Applying Behavioral Economics to the Business of Higher Education

by Roland B. Stark

Keywords: behavioral economics, survey research, market research, enrollment management, causality

"If I'd asked my customers what they wanted, they'd have said a faster horse." Henry Ford

Abstract

Colleges and universities are keenly interested in the factors that motivate students or their parents with respect to key decisions: whether to apply, to enroll if accepted, to stay in school, or to contribute as alumni. Abundant evidence shows that traditional means of identifying these factors tend to produce misleading results. "Derived importance" approaches that draw on principles of behavioral economics fare much better. We demonstrate a variety of such methods and recount a series of applied situations in which they have played an important role in illuminating the reasons for key choices. Techniques discussed include group differences on objective vs. subjective measures; correlation; regression; vignette research including conjoint analysis; and market basket research.

Introduction

Of vital importance to institutions of higher education is an understanding of what drives students to apply to a given set of schools, to enroll at one school among several options, to stay in school or withdraw, and to contribute their time, expertise, and/or money as alumni. Unfortunately, determining the drivers of such key behaviors is often anything but straightforward. One may simply ask:

- What made you apply, or not apply?
- What were your priorities in your choice of an enrollment school?
- What reasons were foremost in your retention/withdrawal decision?
- What drove your decisions as to whether to donate?

We will show how such indicators of “stated importance” can be dramatically misleading and how they can obscure the more relevant and sometimes more subconscious drivers of behavior. This paper contrasts traditional, straightforward procedures for evaluating student priorities in the college decision

---

1 Statistician and Research Consultant, Integrative Statistics, Maynard, MA, USA; roland@integrativestatistics.com. A 2011 version of this paper was co-authored by Tara Scholder, then Vice President for Research at Maguire Associates of Concord, MA, where many of the studies cited here were conducted.
process with a series of more advanced procedures. The latter allow an institution to attain a greater understanding of the drivers of student choice and, as a result, to make smarter decisions. In a crowded marketplace, such an understanding can be a competitive advantage, allowing an Enrollment Management office to craft more sophisticated marketing and engagement strategies and better achieve its strategic, enrollment, image, and financial objectives.

**Stated Importance**

With higher education decisions, the most common way to ascertain the importance of a factor has been simply to ask. For example, one might use a construction such as “Please rate each of the following items for its importance in your decision to enroll at a particular school, using a scale of 1 (Not at All Important) to 5 (Extremely Important).” Alternatively, one could ask research participants to choose up to three operative reasons out of a list of ten possible reasons for their choice. One might even opt for a more involved process such as maximum difference scaling (“max diff”), in which one performs complex calculations on a large set of head-to-head importance ratings for different factors. The validity of all such indicators of “stated importance” rests on at least five assumptions:

1. The true reasons are salient enough to stand out in respondents’ minds.
2. Respondents are self-aware enough to accurately answer the question.
3. They are being honest with themselves.
4. They are being honest with the questioner and not merely answering in what they perceive to be a socially desirable manner.
5. They are otherwise rational in the way they conduct their evaluation.

All of these assumptions have been examined comprehensively via survey and experimental research… and frequently found to be untenable. In fact, “The most important thing that social psychologists have discovered over the last 50 years,” writes University of Michigan psychologist Richard Nisbett, “is that people are very unreliable informants about why they behaved as they did, made the judgment they did, or liked or disliked something” (Nisbett, 2007, 269). In publications beginning with an oft-cited paper from 1977, Nisbett and Timothy Wilson describe a host of experimental examples that undermine faith in accepting respondents’ reasons at face value. For example (Nisbett and Wilson, 2007, 270):

In one study experimenters videotaped a Belgian responding in one of two modes to questions about his philosophy as a teacher: he either came across as an ogre or a saint. They then showed subjects one of the two tapes and asked them how much they liked the teacher. Furthermore, they asked some of them whether the teacher’s accent had affected how much they liked him and asked others whether how much they liked the teacher influenced how much they liked his accent.

Subjects who saw the ogre naturally disliked him a great deal, and they were quite sure that his grating accent was one of the reasons. Subjects who saw the saint realized that one of the reasons they were so fond of him was his charming accent. Subjects who were asked if their liking for the teacher could have influenced their judgment of his accent were insulted by the question.

---

2 The main exceptions involve choices made through special expertise – a pro golfer’s choice of a club, or, no doubt, a statistician’s choice of an analytic method.
Nisbett and Wilson’s work has been joined by several landmark studies by Amos Tversky and Daniel Kahneman, two Stanford colleagues who more than any others helped found the field that came to be known as behavioral economics (see, for example, Tversky and Kahneman, 1982). Kahneman, an experimental psychologist who never took an economics course, won the Nobel Prize in Economics in 2002 for his work related to human decision-making. In later works (summarized in Lewis, 2016), he demonstrated many ways in which people made decisions or judgments based on convenient but flawed heuristics (shortcuts) rather than on rational criteria.

In recent years, experimental psychologists Daniel Gilbert of Harvard (Gilbert, 2006), Gerd Gigerenzer of Germany’s Max Planck Institute (Gigerenzer, 2007), Daniel Ariely of Duke (Ariely, 2010), and 2017 Nobel Prize winner Richard Thaler of the University of Chicago have further exposed the pitfalls of using stated assessments of factors’ importance and of relying on the assumptions underlying such assessments. One of Ariely’s celebrated findings is that the amount of money people are willing to pay for a product can be made to vary considerably, simply by the subtle introduction of a random number (the last two digits of their social security number) into the decision process. This random agent has been shown to be operative regardless of the reasons that respondents volunteer for the dollar amounts they give.  

---

**The Pitfalls of Stated Importance in Higher Education Research**

**Reasons for Withdrawal**

For many college counseling staff members, financial woes leap to mind as a primary reason for student dropout. And indeed many counselors have reported that, when they have asked dropouts what sort of issues have posed problems for them at this school, the most common answer has been “the school does not offer enough financial aid.” Isn’t this conclusive evidence for the importance of this factor? Not for those who have gone farther and asked retained students the same question. It turns out that nearly just as many non-dropouts cite lack of financial aid as a problem! We do not deny that insufficient aid is the operative reason for some dropouts, but if the issue is cited nearly as often by both groups, it simply cannot be the overwhelming reason for withdrawal that it is sometimes believed to be.

**“Deal-Breakers”**

One higher-education survey asked about the preferences of potential applicants to a certain small, religious university located 75 minutes from the nearest major city. The survey asked whether each of 15 hypothetical school characteristics would be a "deal-breaker" (the student would definitely not apply to such a school); a "deal-maker" (the student would very likely apply to such a school); or neither.

There were 95 respondents who reported that they had already applied to the school secretly sponsoring the study. Presumably, close to none of those 95 would have marked as a deal-breaker any of the features commonly associated with that school! In reality, 48% did so. In fact, some of these applicants marked as many as three of this school’s well-known characteristics as deal breakers for their application decision.

---

3 Along the way, researchers have also shown ways in which irrationality serves some important purposes and find that it can actually be capitalized upon by those marketing products from tennis shoes to college education. For example, see Bowman (2010) and Grapentine and Weaver (2009).
Even if a few students had forgotten whether they had in fact applied, and even if some had done a poor job of acquainting themselves with their application school, these results seriously call into question a reliance on stated importance.

Derived Importance

Derived importance involves determining a statistical association to help discern respondent priorities. We will explore five methods of deriving importance: group differences, correlation analysis, regression analysis, vignette research, and market basket analysis.

Group Differences

In deriving importance via group differences, we see, for instance, how a factor distinguishes between accepted students who enroll at a school and those who do not. In the case of one university, accepted students rating the school’s major academic programs as “Excellent” were 67% likely to enroll, but those who rated them as “Very Good,” only 34% likely. One might also describe these in reverse order and say that those enrolling were much more likely to rate the school’s major programs as ‘Excellent’ than non-enrolling students. Either way, we can see that opinions of the school’s major programs function very well as a discriminator.

➔ Aren’t respondents still supplying subjective information? Yes, but rather than trying to account for reasons, here they are merely describing what they see. This school has high-quality major programs, that one has moderately-high-quality, and so on. Describing what exists as a perceived “fact” is cognitively and affectively much simpler than disentangling one’s own priorities, the drivers of one’s choices.
Causality can of course be a slippery thing, and statistical connection does not *necessitate* a causal connection. Still, absent any discovery to the contrary (absent any confounding variables), it seems safe to say that what we see here is not a coincidence and that opinions of this school’s major programs do in fact substantially determine one’s enrollment decision.

**Correlation**

This same phenomenon can be expressed in another, numerically more succinct way, using correlation. Rather than contrasting the yields of various groups of students, defined by their opinions, we can characterize the importance of major programs with reference to the relationship between two variables, using a single correlation coefficient, *r*. This is expressed on a scale from +1 through zero to -1. If the correlation is exactly +1, there is a perfect, positive association between the two variables. If exactly -1, there is a perfect, negative association.

How does correlation work in a real world example? Suppose that ratings of faculty quality are expressed on a scale from 1 (Poor) to 5 (Excellent). Suppose also that we are interested in the likelihood of applying to an institution, a likelihood which students are asked to assess on a scale from 1 (Definitely Not) to 5 (Definitely Will). Figure 2 shows how these two variables might relate and how we could derive the importance of the quality of faculty as it relates to application intention.

---

4 While it may be popular to quote the saying, “Correlation does not imply causation,” this can itself mislead, since the definitions of *imply* include both *suggest* and *necessitate*!

5 For a student to gauge his or her likelihood of applying, even though somewhat more complex than noting perceptions of quality, is still a simpler matter than is elucidating any cause-and-effect relationship.
Figure 2
Assessing Derived Importance via Correlation:

How well does rating of college’s quality of faculty correlate with likelihood of applying there?

![Scatterplot of College's Quality vs. Likelihood of Applying]

Steep, narrow band a sign that Quality Rating and Likelihood of Applying correlate highly (high derived importance). Here, $r = .62$.

“Bucks the trend.”

Figure 2 shows that, for the most part, students assigning a low rating to quality of faculty believe they are unlikely to apply, and those giving a high rating, likely. The two variables correlate so strongly ($r = .62$), then absent any confounding variables, or any evidence of reverse causality, we would infer a high level of importance to the faculty ratings in determining likelihood of applying. Using correlation in this manner allows us to conduct more complex and potentially more revealing analyses, as we shall see.

More Conflicts Between Stated and Derived Importance

Suppose that the results of both approaches were very well aligned: what would this look like in practice? Simply put, alignment would indicate that both methods are equally (and quite highly) valid for the purpose intended. In Figure 3, we portray a fictional example in which both stated and derived methods correspond well and lead to the same conclusions.

Note that in this example we “up-level” our use of correlation, and of the scatterplot as a data graphing technique. Instead of using correlation to gauge the derived importance of a single topic such as faculty quality ratings (as in Figure 2), here we use correlation to determine, for a set of 19 topics, whether there is generally a good match between stated and derived importance. In Figure 3, each point is a topic rather than a student.
Figure 3

**Stated vs. Derived: The Ideal Relationship**

If both are quite valid, they will correspond well and we will see a steep, narrow band of points.

![Graph showing derived importance vs. average stated importance]

**Note:** Each point is a topic.

1. Athletic opportunities
2. Interdisciplinary study
3. Academic reputation
4. Small class size
5. Academic facilities (library, classrooms, computers, etc.)
6. Close contact with faculty
7. Caring faculty and staff
8. Campus safety/security
9. Distance from home
10. Quality of faculty
11. Academic competitiveness
12. Value of education (combination of quality & cost)
13. Students you are easily comfortable with
14. Area surrounding campus
15. Availability of financial aid
16. Internship/co-op opportunities
17. Preparation for career
18. Career services
19. Academic advising and learning support services

In the fictional example shown in Figure 3, we find an excellent overall match, with an *r* of .9. The question is how close real data comes to this level of congruence.

Figure 4, based on real data, shows a sharp counter-example where the match is extremely poor. Note how the cloud of points is nearly circular, indicating a correlation that approaches zero. Clearly, when we find that a feature (Distance from Home) that is lowest on stated importance is actually highest on derived importance, something is terribly amiss, and suspicion is cast on the validity of one or both sets of indicators.

---

6 In the case of Distance from Home we show |r|; *r* is actually -.41.
Similar results to that of Distance from Home have been obtained for many colleges regarding the importance of Parents’ Preference, which is typically downplayed, misleadingly so, by students.

Figure 5 builds on this information. It shows seven scatterplots portraying a selection of stated vs. derived relationships like that of Figure 4. The graphs are drawn from surveys of different colleges’ admitted students, inquirers, or parents of inquirers. Each graph plots stated vs. derived importance for about 20 topics as they relate to an enrollment or application decision. In each case, look for the extent to which the points form a narrow band running from lower left to upper right (good alignment between methods) as opposed to displaying a weaker, more circular or “blob-ish” relationship.

Results in Figure 5 are quite mixed. Colleges A ($r = .30$) and D ($r = .16$) show particularly disappointing patterns which, like those in Figure 4, undermine confidence in stated importance indicators. On the other hand, College B ($r = .66$) and College E, Parents of Prospects ($r = .65$) have a stronger, if not very strong, correspondence. The average $r$ from these seven analyses is .40.\(^7\) This translates into an explained variance between stated and derived indicators, $R^2$, of only 16%. Although we would never expect to explain close to 100%, as no survey indicator is perfectly reliable, shouldn’t we realistically hope to explain three to four times as much as this if both approaches were largely valid?

\(^7\) Averaged using the preferred Fisher transformation method.
Figure 5

Stated vs. Derived Importance: College A, Admitted Students

Stated vs. Derived Importance: College B, Admitted Students

Stated vs. Derived Importance: College C, Prospects

Stated vs. Derived Importance: College D, Inquirers

Stated vs. Derived Importance: College E, Inquirers

Stated vs. Derived Importance: College E, Parents of Prospects

Stated vs. Derived Importance: Graduate School, Prospects
A note on correlations: To be able to find a strong correlation assumes a certain amount of variability in the data, and there are many factors that can reduce such variability. In some cases, the scale used may limit opportunities to find strong correlations. Utilizing a five- rather than a ten-point scale can lead to less variability in feedback. Derived correlations will also be depressed by a lack of normality (bell curve) in the shape of a distribution, as when we have binary variables (such as enrolled/not enrolled) or skewed or bimodal distributions.

Is a lack of variability necessarily an undesirable outcome in and of itself?

Not necessarily. What institution would be disappointed to learn that students uniformly believe it has a strong academic reputation or a great location?

Exploring in Depth the Importance of Financial Aid

Everything should be made as simple as possible, but no simpler. Albert Einstein

Next we describe an analysis that drew on survey and institutional data from students accepted at a major mid-Atlantic university, and from their parents, with 6,300 participants in all. Here, survey ratings were made on a variety of topics, including many of those listed in Figure 3 and 4, as well as several dealing with programs offered by Admissions and Financial Aid Offices.

At the request of this institution, these surveys delved particularly deeply into financial matters. Respondents were asked to compare this university’s aid package with that of other schools. They answered Gabor-Granger-type questions about just how large an increase would have been needed to cause them to switch from the nearest competitor school to the sponsoring school, or vice versa. They completed Van Westendorp-style questions about price points. Finally, a wealth of institutional data were available on students’ financial need, aid received, and the gap between price and aid, both at sponsor and nearest competitor. Clearly, some of the results relied on completely subjective input; some, on completely objective data; and some on a hybrid.

Fourteen distinct findings on the importance of financial aid emerged from the considerable analytic commotion. Some findings were drawn from student data (“S” in Figure 6) and some from parents’ (“P”). Figure 6 arranges these findings along the X-axis of subjective – objective and the Y-axis of higher to lower apparent importance of aid to enrollment decisions, i.e., to yield. Placement along the former was determined by the extent to which findings relied on respondents’ own assessment of the importance of aid in their decision; placement along the latter, by $R^2$, where applicable. Certainly there was room for researcher judgment in graphing these findings; the three points marked with a question mark were particularly difficult to place.

But even allowing for researcher subjectivity, the pattern that emerges is extraordinary. Every single finding at the far left (most subjective data) indicated moderate to high importance of aid; every single finding at the far right (most objective data), low to very low importance. With or without the dotted line that estimates the best fit for the points, the diagonal
shape dominates the graph. Indicator objectivity is strongly and negatively correlated with the importance of aid \((r \sim .75)\).

In fact, arguably five out of six hybrid results pointed to just as low an importance level as the fully objective ones did. This seems encouraging for further research, since even a method that partially tempers subjective assessments with some objective data has some promise of arriving at the right answer. For this university, that answer is that aid is quite a minor factor \((R^2 \sim .03)\), regardless of what students and especially parents will claim. This university learned that it possesses surprisingly little leverage in the form of aid and that in enhancing yield it would do better to focus its efforts on other factors that are stronger drivers.

**Regression**

Regression expands on correlation. Regression is a method that shows how each of a set of predictors correlates with an outcome, while controlling for (holding constant) the other predictors. In some cases, it is desirable to include a variety of predictors in a regression and let them “fight it out” to see which are relatively more important. When the predictors involved are very much independent, such as gender and ethnicity, this can work well. Results become much more “muddy” and complicated to interpret when one uses a set of opinion variables such as quality ratings on a number of college features. Mosteller and Tukey put it mildly when they say (1977, p. 327), “Just dumping in a lot of closely correlated variables, and expecting a fit to the data to tell us, directly and simply, which ones are more important usually expresses unjustified optimism.” However, a carefully designed, sequential or hierarchical\(^8\) approach can show how much certain opinions matter when we control for demographic or financial aid variables.\(^9\) We do best when we control for those things that are most objective; that occur earlier in time; and that tend to be most generative, e.g., socioeconomic status as opposed to brand of toothpaste (Davis, 1985).

One way around this difficulty is to first condense a large number of ratings into a small number of “scales,” “indices,” or “factors.” Through factor analysis, one may find that a dozen or more ratings of a school’s features can be effectively reduced to a few factors each entailing opinions on broad topics such as Academics, Social/Community, and Cost/Aid. If they are reasonably independent from one another, such factors can then be plugged into a regression as described above, perhaps alongside more objective variables, to learn more about their relative importance.

**Vignette Research**

Rather than looking at one factor at a time, as with some of the methods described above, the vignette technique, like regression, looks at several factors simultaneously to derive the importance of each. While this technique is more common in psychological and health care research, it is also useful in education market research. For instance, a prospective student might be asked, “All else being equal, how appealing on a scale of 1 (Not at All Appealing) to 10 (Extremely Appealing) would you find a college that was located in a major city, where the net cost of attendance was $25,000, and that had a very prominent athletic program?”

---

\(^8\) Not to be confused with hierarchical linear modeling or HLM.

\(^9\) For an introduction to the use of correlation, regression, and path analysis to analyze cause and effect, see Stark and Courtney (2015).
The student would then assess the appeal of a series of schools with different combinations of the same three factors (location, cost, and athletic prominence). The level of appeal is used as an outcome variable in conducting an analysis of variance (ANOVA) on the resulting data that allows one to assess the importance of each factor. If the levels of each factor are limited to two (city vs. non-city; $25,000 vs. $20,000; prominent vs. not prominent athletic program), then there is a manageable number of eight (or $2^3$) vignettes to test. With more combinations, one may need to explore experimental design options and evaluate the differences between fractional factorial and full factorial designs.

In the same spirit, *conjoint analysis* is a more specialized and sophisticated subset of the above ANOVA approach. The essential problem addressed by conjoint analysis is: what if there are too many factors with too many levels of each to realistically ask a subject to assess all combinations? To expand upon the example just provided, suppose one wanted to test:

- 3 locations (major city vs. small city vs. country),
- 3 cost levels ($15,000 vs. $20,000 vs. $25,000), and
- 2 athletic levels (prominent vs. not prominent).

To cover all possibilities and allow for a standard ANOVA, one would need to test $3 \times 3 \times 2$ or 18 combinations. Such a long series of scenarios might make a survey participant blanch. The conjoint procedure allows one to choose a representative subset of the 18 that, while short of being conclusive, will give some indication of the relative importance of each of the three factors.

We used a vignette method with conjoint analysis to derive importance in a study of parents of prospective college applicants (Figure 7). In this example, parents assessed combinations involving religious affiliation, cost, selectivity, and enrollment size. The analysis revealed that, though parents were hesitant to admit it in a stated importance context, they highly prioritized a school’s cost when they made their application choice (perhaps as opposed to enrollment choice, as in one analysis above). The lowest cost level tested (blue) drew very positive responses, and the highest cost (yellow), very negative. Selectivity, to a lesser extent, was another area where this approach revealed considerably greater importance than that indicated by stated importance measures. Vignette analysis thus made it possible to bypass some perhaps socially acceptable responses that clouded the real importance of these dimensions.

**Figure 7. Conjoint Analysis**

Does the vignette method (conjoint or otherwise) conclusively demonstrate causality? No, but being a hybrid of subjective and objective approaches, it accomplishes at least two things. It is likely to avoid...
some of the pitfalls that come with completely subjective data, and it may well appeal to stakeholders who prize having some sort of choice-based ratings to analyze.

**Market Basket Research**

Students often spend much time and effort deciding on the colleges and universities to which they will apply. This *set* of schools can provide higher education professionals with a wealth of information. Not only can we discover the specific schools against which our institutions compete for the application interest of a particular student, but we also can learn the preferences and drivers of interest among applicants through the in-depth study of these sets. Analyzing the characteristics and features of the schools in such a set has been called a College Market Basket (CMB) analysis.

Market basket (or affinity) analysis is a modeling technique based upon the theory that if we buy certain items we are more likely to buy certain others. Simply put, we examine products that are purchased together. In the retail world, market basket analysis has confirmed that customers who tend to buy hot dogs also buy buns, and those who buy cold medicine also frequently purchase tissues and (because research retractions are so seldom noticed) orange juice. Market basket analysis can also lead to less obvious associations such as the discovery that at one retail chain beer and diapers were often purchased together—by men.

Applying this technique to enrollment management can reveal insights into application behavior and college preferences, including intended mobility for college, college type and size preference, and price sensitivity. Such an analysis conducted for one large, Catholic, selective, expensive Eastern university uncovered several useful facts about accepted students:

- They had applied almost exclusively to private schools, and tuition was consistently high across a given student’s CMB. However, given that, they chose schools with surprisingly wide ranges of selectivity levels, mean SAT scores, and retention rates.
- Enrolling students’ schools averaged 41 points lower in SAT scores than non-enrolls.
- For most students, no more than half of schools listed were religious schools.

→ Does the CMB method conclusively demonstrate causality? No, but it does bring to light statistical connections based more or less on hard facts. That is, unless a respondent (or agency supplying such information) makes errors in listing application schools.

**Concluding Comments**

We have presented guidelines on incorporating and interpreting valid measures of different factors’ importance in decision-making. We hope this material will be useful to researchers and consumers of research in higher education. Our work confirms over and over again that determining the drivers of key behaviors such as application, enrollment, retention and support shouldn’t be treated as a simple process, but requires a thoughtful and deliberate approach and a varied toolkit. The reward of such extra effort is a greater and more sophisticated understanding of the drivers of decision-making to inform institutional
strategy. Perhaps needless to say, we also feel that these ideas, like so much connected with behavioral economics, have wide applicability in fields beyond education.

**Sources**


