#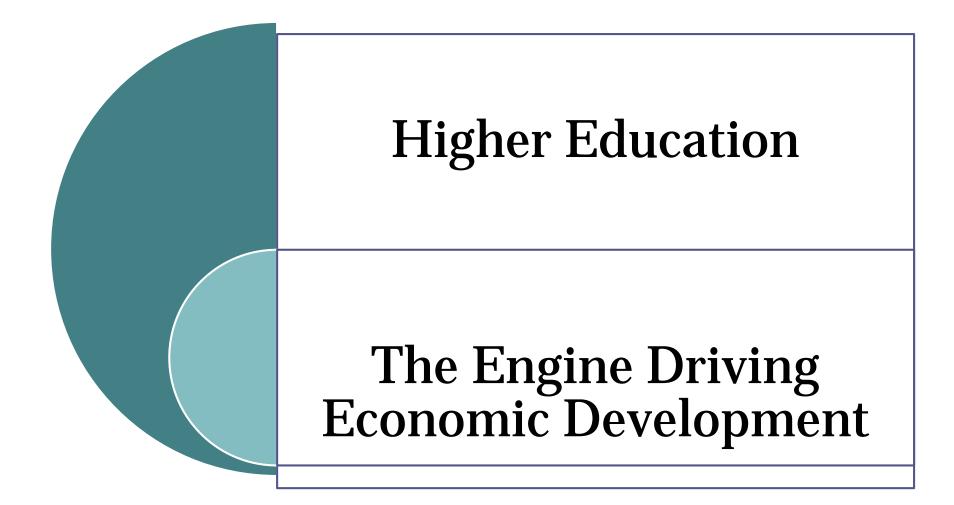 Transparency, Accountability, Information Symmetry and Integrity: Creating Guiding Principles for Developing US Institutional Ratings and Rankings

Gerald W. McLaughlin
DePaul University
Josetta S. McLaughlin
Roosevelt University

2014 IABS
Sydney, Australia

Higher Education

The Engine Driving Economic Development

Purpose

To identify principles and measures currently used for rating and ranking institutions of higher education.

To make recommendations for improving the accountable, information symmetry, and transparency of rankings.

To address questions of data and information integrity in ratings and rankings of institutions.

Importance

Demand for ratings and rankings by students and parents, HE administrators, and politicians is likely to remain constant.

- Rankings should be based on guidelines and principles that build integrity into the process of developing the tool.

Currently available data used at both the national and international levels rely heavily on publishing and wealth as measures of productivity.

- Thus, critics charge that this is at the expense of service and teaching as a public good.

Issues align with CSR expectations of good governance among publishers and producers of the ratings and rankings.

- Publishers who fail to be transparent, fair, and discourage unethical practices itself or among clients inflict harm on the overall system of HE.

Background

- **College and university rankings** are lists of institutions in higher education, ordered by combinations of factors... The subject has produced much debate about rankings' usefulness and accuracy. The expanding diversity in rating methodologies and accompanying criticisms of each indicate the lack of consensus in the field. (College and university rankings, n.d.)

Background

- The discussion gained complexity in the 1980s when the purpose for developing the tools was refocused to support the public's (e.g., consumer) ability to compare institutions and the publisher's ability to provide rankings for financial gain.

Examples – DePaul University

Item Ranked	Source	Ranking	Methodology
Best cafeteria variety in IL	CollegeProwler.com	8/44	Student surveys, student reviews, polls
DePaul Blue Demons overall RPI	TeamRankings.com	277	Mathematical formula for winning percentages. It involves the team in question & the opponents.
Most Dangerous Colleges 2010	TheDailyBeast.com	24/258	Crime statistics
Sexual Health Report Card Rankings 2012	Prnewswire.com*	114/141	student health center representatives, and students on campus.
America's Top Colleges - Forbes 2012	Forbes.com	512/650 (359 in 2010?)	Ratemyprofessor.com 17.5%
Entrepreneurship program	Princeton Review	Top 25 grad and UG	Availability of internships, externships, experiential learning. Student outcomes – own business? >2,000 schools
Part-time MBA	Bloomberg Businessweek.com	42/219	Survey part-time MBAs, also use GMAT, class size and other info.
Contribution to public good	Washingtonmonthly.com	130/281	Began with 1569 schools from IPEDS, equally weighted 3 categories
Service Learning	Usnews.com	Top 25	> 1500 schools. College presidents, chief academic officers, deans of students or admissions.

The Business of Rankings, DePaul Brown Bag, Susan Stachler and Gerry McLaughlin

*http://www.prnewswire.com/news-releases/university-of-illinois-dethrones-columbia-university-to-take-top-spot-in-the-2012-trojan-sexual-health-report-card-rankings-175387711.html?ispopup=y

Robert Morse (*US News and World Report*)

In the early years, the thing that's happening now would not have been imaginable. This idea of using the rankings as a benchmark, college presidents setting a goal of 'We're going to rise in the *U.S. News* ranking,' as proof of their management, or as proof that they're a better school, that they're a good president. That wasn't on anybody's radar. It was just for consumers.

Source: Gladwell, 2011

New Initiatives

- Obama – Postsecondary Institution Rating System

- CHEPS – U-Multirank

Obama Rating Plan

- Washington Post, August 22, 2013

- "President Obama proposed that (by 2015) the federal government rate the nation's schools to hold them accountable for performance and help bring soaring tuition under control."

- "Obama said, his administration will begin **evaluating** colleges on measures such as the average tuition they charge, the share of low-income students they enroll and their effectiveness in ensuring students graduate without too much debt.

- The president also will seek congressional approval — which could prove difficult — to steer more **federal student aid** toward colleges that **score highly** in the ratings."

Postsecondary Institutional Rating System

- The ratings will be based upon such measures as:
    - **Access**, such as percentage of students receiving Pell grants;
    - **Affordability**, such as average cost of attendance, scholarships, and student loan debt; and
    - Outcomes, such as graduation and transfer rates, including those for Pell grant recipients, graduate earnings, and advanced degree attainment of graduates (**Accountability**).
- The Department intends, through these ratings, to **compare** colleges with similar missions and **identify colleges** that do the most to help students from disadvantaged and underrepresented backgrounds, as well as colleges that are **improving their performance**.

Federal Register / Vol. 78, No. 242 / Tuesday, December 17, 2013

Response: Institute for Higher Education Policy (NGO)

- **48 page response to the Secretary of Education request for comments on the Postsecondary Institution Rating System (PIRS)**

  - 1. Critical information on post-college outcomes, completion, cost, and access is key (consumer).
  - 2. Contextualize the information without undue complexity and in relation to the consequences.
  - 3. **Collect better data**, but don't delay providing information now.

IHEP - Attachment (example)

U-Multirank

- **Multidimensional ranking - Going beyond the traditional focus on research excellence**
  - Five dimensions: teaching & learning, research, knowledge transfer, international orientation, regional engagement
  - No composite indicators, no pre-defined weights on individual indicators, single ranking for each indicator
  - Validity, reliability, feasibility of each indicator
- **User-driven ranking**
  - Personalised ranking allows users to rank by their own preferences and priorities on dimensions and indicators ("democratised" approach)
  - Flexible web tool

What are the distinctive features of U-Multirank?

- **Comparing like with like**
  - Link to mapping indicators allowing identification of institutions with similar institutional profiles
- **Multi-level ranking**
  - Combining institutional ranking (whole institutions) and field-based rankings (start with: electrical and mechanical engineering, business studies, physics)
- **Stakeholder-oriented processes**
  - Intensive inclusion of stakeholders in development and continuous refinement of U-Multirank

Standards For Ratings

Data Integrity

Berlin Principles

Standards From Educational and Psychological Tests (Proposed)

Generally Accepted Standards of Data Integrity

Raw Data	Data Sets	Data Source
Accuracy	Sufficiency	Authority
Reliability	Relevancy	Verifiable
Validity	Timeliness	Interpretable
Completeness	Generalizability	Accessible

McLaughlin, Pavelka & McLaughlin (2005): Assessing the Integrity of Web Sites

The 2006 Berlin Principles on Ranking of Higher Education Institutions

Purpose of Rankings

- Respond to demands from consumers for easily interpreted information
- Stimulate competition between institutions of higher education
- Provide some (**?**) of the rationale for the allocation of funds
- Differentiate between types of institutions and different programs
- Contribute to the definition of a quality educational institution within a particular country
- Become part of the fabric of accountability and the work of accreditation agencies

Purposes & Goals

- Be one of a number of diverse approaches to the assessment of higher education
- Be clear about their purpose and their target groups
- Recognize (**?**) the diversity of institutions and take the different missions into account
- Provide clarity about the range of information source and the messages each source generates
- Specify the linguistic, cultural, economic, and historical contexts of the educational systems being ranked
- **? – question marks are ours.**

Source: IHEP, 2007

The Berlin Principles on Ranking of Higher Education Institutions (continued)

Design & Weighing of Indicators

- Be transparent regarding the methodology used for creating the rankings
- Choose indicators according to their relevance and validity (**?**).
- Measure outcomes rather than inputs where possible (**?**).
- Stabilize the weights assigned to different indicators

Collection & Processing of Data

- Pay due attention (**?**) to ethical standards and good practices recommended in the Principles
- Use audited and verifiable data whenever possible (**?**), making data comparable and compatible across institutions
- Include data that are collected with proper procedures for scientific data collection (robust random samples)
- Apply measures of quality assurance to ranking processes themselves.
- Apply organizational measures that enhance the credibility (**?**) of the rankings

? – question marks are ours.

The Berlin Principles on Ranking of Higher Education Institutions (continued)

Presentation of Ranking Results

- Provide consumers with a clear understanding of all the factors used to develop a ranking and offer them a choice in how rankings are displayed. (And have some choice in how the factors are weighted).
- Be compiled in a way that eliminates (**?**) or reduces error in the original data and be published in a way that errors and faults can be corrected.

? – question marks are ours.

Standards for Educational and Psychological Tests

"The applicability of the standards to an evaluation device or method is not altered by the label applied to it (e.g., test, assessment, scale, inventory)." (p. 3, Standards)

Response of the American Psychological Association to the Secretary of Education about the PIRS

"It is our hope that as you move forward with this effort, you will engage experts from our Division on Evaluation, Measurement and Statistics to ensure that any system that is developed is in conformity with **basic standards** of assessment."

Standards for Educational and Psychological Tests, AERA, APA, NCME, 1999

Example Standards for Ratings (*Standards* – AERA, APA, NCME)

Instrument construction, evaluation, and documentation (accountability and transparency)	• Validity • Reliability and Errors of Measurement
Fairness (Integrity of data)	• Fairness in rating and rating use • The rights and responsibilities of rated institutions
Rating Applications (Information symmetry)	• The responsibilities of rating users • Ratings in institutional evaluation and Public Policy

Adapted from the *Standards for Educational and Psychological Tests*, AERA, APA, NCME, 1999

Validity - Description

- Validity is "the degree to which evidence in theory support the interpretation. It begins with "an explicit statement of the proposed interpretation..." Along with this statement is the rationale for the interpretation.
  - Sources of validity evidence include an
    - analysis of the ratings content and the construct is intended to measure by expert judgment of qualified experts,
    - evidence concerning the fit between the construct being rated and the detailed performance of the institutions;
    - an analysis of the relationships between aspects being rated;
    - the relationship between the ratings and other external measures of the construct;
    - the accuracy with which the ratings predict future events; and
    - the degree to which the ratings can be generalized to new situations such as different institutions and/or different countries.

Validity – Example Standards

- The provider of the ratings should present a rationale for each recommended interpretation and use of the ratings together with a comprehensive summary of the evidence in theory bearing on the intended use (1.1)
- The provider of the ratings should set forth clearly how the ratings are intended to be interpreted and used. The population(s) for which the rating methodology is appropriate should be clearly delimited, and the construct that the rating is intended to assess should be clearly described (1.2)
- If validity has not been established for some common interpretation or evidence is inconsistent with the interpretation, that fact should be made clear and users should be cautioned about making unsupported interpretations. (1.3)
- When a validation rest in part on the opinion of expert judges, and observers, or Raiders, procedures for selecting such experts should be fully described. The qualifications should be presented. The description should include any training and instructions and the degree to which the individuals interacted in making their decisions..(1.7)

Reliability - Description

- Reliability refers to the consistency and stability of the measurements obtained through the rating/ranking process.
  - Reliability is based on the objectivity, or the degree to which two perspectives of a characteristic would produce the same conclusion.
  - It also involves the stability of the perspectives over time.
  - Finally it involves the internal consistency of a measurement when multiple indicators are involved.
  - The lack of reliability creates measurement error that are generally viewed as random and unpredictable.
  - They are usually associated with a **computed coefficient**

Reliability – Example Standards

- For each total score, sub score, or combination of scores that is to be interpreted, estimates of relevant reliabilities and standard errors of measurement should be reported. (2.1)
- The standard error of measurement, both overall and conditional begin parenthesis if relevant), should be reported both in raw score or original scale units and in units of each derived score recommended for use and test interpretations. (2.2)
- When selective judgment enters into test scoring, evidence should be provided on both interrater consistency in scoring and within rater consistency over repeated measures(2.10)
- If the definition of variables differs within different populations, separate reliability analyses should be provided for scores produced under each major variation if adequate sample sizes are available(2.18)

Fairness - Description

- A full consideration of fairness would explore the many functions of ratings in relation to its many goals, including the broad goal of achieving equality of opportunity for various cultures.
- It would consider the ratings' technical properties, the way they are reported, and the factors that are validly or erroneously thought to account for patterns of ratings for groups and individuals.
- Two important aspects are fairness as equitable treatment and the rights and responsibilities of institutions being rated.
- Bias due to inappropriate selection of rating variables interacts with cultural differences as well as governmental definitions and unique linguistic interpretations.

Fairness – Example Standards

- When credible evidence indicates that measures, such as research funding and journal publication, differ across cultures, validity evidence should be collected and reported for each culture (7.1).

- When credible research reports that differential metric functioning exists across cultural, ethnic, and or linguistic groups developers of ratings should conduct appropriate studies on validity and reliability to eliminate bias from rating scores (7.3).

- When empirical studies of differential prediction of performance for members of different subgroups are conducted, they should include analyses separately for each group (7.6).

- When ratings are proposed for use as instruments of social, educational, or personal policy, those proposing the rating should fully and accurately inform policymakers of the characteristics of the ratings as well as any relevant and credible information concerning the likely consequences of using the ratings (7.9).

Institutional Rights & Responsibilities - Description

- Fair and equitable treatment of institutions involves providing, in advance of the rating,
  - information about the nature of the rating,
  - the intended use of the scores of the rating, and
  - the confidentiality of the results.
  - Informed consent may be appropriate.
- Responsibilities of those participating include
  - preparing themselves for the appropriate collection and analysis of data,
  - representing themselves honestly, and
  - informing appropriate individuals if they believed the ratings to not adequately reflect institutional characteristics.

Institutional Rights & Responsibilities – Example Standards

- Any information about the measures used in the rating and purposes that is available to any institution should be available to all institutions free of charge and in accessible formats (8.1).

- Institutions being rated should be provided in advance as much information about the rating, the rating process, the intended use of the ratings, the analytic criteria, the rating policies, and confidentiality protection as is consistent with obtaining valid responses (8.2).

- Rated institutions should be aware that any form of cheating is inappropriate and may result in sanctions (8.7).

- When score reporting includes assigning institutions to categories. The least stigmatizing labels, consistent with accurate representation, should always be assigned (8.8).

Responsibility of Users of Ratings - Description

- The appropriate use of ratings and the sound interpretation of their scores are likely to remain primarily the responsibility of the user.
- Institutions, their stakeholders, legislators, policymakers, the media, the courts, and the public at large often yearn for unambiguous interpretations.
- They often attribute major implications to a single factor.
- Ultimately however the use of ratings and interpretation of their scores require an element of professional judgment.

User Responsibilities - Example Standards

- When a rating is to be used for a purpose for which little or no documentation is available, the user is responsible for obtaining evidence of the rating solidity and reliability for this purpose (11.2).
- Responsibility for rating use should be assumed by our delegated only to those individuals who have the training, professional credentials, and experience necessary to handle this responsibility (11.3).
- Rating users should be alert to the possibility of scoring errors; they should arrange for rescoring if individual scores are area data suggest the need for it (11.10).
- If the integrity of a rating institutions scores is challenged, the rating developer, or sponsor, should inform the institution of their relevant rights, including the possibilities of appeal and representation by counsel (11.11).

Use in Policy Making/Program Evaluation - Description

- Rating results are often one important source of evidence for the initiation, continuation, modification, termination, or expansion of various programs and policies.
- The use of ratings can be interpreted to describe the long-term pattern of effects for one or more groups of institutions within a country.
- Ratings can be used to inspire institutions as well as to infer institutional effectiveness.
- The desire to influence ratings to show acceptable institutional performance could lead to inappropriate practices such as modifying data definition, collection and analysis.

Policy Making and Evaluation – Example Standards

- When a rating is used to serve multiple purposes, evidence of technical quality for each purpose should be provided. (15.1).
- In institutional evaluation or policy studies, investigators should complement rating results with information from other sources to generate defensible conclusions based on their interpretation (15.4).
- The integrity of the rating results should be maintained by eliminating practices designed to raise rating scores without improving performance on the construct or domain measured by the rating (15.9).
- Those who mandate rating programs should ensure that the individuals who interpret them to make decisions within the public context are qualified to assume this responsibility and are proficient in the appropriate methods for interpreting rating results (15.13).

Conclusions (operational)

Standards for ratings, based on their purpose, are needed.

The Berlin Principles are a good framework of principles.

Standards need to be developed to support the Principles – like objectives need to be developed to support goals.

*Standards (*AERA, APA, NCME*)* is an excellent starting point to develop unbiased ratings that have accountability, transparency and integrity.

Conclusions (strategic)

Consumer information and accountability may not be compatible.

Accountability to different stakeholders may not be compatible.

Stakeholders include students, governments, academics, institutions, NGO's, and major charities (Gates, Lumina, etc).

We live in most interesting times.

Thank You

Transparency, Accountability, Information Symmetry and Integrity: Creating Guiding Principles for Developing US Institutional Ratings and Rankings

Gerald W. McLaughlin, gmclaugh@depaul.edu
Josetta S. McLaughlin, jmclaugh@roosevelt.edu

2014 IABS, June 2014
Sydney, Australia