Session Number 1793

A Value-Added Perspective on the College Ratings

Bruce R. Thompson
Rader School of Business, Milwaukee School of Engineering

Abstract

In recent years, various models have been developed to measure the quality of educational institutions. One group of models, particularly popularized by the U.S. News and World Report's ratings of colleges and universities, along with specialized programs such as engineering schools, makes use of data such as that on incoming students and resources to rank the institutions. A quite different approach has become increasingly widespread in rating K-12 schools. This approach uses statistical tools to rate schools by their outcomes while controlling for inputs. This paper examines the US News approach through the lens of value-added analysis.

Introduction

College ratings based on models utilizing data have become increasingly popular in the past twenty years. The most financially successful are those published by U.S. News and World Report in its annual fall college guide.[1] These and similar ratings have enjoyed widespread success with parents and students, but have also received considerable criticism.

During the same period, there has been a growth of "value-added" models to evaluate the success of pre-college schools. While these models take various specific forms, they all are motivated by the philosophy that schools should be judged by the value they add rather than the resources used or the quality of the incoming students.

This paper applies a value-added perspective to the college ratings, including the ratings of engineering programs and schools.

A Brief Introduction to Value Added Models

Figure 1 shows a conceptual view of the value-added model. It visualizes a school as a transformation mechanism. Students enter with certain characteristics, including academic skills and various social and economic characteristics, such as family income, their ethnicity, and family stability, all of which can have an impact on academic achievement. They leave the school, it is hoped, with much greater academic skills than when they entered. From the value-added perspective, the greater the student outputs compared to the student inputs the more successful the school.

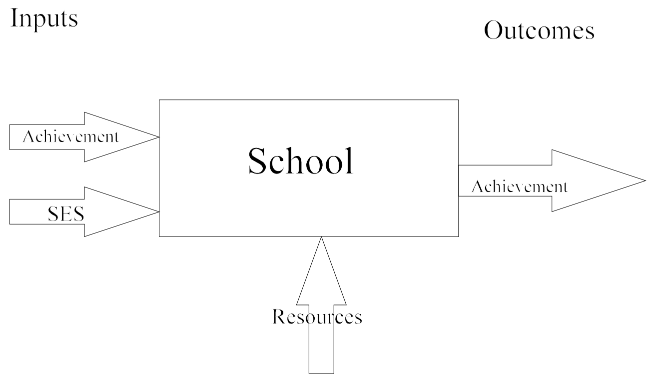


Figure 1. A value-added model

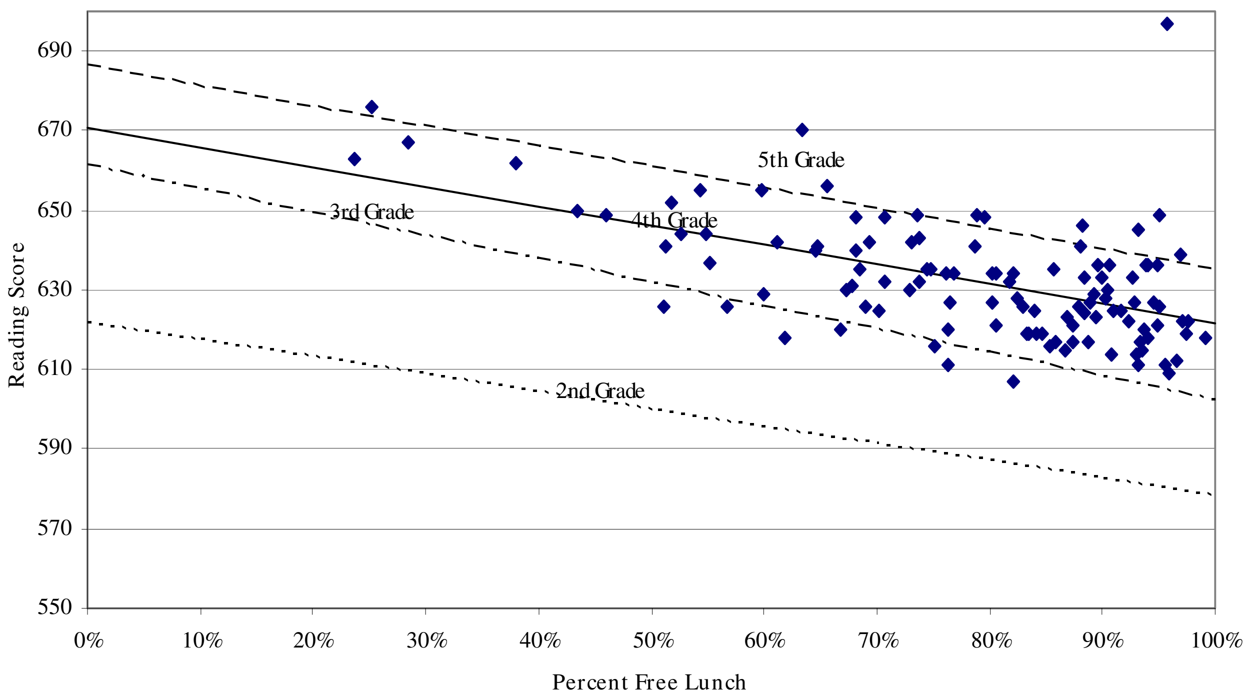
Traditional school ratings at the K-12 level, by contrast, rank schools strictly by outputs. For example, all schools in a state might be compared on the basis of how well their students performed on a state-wide reading test. In practice, such rankings reflect the socio-economic status of the school's students more than the school's contribution.

Figure 2, for example, shows a scatter plot of average performance on a reading test for all schools in an urban district. In this figure, the percentage of students qualifying for free and reduced lunch is a proxy for the average poverty rate in the school. None of these schools serve an especially prosperous population; few have a subsidized lunch rate below fifty percent. Even so, there is a dramatic relationship between poverty and reading achievement. Students in schools where all qualify for free lunch are on average a year behind those in schools where only half qualify.

Starting about forty years ago a series of reports appeared describing the impact of poverty and other socio-economic factors on student achievement.[2] Over the same period research was unable to measure significant effects on achievement from readily-available school characteristics, such as those used for accreditation.[3] This combination of strong effects of family characteristics and weak measurable school effects led many to conclude that student achievement was mostly outside the school's control.

In essence, this conclusion turned the traditional relationship between education and upward mobility on its head. Rather than education as a route out of poverty, elimination of poverty became prerequisite for a good education.

Figure 2. Fourth Grade Reading vs. Free Lunch



Yet when the value added-approach is applied to schools, allowing control for student characteristics, results from schools serving apparently similar populations differ substantially. This implies that something the schools do significantly affects student achievement. That earlier studies missed these differences may imply that the studies were looking at the wrong characteristics. The characteristics that were measured (often measures of resources used) were not those affecting student achievement. It appears instead that the crucial differences between schools reflect differences in curricula, teachers, and leadership.

How are value-added scores calculated? While each model differs somewhat, here is an example:[4]

- Use regression analysis to find relationships between student inputs and school outputs.

- Use the regression equation to predict each school's outputs based on its student characteristics.

- Calculate the gaps between the actual and predicted outputs.

- Use these gaps to get a rating for each school for each output.

- Average each school's ratings for all outputs to get overall school scores.

Generally two kinds of student data are used for the inputs in value-added models: data on student socio-economic status and on previous student achievement. Models may incorporate either or both of these, depending on data availability and the philosophy of the model designers.

- Socio-economic status (SES). The most common SES input is poverty (usually measured by the percentage of students qualifying for free or reduced lunch). Other common measures are ethnic mix and student mobility. Theoretically, there is no limit to the number of inputs used but in practice it may be more practical to use one or two inputs. The more inputs, the higher the burden of collecting and verifying the accuracy of data. Many of the possible inputs correlate strongly with each other, so that it is often difficult to sort out the relative influence of each. In one analysis, rankings of schools based on all available student inputs varied little from those using only poverty[5]. In some circles, however, the use of SES data is regarded with suspicion for fear it will reenforce low expectations for poor and minority students.

- Prior student achievement. Philosophically, rating schools by how much students gain while in the school is very attractive, since it emphasizes that schools are responsible for learning while the student is at the school, not what took place before. In fact, some advocates of using gains in scores argue that their use eliminates the need to consider SES.[6] However, prior student achievement is of limited usefulness in rating elementary schools since testing does not normally start until students have been at the school for several years.

Theoretically, non-student inputs to the school, such as money and other resources or the educational program, could also be treated as inputs in a value-added model. In practice, the models are generally limited to inputs related to students.[7] The result of the model can, in many cases, be used to evaluate the effectiveness of the non-student inputs. For example, do schools with more resources enjoy greater achievement?

The College Ratings and Their Discontents

U.S. News and World Report introduced its college ratings in 1983 as a simple survey of college presidents who were asked to rate peer institutions.[8] Since then the magazine has incorporated a wide variety of data into its ratings. It publishes a separate annual guide to colleges and makes much of its data available for a fee on its web site.[9] The annual college issue is reported to be its single biggest seller. Others have produced their own ratings using variations of the U.S. News approach.[10]

The college ratings have generated a variety of criticisms that fall into several groups:

- That any attempt to reduce what colleges do to a number trivializes education and misses much that is important.[11] This criticism echos complaints by K-12 educators about standardized tests, but fails to address the desire of parents, students, and others for some objective measure of school quality.

- That the ratings are essentially conservative, aimed at confirming already-existing perceptions of which colleges are best. Some suggest this bias is deliberate, reflecting the personal desire of the editors to affirm the excellence of the colleges they attended or a belief that a list confirming expectations is more credible than one that challenges conventional wisdom. Other critics suggest that the conservative bias is simply reflects the limitations of available data.

- That the ratings encourage colleges to behave in ways harmful to students.[12] For example, to lower reported admission rates a college may encourage applications from students who have no chance of acceptance. Or it may emphasize research with little effect on undergraduates. Responding to criticisms that the use of yield data encouraged colleges to place increasing emphasis on early admission where students promise to enroll if admitted, the magazine recently dropped the use of yield data.

- That the ratings mostly reflect resources used, not the learning taking place.[13]

How the college ratings are calculated

Three different lists from U.S. News illustrate much of the variety of data that can be incorporated into the ratings. This section describes the methodology of three of the lists:

1. Undergraduate ratings for universities and colleges. There are four different lists: national universities-doctoral, universities-master's, liberal arts colleges, and comprehensive colleges, but the data used and the methodology are largely the same for each. Table 1 shows the data used along with the weighting factors.

2. Undergraduate engineering programs. This rating is based solely on a survey of deans and senior faculty. A similar approach is used in other specialty programs.

Table 1. Undergraduate Rating Factors

Type	Factor	National Universities (Doctoral) and Liberal Arts Colleges (Bachelor's)	Universities (Master's) and Comprehensive Colleges (Bachelor's)
P	Peer assessment survey	25%	25%
I	Acceptance rate	1.5%	1.5%
I	High school class standing—top 10%	6%	0%
I	High school class standing—top 25%	0%	6%
I	SAT/ACT scores	7.5%	7.5%
R	Faculty compensation	7%	7%
R	Percent faculty with top terminal degree	3%	3%
R	Percent full-time faculty	1%	1%
R	Student/faculty ratio	1%	1%
R	Class size, 1-19 students	6%	6%
R	Class size, 50+ students	2%	2%
O	Average graduation rate	16%	20%
O	Average freshman retention rate	4%	5%
R	Average educational expenditures per student	10%	10%
O	Average alumni giving rate	5%	5%
V	Graduation rate performance	5%	0%
	Total	100%	100%

Type of Measure: I-Input; O-Output, R-Resource; P-Peer survey; V-Value added

Table 2. Graduate Engineering Schools

Type	Factor	Weight
P	Dean survey	25.00%
O	Recruiter survey	15.00%
I	GRE-quantitative	4.50%
I	GRE-analytical	4.50%
I	Acceptance rate	1.00%
R	PhD students/faculty	7.50%
R	MS students/faculty	3.75%
R	Faculty in NAE	7.50%
O	Total doctoral degrees	6.25%
R	Expenditures	15.00%
R	$/faculty	10.00%

3. Graduate engineering schools. Table 2 shows the data used and weighting factors. The magazine offers similar calculations for a number of other graduate programs.

Applying the value-added approach to the college ratings

The factors used in the college ratings can be classified into five categories: output measures, input measures, resources, peer ratings, and value-added measures. Both correlation and regression analysis were used to get an idea of their influence on the final ratings.

- Output measures. For undergraduate college and university scores, these include the graduation rate, freshman retention rate, and the alumni giving rate, for a total of 25% of the scores for liberal arts colleges and research universities and 30% of the scores for the others. No output measures are included in the calculation of the scores for undergraduate engineering program. For graduate engineering programs, the recruiter survey is arguably an output measure so long as the responses are based on actual experience. Total doctoral degrees might also be considered an outcome measure but is problematic since it may reflect program size more than program effectiveness.

- Input measures. For undergraduate liberal arts and research universities, these contribute 16% of the score and include acceptance rate, high school class standing, and SAT/ACT scores. For graduate engineering programs, quantitative and analytical GRE scores and acceptance rates are input measures and together contribute 10% of the overall score.

- Resources. 30% of the undergraduate college and university score comes from resource measures, including faculty compensation, faculty with top terminal degrees, percent of faculty that is full time, student-faculty ratio, classes under 20, classes over 50, and educational expenditures per student. Just under 44% of the score for graduate engineering programs comes from resources, including graduate students per faculty, faculty in the National Academy of Engineering, and expenditures.

- Peer assessment. 25% of the college and university and of the graduate engineering scores, and 100% of the undergraduate engineering scores come from peer surveys. Of all the factors used, peer surveys correlate best with overall score, as shown in Appendix I. As mentioned, the original college ratings were based entirely on peer surveys.

- A value-added measure. In its ratings of liberal arts colleges and research universities, U.S. News has included a value-added measure. The predicted graduation rate based on student inputs and school resources is compared to the actual graduation rate. This counts for 5% of the overall score. The overall score and the graduation performance rate are not correlated as reflected in Figure 3 and Appendix I.

A multiple linear regression analysis of the school scores against the various factors was done for the top tier of liberal arts colleges. This confirmed the strong relationship between peer ratings and the scores. At the 95% confidence level, the relationships between scores and most of the factors were statistically significant. At this level the only non-significant factors were student faculty ratio, the percentage of full-time faculty, and the SAT/ACT scores. However, because most of the factors are heavily collinear with each other, using regression to judge the relative importance of each factor is problematic.

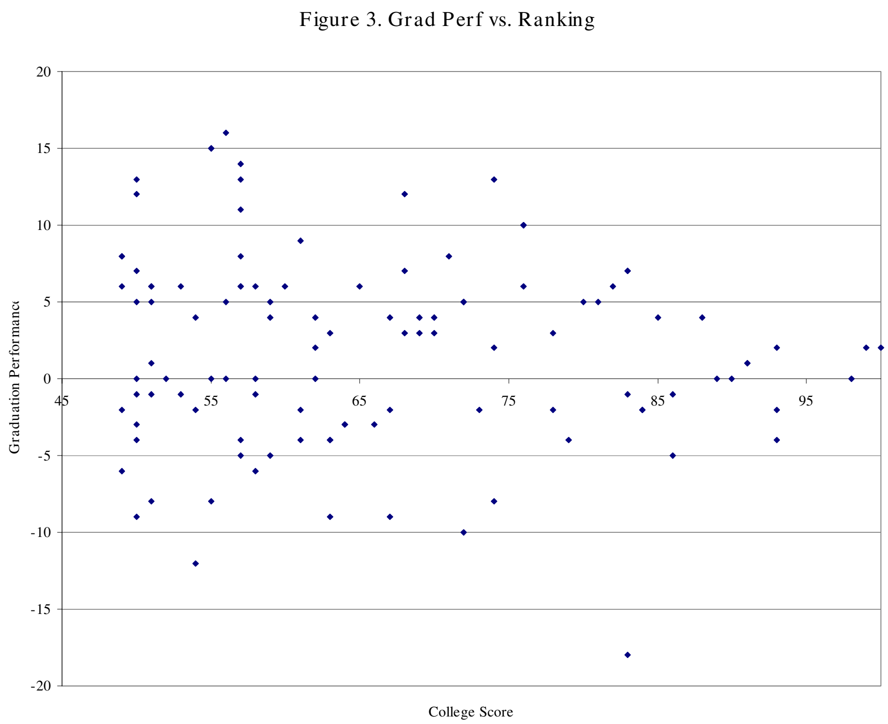


Figure 3. Graduation performance compared to overall score (liberal arts top tier)

As noted there is a strong relationship between peer assessments and the overall score of the colleges. However, as shown in Figure 4, this relationship is not perfect. Colleges whose overall scores were ten points or more below that predicted by their peer assessment had relatively high incoming test scores, but were downgraded in the ratings because of resources that lagged those of other colleges with similar tests scores and peer ratings.

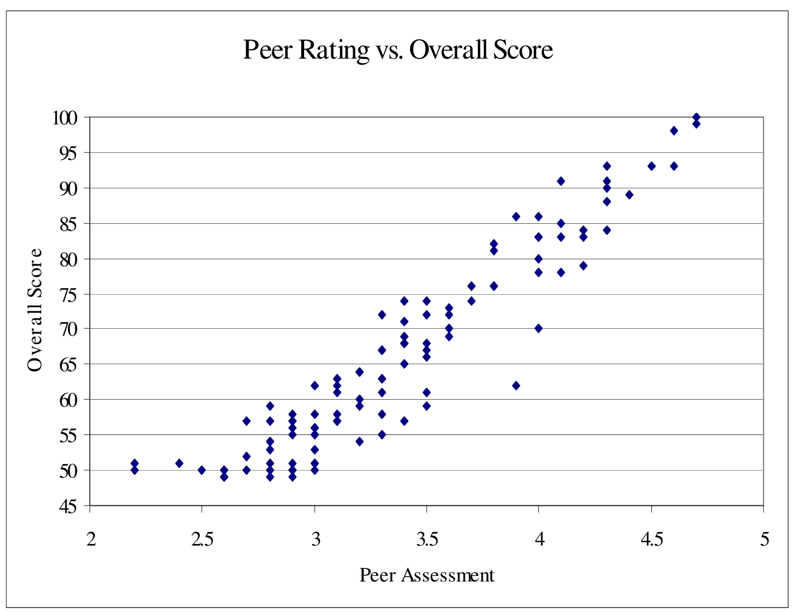


Figure 4. College Overall Scores Compared to Results of Peer Assessments (Liberal Arts Colleges-top tier)

Rating the colleges using value-added

Using the input and output categories from Table 1, it is possible to develop value-added scores for both the national universities and liberal arts colleges. These value-added scores were calculated by regressing the outputs against the measures of student inputs. A total value-added score was obtained by averaging the individual scores for freshman retention, graduation, and alumni contributions. Appendix II shows some of the values obtained.

As suggested by Figure 5, there is little relation between the overall score calculated by U.S. News and the value-added score. The lack of a strong relationship was confirmed by correlations near zero, shown in Appendix III.

Do increased financial strength or greater faculty resources lead to higher value-added scores? As noted neither was incorporated into the calculation. Yet, as shown in Appendix III, the relationships are weak and, if anything, slightly negative. It seems likely that to the extent resources improve outputs, the effect is completely explained by better inputs. In other words, greater resources help attract students with higher incoming test scores and grades which translate into higher graduation rates, better retention, and higher donations, but there is no additional effect from resources.

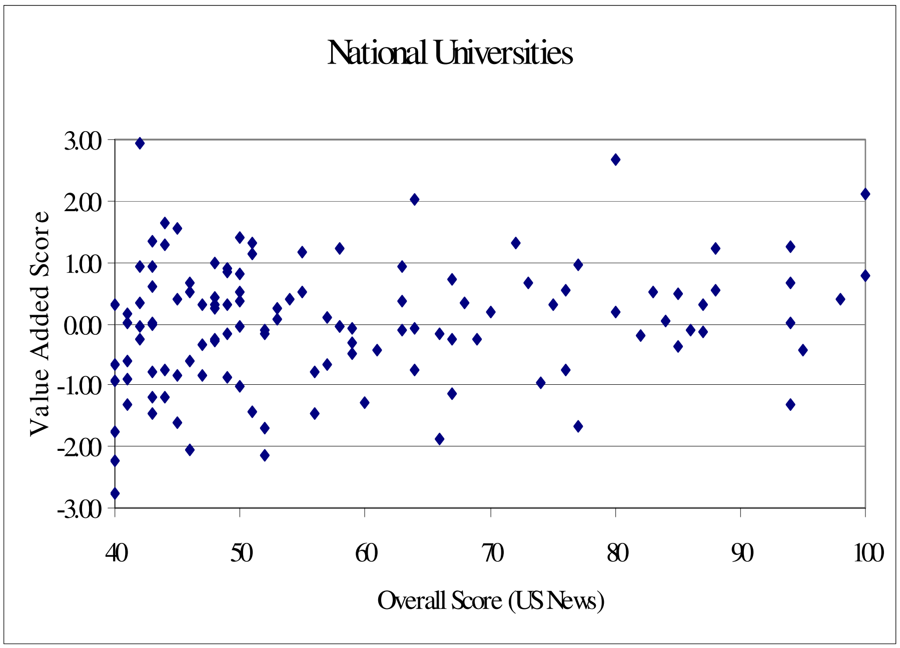


Figure 5. Comparison of value-added and US News scores

Value-added scores were also calculated for the graduate engineering programs. In this case, there was a positive relationship between the value-added scores and U.S. News' overall scores. However, this result should be treated with caution for a number of reasons. First, the range of input values was surprisingly narrow; selection rates and GRE scores varied far less among the graduate engineering schools than among even the top tiers of colleges and universities. Reflecting the narrow range of inputs, the coefficients of determinations of the regressions were low (see Appendix II). As noted earlier, neither of the output measures available may reflect student learning. For example, both may be strongly influenced by program size, so that a high-quality program that deliberately kept its enrollment low would grant low numbers of Ph.D.'s and be less familiar to recruiters.

U.S. News and similar college ratings are additive. Colleges are ranked according to all the factors used (inputs, output, resources, reputation, and even value-added measures). The final rankings reflect weighted averages of the individual rankings for the factors. In a value-added model, by contrast, high inputs raise expectations for outputs. In other words, a college with inputs associated with high outputs has a higher hurdle to overcome.

To illustrate the contrast, consider two colleges. At the first, each of the factors is at the eightieth percentile relative to all colleges. At the second the factors are all at the twentieth percentile. In the additive model, the first college would be placed near the top of the list, while the second would be placed near the bottom. In a value-added model, each of the two colleges would be calculated as making an average gain, one sufficient to keep their students in the same relative position but not move up or down. Thus the value-added model would place both at the fiftieth percentile. This helps explain why there is so little relationship between U.S. News' overall school ratings and its value-added calculation of graduation rate performance.

Conclusion

The data offered by sources such as U.S. News can be valuable to students and their parents. For example, information on average SAT scores can help a potential student judge how well he or she would meet the school's academic challenge. But it seems doubtful that the resulting ratings are good measures of "America's best colleges" if by that is meant those colleges that most effectively educate students.

Potentially, value-added measures of colleges could overturn common perceptions of which colleges and universities are best. The biggest obstacle to their use is the lack of good measures of outcomes that accurately reflect student learning. By contrast, data on inputs, resources, and reputation have been developed to the point that they are reasonably reliable and consistent. Unlike at the secondary school level where state-wide graduation tests are spreading, it seems unlikely there will soon be widely accepted measures of college outcomes.

Lacking outcome measurements, it may be possible to modify a value-added model to tie it to intermediate results that have been shown to be related to outcomes. One possibility is the National Survey of Student Engagement (NSSE) which surveys students concerning the level of academic challenge, active and collaborative learning, student-faculty interaction, enriching educational experiences, and supportive campus environment.[14] Some research indicates the issues covered in this survey are more closely related to outcomes than the factors included in the current college ratings. But to come up with a value-added rating system using the NSSE as a proxy for outcomes requires two changes:

- Release of the NSSE results for most colleges. Currently the results are controlled by the colleges. A few release them (presumably mainly those who score well), but most do not.

- A model, based on sampling, that would relate the NSSE results to educational outcomes.

Once these problems are solved, it is likely we will see a revolution in our view of what are the best colleges.

References

1. usnews.com (2003). America's best colleges. Available on-line at http://www.usnews.com/usnews/edu/college/rankings/rankindex_brief.php.

2. Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966) Equality of educational opportunity. Washington, DC: US Department of Health, Education and Welfare.
White, S.B., Reynolds, P., Thomas, M., & Gitzlaff, N. (1993). Socioeconomic Status and achievement revisited, Urban Education 28, pg328.

3. Saterfiel, Thomas H. (1974) The relationship between student achievement and accreditation variables associated with the student's school. ERIC Number: ED157913.

4. Thompson, Bruce R. (2004). Equitable Measurement of School Effectiveness, Urban Education, 39: 3.

5. Thompson (2004). Op. cit.

6. Sanders, W.L. and Horn, S.P. (1998). Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research, Journal of Personal Evaluation in Education. 12:3, 247-256.
Stone, J. E. (2002), The Value-Added Achievement Gains of NBPTS-Certified Teachers in Tennessee: A Brief Report, College of Education, East Tennessee State University. Available at http://www.education-consumers.com/briefs/stoneNBPTS.shtm
Meyer, R.H. (1997). Value-Added Indicators of School Performance: A Primer, Economics of Education Review. 16:3, pp 283-301.

7. Clotfelter, C., & Ladd, H. (1996). Recognizing and rewarding success in public schools. In H Ladd, Holding schools accountable, Chapter 2. Washington, DC: The Brookings Institution.
Webster, W., Mendro, R., & Almaguer, T. (1993). Effectiveness indices: the major component of an equitable accountability system. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

8. U.S. News (2003). The challenges of striving to measure academic excellence. Available at http://www.usnews.com/usnews/edu/college/rankings/about/challenges.php.

Peck, Don (2003), The selectivity illusion, The Atlantic Monthly. November 2003.

9. The guide is available on-line at http://www.usnews.com/usnews/edu/college/rankings/rankindex_brief.php

10. The Center (2003). The top American research universities. University of Florida. Available at thecenter.ufl.edu/research2003.html.

11. Beemer, Richard (2002). Learning beyond measure. The New York Times, September 17, 2002.

12. Graham, Amy, and Tompson, Nicholas (2001). Broken Ranks: U.S. News' college rankings measure everything but what matters. And universities do not seem to mind. Washington Monthly, September 2001.

Clarke, M. (2002, March 20). Quantifying Quality: What can the *U.S. News and World Report* rankings tell us about the quality of higher education? *Education Policy Analysis Archives*, *10*(16). Available at http://epaa.asu.edu/epaa/v10n16/.

Winter, Greg (2003). Secret of one college's success is aid for academic achievers, *The New York Times*, December 22, 2003.

13. Pascarella, Ernest T. (2001). Identifying excellence in undergraduate education. *Change*, 33:3, pg. 18.

14. The NSSE web site is at http://www.iub.edu/~nsse. This site describes the survey but does not list results for individual colleges. U.S. News lists results for several colleges at http://www.usnews.com/usnews/edu/college/rankings/ranknsse.php.

Confessiore, Nicholas (2003). What makes a college good? *The Atlantic Monthly*, November 2003.

Kuh, George D. (2003). What we're learning about student engagement from NSSE. *Change*, 35:2, pg. 24.

Kuh, George D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change*, 33:3, pg. 10.

Matthews, Jay (2002). A new assessment of student life. *Washington Post*, November 12, 2002.

Appendix I. Correlations between U.S. News' Overall Scores and Some Factors

Part I. Colleges and Universities

	Peer Assessment	*Average Freshman retention rate*	*2002 actual grad. rate*	*2002 overperf./underperf.*	*% of classes under 20*	*% of classes with 50 or more*	*Student/faculty ratio*	*% of full-time faculty*	*SAT/ACT 25th percentile*	*SAT/ACT 75th percentile*	*Freshmen in top 10% of HS class*	*Acceptance rate*	*Financial resources*	*Avg alumni giving rate*
Top Liberal Arts Colleges	0.94	0.79	0.78	-0.10	0.11	0.36	0.58	0.37	0.86	0.85	0.85	0.82	0.76	0.71
Top PhD Universities	0.90	0.86	0.88	-0.04	0.61	-0.06	0.64	0.19	0.60	0.91	0.83	0.84	0.74	0.72

Part II. Engineering Graduate Programs

	Peer assessment	*Recruiter assessment*	*Quant. GRE*	*Analytical GRE*	*Acceptance rate*	*PhD students/faculty*	*NAE membership*	*Research expenditures*	*Research $/faculty*	*PhD's granted*
Graduate Engineering Programs	0.95	0.88	0.54	0.51	0.47	0.74	0.73	0.86	0.59	0.83

Appendix II: Results of Value-Added Models

Output	Adusted R-Squared	Inputs	p-value
Liberal Arts Colleges--Top Tier			
Graduation Rate	0.58	SAT/ACT 25th percentile	0.418
		SAT/ACT 75th percentile	0.507
		Freshmen in top 10% of HS class	0.001
		Acceptance rate	0.004
Retention Rate	0.65	SAT/ACT 25th percentile	0.004
		SAT/ACT 75th percentile	0.069
		Freshmen in top 10% of HS class	0.080
		Acceptance rate	0.002
Alumni Giving Rate	0.33	SAT/ACT 25th percentile	0.438
		SAT/ACT 75th percentile	0.083
		Freshmen in top 10% of HS class	0.046
		Acceptance rate	0.843
National Universities (PhD)--Top Tier			
Graduation Rate	0.70	SAT/ACT 25th percentile	0.509
		SAT/ACT 75th percentile	0.000
		Freshmen in top 10% of HS class	0.000
		Acceptance rate	0.051
Retention Rate	0.77	SAT/ACT 25th percentile	0.716
		SAT/ACT 75th percentile	0.000
		Freshmen in top 10% of HS class	0.000
		Acceptance rate	0.000
Alumni Giving Rate	0.53	SAT/ACT 25th percentile	0.487
		SAT/ACT 75th percentile	0.000
		Freshmen in top 10% of HS class	0.010
		Acceptance rate	0.049
Graduate Engineering Programs			
Recruiter survey	0.22	Quant. GRE	0.992
		Analytical GRE	0.034
		Acceptance rate	0.097
PhD's granted	0.09	Quant. GRE	0.049
		Analytical GRE	0.904
		Acceptance rate	0.652

Appendix III. Value-Added Scores Compared to U.S. News Scores

	Faculty resources rank	*Financial resources rank*	*Overall Score*
Top Liberal Arts	-0.02	0.03	0.24
Top Universities	-0.07	-0.18	0.18
Graduate Engineering Programs			0.75

Biography

BRUCE R. THOMPSON is professor in the Rader School of Business at the Milwaukee School of Engineering. He holds a Ph.D. in mechanical engineering from the University of Pennsylvania, an MBA from the University of Chicago, and a BA in physics from Amherst College. He has served as president of the Milwaukee Board of School Directors and started a software company. His research interests include using statistical models to measure school effectiveness.