What Drives Student Choices?
Applying Behavioral Economics to Higher Education
by Roland Stark and Tara Scholder

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

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 leave, and to contribute their time, expertise, and/or money as alumni. Unfortunately, determining the drivers of such key behaviors is often anything but straightforward. At its simplest, one may ask students why they did or did not apply or enroll and ask alumni why they did or did not provide support. However, stated importance can sometimes lead one to miss more subconscious drivers of behavior, which is better ascertained through more sophisticated analyses.

This paper contrasts straightforward methods of evaluating student priorities in the college decision process with more advanced methods. While both methods have applicability and value in understanding student choice, the addition of more advanced methods often allows one to attain an even greater understanding of student choice and, as a result, make smarter decisions. In a crowded marketplace, an institution’s greater understanding of the drivers of student choice can be a competitive advantage, allowing a college or university to craft more sophisticated marketing and engagement strategies and better achieve their strategic, enrollment, image, and financial objectives.

Stated Importance
The most common way to ascertain the importance of a factor is 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 out of a list of perhaps ten possible reasons for their choice. These are just two of many ways to assess stated importance.

Regardless of the method, the validity of stated importance rests on at least five assumptions:

1. The true reasons are salient enough in the target audience’s mind to stand out.
2. They are self-aware enough to accurately answer the question.
3. They are being honest with themselves.
4. They are being honest with us and not answering in a socially desirable manner.
5. They are otherwise rational in the way they conduct their evaluation.

All of these assumptions have been examined extensively via survey and experimental research and found to be potentially misleading in selected instances. “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 respondents’ rationality and honesty with themselves and others. For example (Nisbett, 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.

Nisbett and Wilson’s paper was followed soon after 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. He demonstrated many ways in which people made decisions or judgments based on convenient but flawed heuristics (shortcuts) rather than on truly rational criteria.

In recent years, experimental psychologists Daniel Gilbert of Harvard (Gilbert, 2006), Gerd Gigerenzer of Germany’s Max Planck Institute (Gigerenzer, 2007), and Daniel Ariely of Duke (Ariely, 2008) have further exposed the dangers of relying on statements about factors’ importance and on relying on the assumptions underlying such statements. One of Ariely’s celebrated findings is that the sum 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. Along the way, these researchers have 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).

Derived Importance

While stated importance is collected by simply asking respondents to assess the importance of a particular product or service attribute, derived importance involves determining the statistical association between performance or evaluations on an attribute and an outcome behavior or a broader performance criterion. Statistical methods then are employed to discern respondent priorities. Within the body of this paper, we will explore three methods of deriving importance: looking at group differences, correlation analysis, and regression analysis. In an addendum, we describe several other methods such as vignette research and market basket analysis.

Group Differences

The first method we will examine involves studying group differences. Deriving importance in this way, we see how a factor distinguishes between, for instance, accepted students who enroll at a school and those who do not. In Figure 1, those rating the school’s major programs ‘excellent’ are 67% likely to enroll, but those who rate them ‘very good,’ only 34% likely. Another way to describe these data is to
say that those enrolling are 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.

**Figure 1**

![Graph showing Derived Importance Assessed Through Group Differences](image)

While causality can of course be a slippery thing and statistical connection does not *necessitate* a causal connection, it seems safe to say that what we see here is not a coincidence and that opinions of the school’s major programs do in fact substantially determine one’s enrollment decision.

**Correlation**

This same phenomenon can be expressed in another, more succinct way, namely the method of correlation. Rather than contrasting the results of the two groups of enrolling and non-enrolling students, for example, we can characterize the importance of major programs with reference to the relationship between two variables using a single number, $r$, expressed on a scale from (usually) +1 through zero to -1. If the correlation is exactly +1, there is a perfect, positive association between the two variables. If the correlation is exactly -1, there is a perfect, negative association.

How does correlation work in a real world example? Suppose quality ratings of faculty are expressed on a scale from 1 (Poor) to 5 (Excellent). Suppose also that we were interested in students at an earlier stage in their decision-making process, that is, the likelihood of applying to an institution 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 interest.

“Although science may be the holding of multiple working hypotheses, the picturing of data allows us to be sensitive not only to the multiple hypotheses we hold, but to the many more we have not yet thought of, regard as unlikely, or think impossible.”

*John Tukey*
Assessing Derived Importance via Correlation:
How well does rating of college’s quality of faculty correlate with likelihood of applying there?

Since assessment of the quality of faculty and likelihood of applying correlate so strongly ($r = .62$), we can infer a high level of importance to the former. Using correlation in this manner allows us to conduct more complex and potentially more illuminating analyses, as we shall see.

In the next section, we show examples in which findings drawn from stated importance contrast with those from derived importance.

Conflicts Between Stated and Derived Importance

Before we discuss conflicts between the methods, suppose that the results of both methods were very well aligned: what would this indicate? Simply put, alignment indicates 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. In this example, we “up-level” our use of correlation and of the scatterplot as a data graphing method. Instead of using correlation to gauge the derived importance of a single topic such as faculty ratings (as in Figure 2), now we use correlation to determine, for each of 19 topics including quality of faculty (see feature #10), whether there is a good match up between stated and derived importance.
Here, it is as if 19 results from analyses such as that of Figure 2 are condensed into one graphic, each plotted along the vertical (y) axis and matched up with its corresponding stated importance on the horizontal (x) axis. In the fictional example shown in Figure 3, we find an excellent overall match up, with an \( r \) of .89. The question is how close real data comes to this level of congruence.

**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 ideal relationship between stated and derived 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

Figure 4 shows a sharp counter-example where the match up is extremely poor. The plot shows real data drawn from a survey of 148 college-bound students. Note how the cloud of points is nearly circular, indicating a correlation \( (r) \) that approaches zero. Clearly, when we find that a feature lowest on stated importance is actually highest on derived importance (see feature 9 - Distance from home), something is amiss, and suspicion is cast on the validity of one or both sets of indicators. Similar results have been obtained for many colleges regarding the importance of parents’ preference, which is typically downplayed (apparently, misleadingly so) by students.
Figure 5 shows seven additional graphs portraying a selection of stated vs. derived relationships. The graphs are drawn from 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 as opposed to displaying a weaker, more circular or “blobbish” relationship.

Results in Figure 5 are quite mixed. Colleges A ($r = .30$) and D ($r = .16$) show particularly disappointing results 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 strong correspondence. For all of these data sets, the average $r$ is .40 (using the preferred Fisher method). This means that the average explained variance between stated and derived indicators is only $0.40^2$ or 16%. Although we would never expect to explain 100%, as no survey indicator is perfectly reliable, should we realistically hope to see 40% or 50%, as with a correlation ($r$) of .6 or .7?
Figure 5

Stated vs. Derived Importance: College A, Admitted Students

Derived Importance: r of Feature's Quality with Enrollment Decision

Stated Importance (range: 1 to 5)

Stated vs. Derived Importance: College B, Admitted Students

Derived Importance: r of Feature's Quality with Enrollment Decision

Stated Importance (range: 1 to 3)

Stated vs. Derived Importance: College C, Prospects

Derived Importance: r of Feature's Quality with Application Decision

Stated Importance (range: 1 to 5)

Stated vs. Derived Importance: College D, Inquirers

Derived Importance: r of Feature's Quality with Application Decision

Stated Importance (range: 1 to 5)

Stated vs. Derived Importance: College E, Inquirers

Derived Importance: r of Feature's Quality with Application Decision

Stated Importance (range: 1 to 5)

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

Derived Importance: r of Feature's Quality with Application Decision

Stated Importance (range: 1 to 5)

Stated vs. Derived Importance: Graduate School, Prospects

Derived Importance: r of Feature's Quality with Application Decision

Stated Importance (range: 1 to 5)
A strong correlation assumes 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 ten- or seven-point scale rather than a three- or five-point scale can lead to more variability in feedback and responses. 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.

In Figure 6, we present a scheme for interpreting charts like those in Figures 3-5, where stated and derived importance findings are compared for a group of topics. Points along the main diagonal can be trusted based on both methods. Those in the upper left are probably cases where the derived findings are more trustworthy. Any points in the lower right are more difficult to interpret as there are many technical, statistical reasons why derived figures could be spuriously low, including low variability in one or both measures or a lack of normality.

**Figure 6**

**Stated vs. Derived Importance Scatterplots: Interpreting Results**

<table>
<thead>
<tr>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust Derived</td>
<td>?</td>
</tr>
</tbody>
</table>

**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 factors in a regression and let them “fight it out” to see which are revealed to be relatively more important. When the factors 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 features. As Mosteller and Tukey put it (1977, p. 327), “Just dumping in a lot of closely correlated variables, and

*Is a lack of variability necessarily an undesirable outcome?*

No, not necessarily. What institution would be disappointed to learn that students uniformly believe it has a high academic reputation, is prestigious, or in a great location? In some cases, consistency is a good thing.
expecting a fit to the data to tell us, directly and simply, which ones are more important usually expresses unjustified optimism.” However, a careful, sequential or hierarchical approach can show how certain opinions matter when controlling for demographic or financial aid variables (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 can find that a dozen or more ratings of a school’s quality on different features can be effectively reduced to a few factors describing 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 the regression described above, perhaps alongside or following demographic variables, to learn more about their relative importance.

**Concluding Comments**

Our review of the literature on stated vs. derived importance and research experience favors the use of a combination of the two techniques. Some authors, like Nisbett, argue for researchers to eliminate altogether any efforts to elicit stated importance. Internal results, while not quite as categorical, do confirm the superiority of derived importance in selected instances. In this paper, we have presented a scheme for interpreting results from stated and derived importance measures which we hope will serve as a useful guide to other researchers and non-researchers. Our research confirms over and over again that determining the drivers of key behaviors such as application, enrollment, retention and support is not a straightforward process, but requires a thoughtful and deliberate approach and understanding of the strengths and weaknesses of both methods. The reward of such extra effort, then, is an even greater and more sophisticated understanding of the drivers of decision making to inform institutional strategy.

**Sources**


Addendum:
Two Additional Methods to Derive Importance

Vignette Research

Rather than looking at one factor at a time, another technique looks at several factors simultaneously to derive the importance of each. While such techniques are more common in psychological or health care research, they are also useful in education surveys. 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 cost of attendance to you was $25,000, and that had a very prominent athletic program?”

The student would then assess the appeal of a series of schools with different combinations on the same three factors (location, cost, and athletics). 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.

Conjoint analysis is a more specialized and sophisticated subset of the above ANOVA approach to analyzing vignette testing data. 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 3x3x2 (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.

Figure 7

Conjoint Analysis: Preference Within Dimensions

Catholic Prospective Parents

![Conjoint Analysis Chart](image)
Maguire Associates recently used a vignette method to derive importance in a study with prospective parents (see Figure 7). In this example, the parents were assessing the relative importance of religious affiliation, cost, selectivity, and enrollment size. The analysis revealed that parents are hesitant to admit, in a stated importance context, the extent to which they weigh a school’s cost when they make their application choice. Of the three levels of cost tested, one (the lowest cost tested) drew very positive responses (blue) and one (the highest cost tested) drew very negative responses (yellow), while participants were neutral about the mid-cost option. Selectivity, to a lesser extent, was another such area where this approach revealed considerably greater importance than that indicated by stated importance measures. Vignette analysis thus made it possible to get around socially acceptable responses that clouded the real importance of these dimensions to parental preferences.

**Market Basket Research**

Students often spend a lot of 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, we also can learn the preferences and drivers of interest among applicants through the in-depth study of the students’ college application sets. This is accomplished by analyzing the characteristics and features of the schools that are included among a student's application set or his/her College Market Basket (CMB).

Market basket (or affinity) analysis is a modeling technique based upon the theory that if we buy a certain group of items we are more (or less) likely to buy another group of items. Simply put, it is a technique that seeks to find the relationships between purchases. The idea behind market basket analysis is simple: examine products that are purchased together. In the retail world, for example, there are natural product and service affinities. Market basket analysis would confirm that customers who tend to buy hot dogs also buy buns and those who buy cold medicine also frequently purchase tissues and orange juice. Market basket analysis can also lead to less obvious associations such as the discovery that beer and diapers were often purchased together—by men—at one retail chain.

Applying the market basket analytic technique to enrollment management, we can think of a student’s application set as a market basket of sorts—the College Market Basket. Analysis of the makeup of a CMB reveals important insights into application behavior and college preferences, including intended mobility for college, college type and size preference, and indications of price sensitivity. Analysis of the CMB of applicants, for example, can lead to a greater understanding of:

- **School affinities** – Likelihood of two colleges or types of colleges being in a student’s application set. For example, if a student includes one single-sex college in his/her CMB, how likely is he/she to include a second?

- **Application driver characteristics** – Identification of the features or characteristics that drive students to include the schools they do in their application set. For example, how much of a driver is distance from home, college type, enrollment size, and cost?

- **Affinity clusters** – Classification of students into subgroups based on the composition of their application set. For example, what demographic, preferential, and attitudinal differences exist between students by College Market Basket composition?