U.S. News & World Report College Rankings: Modeling Institutional Effects on Organizational Reputation

MICHAEL N. BASTEDO
University of Michigan

NICHOLAS A. BOWMAN
University of Notre Dame

Processes of certification and evaluation are some of the most powerful institutional forces in organizational fields, and in the higher education field, rankings are a primary factor in assessing organizational performance. This article explores the institutional effects of the *U.S. News & World Report* undergraduate rankings on the reputational assessments made by senior administrators at peer universities and liberal arts colleges. In the estimation of structural equation models, we found that published college rankings have a significant impact on future peer assessments, independent of changes in organizational quality and performance and even of prior peer assessments of reputation.

In 1983, when *U.S. News & World Report* published its first set of college rankings, Oberlin College ranked fifth among the nation's liberal arts colleges. This was hardly surprising as the ranking was based entirely upon reputation among its peers, and Oberlin had a long and storied history. It had a particularly strong reputation for placing students in top graduate programs, for decades ranking first among liberal arts colleges in the number of students who ultimately earned the doctorate (National Science Foundation 2006). Four years later, in 1987, Oberlin remained in the fifth position; the next year Oberlin was stunned to fall completely out of the top 10. Beginning in 1988, the *U.S. News* rankings methodology began to include quantitative measures of student and institutional characteristics that it believed were more objective measures of quality, and consequently peer reputation became a progressively smaller part of the overall ranking. Oberlin did not score at the top of these new measures, and its ranking suffered. By the mid-1990s, Oberlin was in

Electronically published December 18, 2009

American Journal of Education 116 (February 2010)

0195-6744/2010/11602-0001$10.00

serious danger of falling out of the top 25, which *U.S. News* defined as the "top tier" of liberal arts colleges. More curiously, Oberlin's peer reputation suffered during the period as well, over time moving closer and closer to its overall ranking, despite having changed little over the years. Was Oberlin no longer as good as it thought it was? Or were the rankings damaging its reputation among college leaders, even while its quality remained relatively unchanged?

Over the past decade, studying the effects of the rankings published in the *U.S. News & World Report* has become a minor industry (Bowman and Bastedo 2009; Elsbach and Kramer 1996; Griffith and Rask 2007; Martins 2005; McDonough et al. 1998; Meredith 2004; Monks and Ehrenberg 1999; Volkwein and Sweitzer 2006). The research has focused on two major areas. First, researchers have been interested in understanding the effects that rankings have on student behavior, especially college selection and choice. A smaller number of researchers have been interested in the impact of rankings on organizational identity and change. What has been underestimated, at least empirically, is the impact of college rankings on the organizational field of higher education.

Processes of certification and evaluation are some of the most powerful institutional forces in organizational fields (DiMaggio and Powell 1983; Dornbusch and Scott 1975; Espeland and Sauder 2007; Guler et al. 2002; Rao 1994; Sauder 2008; Sauder and Espeland 2009; Zuckerman 1999). At an individual level, evaluation is effective when it is perceived to be central to the work of the organization and when the organization is capable of influencing the evaluation. At an organizational level, the use of indirect indexes of organizational quality is one of the primary mechanisms of establishing organizational reputation (Perrow 1961). In the field of higher education organizations, prestige is one of the most important factors in assessing organizational performance, and the *U.S. News* rankings are the most prominent assessment of that performance. Even as administrators decry the influence of college rankings, these rankings become increasingly legitimate and nearly impossible to ignore, particularly for elite universities and liberal arts colleges. The rankings are thus a powerful force in shaping organizational decision making and identity (Elsbach and Kramer 1996; Espeland and Sauder 2007).

What is less well understood are the institutional effects that college rankings and other forms of certification have at a cognitive level, where the form of certification is an increasingly legitimate and unquestioned proxy for orga-

---

MICHAEL N. BASTEDO is associate professor of education at the Center for the Study of Higher and Postsecondary Education at the University of Michigan. NICHOLAS A. BOWMAN is a postdoctoral research associate in the Center for Social Concerns at the University of Notre Dame.

nizational reputation. We explored the institutional effects of the *U.S. News* undergraduate rankings on the reputational assessments made by the deans and presidents of peer universities and liberal arts colleges. To measure institutional effects, we examined the impact of early *U.S. News* rankings on future peer assessments of reputation while controlling for prior peer assessments and measures of instructional quality and organizational performance. In the estimation of a structural equation model, we have found that published college rankings have a significant impact on future peer assessments, independent of changes in organizational quality and performance and even of prior peer assessments of reputation. Prior peer assessments do not have the same effect on rankings, however, demonstrating the power of external assessment in institutional fields. We are thus able to demonstrate empirically how *U.S. News* rankings influence the organizational field of higher education.

Research Questions

Given the oft-stated importance of college rankings in higher education, there are surprisingly few studies of their impact on organizational behavior. Elsbach and Kramer (1996) found that initial rankings can create an organizational crisis of legitimacy, and McDonough et al. (1998) found that colleges are engaging in "reputation games" surrounding their placement in the rankings. In a study of law school reactions to the rankings, Espeland and Sauder (2007) found that the rankings have an inevitability that is interpreted as a self-fulfilling prophecy that drives adherence to the ranking norms. In business firms, competition for certification and reputation has been found to have strong effects in organizational fields as dissimilar as day care centers (Baum and Oliver 1992), the early automobile industry (Rao 1994), voluntary social service agencies (Singh et al. 1986), modern multinational firms (Guler et al. 2002), and even college basketball tournaments (Washington and Zajac 2005).

A rational choice perspective would argue that reputation is a function of information exchange and organizational performance as prestige in higher education is largely a function of instructional resources and financial performance (Brewer et al. 2001). As a result, we would not expect early rankings in the organizational field to have a disproportionate impact, and any differences between rankings and reputation would largely represent a time lag between changes in quality and expert recognition of those changes. The literature on signaling helps us to understand this process (Podolny 1993; Spence 1973). It is relatively well known that there can be substantial lags in time between shifts in quality and their perception in the organizational environment and that shifts in quality are often relatively unobservable to other organizations in the field. These organizations are likely to have good infor-

mation about only a small number of close competitors, and as a result they are likely to use other measurable indicators of financial and instructional health to gauge quality shifts.

A neo-institutional approach would see early rankings as crucially important to the shaping of cognitive perceptions of organizational prestige. In the field of higher education, the *U.S. News* organization serves as an ostensibly independent evaluator of college and university quality. Neo-institutional theory is a powerful approach to understanding these field-level dynamics (DiMaggio and Powell 1983; Schneiberg and Clemens 2006), particularly as they are connected to cognitive-level understandings that are highly legitimate and unquestioned (Espeland and Sauder 2007; Scott 2008). Within an institutional theory framework, evaluations within an organizational field—if they are perceived to be legitimate, important, and malleable over time—will have effects on future assessments of reputation independent of organizational quality and performance. Further, as rankings are driving professional assessment of reputation, early peer reputation assessments will not influence future college rankings.

Question 1: Controlling for changes in instructional quality and financial performance, as well as early peer assessments, do early college rankings have a significant and concomitant influence on future peer assessments of reputation?

Question 2: Controlling for changes in instructional quality and financial performance, as well as early college rankings, are early peer assessments reputation significantly associated with future college rankings?

Institutional theory would also predict that, when certifications are highly legitimate, the formal structure of the evaluation can have institutional effects that are independent of prior reputation and organizational performance (Meyer and Rowan 1977). As the form of the evaluation becomes increasingly familiar and integrated in the thinking of participants, the form attains a legitimacy that is as powerful as the substance of the evaluation itself (Bowman and Bastedo 2009; Leung 2007; Rao et al. 2003, 2005; Zuckerman 1999). The form itself asserts a categorical membership that defines the identity of organizations, which in turn has consequences for resources and ultimately organizational survival (Bastedo and Bowman 2009; Hannan et al. 2007). These effects are enhanced in organizational fields where inputs and outputs are highly ambiguous and difficult to assess; thus research has demonstrated similar effects in fields such as French gastronomy (Rao et al. 2005), Italian winemaking (Leung 2007), and securities analysis for American stock investors (Zuckerman 1999).

In the case of the *U.S. News* rankings, changes in ranking can be accompanied by changes in tier, which are roughly qualitative groupings by quartile. Thus a small change in ranking can lead to complete change of category as

a college moves from one tier to another. In addition, *U.S. News* has provided us with a natural experiment; there have been periods where the magazine has changed how tiers are calculated, so that changes in tier placement are completely unrelated to changes in ranking or the indicators. Institutional theory predicts that tier level will influence not only future peer assessments but also changes in tier on future peer assessments of reputation, because the rankings and tier placements are so legitimate that they alter the internal assessments of reputation made by college leaders.

Question 3: Controlling for changes in instructional quality and financial performance, as well as early peer assessments, does tier level within the *U.S. News* rankings have a significant and concomitant influence on future peer assessments of reputation?

Question 4: Controlling for changes in instructional quality and financial performance, as well as early peer assessments and tier level, do changes in *U.S. News* tier level have a significant and concomitant influence on future peer assessments of reputation?

These research questions are offered net of other organizational influences, including instructional quality and financial performance, which are used as objective indicators of quality in college rankings. Importantly, the indicators being reported are entirely transparent to the magazine's audience, allowing rankings consumers to make their own judgments of institutional quality and to gauge shifts in that quality over time.

Method

The data were taken from print editions of the "America's Best Colleges" issues of *U.S. News & World Report*. To examine the impact of overall ranking, we chose to use the longest time period available. This included information from the top 25 liberal arts colleges and the top 25 national universities in 1989, the first year in which overall rankings and peer assessment ratings were both published. Information from 1995, 2000, and 2006 for these same institutions was also collected; in 2006, all of these top liberal arts colleges were still ranked in the top 30 and all of the national universities were still ranked in the top 27.

To examine the effect of tier placement, we were not able to use data from the same years, because *U.S. News* only ranked the top 25 institutions before 1990, and peer assessment ratings for institutions below the top 25 were not provided until 1992. At this point, institutions were grouped into four tiers, with Tier 1 as the highest and Tier 4 as the lowest, and institutions in Tiers 2–4 were listed within each tier in alphabetical order. The final year in which tier information was provided in the print edition of *U.S. News* was 2000. To

determine the impact of tier placement and tier change, we collected data from the longest possible time period during which relevant information for all institutions was available, from 1992 to 2000. Institutions were included in this sample if there were complete data on all variables ($n = 119$ for liberal arts colleges, $n = 168$ for national universities). The number of liberal arts colleges listed in each tier changed between 1992 and 2000, which will be discussed in more detail below.

Measures

Dependent variable.—In the models that directly tested questions 1, 3, and 4, the dependent variable was the peer assessment rating for the most recent year available. The way in which peer assessment ratings was measured varied over time. Before 1998, deans and presidents were asked to arrange institutions within the same organizational type into quartiles. The ratings of these respondents were compiled and then converted into peer assessment ratings, in which the top institution received a 1, the next received a 2, and so forth. After 1998, deans and presidents rated institutions on a 1 (marginal) to 5 (distinguished) scale, and a peer assessment rating was computed by calculating the mean response of participating deans and presidents. Despite this measurement change, the correlations between earlier and later ratings are extremely strong (e.g., for peer assessment ratings of liberal arts colleges in 1989 and 2006, $r = -.86$, and for liberal arts colleges in 1992 and 2000, $r = -.96$).

Independent variables.—For all models, early overall ranking was included, in which the top institution received a ranking of 1, the next institution received a 2, and so forth. (In addition, for models that tested question 2, overall ranking in 2006 was used as the dependent variable.) *U.S. News* computes the overall college rankings using a number of factors. Peer assessments of institutional reputation constitute the largest portion (25 percent of the overall score), and other factors include student retention (20 percent), faculty resources (20 percent), student selectivity (15 percent), financial resources (10 percent), graduation rate performance (5 percent), and alumni giving rate (5 percent). The factors are weighted according to those proportions and combined to create an overall score for each institution. The final scores are rescaled so that the top institution receives a score of 100 and the other schools' scores are a proportion of the top score. Institutions are then ranked according to their overall score. If two or more schools receive the same score, then they receive the same rank (i.e., there is no tiebreaker).

To compare the top 25 institutions in 1989 and 2006, we used four measures to compute the control variable: graduation and retention rank, faculty resources rank, selectivity rank, and financial resources rank. For all of these

variables, the top school received a 1, the next-highest school received a 2, and so forth. To determine whether institutions had improved or regressed over time relative to other schools, the 2006 rank for each variable was subtracted from the 1989 rank, with improvements indicated by negative scores. For example, if a school was ranked number 20 in selectivity in 1989 and number 10 in 2006, its selectivity change score would be $10 - 20 = -10$. Because there were so few cases for these top institutions, an overall change score was computed by calculating the mean of the four change scores for graduation and retention rank, faculty resources rank, selectivity rank, and financial resources rank; this variable was used to control for changes in overall institutional quality.

Most of the data used to compute the change scores was only published for institutions in the top tier. Therefore, to compute similar change scores for all national universities and liberal arts colleges in 1992 and 2000, freshmen retention rate, graduation rate, proportion of freshmen in the top 10 percent of their high school class, and acceptance rate were used. To compute change scores for each of the four control variables, the 1992 score was subtracted from the 2000 score (see app. A in the online version of this article). In addition, dummy-coded variables to indicate tier level in 1992 were created. Finally, change in tier level was calculated by subtracting tier level in 2000 from tier level in 1992; therefore, positive values indicate an improvement in tier (e.g., moving from Tier 3 to Tier 2), and negative values indicate movement into a less prestigious tier. To ensure that the relative variances among all variables were approximately equal for the analyses described below, all continuous variables were standardized with a mean of zero and a standard deviation of one.

Analyses

The analyses were designed to address three separate yet related questions. First, do overall rankings influence future peer assessment ratings? Second, does tier level influence future peer assessment ratings? Third, does switching between tiers result in changes in future peer assessment ratings? Because the correlations between two independent variables—peer assessment ratings and overall rankings—were extremely high (for liberal arts colleges, $r = .78$ in 1989 and $r = .75$ in 2006), structural equation modeling (SEM) was employed. SEM accounts for high multicollinearity by incorporating the correlations between independent variables into the model, which in turn yields accurate estimates of the unique relationships between the independent and dependent variables (Byrne 2006; Kline 2005). When modeling the impact of college rankings on other indicators, some previous researchers (e.g., Griffith and

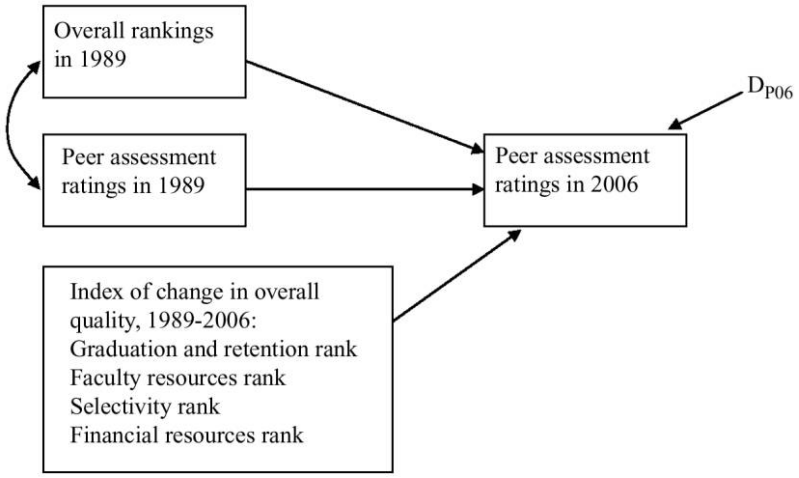

FIG. 1.—Structural equation model for top 25 institutions

Rask 2007; Volkwein and Sweitzer 2006) have used ordinary least squares (OLS) multiple regression analyses with several measures of institutional prestige as independent variables; however, OLS multiple regression is not robust to high multicollinearity and can yield inaccurate results under these circumstances (Pedhazur 1997). Furthermore, in our analyses, we intended to examine and potentially rule out two alternative explanations that are not addressed in previous research: (*a*) that changes in peer assessment ratings merely reflect changes in institutional quality over time and (*b*) that the relationship between previous overall rankings and future peer assessment scores is merely a statistical artifact.

Using the SEM software program EQS 6.1 for Windows, we analyzed covariance matrices of the data using maximum likelihood estimation. All variables were observed, so it was not necessary to perform confirmatory factor analyses or to create measurement models. For the top 25 institutions, structural equation models were created to analyze the effect of 1989 overall rankings, 1989 peer assessment ratings, and change in institutional quality on 2006 peer assessment ratings (question 1). Separate analyses were performed for liberal arts colleges and national universities. In addition, since the proportion of cases to parameters for these analyses is less than 5 : 1 (Kline 2005), an additional SEM model was created that combined liberal arts colleges and national universities. To account for the high multicollinearity between peer assessment ratings and overall rankings, the models included a correlational path between 1989 overall rankings and 1989 peer assessment ratings (see fig. 1).

Additional analyses were also performed to determine when the influence of

peer assessment ratings was strongest. These analyses used early peer assessments, early college rankings, and change in quality as independent variables, with later peer assessment ratings as the dependent variable. The years included 1989–95, 1995–2000, and 2000–2006. Separate models for national universities, liberal arts colleges, and all institutions were examined.

Since overall rankings are composed (in part) by peer assessment ratings, it is possible that there will be an artifactual relationship between overall rankings and peer assessment ratings, which is potentially problematic for the analyses of the top 25 institutions. That is, a significant relationship may result not from a causal impact of overall rankings on future peer assessment ratings but from the way in which overall rankings are computed. As a result, we created additional structural equation models—one for national universities, one for liberal arts colleges, and one that combined both types of institutions—to rule out this possibility. These models used the same independent variables but used 2006 overall ranking as the dependent variable. If high-level college administrators used overall rankings as a basis for forming judgments about peer institutions, then we would expect that overall rankings cause changes in future peer assessment ratings, but we would not expect peer assessment ratings to "cause" changes in future overall rankings (question 2). Similar analyses with overall ranking as the dependent variable were also conducted for the shorter time intervals (i.e., 1989–95, 1995–2000, and 2000–2006). These "artifactual" models are designed to provide a more rigorous test of causality arguments by formally testing additional models in which we do not expect to reject the null hypothesis. Moreover, this method of accounting for multicollinearity is superior to using change in peer assessments as a dependent variable, since using a change variable would ignore potential ceiling effects of peer assessment ratings for the top institutions (Nunnally 1982).

For the models that examined the influence of tier level on future peer assessment ratings (question 3), the independent variables included peer assessment ratings in 1992; dummy-coded variables to indicate whether an institution was listed as Tier 2, Tier 3, or Tier 4 in 1992 (with Tier 1 as the referent group); and changes in graduation rate, freshmen retention rate, proportion of students in top 10 percent of high school class, and acceptance rate between 1992 and 2000. The dependent variable was peer assessment ratings in 2000. Separate analyses were performed for liberal arts colleges and national universities. As with the previous models, there was multicollinearity among the independent variables, but it was often unclear in advance which variables would have high or moderate intercorrelations. Because our theoretical interests rested solely with the effect of the independent variables on the dependent variable, we added correlational paths between two variables whenever doing so would result in a substantial reduction ($p < .01$) of the chi-square statistic (Kline 2005).

Two additional models were analyzed to determine whether a change in tier, independent of initial tier placement, would contribute to changes in future peer assessment ratings (question 4). These two models included the same independent variables as the tier-level models but had one additional variable that indicated change in tier level between 1992 and 2000. As with the previous tier-based models, separate analyses were conducted for national universities and liberal arts colleges, and correlational paths between independent variables were added when they significantly improved model fit.

Assumptions of Structural Equation Modeling

Preliminary analyses suggested that the assumptions of SEM had generally been met. For the top 25 institutions, there were no outliers on any variable where a value was greater than three standard deviations away from the mean. In contrast, there were several such cases for the tier-based models for liberal arts colleges and national universities. These outliers only occurred for variables that controlled for change in institutional quality over time, such as change in freshmen retention rate. Preliminary analyses showed that SEM results did not vary substantively when these cases were excluded, so they were kept in the model. The univariate normality—as gauged through skew and kurtosis statistics—was acceptable for variables in all models. There was only one variable in one model in which the kurtosis statistic was greater than that recommended by Kline (2005),[1] and skew statistics for all variables in all models were within the recommended range (for univariate kurtosis and skew statistics for all variables, see app. B in the online version of this article).

In addition, multivariate kurtosis statistics were computed to indicate whether particular cases might have had an unusually high impact on the SEM results. For this purpose, EQS 6.1 provides Mardia's coefficient and normalized estimates. Absolute values greater than 3 represent a nontrivial degree of multivariate kurtosis, but modeling statistics may not be affected until kurtosis values reach an absolute value of 5 or higher (Bentler 2006). In the current study, these statistics were excellent for models of the top 25 institutions and the liberal arts tier models, with all normalized estimates having absolute values less than 2, but these statistics were larger for the tier-level and tier-change models for national universities (normalized estimates of 7.9 and 7.5, respectively). Additional analyses suggested that these coefficients were not the product of a single substantial outlier, so removing a single problematic case could not reduce the multivariate kurtosis. As such, the results of the national university tier-based models should be compared with the corresponding models for liberal arts colleges before any strong conclusions can be drawn.

Furthermore, SEM analyses assume that the variances of all variables are roughly similar; for instance, the model would be distorted if one variable had a standard deviation of 0.5 and another had a standard deviation of 50. Differences among relative variances can become problematic if the standard deviations of any two variables differ by a factor of 10 or greater (Kline 2005). As noted earlier, the continuous variables in the model were standardized to remedy this potential concern.

Finally, although SEM can account for high multicollinearity, Kline (2005) suggests that correlations above .85 between any two variables in the model might indicate that these variables are redundant and/or are measuring the same construct. In the current study, very high correlations were found between peer assessment ratings in 1992 and peer assessment ratings in 2000 for national universities ($r = -.94$) and for liberal arts colleges ($r = -.96$). While Kline cautions against the use of variables that are very highly correlated, this extreme correspondence works against our predictions for questions 3 and 4, since we expect factors other than previous peer assessment ratings to affect future peer assessment ratings. In addition, variance inflation factors (VIF) for all variables were computed to determine whether any single variable was redundant. Calculating VIF provides information beyond that obtained through Pearson correlations, because VIF can identify a variable that is not very highly correlated with one other variable yet the vast majority of its variance is already captured through a combination of other variables (Kline 2005). The VIF statistic is computed with the formula $1/(1 - R^2)$, where R^2 represents the squared multiple regression of a given variable regressed on all other variables in the model, with values greater than 10 suggesting that a variable may be redundant. The only variables in any model with VIFs greater than 10 were peer assessment ratings in 1992 and peer assessment ratings in 2000 (this was the case for national universities and liberal arts colleges in both tier-level and tier-change models). Once again, this high level of correspondence works against our expectations in questions 3 and 4; therefore, no adjustments to the models were made.

Results

Top 25 Institutions

Among liberal arts colleges, 1989 overall rankings successfully predicted 2006 peer assessment ratings ($\beta = -.49$ $p < .001$).[2] The Bentler-Bonnet normed fit index (NFI), the comparative fit index (CFI), and the chi-square statistic (χ^2), which are three commonly used goodness-of-fit measures, were all excellent for this model (NFI = .994, CFI = 1.000, $\chi^2(2) = .48$, NS).[3] Furthermore,

TABLE 1

Maximum Likelihood Parameter Estimates of Direct Effects for Two Structural Equation Models Explaining Peer Assessment Ratings in 2006 among Top 25 Institutions

DIRECT EFFECT ON 2006 PEER ASSESSMENT	LIBERAL ARTS COLLEGES			NATIONAL UNIVERSITIES		
	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
1989 overall rankings	−.517***	.107	−.488	−.238*	.104	−.224
1989 peer assessment ratings	−.517***	.107	−.488	−.804***	.104	−.755
Change in quality	−.254***	.067	−.240	−.177**	.063	−.166

* $p < .05$.
** $p < .01$.
*** $p < .001$.

these findings were replicated for national universities (NFI = .911; CFI = .928; $\chi^2(2) = 7.812$, $p < .03$; $\chi^2/df = 3.91$); that is, overall rankings in 1989 were a significant predictor of peer assessment ratings in 2006 ($\beta = -.22$, $p < .03$), even when controlling for previous peer assessment ratings and changes in institutional quality (see table 1). In addition, when all institutions were included in the same model (NFI = .953; CFI = .968; $\chi^2(2) = 5.593$, $p = .06$; $\chi^2/df = 2.80$), the independent effect of overall rankings on future peer assessment ratings persisted ($\beta = -.31$, $p < .005$; see table 2). In all three models, improvements in institutional quality led to higher peer assessment ratings (all p's $< .005$). In sum, these patterns are quite consistent with our expectations for question 1.

Additional analyses suggest that most of the impact of college rankings occurs in the earlier years of the rankings. Specifically, overall rankings in 1989 have a significant effect on peer assessment ratings in 1995 for all institutions ($\beta = .26$, $p = .01$) and for liberal arts colleges ($\beta = .40$, $p < .01$). Moreover, overall rankings in 1995 significantly affect peer assessment ratings in 2000 for national universities ($\beta = .13$, $p < .04$), liberal arts colleges ($\beta = .27$, $p < .01$), and all institutions ($\beta = .17$, $p < .08$). However, overall rankings in 2000 did not significantly predict peer assessments in 2006.

Importantly, as predicted, the artifactual models did not display a significant relationship between 1989 peer assessment ratings and 2006 overall rankings. Specifically, the paths between these two variables for liberal arts colleges ($\beta = -.00$, NS), for national universities ($\beta = .10$, NS), and both institutional types combined ($\beta = .03$, NS) were not significant. The goodness-of-fit measures for the liberal arts college (NFI = .993; CFI = 1.000; $\chi^2(2) = .48$, NS), national

TABLE 2

Maximum Likelihood Parameter Estimates of Direct Effects for a Structural Equation Model Explaining Peer Assessment Ratings in 2006 in Top 25 Institutions Combined

Parameter	Unstandardized	SE	Standardized
1989 overall rankings	−.339**	.119	−.310
1989 peer assessment ratings	−.629***	.119	−.575
Change in quality	−.298***	.073	−.272

** $p < .01$.
*** $p < .001$.

university (NFI = .894; CFI = .913; $\chi^2(2) = 7.812$, $p < .03$; $\chi^2/df = 3.91$), and combined models (NFI = .956; CFI = .970; $\chi^2(2) = 5.593$, $p = .06$; $\chi^2/df = 2.80$) ranged from acceptable to excellent. Furthermore, none of the effects for the shorter time periods (e.g., 1989 peer assessments predicting 1995 overall rankings) were significant. These findings provide additional support for the notion that overall rankings have a causal impact on future peer assessments of reputation.

Tier Systems

Initial tier level.—After adding correlational paths among the independent variables, the model for liberal arts colleges fit the data well (NFI = .943; CFI = .975; $\chi^2(21) = 35.529$, $p < .03$; $\chi^2/df = 1.69$). As predicted for question 3, even when controlling for peer assessment ratings in 1992 and changes in institutional quality, Tier 2 and Tier 3 institutions had lower peer assessment ratings in 2000 than did Tier 1 institutions ($\beta = -.16$, $p < .001$, and $\beta = -.15$, $p < .005$, respectively). However, somewhat surprisingly, there was no difference in 2000 peer assessment ratings between Tier 4 and Tier 1 institutions ($\beta = -.06$, NS; see table 3). This exact pattern was replicated for national universities (NFI = .952; CFI = .976; $\chi^2(20) = 38.090$, $p < .01$; $\chi^2/df = 1.90$). All else equal, universities that were labeled as Tier 2 or Tier 3 in 1992 received significantly lower peer assessment ratings in 2000 than did Tier 1 universities ($\beta = -.17$, $p < .001$, and $\beta = -.09$, $p < .03$, respectively), but there was no such difference between Tier 4 and Tier 1 universities ($\beta = -.02$, NS).

Furthermore, the control variables that adjusted for change in institutional quality did little to improve the models. For liberal arts colleges, none of the four variables significantly predicted future peer assessment ratings. For national universities, two of the four predictors were significant: increases in

TABLE 3

Maximum Likelihood Parameter Estimates of Direct Effects for Two Structural Equation Models Explaining Peer Assessment Ratings in 2000 in Tiers 1–4

	LIBERAL ARTS COLLEGES			NATIONAL UNIVERSITIES		
DIRECT EFFECT ON 2000 PEER ASSESSMENT	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
1992 peer assessment ratings	−.880***	.051	−.863	−.971***	.046	−.904
Change in freshmen retention rate	.027	.027	.026	.050*	.025	.050
Change in graduation rate	.051	.026	.050	−.068*	.027	−.061
Change in freshmen in top 10% of high school class	−.019	.025	−.019	.014	.025	.014
Change in acceptance rate	.030	.025	.029	−.006	.024	−.007
Tier 4 in 1992	−.198	.153	−.064	−.054	.126	−.020
Tier 3 in 1992	−.334**	.111	−.146	−.207*	.092	−.089
Tier 2 in 1992	−.364***	.079	−.161	−.366***	.070	−.170

* $p < .05$.
** $p < .01$.
*** $p < .001$.

freshmen retention rates led to higher peer assessment ratings, but increases in graduation rates were negatively associated with future peer assessments of reputation.

Changes in tier level.—Controlling for peer assessment in 1992, tier level in 1992, and changes in institutional quality, change in tier level had a significant effect on peer assessment ratings in 2000 for national universities ($\beta = .05$, $p < .05$) and a marginally significant effect for liberal arts colleges ($\beta = .05$, $p < .06$; see table 4). These results are quite consistent with our predictions for question 4. For both national universities and liberal arts colleges, the models fit the data well (NFI = .925; CFI = .955; $\chi^2(27) = 63.045$, $p < .001$; $\chi^2/df = 2.34$; and NFI = .920; CFI = .958; $\chi^2(28) = 54.250$, $p < .005$; $\chi^2/df = 1.94$, respectively). The same patterns for changes in institutional quality noted in the previous tier-level models were also found for the tier change models.

Limitations

There are several limitations in this study that are worth mentioning. First, only liberal arts colleges and national universities are examined, because the classification systems for what are now known as master's universities and comprehensive colleges have changed substantially over time. As a result, any attempt to track these institutions longitudinally would yield a great deal of missing data. Second, because we wanted to ensure that we maintained independent observations within our sample, we included only two years—one start point and one end point—in our analyses. Third, we were not able to use the earliest possible version of the rankings because not enough information was provided in the *U.S. News* tables. Specifically, 1988 was the first year in which a variety of measures were used to compute the rankings (previously, the college "rankings" were determined exclusively through a biannual survey), but peer assessment ratings were not made available until the following year. In addition, schools beyond the top 25 were first listed in *U.S. News* in 1990, but peer assessments were not provided for these schools until 1992. Therefore, we used the earliest and most recent years that were available from these data.

Discussion

Overall, the results were quite consistent with predictions drawn from institutional theory. For national universities and liberal arts colleges, future peer assessments of reputation are substantially influenced by (*a*) overall rankings, (*b*) tier level, and (*c*) changes in tier level. Moreover, these effects are found

TABLE 4

Maximum Likelihood Estimates of Direct Effects for Two Structural Equation Models Explaining Peer Assessment Ratings in 2000 in Tiers 1–4

	LIBERAL ARTS COLLEGES			NATIONAL UNIVERSITIES		
	Unstandardized	SE	Standardized	Unstandardized	SE	Standardized
1992 peer assessment ratings	−.844***	.054	−.826	−.943***	.048	−.872
Change in freshmen retention rate	.022	.026	.021	.053*	.024	.052
Change in graduation rate	.041	.026	.040	−.071**	.026	−.063
Change in freshmen in top 10% of high school class	−.022	.025	−.021	.009	.025	.009
Change in acceptance rate	.041	.025	.040	.002	.024	.003
Tier 4 in 1992	−.322^{+}	.167	−.104	−.168	.137	−.060
Tier 3 in 1992	−.418***	.118	−.183	−.290**	.099	−.125
Tier 2 in 1992	−.421***	.083	−.186	−.394***	.070	−.181
Change in tier level	.049^{+}	.026	.049	.050*	.025	.050

$^{+}$ $p < .10$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

even when controlling for previous peer assessments of reputation and other relevant factors. This lends strong support for the idea that the *U.S. News* rankings have institutional effects in the field of higher education. Due to controls for transparent measures of instructional quality and performance, we can also reject the hypothesis that differences in rankings and reputation derive from a time lag between quality shifts and expert recognition. Even higher education experts, who might normally be expected to have relatively stable assessments of reputation over time, are substantially influenced by rankings that many of them ostensibly disdain. These effects are robust across both national universities and liberal arts colleges despite strong differences in organizational size, network structure, and missions between these fields.

The differences between top-tier and lower-tier institutions are fairly straightforward to interpret. Colleges and universities that are initially perceived as reputable suffer a substantial setback if they are not recognized as being in *U.S. News*'s "top tier"; furthermore, this negative effect is most pronounced for those institutions that are closest to—but not quite within—this upper echelon. In contrast, institutions in the bottom tier are generally not mentioned in the same context as the nation's top colleges and universities. The mere inclusion of these less prestigious institutions in the *U.S. News* national rankings bestows reputational benefits.

These results demonstrate that the formal structure of external evaluations serves an institutional purpose similar to the formal structure of organizations (Meyer and Rowan 1977). Looking at the rankings as simply pure numbers highly underestimates the power generated by the rankings as an organized form. We can thus expect the organized form of the rankings to impose a discipline upon ranked organizations in very specific ways that shape organizational responses through decoupling and other measures (Espeland and Stevens 1998; Sauder and Espeland 2009). Rankings thus become a form of interorganizational dependency that result in a multitude of organizational behaviors including bridging and co-optation techniques (Bastedo and Bowman 2009). We can see this most clearly through the multiple ways in which institutions seek to manipulate the data provided to *U.S. News* (Sauder and Fine 2008; Stevens 2007) and organize to oppose the rankings on principle (Thacker 2005).

As many in higher education believe that the *U.S. News* rankings are shallow or even misleading (McDonough et al. 1998; Thacker 2005), and acknowledging the widespread evidence that rankings have significant effects on college choice at both undergraduate and graduate levels (Bowman and Bastedo 2009; Martins 2005; Monks and Ehrenberg 1999; Sauder and Lancaster 2006), the additional institutional effects of *U.S. News* on expert opinion are particularly significant. From a methodological perspective, this effect may distort the quality of the reputation scores that are used in calculating the rankings. The

more that the reputations of colleges are affected by the ranking, despite other evidence to the contrary, the more the *U.S. News* rankings become college reputation. Reputation scores are thus likely to have decreased value over the long run.

The impact of rankings on reputation has other effects. Differences in rankings over time are vanishingly small and represent only minute differences in the measures of performance used by *U.S. News*. Yet even experts interpret these differences as significant differences in university reputation. Students, in turn, rely upon these rankings to make college choices, affecting primarily the admissions indicators that form the majority of the *U.S. News* ranking (Bowman and Bastedo 2009; Monks and Ehrenberg 1999). The result is a Matthew effect as colleges who are already privileged in the field make additional gains while their competitors—who are essentially identical on most meaningful indicators—fall further behind (Merton 1968). Thus higher education becomes a winner-take-all market where marginal differences in performance lead to large differences in reputation and resources (Frank and Cook 1996). When prestige is academic currency, the result is a "positional arms race," where colleges spend significant resources to attract students who differ only marginally on indicators of quality.

It will be fascinating to see if the effects demonstrated here persist over time. In recent years, there has been a proliferation of college rankings by other magazines, external organizations, and university institutes, both in the United States and around the world (Institute for Higher Education Policy 2007). Although the *U.S. News* rankings remain the most prominent in the United States, there is emerging competition that seeks to improve the fit between rankings and student concerns and thus is highly diverse and differentiated. Rankings from *Washington Monthly* and *Mother Jones* focus on institutions that serve the public good; rankings from The Center for Measuring University Performance focus on measuring research productivity; and rankings from *The Advocate* and *Black Enterprise* rank the best colleges for gay students and black students, respectively. Jiao Tong University in Shanghai has made its name by providing a single ranking of world universities, and European university leaders are developing their own ranking. Economists are seeking to create unobtrusive measures of college ranking based upon revealed student preferences (Avery 2005). The *U.S. News* rankings themselves have become increasingly differentiated, allowing virtually unknown colleges to call themselves the *U.S. News* leader among "Southern baccalaureate colleges" (Ouachita Baptist University) or "Midwestern master's universities" (Creighton University). With so many versions of college rankings entering the public consciousness, conceptions of college reputation are likely to become increasingly multidimensional.

Notes

1. Kline (2005) suggests that skew statistics less than 3 and kurtosis statistics less than 10 are satisfactory. He also notes that there is little consensus for a value that represents a potentially high level of kurtosis; various authors have suggested values ranging from 8 to 20. The kurtosis statistic for change in freshmen retention rate for national university tier models was 11.97.

2. Because lower values reflect better overall rankings in the *U.S. News* system (e.g., 1 is best possible ranking, 2 is next highest, etc.), negative beta coefficients indicate that better overall rankings are associated with higher future peer assessment ratings.

3. In general, adequate values for SEM goodness-of-fit measures are at least .90 for NFI and CFI and a ratio of χ^2 to *df* less than 5.0 (Byrne 2006).

References

Avery, Christopher, Mark E. Glickman, Caroline Minter Hoxby, and Andrew Metrick. 2005. "A Revealed Preference Ranking of Colleges and Universities." Working Paper no. W10803, National Bureau of Economic Research, Cambridge, MA.

Bastedo, Michael N., and Nicholas A. Bowman. 2009. "College Rankings as an Interorganizational Dependency: Establishing the Foundation for Strategic and Institutional Accounts." Working paper, School of Education, University of Michigan.

Baum, Joel A. C., and Christine Oliver. 1992. "Institutional Embeddedness and the Dynamics of Organizational Populations." *American Sociological Review* 57 (4): 540–59.

Bentler, Peter M. 2006. *EQS 6 Structural Equations Program Manual*. Encino, CA: Multivariate Software, Inc.

Bowman, Nicholas A., and Michael N. Bastedo. 2009. "Getting on the Front Page: Organizational Reputation, Status Signals, and the Impact of *U.S. News & World Report* Rankings on Student Decisions." *Research in Higher Education* 50 (5): 415–36.

Brewer, Dominic J., Susan M. Gates, and Charles A. Goldman. 2001. *In Pursuit of Prestige: Strategy and Competition in U.S. Higher Education*. Somerset, NJ: Transaction.

Byrne, Barbara M. 2006. *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: Erlbaum.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147–60.

Dornbusch, Sanford M., and W. Richard Scott. 1975. *Evaluation and the Exercise of Authority*. San Francisco: Jossey-Bass.

Elsbach, Kimberly D., and Roderick M. Kramer. 1996. "Members' Responses to Organizational Identity Threats: Encountering and Countering the *Business Week* Rankings." *Administrative Science Quarterly* 41 (3): 442–76.

Espeland, Wendy Nelson, and Michael Sauder. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113 (1): 1–40.

Espeland, Wendy Nelson, and Mitchell L. Stevens. 1998. "Commensuration as a Social Process." *Annual Review of Sociology* 24 (1): 313–43.

Frank, Robert H., and Philip J. Cook. 1996. *The Winner-Take-All Society*. New York: Penguin.

Griffith, Amanda, and Kevin Rask. 2007. "The Influence of the *U.S. News & World*

Report Collegiate Rankings on the Matriculation Decision of High-Ability Students, 1995–2004." *Economics of Education Review* 26 (2): 244–55.
Guler, Isin, Mauro F. Guillen, and John Muir MacPherson. 2002. "Global Competition, Institutions and the Diffusion of Organizational Practices: The International Spread of ISO 9000 Quality Certificates." *Administrative Science Quarterly* 47 (2): 207–32.
Hannan, Michael T., Laszlo Polos, and Glenn R. Carroll. 2007. *Logics of Institutional Theory: Audiences, Codes, and Ecologies*. Princeton, NJ: Princeton University Press.
IHEP (Institute for Higher Education Policy). 2007. *College and University Ranking Systems: Global Perspectives and American Challenges*. Washington, DC: IHEP.
Kline, Rex B. 2005. *Principles and Practice of Structural Equation Modeling*. New York: Guilford.
Leung, Ming D. 2007. "Category Cognition: Audience Recognition of Winemaker Categories in the Barolo and Barbaresco Winemaking District." Paper presented at the annual meeting of the American Sociological Association, New York, August.
Martins, Luis L. 2005. "A Model of the Effects of Reputational Rankings on Organizational Change." *Organization Science* 16 (6): 701–20.
McDonough, Patricia M., Anthony Lising Antonio, Marybeth Walpole, and Leonor Xochitl Perez. 1998. "College Rankings: Democratized College Knowledge for Whom?" *Research in Higher Education* 39 (5): 513–37.
Meredith, Marc. 2004. "Why Do Universities Compete in the Rankings Game? An Empirical Analysis of the Effects of the *U.S. News & World Report* College Rankings." *Research in Higher Education* 45 (5): 443–61.
Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159 (3810): 56–63.
Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340–63.
Monks, James, and Ronald G. Ehrenberg. 1999. "The Impact of *U.S. News & World Report*: College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions." Working Paper no. 7227, National Bureau of Economic Research, Cambridge, MA.
NSF (National Science Foundation). 2006. *U.S. Doctorates in the 20th Century*. Washington, DC: National Science Foundation.
Nunnally, Jum C. 1982. "The Study of Human Change: Measurement, Research Strategies, and Methods of Analysis." In *Handbook of Developmental Psychology*, ed. Benjamin B. Wolman. Englewood Cliffs, NJ: Prentice-Hall.
Pedhazur, Elazar J. 1997. *Multiple Regression in Behavioral Research*. New York: Wadsworth.
Perrow, Charles. 1961. "Organizational Prestige: Some Functions and Dysfunctions." *American Journal of Sociology* 66 (4): 335–41.
Podolny, Joel M. 1993. "A Status-Based Model of Market Competition." *American Journal of Sociology* 98 (4): 829–72.
Rao, Hayagreeva. 1994. "The Social Construction of Reputation: Certification Contests, Legitimation, and the Survival of Organizations in the American Automobile Industry, 1895–1912," in "Competitive Organizational Behavior," special issue, *Strategic Management Journal* 15 (Winter): S29–S44.
Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand. 2003. "Institutional Change in Toque Ville: Nouvelle Cuisine as an Identity Movement in French Gastronomy." *American Journal of Sociology* 108 (4): 795–843.
Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand. 2005. "Border Crossing: Bricolage and Erosion of Categorical Boundaries in French Gastronomy." *American Sociological Review* 70 (6): 968–91.
Sauder, Michael. 2008. "Interlopers and Field Change: The Entry of *U.S. News* Rankings into the Field of Legal Education." *Administrative Science Quarterly* 53 (2): 209–34.

Sauder, Michael, and Wendy Nelson Espeland. 2009. "The Discipline of Rankings: Tight Coupling and Organizational Change." *American Sociological Review* 74 (1): 63–82.
Sauder, Michael, and Gary Alan Fine. 2008. "Arbiters, Entrepreneurs, and the Shaping of Business School Reputations." *Sociological Forum* 23 (4): 699–723.
Sauder, Michael, and Ryon Lancaster. 2006. "Do Rankings Matter? The Effect of *U.S. News & World Report* Rankings on the Admissions Process of Law Schools." *Law and Society Review* 40 (1): 105–34.
Schneiberg, Marc, and Elisabeth S. Clemens. 2006. "The Typical Tools for the Job: Research Strategies in Institutional Analysis." *Sociological Theory* 24 (3): 195–227.
Scott, W. Richard. 2008. *Institutions and Organizations*. 3rd ed. Thousand Oaks, CA: Sage.
Singh, Jitendra V., David J. Tucker, and Robert J. House. 1986. "Organizational Legitimacy and the Liability of Newness." *Administrative Science Quarterly* 31 (2): 171–93.
Spence, A. Michael. 1973. "Job Market Signaling." *Quarterly Journal of Economics* 87 (3): 355–74.
Stevens, Mitchell L. 2007. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.
Thacker, Lloyd. 2005. *College Unranked: Ending the College Admissions Frenzy*. Cambridge, MA: Harvard University Press.
Volkwein, J. Fredericks, and Kyle V. Sweitzer. 2006. "Institutional Prestige and Reputation among Research Universities and Liberal Arts Colleges." *Research in Higher Education* 47 (2): 129–48.
Washington, Marvin, and Edward J. Zajac. 2005. "Status Evolution and Competition: Theory and Evidence." *Academy of Management Journal* 48 (2): 282–96.
Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104 (5): 1398–1438.