High Educ
DOI 10.1007/s10734-010-9339-1

Anchoring effects in world university rankings: exploring biases in reputation scores

Nicholas A. Bowman · Michael N. Bastedo

Abstract Despite ongoing debates about their uses and validity, university rankings are a popular means to compare institutions within a country and around the world. Anchoring theory suggests that these rankings may influence assessments of institutional reputation, and this effect may be particularly strong when a new rankings system is introduced. We test this possibility by examining data from the first 3 years of the *Times Higher Education Supplement* (*THES*) world university rankings. Consistent with an anchoring hypothesis, the initial *THES* rankings influenced peer assessments of reputation in subsequent surveys, but second-year rankings were not related to changes in reputation in the third year. Furthermore, as expected, early peer assessment ratings were not associated with changes in future rankings. These findings provide strong evidence for an anchoring effect on assessments of institutional reputation. We discuss the usefulness of these peer assessments, along with ways in which reputational surveys can be improved.

Keywords University rankings · Higher education · Institutional reputation · Anchoring effects · Global competition

Introduction

The twentieth century has witnessed the massive expansion of higher education around the world, with enrollments growing 200-fold from 1900 to 2000 (Schofer and Meyer 2005). Unsurprisingly, higher education attainment has also become increasingly

N. A. Bowman (✉)
Center for Social Concerns, University of Notre Dame, 164 Geddes Hall, Notre Dame, IN 46556-4633, USA
e-mail: nbowman@nd.edu

M. N. Bastedo
Center for the Study of Higher and Postsecondary Education, University of Michigan, 610 E University Ave, SEB 2117, Ann Arbor, MI 48109-1259, USA

Published online: 01 May 2010

important for job market outcomes. One consequence of the increasing importance of higher education has been the development of ranking systems in countries and regions, as the public seeks to differentiate institutions from one another in a growing, complex market (Dill and Soo 2005). The development of individual rankings systems has become a contested arena of its own, as international league tables like the Jiao Tong University rankings and the *Times Higher Education Supplement* (*THES*) rankings compete for dominance (Institute for Higher Education Policy 2007; Usher and Savino 2006).

Institutional leaders see rankings and league tables influencing organizational mission, strategy, personnel, recruitment, and public relations—in short, pervading nearly every aspect of the enterprise (Hazelkorn 2007, 2008). Rankings seem to have a particularly strong influence on decision making in professional schools and other postgraduate programs (Elsbach and Kramer 1996; Espeland and Sauder 2007; Sauder and Espeland 2009; Sauder and Fine 2008). Rankings have been shown to drive resources from external providers in the U.S. when the decision makers are insiders who are sensitive to the prestige hierarchy of higher education (Bastedo and Bowman in press). Rankings are also increasingly used as a policy instrument to assess the performance of institutions by government agencies (Salmi and Saroyan 2007; Sponsler 2009).

We have increasing knowledge of the impact of rankings on students, particularly in the U.S., where institutional data on student behavior is far more available for analysis (Griffith and Rask 2007; Meredith 2004; Monks and Ehrenberg 1999). Getting onto the "front page" of the rankings—in other words, getting into the top category or group—has been shown to have a strong impact on students applying to all selective universities (Bowman and Bastedo 2009). As a result, institutions that move up in the rankings, and especially those who make the top group, see significant improvements in the class rank and standardized test scores of their first-year students. In addition, the schools have more applications, higher yield rates, and thus lower acceptance rates.

Prior research conducted in the U.S. has revealed that rankings affect not only institutional and student behavior, but also the assessments of institutional reputation and prestige made by faculty and institutional leaders (Bastedo and Bowman 2010). Because institutional actions related to quality are relatively opaque—it is hard to know what even your closest competitors are doing to improve academic performance—quality is a highly ambiguous measure that is vulnerable to external influences (Podolny 1993). There are also substantial time lags between changes in academic quality and reputational assessments, and many changes in reputation are unrelated to quality, making it difficult to determine the difference between changes in quality and changes in simple prestige. As a result, over time rankings increasingly *become* reputation, rather than reputation being an independent indicator that rankings can use to assess changes in quality (Bastedo and Bowman 2010; Stake 2006).

Thus, studying the process by which these reputational assessments are made is particularly crucial. This article examines the impact of rankings on the reputational assessments made by insiders in the higher education field. We take advantage of a natural experiment that occurred with the introduction of the *Times Higher Education Supplement* world university rankings to consider the effect of "anchoring" on the decision-making processes of reputational arbiters. As predicted by decision-making theory, we find that anchoring effects exert a substantial influence on future reputational assessments, raising substantial concerns about the validity and reliability of the reputational scores used by ranking agencies to sort and stratify institutions by quality.

Anchoring effects in judgment processes

When making judgments for which the answer is ambiguous, most people will start with a particular value that is available to them, and then adjust their final judgment accordingly. This phenomenon is known as the anchoring effect (also the anchoring-and-adjustment heuristic). In their pioneering study, Tversky and Kahnemann (1974) asked participants whether the percentage of African nations who are in the United Nations is higher or lower than a certain number that had been arbitrarily determined by participants' spinning a "wheel of fortune." The wheel was designed so that the value would either be high (65%) or low (10%). After making this initial judgment, participants were asked to guess what percentage of African nations were actually members of the U.N. People who initially considered the higher value, on average, gave much higher percentages than those who initially considered the lower value (45 vs. 25%, respectively). According to anchoring theory, people use the starting value to inform their judgments, and then they adjust (insufficiently) from this value when making their final judgment, even when the starting value is entirely random.

Subsequent research has shown that anchoring effects occur in a wide variety of contexts, including assessing real estate values (Northcraft and Neale 1987), estimating average car prices (Englich 2008), judging truthfulness in lie detection (Zuckerman et al. 1984), estimating the height of Hummer SUVs (Janiszewski and Uy 2008), predicting spouses' behavior (Davis et al. 1986), and even making criminal sentencing decisions (Englich et al. 2006). It seems logical that people making uninformed guesses would anchor their judgments on some starting value that may be informative. Indeed, anchoring effects operate largely by bringing to mind knowledge that is consistent with the anchor point (see Mussweiler and Strack 1999, 2000). For example, if a home buyer sees a relatively high value for the listing price of a house (Northcraft and Neale 1987), she would be more likely to think about the positive aspects of the property and weigh these in her judgments than if she saw a lower listing price.

However, these influences are not limited to novices or laboratory settings. Anchoring effects are observed in a variety of real-world settings (Chapman and Bornstein 1996; Englich 2006; Englich et al. 2006; Janiszewski and Uy 2008; Mussweiler et al. 2000; Northcraft and Neale 1987). Moreover, experts' judgments are also subject to anchoring effects, such as judges and prosecutors making sentencing decisions (Englich et al. 2006; Englich and Mussweiler 2001), auditors estimating corporate fraud (Joyce and Biddle 1981), real estate agents assessing home values (Northcraft and Neale 1987), and car mechanics and dealers estimating used car values (Mussweiler et al. 2000). Importantly, these biased judgments occur even when the anchoring values are clearly arbitrary (Englich et al. 2006; Tversky and Kahnemann 1974).

It seems quite likely that anchoring effects would occur for raters who are asked to name the top universities throughout the world. Specifically, the *THES* world university rankings could serve as an anchoring point for making initial judgments. Academics who see Harvard University and the University of California, Berkeley, at the top of the 2004 rankings might think about the various positive qualities of these institutions. When they are asked to provide a list of top universities in their field, they will consider these positive attributes and therefore be more likely to name Harvard or Berkeley than if they had not seen the rankings. In practice, some raters might use the rankings to provide a list of potential universities that they could name, whereas other raters might not have seen the rankings within the past few weeks or months. The effects of anchoring should be apparent in both instances; in fact, Mussweiler (2001) demonstrated that anchoring effects are

equally strong when people make judgments either immediately after exposure to an anchor or a full week after seeing the anchor.

Some indirect evidence suggests anchoring effects may be present in college rankings. Bastedo and Bowman (2010) found that *U.S. News* undergraduate rankings predict changes in peer assessments of reputation, but these effects only occurred during the earlier years of the rankings (late 1980s–1990s), not in more recent years. The format of these rankings has changed substantially since the early years (in fact, the first three rankings were based solely upon a peer reputation survey), so it is not possible to conduct a definitive analysis of anchoring effects. In addition, Stake (2006) found that *U.S. News* law school rankings predict changes in academic reputation during the following year and that reputation and rankings have become more closely aligned over time. Some of the observed effects in that study were more pronounced in the earlier years, but this evidence is also indirect at best.

Present study

This study examines whether and under what conditions the *THES* world university rankings affect peer assessments of reputation. The early *THES* rankings constitute an interesting natural experiment, because (a) they were instantly popular, which means the first-year rankings served as a salient anchoring point for future peer assessments, and (b) the methodology for measuring peer assessments and overall rankings was virtually identical during the first few years of the survey, which allows for meaningful comparisons and analyses across these years. Moreover, world rankings may be particularly susceptible to anchoring effects, because academic raters generally have less knowledge about institutions worldwide than in their own country. *THES* rankings are calculated on the basis of overall scores: 40% of the overall score comes from peer review scores (i.e., peer assessments of reputation), 20% from research productivity (i.e., the frequency that research articles from an institution are cited by other scholars), 20% from academic quality (i.e., faculty-student ratio), 10% from employer ratings of the institutions whose graduates they prefer to hire, 5% from the proportion of staff who come from other countries, and 5% from the proportion of international students.

Three hypotheses follow from the preceding discussion. First, because the *THES* rankings likely serve as an anchor for academic raters' judgments, overall rankings will be positively associated with future peer assessments of reputation, controlling for earlier peer assessments. Second, early peer assessments of reputation will not be positively related to subsequent overall rankings. This hypothesis is important, because the causal argument for the impact of rankings would be more strongly supported if rankings contribute to subsequent peer assessments, but not vice versa. Third, because the first year of the rankings likely serves as a particularly informative source (i.e., by providing a novel formalized hierarchy), the effects of 2004 overall rankings on future peer assessments of reputation will be stronger than the effects of overall rankings in subsequent years.

Method

Data source

All data were obtained from the *THES* world university rankings. Only the first 3 years of the rankings were used (i.e., 2004, 2005, and 2006), because the formula for computing the

peer review score changed substantially in 2007. The final sample included the 197 universities that were listed in the top 200 for more than 1 year. Europe had the most universities of any region ($N = 83$, 42%), followed by North America ($N = 68$, 35%), Asia ($N = 33$, 17%), and Australia/New Zealand ($N = 13$, 7%).

Measures

The primary variable of interest was peer assessments of reputation, which was computed from the *THES* peer review score. This score was based on a survey of academics from around the world. The raters reported their general area of study (science, technology, medicine, social science, or arts and humanities), and they provided a list of up to 30 universities that they considered to be leaders in this area. This index had a maximum of 1,000 in 2004, though the highest score that any institution received was 665. In 2005 and 2006, the index was scaled so that the top university received a score of 100. To provide comparability across samples for our analysis, scores from 2004 were all divided by 6.65 so that the top university in this year also had a score of 100.

The peer review scores in 2005 and 2006 were calculated using ratings from multiple years; specifically, the 2005 peer review score contained academics' ratings that were made in 2004 and 2005, and the 2006 score contained ratings from 2004, 2005, and 2006. Because the primary interest of this study was to examine changes in peer assessments, the 2005 and 2006 indices were adjusted to approximate the ratings that were given only in the most recent year. For example, to estimate the ratings that occurred in 2005, the 2005 *THES* peer review score was multiplied by 2 and then the score from 2004 was subtracted from this total. Thus, if a university had *THES* peer review scores of 70 in 2004 and 75 in 2005, then its adjusted peer review score for 2005 would be $(75 \times 2) - 70 = 80$. A similar adjustment was made in 2006 by multiplying the 2006 *THES* peer score by 3, then subtracting twice the 2005 *THES* score (because the 2005 *THES* index contains ratings from 2004 to 2005). These adjustments assume that the number of new raters is constant across all 3 years; the limited information available suggests that this is approximately true, although there are probably slightly more raters in more recent years. Because 2004 was the first year of the rankings, no adjustments to the peer review score were necessary for that year.

The overall world university ranking was also included. For ease of interpretation, this variable was reverse-coded so that higher values reflect better rankings. In addition, a dummy-coded variable indicated whether a university was located in the United States. Finally, for inclusion in the structural equation models, the continuous variables were standardized with a mean of 0 and a standard deviation of 1.

Analyses

To examine how world university rankings affect peer assessments of reputation, structural equation modeling (SEM) was used. The use of SEM is preferable to ordinary least squares multiple regression in this study for two reasons. First, SEM is ideal for modeling the process through which several variables affect (or are related to) one another. It uses a single analysis to perform essentially the same task that requires several regressions in a traditional path analysis. SEM also provides goodness-of-fit indices that indicate the degree to which the data fit the conceptual model. Second, this study uses several variables that are highly correlated with one another. Whereas high multicollinearity among independent variables is quite problematic in multiple regression analyses (Pedhazur 1997), SEM can

incorporate these relationships into the model, which thereby produces more accurate estimates of the unique variance explained by the predictor variables (Byrne 2006; Kline 2005).

The SEM software program EQS 6.1 for Windows was used to analyze covariance matrices of the data with maximum likelihood estimation. Three structural equation models were analyzed, which contained data from (1) 2004 and 2005, (2) 2004 and 2006, and (3) 2005 and 2006. Other than the inclusion of different years in the analyses, the three models were identical (for an overview of the models, see Fig. 1). The paths from peer assessments to overall rankings in the same year reflected the fact that overall rankings are partially defined by peer review scores. Because the sample was fairly small by SEM standards, the models used observed (not latent) variables; as a result, it was not necessary to compute measurement models.

We expected to find a direct effect of overall ranking on subsequent peer assessments. We also anticipated that this effect would be greatest in the two models that contained data from 2004, since the effects of rankings should be strongest in the first year. Moreover, to ensure that these expected effects were not a by-product of the strong association between peer reviews and overall rankings, a path from peer assessments to subsequent overall rankings was included. If there is a direct effect of overall rankings on later peer reviews, but no effect of peer reviews on later overall rankings, then the alternative explanation for the findings would be refuted, and the evidence for a causal link between rankings and peer assessments would be substantially strengthened.

Preliminary analyses showed that univariate skewness and kurtosis statistics for all variables were at or below the recommended values of 3 and 10, respectively, for SEM analyses (Kline 2005). For all models, variance inflation factors for all variables were well below the recommended 10:1 ratio. Several common measures were used to assess goodness-of-fit: the normed fit index (NFI), the non-normed fit index (NNFI), the comparative fit index (CFI), and the ratio of chi-square to degrees of freedom (χ^2/df). For adequate model fit, the first three indices should be at least .90, and the ratio of chi-square to degrees of freedom should be less than 5 (Bentler and Bonett 1980; Bollen 1989; Hu and Bentler 1999). Using the chi-square ratio as a means of assessing fit is preferable to using the significance test of chi-square analyses, because chi-square values will often be significant with larger samples, even when other fit indices are excellent. As shown in Table 1, the data fit the models quite well; in all three models, the NFIs and CFIs were .99, the NNFIs were at least .93, and the χ^2/df ratios were less than 4.

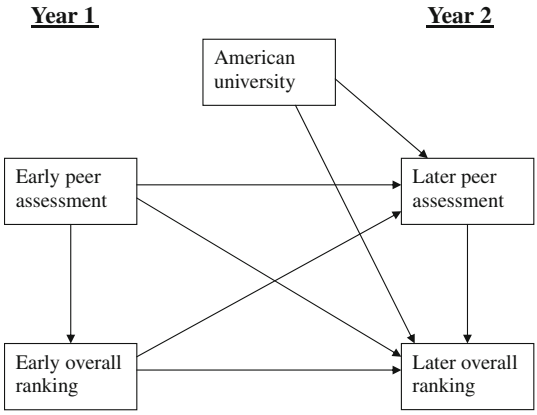

Fig. 1 Conceptual figure for structural equation models

Table 1 Goodness-of-fit indices for all structural equation models

Fit index	Model		
	2004 → 2005	2004 → 2006	2005 → 2006
Normed fit index (NFI)	.99	.99	.99
Non-normed fit index (NNFI)	.93	.94	.94
Comparative fit index (CFI)	.99	.99	.99
Chi-square to degrees of freedom (χ^2/df)	3.70	2.68	3.23
Chi-square statistic	7.39*	5.36	6.47*

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

Limitations

Some limitations in this study should be noted. First, as discussed in more detail below, the initial *THES* rankings were conducted the year after the Shanghai Jiao Tong University released its initial world university rankings; therefore, some of the initial *THES* raters may have been influenced by this alternative ranking system. However, any potential effect of the first year of the Jiao Tong rankings would be consistent with our argument about the influence of college rankings. Second, the *THES* rankings are more popular in Europe and Asia than in the United States, where the *U.S. News & World Report* rankings are most dominant. Thus, the effect of rankings on reputation is likely greater among European and Asian raters than among American raters. Because the THES rankings only report average reputation scores for all raters, we could not explore this possibility directly. However, this limitation *reduces* the likelihood of finding an overall anchoring effect, making such a finding more credible despite the limitation. Third, because the formula for the *THES* peer assessments was changed after 2006, we were not able to conduct analyses after the third year of the *THES* rankings. The immediate popularity of these rankings and the consistency of the early peer assessment formulas allow us to explore relevant dynamics within the first 3 years.

Results

As predicted, for models with 2004 indicators predicting 2005 outcomes, overall rankings had a strong effect on subsequent peer assessments of reputation, $\beta = .35$, $p < .001$ (see Fig. 2). Moreover, U.S. universities experienced a decline in peer assessments relative to other universities, $\beta = -.15$, $p < .005$. However, peer assessments were not related to future overall rankings, $\beta = -.08$, *ns*. The same results were found for 2004 indicators predicting 2006 outcomes: Overall rankings contributed positively to future peer assessments of reputation, $\beta = .20$, $p < .01$, American universities dropped in peer assessments, $\beta = -.19$, $p < .001$, and there was no relationship between peer assessments and future rankings, $\beta = -.08$, *ns* (see Fig. 3). In contrast, none of these relationships held in the third model. As shown in Fig. 4, 2005 overall rankings did not significantly predict 2006 peer assessments, $\beta = .01$, *ns*, and U.S. universities did not drop in peer assessments relative to other institutions, $\beta = -.07$, *ns*. Furthermore, 2005 peer assessments were negatively related to 2006 overall rankings, $\beta = -.34$, $p < .001$.

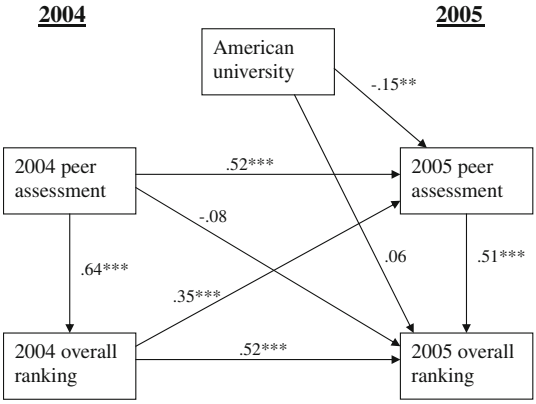

Fig. 2 Standardized coefficients for structural equation model predicting outcomes in 2005. *Note*. To simplify presentation, variances and disturbances are not shown. * $p < .05$ ** $p < .01$ *** $p < .001$

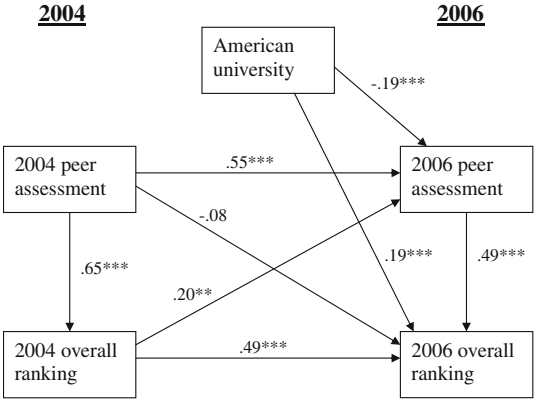

Fig. 3 Standardized coefficients for structural equation model predicting outcomes in 2006. *Note*. To simplify presentation, variances and disturbances are not shown. * $p < .05$ ** $p < .01$ *** $p < .001$

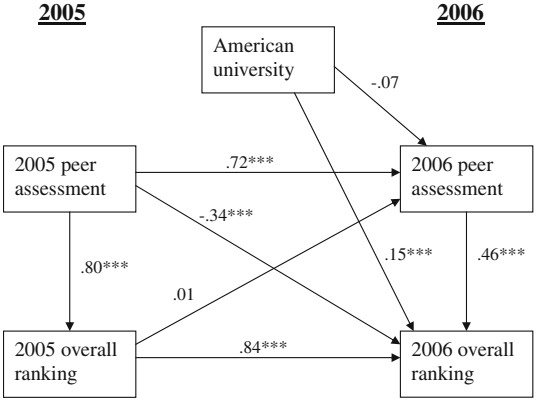

Fig. 4 Standardized coefficients for structural equation model predicting outcomes in 2006. *Note*. To simplify presentation, variances and disturbances are not shown. * $p < .05$ ** $p < .01$ *** $p < .001$

Discussion

Consistent with the first hypothesis, being ranked highly in the world university rankings contributes to increased reputation. We must conclude that academics across the world are influenced—whether consciously or unconsciously—by external assessments of their

institutions. Although previous work has shown that high-level administrators at American colleges and universities are substantially affected by college rankings (Bastedo and Bowman 2010), this study provides the first glimpse into the impact of world rankings on worldwide perceptions. The present effect of world rankings is perhaps more impressive not only because these rankings cover a larger number of possible institutions, but also because the raters are making judgments about their own area of study. That is, academics' perceptions of the top institutions in their field are affected by the world rankings that do not differentiate among fields. Thus, the world rankings arguably do not provide a valid basis of judgment for field-specific ratings (though institutes of technology may be an exception). However, this effect is quite consistent with selective accessibility models of anchoring (Mussweiler and Strack 1999, 2000). After examining the list of universities, academics are likely to think about the positive qualities of the top institutions. On average, academics probably know more about their own field than other fields; as a result, the overall rankings will likely bring to mind attributes that are field-specific, which then affects raters' final decisions.

The evidence for a causal link between world rankings and subsequent peer assessments of reputation is bolstered by the lack of an effect of 2004 peer assessments on subsequent overall rankings, which is consistent with the second hypothesis. These non-significant findings strongly suggest that the expected link between 2004 overall rankings and subsequent peer assessments is not merely driven by high correlations among several variables in the model. Surprisingly, 2005 peer assessments are negatively related to 2006 overall rankings; this pattern is the opposite of what one might expect with these two measures. Follow-up analyses reveal that this significant link is actually a suppressor effect that stems from including both 2005 and 2006 peer assessments in the same model; in other words, this negative effect does not occur if 2006 peer assessments are removed from the model.

Importantly, only the first year of world rankings has an impact on subsequent peer assessments of reputation, which is consistent with the third hypothesis and with an anchoring explanation. In 2004, academics who completed the peer ratings could not use the world university rankings to inform their opinions, because the first rankings had not yet been published. However, the introduction of the *THES* world rankings into popular culture provided an anchoring point for making judgments about institutional reputation. The first year of the rankings is particularly crucial, because this inaugural list provides a novel and salient point from which one can start making judgments. In contrast, the second year of the rankings might be informative only as an indicator of upward or downward trends in overall ranking for raters who are diligent enough to track these trends.

The powerful effects of the introductory year of the rankings are also apparent in other parts of the model. For instance, the link between peer assessments of reputation in the first year and those in the second year ($\beta = .52$, see Fig. 2) is much weaker than correspondence between peer assessments the second and third year ($\beta = .72$, see Fig. 4). In other words, consistent with an anchoring perspective, peer assessments were solidified after the first year of *THES* rankings became publicly available. The relationship between peer assessments and overall rankings within the same year is also much stronger in 2005 ($\beta = .82$, see Fig. 4) than in 2004 ($\beta = .64$, see Fig. 2), which further implies that academics shifted their ratings to correspond more closely with the overall rankings.

The role of anchoring in making judgments may also explain why peer assessments of American universities dropped after 2004. According to the *THES* rankings and widely shared perceptions, the United States has the top colleges and universities in the world (in the first year of the *THES* rankings, nine of the top 15 schools were from the U.S.), but they were then placed in a direct comparison with European, Asian, and other North

American institutions. In addition, more European than American schools appeared in the top 200 and the top 100 universities in 2004. Because academics may have begun to see universities throughout the world—or at least those in North America, Europe, Asia, and Australia—as being on the same playing field, they tended to downgrade the relative prestige of American institutions. Consistent with an anchoring explanation, this same pattern does not occur in the model in which 2005 indicators predict 2006 outcomes.

The presence of the Academic Ranking of World Universities (ARWU) by Shanghai Jiao Tong University might further explain some of these dynamics. The ARWU rankings began in 2003, 1 year before the *THES* rankings were introduced. Some of the raters for the first year of the *THES* rankings may have seen the ARWU rankings before making their judgments. The ARWU rankings provide an even greater advantage to American universities than the *THES* rankings; in 2003, 13 of the top 15 schools in the ARWU were from the U.S. (Shanghai Jiao Tong University 2003). As a result, some of the initial *THES* raters may have been influenced by the ARWU ratings and gave U.S. institutions very high marks. Thus, the downgrading of American institutions' peer assessments after 2004 may constitute both the salience of *THES* rankings as a meaningful anchoring point after the first year and the use of ARWU rankings by some *THES* raters before the first year.

Conclusion and implications

One of the many uses of rankings is to provide useful information to consumers as they make decisions about college choices (Ehrenberg 2003; Marginson 2007). For many years, there has been concern about the use of rankings based on equity concerns (Cremonini et al. 2008; McDonough et al. 1998), inappropriateness of the instrument for external decision making (Dill and Soo 2005; Hazelkorn 2008; Salmi and Saroyan 2007) and substantial concerns about the validity and reliability of rankings (Bowden 2000; Kroth and Daniel 2008; Tight 2000; Turner 2005; Van Dyke 2005). Rankings, in short, have been under attack on many fronts for a long time, with the reputation assessments of the rankings coming under particular criticism.

In some ways, this research provides fuel for many of these concerns. Although we only analyzed data from the *Times Higher Education Supplement*, it seems highly likely that anchoring effects are robust across ranking schemes worldwide, as this is based on a well-established psychological effect. When also considering the fact that the differences between rankings and reputation are becoming vanishingly small over time (Bastedo and Bowman 2010; Stake 2006) and that rankings are highly stable over time (Bowman and Bastedo 2009; Usher and Savino 2006), it is difficult to maintain the fantasy that reputational scores are independent from the rankings themselves. It would take a massive, discontinuous change in academic quality to notably influence reputation scores in any given year. Nearly always, the causal chain is that rankings change in response to shifts in their particular indicators (e.g., faculty-student ratio), and reputations shift in response to rankings. But clearly, rankings drive reputation, not the other way around.

Because reputational assessments are quite susceptible to anchoring effects, and because peer assessments of reputation are strongly correlated with other rankings indicators (Volkwein and Sweitzer 2006), reputation scores may add relatively little value to university rankings systems. So what, then, is the purpose of including reputation surveys in rankings formulas? From our perspective, the inclusion of reputation largely serves to maintain the status quo, establishing the credibility of the rankings and ensuring stability in results over time. Some of the oldest Western universities, such as Oxford, Cambridge, and

Harvard, generally receive the highest peer assessment ratings. Although removing reputation measures from the rankings calculations would likely do little to alter the pecking order, any changes would come at the expense of these schools and in the favor of (relative) upstarts.

Any changes at the top of the hierarchy are bound to gain substantial, largely negative attention. For example, when the California Institute of Technology (CalTech) jumped from #9 to #1 in 1999 after *U.S. News and World Report* made a seemingly small change in their undergraduate rankings formula, the substantial public and institutional backlash led *U.S. News* to return to their previous methods in the following year. As Gottlieb (1999) suggests, even small changes tend to garner substantial media interest—and are perhaps even driven by the desire for such attention—but organizations that conduct rankings must also be concerned about the face validity of their results. Successful ranking systems must therefore walk a line between showing too few changes in rankings over time, which leads to disinterest, and generating too much change, which leads to illegitimacy. Ordinal rankings also create precise distinctions between schools with very similar overall scores, so many smaller rankings shifts will occur simply as the product of minor fluctuations in institutional indicators.

Many institutions and departments go to great lengths to maintain or increase their position in the college rankings, operating under the assumption that these rankings have substantial consequences. The current study, however, suggests that some key reputational effects occur only in concert with the introduction and widespread use of a particular rankings system. Thus, an organization that conducts college rankings may make laudable efforts to improve its initial formula, but these changes may not undo any reputational shifts that occurred in the introductory year(s). Once reputational assessments are formed, they are often quite difficult to change without specific evidence to the contrary.

Engineering effective reputational surveys is a difficult proposition. As mentioned earlier, one of the major issues is that respondents are asked to rate colleges about which they have little first-hand knowledge, and there are long time lags between changes in quality and subsequent reputation. Therefore, one solution would be to ask respondents only to rate universities about which they have deep knowledge. Unfortunately, this will likely generate a conflict of interest: Universities generally have the deepest knowledge of their closest competitors, and these institutions compete for higher rankings. Indeed, manipulation of reputational surveys may be a major problem with some college rankings systems. Using Freedom of Information Act requests for public documents, newspapers have published the *U.S. News and World Report* undergraduate peer assessment surveys of several university presidents and provosts, and it appears that their responses were narrowly tailored to improve the relative status of their own institution (Crabbe 2009; Lee 2009). Although *U.S. News* says it is aware of this phenomenon and that it removes surveys that show evidence of manipulation, they have refused to say how they inspect for manipulation, how they redact their files for inspection, or even how pervasive the problem may be.

Therefore, an effective system must address these concerns in a systematic way. An organization that conducts college rankings must have specific criteria for identifying survey responses that show evidence of manipulation, and it must ensure that informants have deep knowledge of the institutions they rank. For instance, faculty members could provide valid reputational assessments *if* they rank programs within their field or subfield, because they are highly knowledgeable about their own discipline, and the overall ranking will become less meaningful as a potential anchoring point. Moreover, sampling a large sample of faculty from many diverse institutions and from a wide range of academic fields

would improve the validity of the survey results. Because reputational surveys are likely here to stay, it is important to use empirical research to make them as informative and unbiased as possible.

References

Bastedo, M. N., & Bowman, N. A. (2010). The U.S. News and World Report college rankings: Modeling institutional effects on organizational reputation. *American Journal of Education, 116*, 163–184.

Bastedo, M. N., & Bowman, N. A. (in press). College rankings as an interorganizational dependency: Establishing the foundation for strategic and institutional accounts. *Research in Higher Education.*

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588–606.

Bollen, K. A. (1989). *Structural equation models with latent variables*. New York: Wiley.

Bowden, R. (2000). Fantasy higher education: University and college league tables. *Quality in Higher Education, 6*(1), 41–60.

Bowman, N. A., & Bastedo, M. N. (2009). Getting on the front page: Organizational reputation, status signals, and the impact of U.S. News and World Report on student decisions. *Research in Higher Education, 50*, 415–436.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Chapman, G. B., & Bornstein, B. H. (1996). The more you ask for, the more you get: Anchoring in personal injury verdicts. *Applied Cognitive Psychology, 10*, 519–540.

Crabbe, N. (2009, June 17). On survey, Machen rates UF with Harvard, other Fla. schools low. *The Gainesville Sun*. Retrieved April 2, 2010, from http://bit.ly/y9AAb.

Cremonini, L., Westerheijden, D., & Enders, J. (2008). Disseminating the right information to the right audience: Cultural determinants in the use (and misuse) of rankings. *Higher Education, 55*, 373–385.

Davis, H. L., Hoch, S. J., & Ragsdale, E. K. E. (1986). An anchoring and adjustment model of spousal predictions. *Journal of Consumer Research, 13*, 25–37.

Dill, D. D., & Soo, M. (2005). Academic quality, league tables, and public policy: A cross- national analysis of university ranking systems. *Higher Education, 49*(4), 495–533.

Ehrenberg, R. G. (2003). Reaching for the brass ring: The U.S. News & World Report rankings and competition. *Review of Higher Education, 26*, 145–162.

Elsbach, K. D., & Kramer, R. (1996). Members' responses to organizational identity threats: Encountering and countering the *Business Week* rankings. *Administrative Science Quarterly, 41*, 442–476.

Englich, B. (2006). Blind or biased? Justitia's susceptibility to anchoring effects in the courtroom based on give numerical representations. *Law & Policy, 28*, 497–514.

Englich, B. (2008). When knowledge matters—differential effects of available knowledge in standard and basic anchoring tasks. *European Journal of Social Psychology, 38*, 896–904.

Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the court room. *Journal of Applied Social Psychology, 31*, 1535–1551.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin, 32*, 188–200.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology, 113*, 1–40.

Gottlieb, B. (1999, September 1). Cooking the school books: How U.S. News cheats in picking its "best American colleges." *Slate*. Retrieved November 19, 2009, from http://www.slate.com/id/34027/.

Griffith, A., & Rask, K. (2007). The influence of the U.S. News and World Report collegiate rankings on the matriculation decisions of high-ability students: 1995–2004. *Economics of Education Review, 26*, 244–255.

Hazelkorn, E. (2007). The impact of league tables and ranking system on higher education decision making. *Higher Education Management and Policy, 19*, 1–24.

Hazelkorn, E. (2008). Learning to live with league tables and ranking: The experience of institutional leaders. *Higher Education Policy, 21*, 193–215.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Institute for Higher Education Policy (IHEP). (2007). *College and university ranking systems: Global perspectives and American challenges*. Washington, DC: Author.
Janiszewski, C., & Uy, D. (2008). Precision of the anchor influences the amount of adjustment. *Psychological Science, 19*, 121–127.
Joyce, E., & Biddle, G. (1981). Anchoring and adjustment in probabilistic inference in auditing. *Journal of Accounting Research, 19*, 120–145.
Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
Kroth, A., & Daniel, H.-D. (2008). International university rankings: A critical review of the methodology. *Zeitschrift fur Erziehungswissenschaft, 11*, 542–558.
Lee, S. (2009, August 19). Reputation without rigor. *Inside Higher Ed*. Retrieved April 2, 2010, from http://www.insidehighered.com/news/2009/08/19/rankings.
Marginson, S. (2007). Global university rankings: Implications in general and for Australia. *Journal of Higher Education Policy and Management, 29*, 131–142.
McDonough, P. M., Antonio, A. L., Walpole, M., & Perez, L. X. (1998). College rankings: Democratized college knowledge for whom? *Research in Higher Education, 39*, 513–537.
Meredith, M. (2004). Why do universities compete in the rankings game? An empirical analysis of the effects of the U.S. News and World Report college rankings. *Research in Higher Education, 45*, 443–461.
Monks, J., & Ehrenberg, R. G. (1999). *The impact of U.S. News & World Report college rankings on admissions outcomes and pricing policies at selective private institutions* (Working Paper #7227). Washington, DC: National Bureau of Economic Research.
Mussweiler, T. (2001). The durability of anchoring effects. *European Journal of Social Psychology, 31*, 499–509.
Mussweiler, T., & Strack, F. (1999). Comparing is believing: A selective accessibility model of judgmental anchoring. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 10, pp. 135–167). Chichester, UK: Wiley.
Mussweiler, T., & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks. *Journal of Personality and Social Psychology, 78*, 1038–1052.
Mussweiler, T., Strack, F., & Pfeiffer, T. (2000). Overcoming the inevitable anchoring effect: Considering the opposite compensates for selective accessibility. *Personality and Social Psychology Bulletin, 26*, 1142–1150.
Northcraft, G. B., & Neale, M. A. (1987). Experts, amateurs, and real estate: An anchoring-and-adjustment perspective on property pricing decisions. *Organizational Behavior and Human Decision Processes, 39*, 84–97.
Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Wadsworth.
Podolny, J. M. (1993). A status-based model of market competition. *American Journal of Sociology, 98*, 829–872.
Salmi, J., & Saroyan, A. (2007). League tables as policy instruments: Uses and misuses. *Higher Education Management and Policy, 19*, 1–39.
Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: Tight coupling and organizational change. *American Sociological Review, 74*, 63–82.
Sauder, M., & Fine, G. A. (2008). Arbiters, entrepreneurs, and the shaping of business school reputations. *Sociological Forum, 23*, 699–723.
Schofer, E., & Meyer, J. W. (2005). The worldwide expansion of higher education in the twentieth century. *American Sociological Review, 70*, 898–920.
Shanghai Jiao Tong University (2003). *Academic ranking of world universities—2003*. Retrieved November 23, 2009, from http://www.arwu.org/ARWU2003.jsp.
Sponsler, B. (2009). *The role and relevance of rankings in higher education policymaking*. Washington, DC: Institute for Higher Education Policy.
Stake, J. E. (2006). The interplay between law school rankings, reputations, and resource allocations: Ways rankings mislead. *Indiana Law Journal, 82*, 229–270.
Tight, M. (2000). Do league tables contribute to the development of a quality culture? Football and higher education compared. *Higher Education Quarterly, 54*(1), 22–42.
Turner, D. R. (2005). Benchmarking in universities: League tables revisited. *Oxford Review of Education, 31*(3), 353–371.
Tversky, A., & Kahnemann, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1130.

Usher, A., & Savino, M. (2006). *A world of difference: A global survey of university league tables*. Toronto, ON: Educational Policy Institute.
Van Dyke, N. (2005). Twenty years of university report cards. *Higher Education in Europe, 30*(2), 103–125.
Volkwein, J. F., & Sweitzer, K. V. (2006). Institutional prestige and reputation among research universities and liberal arts colleges. *Research in Higher Education, 47*, 129–148.
Zuckerman, M., Koestner, R., Colella, M. J., & Alton, A. O. (1984). Anchoring in the detection of deception and leakage. *Journal of Personality and Social Psychology, 47*, 301–311.