ASSESSING THE CONSTRUCT VALIDITY OF THE GLOBAL 100 SUSTAINABILITY RANKING FOR SCHOOLS OF BUSINESS

Gerald W. McLaughlin*
Josetta S. McLaughlin

Abstract: Colleges of business rankings purport to address relative performance on programs such as sustainability. The primary criticism of rankings is that providers have not established reliability or validity of the ranking. This study examines whether *The Global 100* sustainability ranking is sufficiently unique to claim that it is based on attributes not used for non-sustainability ranking (divergent validity) and whether it is appropriately related to independent characteristics expected to measure this attribute (convergent validity).

Keywords: construct validity; sustainability; ranking

INTRODUCTION

Institutional rankings of colleges and universities are increasingly visible to important stakeholders, including current students and their parents. Rankings have proliferated since the 1950s and publication of such rankings as *US News and World Report*'s "America's Best Colleges" (Myers & Robe, 2009). This was followed by publication of business school rankings and, more recently, by ranking based on more specific content areas such as sustainability (See, for example, *Bloomberg Businessweek*'s "B-Schools: Social and Environmental Rankings"). These rankings have frequently been interpreted by the public as a proxy for the quality of the institution, college or programs being ranked (McGuire, 1995).

One constant criticism throughout the history of ranking systems to assess quality has been that providers have not met the required standards for reliability and construct validity (Myers & Robe, 2009; McGuire, 1995; Stuart, 1995). Validity of rankings is, in fact, one of the most important issues surrounding reputational studies, especially given that invalid rankings can unfairly damage an institution's reputation and can undermine its competitiveness (McGuire, 1995). Some providers of rankings attempt to deal with this problem by providing caveats or disclaimers on their websites and in their published reports. For example, the Aspen Institute,

* Author contact information:
Gerald W. McLaughlin: gmclaugh@depaul.edu • DePaul University, Chicago, IL 60605 USA
Josetta S. McLaughlin: jmclaugh@roosevelt.edu • Walter E. Heller College of Business Administration, Roosevelt University, Schaumburg, IL 60173 USA

developer of the Beyond Grey Pinstripes (BGP), notes that for their top sustainability programs, e.g., *The Global 100,* "[O]ther than obvious errors, staff at Aspen CBE do not attempt to assess the validity of the self-reported data." Center for Business Education (CBE) staff rely on pledges of honesty and accuracy by the submitter and transparency of data online" (*"The Global 100,"* n.d.) Unfortunately, though their attempt to provide a caveat is to be commended, this approach to creating a database can be questioned on many grounds, to include lack of full disclosure, rater bias, halo effects, and general issues related to data integrity (Mallette, 1995; Stuart, 1995).

Providers of rankings have historically made considerable efforts to address criticisms around construct validity (Morse & Gilbert, 1995). However, the traditional scholarship of validity would suggest that they need to go beyond simply noting such things as not attempting to assess the validity of self-reported data to asking critical questions concerning the reliability and validity of rankings. We address this need by examining *The Global 100* within the context of construct validity. The analysis addresses whether the ranking is sufficiently unique to claim that it is based on attributes not previously used for ranking of institutions on other dimensions (divergent validity) and whether the ranking is appropriately related to other independent institutional characteristics expected to measure the attribute of interest, i.e., sustainability (convergent validity). The need to address validity for business school rankings that make claims about sustainability is important since rankings are increasingly being used to strategically market the uniqueness of sustainability programs.

Background on Ranking in Higher Education

Early US college rankings followed efforts by European scholars in the early 1900s to study the eminent members of society (Myers & Robe, 2009). Though scholars did not suggest use of rankings as a tool for measuring institutional quality, US scholars produced rankings with the aim of using data and rankings for this purpose. Similar to the Europeans, the first US rankings, published by James McKeen Cattell, were primarily interested in the study of eminent men (Myers & Robe, 2009). Rankings would not, however, become a predominant method for producing reputational rankings as a proxy for institutional quality until the late 1950s with the work of Australian geographers Albert Somit and Joseph Tanehaus and their ranking of American geography departments (Myer & Robe, 2009). This was followed by *the Cartter Report: 1966 Assessment of Quality in Graduate Education,* which ranked 29 disciplines. the Roose and Andersen 1970 study on graduate programs, and the National Academy of Sciences *Assessment of Research-Doctorate Programs in the United States* (Myer & Robe, 2009). A parallel stream of research on reputational rankings for undergraduate programs began with a 1957 Chicago Sunday Tribune article that "ranked the top 10 universities, coeducational colleges, men's colleges, and women's colleges according to their undergraduate quality" (Myers & Robe, 2009, p. 15). Other rankings that followed included *The Gourman Report* (beginning in 1955 and ending in 1997) and a 1981 pilot study by Solmon and Astin that covered four states and a list of 80-150 departments in 7 fields (Myers & Rober, 2009).

Reputational rankings did not garner the attention of the public until 1983 when they were published by *US News and World Report*. Robert Morse, Director of Data Research for "America's Best Colleges" noted in 1995 that "the main users… are prospective students and

their parents", noting that an academic document would be of limited interest to their readers (Morse & Gilbert, 1995). Morse met extensively with college administrators and in response made incremental changes to the ranking over time. For example, in 1988, the reputational component of the ranking was reduced to 25 percent while objective input and output data made up the remaining 75 percent. Other popular magazine rankings include *Money*'s "America's Best College Buys" and the Princeton Review's *The Best 368 Colleges*. Other producers, especially with respect to ranking of business programs, include (but are not limited to) *Bloomberg Businessweek, Forbes, the Wall Street Journal, The Economist*, and *The Financial Times*. Global rankings of note include The Shanghai Jiao Tong University's "Academic Rankings of World Universities", and the *Times Higher Education* of England (results published by *US News and World Report* in 2008). (Myers & Robe, 2009)

Ranking based on sustainability is one of the newest rankings published by popular magazines; it has been reluctantly addressed by the popular press due, in part, to the definitional ambiguity of the term (Conversation with Morse, 2009). By contrast, special interest not-for-profit groups have stepped forward to collect data that have in turn been used to grade and rank educational institutions on their sustainability practices. All claim to provide unique rankings; all tend to operate under a business model that rewards them for the visibility and voice in arguing the credibility of their rankings. They all share in common the criticism surrounding methodological questions that have been debated for decades, especially with respect to reliability and construct validity.

The use of rankings by prospective students and their parents has not gone unnoticed by college and university administrators that have increasingly adopted more marketing-oriented enrollment management models to manage enrolment at higher education institutions (Hunter, 1995). Many institutions began using the rankings found in publications such as *Peterson's Guide, U.S. New & World Reports, Business Week, and The Financial Times* to promote the institution, first in marketing mailers and promotional literature and more recently on websites. Colleges of business followed this same pattern, with many now using business school rankings to publicize the quality of their programs. Critics charge that this pressure to look good in rankings has led institutions to look for strategies to improve rankings and thus enhance university, college, or program visibility (Glenn, D., 2010; Jaschik, 2010). By contrast, few academicians critically review the construct validity of the rankings and examine the methodological rigor used by the organizations publishing the ranking. This problem is becoming more relevant to schools and colleges of business as rankings move from the traditional institutional rankings to college-level rankings to rankings focusing on new areas of study. The rankings appear to have two primary intents – to market a program (or institution) and to encourage a shift in societal values concerning the need for sustainable development. Uses of rankings include enhancing university, college or program visibility, benchmarking using comparator groups, and increasing awareness among stakeholders of particular areas of performance. As noted earlier, this paper focuses specifically on ranking to bring attention to how a school or college of business performs on "sustainability".

Ranking on Sustainability

Societal interest in implementation of sustainable practices has spurred the new round of rankings. Visible sustainability ranking initiatives include *The Princeton Review's Guide to 286 Green Colleges (*published by The Princeton Review, Inc. in partnership with the U.S. Green Building Council), *The College Sustainability Report Card* (published by The Sustainable Endowments Institute), The *Bloomberg Businessweek's* "B-Schools: Social and Environmental Rankings" (developed using the Aspen Institute's biannual "Beyond Grey Pinstripes" (BGP) research survey and subsequent rankings), and *The Global 100* (based on BGP and published by The Aspen Institute). *The Princeton Review's Guide to 286 Green Colleges* provides a qualitative and quantitative guide to a school's performance as an "environmentally-aware institution". The three concerns addressed by the ranking are 1) whether the students have a campus quality of life that is both healthy and sustainable, 2) how well an institution is preparing its students for employment in the green energy economy of the 21st century, as well as for citizenship in a world now defined by environmental concerns and opportunities, and 3) the degree to which an institution's policies are environmentally responsible (*The Princeton Review's Guide to 286 Green Colleges*. 2010-2011). The data collection instrument collects responses to ten survey questions that focus on sustainability performance. By contrast, *The College Sustainability Report Card* collects data at the institutional level through four surveys administered to more than 300 institutions. The surveys cover campus operations, dining services, endowment investment practices, and student activities to rank performance on Administration, Climate Change and Energy, Food and Recycling, Green Building, Student Involvement, Transportation, Endowment Transparency, Investment Priorities, and Shareholder Engagement.

The *Bloomberg Businessweek's "*B-Schools: Social and Environmental Rankings " is developed using the Aspen Institute's biannual 'Beyond Grey Pinstripes' research survey and subsequent rankings. The Aspen Institute worked with professional accrediting agencies AACSB and EFMD to identify a roster of approximately 600 accredited full-time MBA programs that could be used to highlight programs that integrate social and environmental stewardship issues into the curriculum. From these 600 programs, Beyond Grey Pinstripes ranks and publishes the top 100 schools as *The Global 100* based on data representing 1) availability of relevant courses (e.g., number of courses offered that contain social, environmental or ethical content), 2) student exposure (e.g., teaching hours covering content and student enrollment in these courses), 3) relevant courses on for-profit impact (e.g., number of courses that demonstrate both relevance and also address the intersection of social and environmental issues), and 4) faculty research (e.g., number of scholarly articles containing some environmental or ethical content being published in peer-reviewed business journals).

As with early criticisms of overall institutional rankings, the most frequent concern regarding newer rankings is that the ranking methodology may be flawed; critics continue to question the reliability and validity of the product. The same concerns found in the literature on institutional reputational ranking permeate the climate of sustainability ranking -- rater bias, halo effects, timing lags, design flaws (Stuart, 1995), the integrity of the data submitted by institutions, failure to meet the standard of full disclosure of ranking methodologies, and the validity of rankings.

Construct validity, the focus of this research, strikes close to the heart of this unease over rankings. This form of validity is "the extent to which a ranking system may be said to measure a theoretical construct such as education quality" (McGuire, 1995). The issues surrounding the construct validity of institutional reputational rankings have been widely discussed but never satisfactorily demonstrated. With respect to ranking on sustainability, the question is whether it has been shown that sustainability rankings are really based on a true measure of sustainability activity. Because the reliability of rankings is a precursor to considering construct validity, it too should be assessed and reported for any ranking of institutions, colleges or programs on sustainability.

The following analysis examines the construct validity of business school rankings and whether the rankings are reliable, i.e., demonstrate objectivity, internal consistency and stability. Claims of uniqueness will be evaluated by comparing business school and college reputational rankings to rankings on the business school attribute of interest, i.e., sustainability, and by controlling for other institutional-level characteristics or attributes.

CONSTRUCT VALIDITY DEFINED

Validity refers to "the degree to which something does what it is intended to do" (Howard, McLaughlin & McLaughlin, 1989). With respect to ranking, we define validity as an ordering that can be properly interpreted and generalized to provide an understanding of the construct of interest. Lack of validity can be a result of systematic bias in the choice of measures and can be limited by a lack of reliability in the data (Howard, et al., 1989).

Construct validity thus concerns the degree to which a ranking builds on the construct that it purports to measure. Such constructs are indirectly rather than directly measured, e.g., happiness, institutional quality, using multiple measures. A primary threat to this type of validity is the failure on the part of the analyst to properly define and operationalize the construct. Operationalization is suspect absent a definition and a comprehensive understanding of what and how the variables represent the construct of interest. As noted by Yaniv, "clarity and validity are inseparable" (2011, p. 590). Establishing construct validity thus requires answers to questions concerning "what properties must characterize a definition in order for that definition to be … useful for cumulative findings?" (Yaniv, 2011: 590).

The classic approach to confirming construct validity has been to establish three sub-categories of validity – convergent validity, divergent validity, and nomological validity (Campbell & Fiske, 1959; Peter, 1981; Howard et al., 1989; Trochim, 2006). Convergent validity concerns whether the measures are appropriate for providing a convergence of understanding on the construct of interest. Divergent validity concerns whether the construct purported to be measured differentiates it from other constructs. Nomological concerns whether the construct adequately confirms or disconfirms key beliefs about what is being measured (Howard et al., 1989). These concerns are beyond the scope of this research.

It is important to note that most social science researchers (to include business scholars) use a relationalist approach to construct validity. They develop a sense of what a construct represents

(more or less), attempt to operationalize what the construct should look like based on direct evidence, and then collect data that supports a theoretical view of how the construct should "behave" given the measures used to describe it. The treatment of construct is thus derived from relationships among indirect measures of a construct of interest that cannot be directly measured. To date, we believe that there is no consensus on what it means for a school or college of business to rank high on sustainability.

Purpose and Methods

This study is designed to examine whether the rankings of business schools/colleges found in *The Global 100* are appropriately measuring what it purports to measure. To address this question, the focus will be on the construct validity of the rankings and will address the following questions: 1) Can *The Global 100* ranking be shown to possess adequate reliability to have construct validity? 2) Is *The Global 100* ranking sufficiently unique to claim that it is based on attributes not previously used for ranking (divergent validity)? 3) Is *The Global 100* ranking appropriately related to other independent institutional characteristics expected to measure this attribute (convergent validity)?

Data Sources

Multiple data sources are used in this study. Data on sustainability rankings for business programs were downloaded for *The Global 100* (Beyond Grey Pinstripes data). Data on schools and colleges of business-level rankings were downloaded from *The Financial Times* and *U.S. News & World Report*. Institutional characteristics for U.S. schools listed in *The Global 100* were downloaded from the *Integrated Postsecondary Education Data System (IPEDS),* a "system of interrelated surveys conducted annually by the U.S. Department's National Center for Education Statistics (NCES)." Information covers institutions that accept federal student financial aid. The institutions are required to "report data on enrollments, program completions, graduation rates, faculty and staff, finances, institutional prices, and student financial aid." Variables downloaded for this study include degrees conferred by curriculum, degrees conferred by level of degree, expenditure and revenue patterns, institutional size, institutional staffing. Comparable databases for schools and colleges of other nations were not available. International data were used in analysis of rankings but could not be included in analysis requiring institutional-level data such as that provided through the US IPEDS database.

Methods

Reliability was evaluated using Conbach's Alpha and factor analysis. Construct validity was evaluated through exploration of divergent validity and convergent validity. In order to show that *The Global 100* sustainability ranking is unique, the sustainability ranking must be sufficiently unique from other over-all reputational rankings of schools and colleges of business, such as found in *The Financial Times* and the *US News & World Report* (divergent validity). Both US and international schools and colleges of business were used in this analysis. In order to show the expected general relationships between the sustainability ranking and those attributes of an institution which logically should be associated with performance or achievement on the

construct, i.e., sustainability, institutional-level IPEDS data on attributes were entered into the analysis (convergent validity). Due to lack of comparable data, only US schools and colleges of business were used in this part of the analysis. These variables are: 1) selectivity, 2) size of graduate program in various curricula, e.g., business, agriculture, law, medical sciences programs, based on degrees conferred, 3) percentage of full-time faculty/tenure-track faculty/tenured faculty, 4) institutional wealth, e.g., endowments, debt loads, 5) expenditures on student services, instructional activity, research, service, and academic support, and 6) percentage of specific expenditures relative to overall instructional costs.

Procedures

Data from selected websites – *The Global 100, The Financial Times, US News and World Report* -- were identified, downloaded, and merged. Institutional-level IPEDS data were then added to the database. Once the databases were cleaned, the following analyses were conducted:

Reliability

Cronbach's Alpha was computed to assess the internal consistency of the component scales. The relationship of the rankings over time was then analyzed to evaluate the stability of the construct. Next the association between the component scores was analyzed to determine the degree to which there is a common factor/institutional ranking. [Note: Sixty-eight of the institutions are U.S. institutions. In addition, some 32 institutions are identified as notable by "Beyond Grey Pinstripes" but are not ranked. There are subscales for the top 100 which are component ranks.]

Construct Validity

To assess divergent validity, the correlation between sustainability rankings found in *The Global 100* and school or college-level rankings, i.e., *The Financial Times* and *US News and World Report* (2011 data), were analyzed using Spearman rank correlations. To assess convergent validity, correlations and comparison of frequencies were used to identify the IPEDS variables that are associated with the rankings.

RESULTS AND DISCUSSION

Reliability is traditionally measured by objectivity, internal consistency, and stability. The ability to measure objectivity is traditionally done with measures of inter-rater reliability. Since we did not have appropriate data from the Aspen Institute, we were not able to calculate this measure.

A second way to measure reliability is to compute the internal consistency. This was done by computing Cronbach's Alpha. The Alpha for 2007 was .789 and the Alpha for 2009 was .549. This is sufficient to conclude that the sum of the ratings has sufficient internal consistency. The third type of reliability is the stability of the rating.

Data for *The Global 100* ranking were available for 2007 and 2009. The stability of the Overall Ranking was .731 ($p<.01$) between 2007 and 2009. This was based on a rank order correlation for 76 institutions that were in both rankings. This appears to indicate adequate stability.

The correlations between the components, i.e., the 4 subscales on which *The Global 100* rankings are based, are shown in Table 1. The diagonals represent the stability of the components. The correlations in the off diagonals are the correlations of the components with other components in the other year. [One aspect that does not show in this analysis is that only 76 of the institutions were on both lists. This produced missing data but there did not seem to be a rational way to estimate the score so they were not included in this analysis.]

Table 1: Stability of Components of *The Global 100* over Time

	Relevant Courses 07	Student Exposure 07	Course / For-Profit-Impact 07	Faculty Research 07
Relevant Courses 09	**.809****	*.470***	*.711***	.284*
Student Exposure 09	.332**	**.448****	.249*	.029
Course/For-Profit-Impact 09	.532**	.309**	**.630****	.339**
Faculty Research 09	.238*	.171	.231*	**.612****

* P < .05, ** P < .01, N = 76

The stability of the four scales ranged from .809 for *Relevant Courses* to .448 for *Student Exposure. Course /For-Profit Impact* was .630 and *Faculty Research* was .612. All of these stabilities are statistically significant ($p<.01$). Again, this indicates adequate stability. This matrix also has some of the characteristics of the Multi-Trait Multi-Method analysis (Campbell & Fiske, 1959). In such an analysis, the four components should have a higher correlation *with themselves* at a different point of time than they do with other components when viewed at the same point in time. The lowest correlation should be *with other components* at a different point in time. It should be noted that the scales for *Student Exposure 07* and *Course/ For-Profit Impact 07* unexpectedly correlated higher with the scale *Relevant Courses 09* (.470, .711) than with their respective scales in 2009 (.448, .630).

The internal consistency of the components was investigated further using Principal Components, with the "ones" in the diagonal. Three factors were extracted based on the Scree test of Eigen values and the loadings on the principal components. These three factors accounted for 78% of the variance. They were rotated using a Varimax Rotation to facilitate interpretation. The results are shown in Table 2.

One caveat is that the loading of *Relevant Courses* and *Course/ For-Profit Impact* on the same factor suggests that these two components are measuring essentially the same aspect of sustainability. This suggests that using both measures is double weighting this component.

Table 2: Factor Analysis of Scale Components

Item	Factor 1	Factor 2	Factor 3
07 Student Opportunity	**.787**	.182	.399
07 Student Exposure	.365	.075	**.763**
07 Course Content	**.868**	.136	.121
07 Faculty Research	.283	**.855**	-.097
09 Relevant Courses	**.856**	.103	.275
09 Student Exposure	.099	.011	**.858**
09 Course Impact/For-Profit Impact	**.807**	.068	.061
09 Faculty Research	.007	**.900**	.175

For divergent validity, the ranking on sustainability should not be a duplicate of a reputational ranking. In order to test the divergent validity of *The Global 100*, two other rankings of Business programs were used: *US News and World Report* and *The Financial Times*. Since all three are rankings, Spearman Rank Correlations were run. Results are in Table 4.

Table 3: Spearman Rank Correlations of Rankings

	The Financial Times	The Global 100	US News & World Report
The Financial Times	1.000 [n=55]	.303 [n = 31]	.794** [n = 55]
The Global 100	.303 [n = 31]	1.000 [n = 41]	.498** [n = 41]
US News and World Report	.794** [n = 55]	.498** [n = 41]	1.000 [n= 99]

Spearman Rank Order Correlations, * P < .05, ** P < .01

Results indicate that *US News & World Report* and *The Financial Times* rankings are rather similar. *The Global 100* ranking is rather different from *The Financial Times*. The *US News & World Report* also has some similarity with *The Global 100*. Overall, these results showing divergence indicate the rankings are likely measuring different construct.

Table 4: Chi Square Analysis. Ranking on *The Global 100* & Carnegie Basic Category

Presence in Global 100	Other	Mid Masters	Large Masters	Doctoral Research	Doc High Research	Doc Very High Resch	Total
No	99.0%	98.9%	94.7%	87.5%	77.9%	56.4%	89.2%
Mentioned	0%	.6%	2.7%	5.6%	4.2%	9.9%	3.2%
Ranked in 100	1.0%	.6%	2.7%	6.9%	17.9%	33.7%	7.6%
Total	96	176	72	72	95	101	878

Convergent Validity. As noted earlier, the idea behind convergent validity is that the ranking should have reasonable relationships with other rankings or measures that are believed to represent the construct of interest. We approach this idea by proposing that sustainability as defined in *The Global 100* should be related to the presence of STEM, health sciences, and engineering programs at the institution. The Carnegie Commission 2000 Basic categorizations, which range from Doctoral/Very High Research to Associate Degrees in Community Colleges,

were used to categorize institutions. Categories were collapsed into 6 levels by combining all categories of the 885 institutions that were not in the Mid-Sized Masters Programs or above into a category called "Other".

Table 5: Convergence with Institutional Characteristics

Finance	*r*	**UG Characteristics**	*r*
Average Faculty Salary	.417	6 Yr Grad Rates (Cohort)	.329
Instructional & Academic Support/FTE Student	.396	Freshman Retention Rates	.324
Tuition & Fee and State Revenue/FTE Student	.330	FR Applicants as % UG	.283
Net Tuition + State Dependency as % Core Revenues	-.309	% Cohort with Pell Grants	-.205
Institutional Support/FTE Student	.241	%UG 25 or older	-.194
Endowment/FTE Student	.214	UG (IS) Tuition and Fees	.179
Student Services $/FTE Student	.134	%UG Entering in FT FT DS Cohort	.170
Institutional Type		**Curriculum**	
*Basic Rated (Carnegie Category)	.397	Business Mgmt & Mktg MS	.317
Very High Research (1=yes; 0=no)	.386	MS/PhD STEM & Health Sciences as % Degrees	.285
Research & Service Expend. $/FTE Faculty	.364	Educ/Leisure/Family as % UG Bachelors	-.232
FTE Students	.306	Other STEM as % UG Bachelors	.224
Masters (1=yes; 1=no)	-.279	Education as % Graduate	-.201
First Prof and PhD's as % Degrees	.268	Engineering as % UG Bachelors	.179
Mid Masters (1=yes; 0=no)	-.148	% FTE Staff as Faculty	-.168

($p<.05$)

A composite variable was constructed after looking at the relationship of the Carnegie Basic Category and the presence of the institution on *The Global 100*.
Basic Rated: 6.00 = "Doctoral/Very Hi Research"; 5.00 = "Doctoral/Hi Research"; 4.00 = "Doctoral/Research"; 3.00 = "Large Masters"; 2.00 = "Mid Masters"; 1.00 = "Other"

Two analyses were conducted. The first examined the mean Rank of those on the list for 100. A One-Way ANOVA -- based on the exclusion of the Other Category -- was conducted. The result showed no significant differences in the standardized scores of the ranks ($F = 1.01$, $df = 4,61$, $p>.05$). This means that for the institutions on the list, their basic type is not related to the rank.

A second test used Chi Square with the data recoded as follows: "2" for institutions that were on *The Global 100* list, "1" for those that were mentioned, and "0" for neither of the above. This analysis was based on 885 institutions who had graduated 10 or more Masters of Business in 2009-2010 and were certified by regional accrediting agencies. The Chi Square from the second analysis was highly significant ($\chi = 171.3$, $df = 10$, $p<.001$). The results are shown in Table 4.

The results indicate that being considered or ranked is strongly related to offering PhD's, and being more likely to engage in funded research, i.e., Doctoral High Research or Doctoral Very High Research. What is not shown in the analysis is how this might be related to AACSB membership, whether this is in turn related to a higher resource base to support sustainability activities, and whether this directly impacts inclusion in *The Global 100.*

An analysis for convergence with institutional characteristics was performed using the sample of 885 institutions previously described and used with recoded criteria for *The Global 100* (2 = Top 100, 1 = Mentioned, 0 = other). Pearson correlations were computed for a set of 48 institutional characteristics. The results are shown below as Table 5 and are grouped by Finance, Institutional Type, UG Characteristics, and Curriculum.

CONCLUSIONS

The Global 100 ranking on sustainability appears to demonstrate acceptable internal consistency and to have adequate internal consistency, though it did better in 2007 than 2009. One concern is that the ranking uses multiple dimensions, two of which appear closely related. Nevertheless, the ranking appears to be sufficiently unique from business program reputational rankings that focus on constructs other than sustainability and that purport to measure quality. When institutional characteristics are added to the analysis, the ranking is sensitive to curriculum (which is expected) but also sensitive to institutional characteristics such as higher selectivity, socio-economic status, and the wealth of the institution (as represented by higher faculty salaries). While we anticipated that the sustainability ranking of business programs would be sensitive to the presence of STEM, health science, and engineering programs that were located in the institution, the analysis could not demonstrate this attribute. However, as expected, ranking among *The Global 100* is related to location of the business program in an institution with high-level research and doctoral programs. These findings clearly merit further investigation. For example, to confirm that *The Global 100* truly represent the schools and colleges of business that are leading in creation of higher education sustainability programs, issues surrounding resource availability, self-reported non-audited data, and the impact of non-business programs should be further explored.

As the Aspen Institute and other ranking organization move forward in further refining sustainability ranking, additional issues that should be resolved. First, though a description of the methodology is provided by most organizations, additional and full disclosure would improve the credibility of any ranking. (This is especially important with respect to inter-rater reliability, the use of fuzzy language (e.g. make "most" of the data available on-line), and descriptions of "smoothing" techniques.) Second, definitional objectivity and clarity are needed both for the term sustainability and for the categories used to operationalize the term. Third, the validity of syllabi for the purpose of sustainability course content that can be used in ranking must be demonstrated. Fourth, the issues surrounding the validity of self-reported data must continue, to include the use of third-party auditors to insure data integrity. (Schools would rather have a high ranking rather than a low ranking.) These recommendations would address many of the general threats to construct validity -- insufficient job of defining what is meant by the construct, insufficient capture of the breadth of the construct (related to the concept itself),

insufficient identification of required scope of measures, interaction of treatments leading to the results reported, interaction of testing and treatment, use of a label that is not a good description for what is observed, and researcher bias.

In the end, the unanswered questions -- how should the term "sustainability" be defined by higher education institutions and who should answer this question—will remain as obstacles to validating what it means to be a "Sustainable University." Furthermore, demonstrating construct validity in rankings is important since institutions, colleges, schools, and programs are increasingly using rankings to benchmark performance against referent institutions. Benchmarking has become a dominant methodology for determining performance relative to "Best Practices" in higher education. The danger is that sustainability will be defined by ranking organizations on criteria that are too narrow, i.e. as environmental sustainability or justice, thus distorting rankings and the broader intent of sustainability initiatives. Because schools and colleges of business are expected to be a major contributor to identifying solutions that address sustainable development, developing sustainability indicators that can be used for rankings and benchmarking is critical. Furthermore, reporting standards, indicators, and rankings must be validated. The same standards for transparency, accountability, and metrics that are demanded of corporations for reporting purposes should be the standard that schools and colleges of business apply to their institutions.

REFERENCES

B-Schools: Social and environmental rankings. n.d.. *Bloomberg Businessweek*.

Campbell, D.T. & Fiske, D.W. 1959. Convergent and discriminant validation by multi-trait-multimethod matrix. *Psycol. Bulletin*. 56: 81-105.

Glenn, D. 2010, December 3. How 3 graduate deans are putting the NRC rankings to use. *The Chronicle of Higher Education.*

Howard, R. D., McLaughlin, G.W., & McLaughlin, J.S. 1989. Bridging the gap between the data base and user in a distributed environment. *CAUSE/EFFECT. 12* (2). 19-25.

Hunter, B. 1995, College guidebooks: Background and development. In *Evaluating and Responding to College Guidebooks and Rankings. NDIR. 88.* Jossey-Bass Publishers.

Integrated Postsecondary Education Data System. n.d. http://nces.ed.gov/ipeds/about/

Jaschik, J. (2010, December 6). Using the rankings. *Inside Higher Ed*. http://www.insidehighered.com/news/2010/12/06/usnews

McGuire, M. D. 1995 Validity issues for reputational studies. In *Evaluating and Responding to College Guidebooks and Rankings. NDIR. 88.* Jossey-Bass Publishers.

Methodology. n.d. Beyond Grey Pinstripes. The Aspen Institute Center for Business and Education. http://www.beyondgreypinstripes.org/about/methodology.cfm

Morse, R.J. & Gilbert J. 1995. Publishers' perspectives: Magazines. In *Evaluating and Responding to College Guidebooks and Rankings. NDIR. 88.* Jossey-Bass Publishers.

Myers. L. & Robe, J. 2009. College rankings: History, criticism and reform. Center for College Affordability and Productivity. Washington, DC.

Peter, P.J. (1981). Construct validity: A review of the basic issues and marketing practices. *Journal of Marketing Research 18*. 133-145.

The Princeton Review's Guide to 286 Green Colleges. (n.d.) The Princeton Review.

Stuart, D.L. 1995. Reputational rankings: Background and development. In *Evaluating and Responding to College Guidebooks and Rankings. NDIR. 88.* Jossey-Bass Publishers.

The Financial Times. n.d. Retrieved December 1, 2010 from http://rankings.ft.com/businessschoolrankings/global-mba-rankings

The College Sustainability Report Card. (n.d.). The Sustainable Endowments Institute.

The Global 100. n.d. The Aspen Institute. Retrieved December 1, 2010 from http://www.beyondgreypinstripes.org/rankings/index.cfm

Trochim, W.M.K. 2006. The multitrait-multimethod matrix. Retrieved June 12, 2011 from *http://www.socialresearchmethods.net/kb/mtmmmat.php*.

U.S. News & World Report. n.d. America's best colleges. Retrieved Dec. 1, 2010 from http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-business-schools/items).

Yaniv, E. 2011. Construct clarity in theories of management and organization. *The Academy of Management Review, 36* (3). 590-592.