ✓
 - Cost ✓
 - Transparency

Transparency

- Transparency is money
- Transparency is time
- Given the complexity and political sensitivity of such a project all stakeholders need to commit to as great a degree of transparency as is technically possible
 - This means making sure that there is no “black box”
 - This means lots of disclosed tasks

Recommendations (or what would success look like)

- The system would be designed to provide reports at the institutional level
- The purpose of these reports would be to help institutions (and the national system which support them) to engage in processes of continual improvement

Recommendations

- Assessment would be computer delivered
- Multiple languages
 - How to optimize so as to serve the needs of affluent and less affluent systems?
- 4-6 hour test
- constructed as 8-12 30 min. blocks
 - NAEP like design
 - Each student would take 2-3 blocks (1.5 hours) twice a year

Recommendations

- Students would be required to take the test but would get no individual feedback
 - Initial testing would be required as part of the registration process
 - Final testing would be part of graduation
 - Like returning you library books
- All participating institutions would have to agree to the same or similar ground rules

Recommendations

- Assessment would contain
 - Background questionnaire
 - Blocks of multiple choice verbal and math ability items for linking
 - Verbal and quantitative critical thinking tasks
 - mixture of higher order multiple choice
 - constructed response
 - engaging and innovative
 - Subject specific tasks for student's declared or intended major for selected majors
 - foundational knowledge
 - skill at problem solving in the domain
 - Items routed on the basis of answers to background questions



Thanks

Thomas Van Essen

tvanessen@ets.org