2016

Proposal of a New Method to Classify Higher Education Institutions Instead of Rankings

Kate D. Elam
Otterbein University, kate.elam@otterbein.edu

Proposal of a New Method to Classify Higher Education Institutions Instead of Rankings

Kate Elam
Department of Business, Accounting, and Economics
Otterbein University
Westerville, Ohio 43081

01 April 2016

Submitted in partial fulfillment of the requirements for
Graduation with Distinction

Advisory Committee:

Michael A. Levin, Ph.D. Distinction Advisor	________________________ Advisor's Signature
Sonja Martin Poole, Ph.D. Second Reader (or Co-advisor if applicable)	________________________ Second Reader's Signature
Meredith Meyer, Ph.D. Distinction Representative	________________________ Distinction Rep's Signature

Table of Contents

Table of Figures

Abstract

Purpose – The objective of this paper is to propose a new method for classifying higher education institutions that differs from traditional rankings. The proposed model segments colleges and universities based on like attributes and describes the groups according to their differentiating features.

Research Method – Data from several rankings publications was collected and simplified for analysis. Several observations (i.e. schools) were eliminated because they did not fit the sampling frame and/or did not include a sufficient amount of information. Duplicate variables (i.e. attributes) were also eliminated. First, exploratory factor analysis was applied to reduce the number of variables being examined. Second, cluster analysis was employed to segment the observations.

Findings – Exploratory factor analysis and cluster analysis revealed five clusters of schools based on seven main underlying factors. Several post-hoc analyses determined that the EFA and cluster models were stable. These analyses confirmed that the constructs measured are in fact distinct, as are the five generated clusters.

Practical Implications – The five-cluster model generated from this study has practical and beneficial applications for both higher education managers and prospective college students and their parents. The insights gained from the model can help colleges and universities more effectively target students in their marketing efforts, and prospective students and their parents can make better-informed decisions about college plans.

Originality/Value – This paper is the first to propose the use of EFA and cluster analysis to segment higher education institutions. The validity of traditional college rankings has been questioned in recent years, and the model proposed in this paper solves many of the inherent problems with conventional rankings.

Proposal of a New Method to Classify Higher Education Institutions Instead of Rankings

American universities have long used rankings to compare or benchmark their performance and status against other universities (Altbach, 2012). Websites and magazines generate college rankings as information that could be considered essential to students who are making decisions about their higher education plans. These rankings vary based on what attributes or factors are included in the rankings and how these attributes or factors are weighted (see Appendix A). These rankings have also contributed to the perception that American higher education has become commercialized. In such rankings, "the higher the position that a university occupies, the more its brand is recognizable…Therefore the ranking of universities has a dimension of pure consumerism" (Hejwosz, 2010). Since U.S. News & World generated its first rankings in 1983, other organizations such as Fiske, the Princeton Review, and The Washington Monthly now offer rankings (Tierney, 2013).

While college rankings generate newsstand sales ("Ranking affects the financial resources of public colleges," 2016) and traffic for websites ("U.S. News breaks online traffic record," 2013), at least six problems remain with these rankings. One, it is difficult to compare different attributes or factors for schools. For example, public and private colleges lack equal measures for comparison (Anderson, 2015). Two, prospective students would struggle to find a school that provides a good personal fit because long rankings typically include many different types of schools ("What's wrong with college rankings?"). Three, data provided by schools could be unreliable if there is a disconnect or misunderstanding about what is being measured. Reputational factors contribute to the ambiguous nature of the data that compose rankings (Altbach, 2012). Four, the variables included in rankings make qualitative comparisons nearly impossible to make. For example, one of the major functions of any institution of higher education—teaching—is often ignored, yet ways to measure such a factor have yet to be developed. Coupled with this inability to objectively measure qualitative factors, variables like research frequently earn more attention because they have more clear-cut sets of measures (Altbach, 2012). Five, the variables on which various rankings are based are sometimes unclear. The variables "…vary in quality, focus, and in the specifics of their methodologies," (Altbach, 2012). Therefore, factors that should be considered important can be disputed between different publications. Six, higher education institutions have used the rankings systems to influence their position on popular publications' lists. Knowing of the measures used to compute particular rankings, schools may allocate resources in an attempt to improve their rank. Altbach (2012) describes, "…in the age of globalization it is easier for academic institutions to leapfrog in the lists with thoughtful planning and adequate resources." Schools such as Tulane University, Bucknell University, Claremont McKenna College, Emory University, and George Washington University have all admitted to submitting incorrect test scores to U.S. News and/or overstating the high school rankings of their incoming freshman (Anderson, 2013).

All of these outward issues are a result of the underlying problem with rankings—they are on an ordinal scale. Ordinal scales show the degree to which a particular measurement exists. They do not, however, represent the absolute size of the values nor indicate any differences between values (Lehmann et al., 1998). In the case of college rankings, the assumption is that a more highly ranked school is better than a lower ranked school. But, being on an ordinal scale, there is no way to know how much "better". When classified using a ranking, fundamentally different schools (e.g. public vs. private) are compared on the same dimensions. Colleges with a higher rank may be interpreted as better schools, when it is possible that they are actually very different in their characteristics.

Grouping colleges by similar or like attributes or factors could resolve the issues associated with rankings because groups are formed by the similarity of the observations or colleges based on a set of attributes or factors (Peterson, 2000). While several analytic approaches to form groups exist, this paper argues for the application of exploratory factor analysis (EFA). The primary purpose of EFA is to define the underlying structure among the variables being analyzed by grouping sets of variables that appear highly interrelated (Hair et al., 2010).

To analyze the data from college rankings, EFA is applied to reduce the number of variables considered to compare schools and define groups of similar characteristics as broader, descriptive factors. EFA has popularly been used for social, behavioral, and psychological applications. Additionally, it has been utilized in marketing to define market segments of industries and companies. However, the use of EFA has not been seen in the context of higher education. Therefore, the objective of this paper is to propose EFA as a method for classifying schools based on their respective attributes, rather than ranking all colleges and universities together on one ordinal scale.

Literature Review

As part of the theory of perfect competition, consumer demand for a given product is considered to be homogeneous (Hunt, 2000). Equilibrium-seeking economic thought argues for the role of perfect information by consumers and firms. Furthermore, this information is regarded as costless and widely available. Hence, consumers know about all products within a given market and attempt to maximize self-interest.

In perfect competition, a firm’s ultimate objective is profit maximization. The assumption of perfect competition is that every firm has the ability to obtain perfect and costless information in order to achieve profit maximization. As a result, competing firms quickly reach equilibrium as they can easily adapt to consumer demands. At the same time, it is assumed that buyers are able to obtain perfect and costless information about firms. Thus, consumers' decisions are straightforward and predictable from a firm's perspective.

Perfect competition struggles to explain competition in the higher education industry. For the purpose of this paper, the higher education industry is limited to schools that offer at least a bachelor's degree and operate with a non-profit view. In this higher education context, consumers hold asymmetric information in regards to the schools within the higher education market. Information asymmetry exists when one entity in a transaction holds more or values differently information compared to the other party. Typically the seller—in this case each college or university—is the party that holds more information than the buyer ("Asymmetric Information").

When colleges and universities hold more information than their buyers (i.e. students), they effectively participate in price discrimination by awarding different financial aid packages to different students. In order for price discrimination to occur, firms must have some market power to alter their prices for some customers, and it must not be possible for consumers to resell the product to others. The higher education market fits this definition of price discrimination, as 1) institutions can easily lower net tuition costs to certain students by offering varying amount of scholarships, grants, and loan opportunities, and 2) it is impossible for low-price students to resell their enrollment to high-price students (Lawson & Zerkle, 2006). Consequently, it is difficult for any student to know the true cost of attendance at any given institution because no two students will be offered exactly the same net tuition cost.

The result is that buyers (i.e. students) in the higher education market have imperfect information about the firms (i.e. colleges and universities) in the market, meaning the market is not in fact perfectly competitive. Colleges and universities compete and differentiate themselves among one another in terms of quality, major offerings, location, sports teams, etc. Yet, all institutions of higher education offer essentially the same product—a college degree (Lawson & Zerkle, 2006).

The Theory of Monopolistic Competition (Chamberlin, 1933) could offer an explanation of competition as it exists in the higher education market. Chamberlin described a situation he observed where a market appears to be perfectly competitive—where many buyers and sellers participate in the economy—yet certain external factors (e.g., buyer preferences) skew the model from perfect competition (see Appendix B). Additionally, Chamberlin coined the term "product differentiation" to describe the efforts of sellers to distinguish their products from others by making them more attractive to targeted buyers (Chamberlin, 1933).

Under monopolistic competition, firms will seek to maximize profits in the short-run (see Appendix C). In the long-run, profits for all firms will reach equilibrium at zero, as the model fluctuates to reflect changes in the market ("Chamberlin's monopolistic competition") (see Appendix D). In higher education, revenue in the form of charitable donations affects profits in the long-run. Institutions are increasingly reliant on donations as a source of long-term finance. Yet, factors like institutional reputation, scholastic aptitude of students, and career choices of graduates heavily impact alumni donations. A college considered high achieving by these measures receives more funding through alumni gifts. As a result, more prestigious colleges will be perceived as higher quality than competing institutions (Cochi Ficano & Cunningham, 2002).

The Theory of Monopolistic Competition could explain how universities compete for students. All institutions of higher education offer inherently a similar product—a degree that reflects some level of knowledge—while simultaneously trying to appeal to the preferences of their potential buyers (e.g., students). Schools differentiate themselves from competition schools through a variety means such as reputation, campus environment, and program offerings.

College rankings can reinforce this view of reputation because the rankings establish a relationship between schools (Altbach, 2012). For example, Harvard is considered a top school because ranking services such as *U.S. News and World Report* historically place Harvard within the top 10 of all universities. Indeed, college rankings communicate information to consumers about the features of colleges and universities. However, each source reports on different attributes for each school (see Appendix A). These inconsistent rankings become problematic when readers take the rankings at face value and are not truly aware of what each ranking measures and how well it does so (Altbach, 2012).

Ranking services incorporate a variety of attributes to determine a list of universities. While this approach could be simple enough, the ranking services are also competing with each other for readers and clicks. As a consequence, each ranking service will incorporate different attributes and weight differently these attributes. For example, Fiske and The Princeton Review have each created guidebooks that rate a number of schools on certain attributes. The *Fiske Guide to Colleges* defines itself as "the definitive college guide of its type" that covers specific aspects of "the best and most interesting institutions in the nation." *Fiske* touts its position as "the most authoritative source of information for college-bound students and their parents" to stand out among other college guides. *The Princeton Review* attempts to differentiate its college guide by listing what it considers to be "The Best 380 Colleges," rather than creating a ranking list of all included schools. The producers of the top 380 schools list state the purpose of the list being "…a resource that would help give applicants up-to-date stats and fact, plus 'straight from the campus' student feedback…to find the schools best for them."

The competition among ranking services creates inconsistency in information available to students and their parents, or consumers. For instance, U.S. News & World Report's college rankings do not take into account measures of educational quality such as teaching or career outcomes upon graduation. A large portion of the elements used to construct the U.S. News rankings is based on reputational measures, which causes more prestigious schools to stand out over others, regardless of any particular characteristics. As a consequence of this competition, consumers have information asymmetry because the ranking services provide various types of information that is interpreted and reported differently.

Further, the use of rankings does not provide prospective students with enough knowledge to assess which schools may be a good personal fit. Highly ranked schools may not necessarily be a suitable option for every student in terms of academic environment, social environment, or athletic and other extracurricular opportunities. Consumers once again have asymmetric information in this regard, as it is difficult to truly distinguish between colleges based on what they individually consider important.

Another issue with rankings is that they attempt to compare attributes of dissimilar schools. Therefore, different schools may dispute which factors should be considered important. In college rankings, schools are compared to one another based on everything from athletics, to facilities, to location. Forcing all of these different factors into a single, mono-dimensional ranking is not possible ("What's wrong with college rankings?," 2015). Additionally, factors related to research (research funding, publications, Nobel prizes, etc.) tend to be heavily weighted in most rankings. Research typically has a greater influence on a school's place on a ranking because it has the most clear-cut set of measures and also has the highest prestige (Altbach, 2012).

With objective measures prevailing over more subjective ones, many colleges and universities have used this information to influence their position on popular publications' lists. With rankings, there is little room for qualitative comparison. Therefore, statistics that can be measured unambiguously, such as test scores and acceptance rates, are given more attention and seen as areas that can be changed to improve a school's ranking. For example, schools have been found to reject or waitlist top applicants who appear to be "overqualified." This approach is used because admissions committees think the applicant will be accepted and choose to attend a more prestigious institution. By sending out fewer acceptances in total, a school that practices this tactic will appear more selective and, as a result, boost their rank ("What's wrong with college rankings?," 2015).

To resolve these problems created by rankings and ranking services, EFA could be used. Exploratory factor analysis (EFA) is a statistical technique used to evaluate a large number of variables in an analysis. The primary purpose of EFA is to define the underlying structure among the variables being analyzed. EFA groups together sets of variables that are highly interrelated (Hair et al., 2010). EFA is used as a data reduction technique, as similar variables are grouped together into broader factors. The resulting broader factors obtained through EFA contain a significant amount of the information represented by the smaller individual variables (Peterson, 2000).

Exploratory factor analysis (EFA) has been widely used in behavioral research and social psychology. In the context of marketing, factor analysis has been used by Jennifer L. Aaker to develop the idea of brand personality, which is defined as the set of human characteristics associated with a brand. Gary L. Frazier & Walfried M. Lassar utilized EFA to describe the decisions manufacturers make with regard to the distribution of their products, based on consumers' interactions with various consumer products. Factor analysis has also been employed in many experiments to identify and explain market segments (Peterson, 2000).

Even with its many applications in other fields, EFA has not been utilized in the context of higher education. Therefore, the research question of interest in this paper is: Can colleges and universities be grouped together based on their specific attributes? EFA will be used to identify which factors relate schools to one another as well as determine how closely related they are.

Research Method

The sampling frame for this research project was limited to public and private non-profit universities that offered at least a bachelor's degree. The initial census contained 1,729 variables and 7,793 observations (e.g., universities). The initial census was reduced by eliminating the following types of institutions: (a) community colleges, (b) technical and trade schools, (c) for-profit institutions, (d) schools that offer distance-only education, (e) institutions that no longer operate, and (f) schools that do not grant primarily 4-year degrees. Additionally, schools identified as Associate's or Special Focus institutions by their Carnegie Basic classification were excluded because these schools did not fit the sampling frame. Finally, colleges not mentioned by Washington Monthly and/or The Brookings Institution were removed because these schools lacked a sufficient amount of information related to the 1,729 variables. The final census included 1,175 universities, or observations.

Initially, 1,729 variables were collected from several sources, including U.S. News & World Report, Forbes, and the Princeton Review (see Appendix A).

Duplicate variables were removed from the dataset. Correlation analysis was conducted on the dataset to remove variables that showed higher degree of correlation (e.g., $r > .8$ and $r < -.8$). Remaining categorical variables were recorded as dummy variables. Several variables were removed because such variables did not include many if not most universities (see Appendix E). The final dataset contains dummy variables and continuous variables (see Appendix F).

To resolve the research question for this paper – Can colleges and universities be grouped according to their characteristics? – the following procedure was followed as discussed by Lilien, Rangaswamy, and De Bruyn (2013). Traditional segmentation using this approach consists of five steps: 1) reduce the data, 2) develop measures of association, 3) identify and remove outliers, 4) form segments, and 5) profile segments and interpret results. Lilien et al. (2013) describes how large datasets with wide variety of items often measure similar or interrelated constructs. Removing overlapping variables prior to analysis is necessary because such variables can prevent detection of the segment structure in the data. Factor analysis is the recommended method for reducing a large number of segmentation basis variables. Using the simplified set of factors, cluster analysis identifies a structure within a body of data. Finally, the segments generated through cluster analysis must be interpreted and defined through discriminant analysis, "…which seeks combinations of descriptor variables that best separate the clusters or segments," (Lilien et al., 2013, p. 89).

The dataset used for exploratory factor analysis included 99 variables and 1175 observations (e.g., universities). EFA is a data reduction technique that allows to uncover unobservable structure to the dataset. SPSS (IBM SPSS Statistics for Windows, Version 19.0) was used to conduct the EFA using a varimax rotation and a principal component analysis.

Consistent with theory, 77 variables were removed for cross-loading and low loadings (e.g., $r < .6$). The final model generated through EFA included 7 factors based on 22 of variables (see Appendix G). The dataset was further refined to include only the variables identified through EFA.

This refined dataset was used to perform cluster analysis to group the 1,175 universities. Marketing Engineering (Marketing Engineering for Excel Software) was used to create the cluster by including 17 segmentation variables and 25 discriminant variables (see Appendix H). Segmentation variables form the clusters and discriminant variables describe the clusters.

The initial cluster analysis relied on K-Means clustering technique. This approach partitions objects into K groups or clusters. The clusters are separated so that each object belongs to only one cluster and each cluster contains at least one variable. In K-means clustering, goodness of fit is determined by the distance between objects in a particular cluster as well as the distance between clusters. Comparing several values of K according to their visual and conceptual interpretability can indicate an appropriate cluster solution (Lehmann et al., 1998, p. 575). The final solution of five groups was then verified through Hierarchal clustering technique. Hierarchical clustering is used to separate observations (in this case, individual schools) into groups based on similarity. Rather than producing one set of K clusters, hierarchical clustering builds a hierarchy of possible solutions. Most hierarchical approaches are agglomerative, where each observation begins as its own cluster, and then the objects are sequentially combined, or agglomerated, according to similarities. This process can be seen in the dendogram for the cluster solution. In a hierarchical cluster model, the two most similar objects, or clusters, are combined first to form a new cluster. The next two most similar objects are combined next to form another cluster, and so on. The appropriate number of clusters can be determined by examining the distance between the groups. The primary application of cluster analysis in marketing is market segmentation (Lehmann et al., 1998, p. 575). Therefore, the eventual use of hierarchical clustering is appropriate for segmenting the higher education market.

Analysis

The first step in generating segments of schools based on attributes was factor analysis to reduce the number of dimensions being analyzed. This step was simplified the final model, which reduced the initial model from 99 variables to 27. The data for these 27 variables for all 1,175 observations (schools) was then analyzed in cluster analysis.

To determine whether the EFA solution was good, several post-hoc analyses were conducted. One, only factors with eigenvalues greater than 1.0 should be kept in the model (Hair et al., 2010, p. 109). Two, the model should contain enough factors that meet an acceptable percentage of variance explained, typically at least 60% (Hair et al., 2010, p. 109). Three, factor loadings of greater than ±0.5 are generally considered necessary for developing a suitable model (Hair et al., 2010, 117). Additionally, a large Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and a statistically significant value for Bartlett’s test of sphericity indicate that sufficient correlations are present among the variables (Hair et al., 2010, p. 104). The EFA solution derived in the first step of analysis met all standards for a stable factor model (see Appendices G, I, J).

In the second and final step of segmenting the schools, cluster analysis of all 1,175 schools revealed five distinct segments (see Appendix K). Furthermore, a discriminant analysis of these five clusters applied several categorical variables to the cluster solution to determine appropriate descriptors for each cluster.

A suitable cluster analysis solution was determined based on a confusion matrix provided for the five-cluster solution (see Appendix L). The given hit rate from the confusion matrix indicates that 65.79% of the observations were correctly classified by the discriminant function. That is, about two-thirds of the schools present in the analysis can accurately be placed into one of five groups based on the included categorical variables.

The cluster model is considered suitable because it is generalizable. That is, the analysis results can be said to represent all the types of schools in the study. Because the analysis included a census rather than a sample of the population, sampling error in the model is considerably reduced. Generalizability is established by measuring the degrees of freedom of an analysis. The degrees of freedom are a comparison of the number of observations to the number of variables, so the larger the degrees of freedom, the more generalizable the analysis results (Hair et al., 2010, p. 176). The degrees of freedom calculated through EFA are sufficient for establishing good generalizability (see Appendix J).

In addition to the above-mentioned benchmarks of a stable cluster solution, the final clustering of schools also achieves discriminant validity. Discriminant validity is defined as the "…extent to which a construct is truly distinct from other constructs…" (Hair et al., 2010, p. 669). Discriminant validity also verifies that individual factors should represent only one latent construct (Hair et al., 2010, p. 688). Therefore, cross loadings indicate that the model has not achieved discriminant validity. The established seven-factor EFA model shows high factor loadings and no cross loading, and therefore can be considered an appropriate model of the variables in the dataset.

The established five-cluster solution also illustrates construct validity, which is the "…extent to which a set of measured variables actually represents the theoretical latent construct those variables are designed to measure (Hair et al., 2010, p. 669). Construct validity is also established by examining factor loadings, which should be above 0.7 (Hair et al., 2010, p. 673). The percentage of variance extracted is another measure of construct validity. This measure describes what portion of the model's variance is explained variance rather than error variance. Generally, a value of at least 0.5, or 50%, suggests that the factors are adequate for explaining the variables. In this study's model, approximately 78% of the model is explained variance (see Appendix I).

By achieving both discriminant validity and construct validity, the model demonstrates that the seven factors defined through EFA are in fact distinct from one another, as are the five groups segmented out through cluster analysis.

Cluster 1, which includes 15.7% of the schools in the dataset, contains private, liberal arts institutions that are not likely to be focused toward minority groups. These schools are likely to be featured in publications based largely on prestige, such as *U.S. News & World Report*'s college rankings and Global Language Monitor's TrendTopper MediaBuzz list of top schools according to their reputation and brand equity.

Cluster 2 contains 7.7% of the schools and includes large, public universities. The schools in this group are business accredited, have law and medical programs, and present several opportunities for research. They also field NCAA Division I football programs.

Cluster 3 includes the bulk of schools in the dataset (43%) and can be considered the opposite of Cluster 2. The schools in this group have small campuses and are a part of NCAA Division III.

Cluster 4, comprised by 5.5% of the schools, represents institutions with undergraduate programs only. These schools do not offer advanced professional degrees (i.e. law or medical programs), and they are less likely to be included in publications based on measures of prestige, such as *U.S. News*.

Cluster 5, containing 28.1% of the schools, are those schools that are more likely to have some sort of specialty focus (e.g. religious, race, or gender-based). Though they're less often recognized by publications that measure schools based on reputation, the schools in this group offer top Master's programs.

Discussion

The primary objective of this paper was to determine whether schools could reasonably be grouped according to a set of characteristics or attributes. Traditional college rankings are problematic due to the inherit drawbacks associated with an ordinal scale. Grouping schools would resolve these inherit drawbacks. Using data collected on a variety of U.S. colleges and universities, exploratory factor analysis and cluster analysis were conducted to uncover an underlying structure that would describe the differences in the schools. The final model revealed that schools can indeed be segmented into well-defined groups, and can be described based on their characteristics.

The nature of an ordinal scale presents numerous problems for college rankings. Grouping schools according to their specific attributes solves the inherent problems of traditional rankings. The five-cluster model presented in this paper eliminates the one-dimensional perspective from which conventional college rankings are produced. This new model functions so that similar schools will cluster together based on their data, such as enrollment stats, program offerings, type of campus and location, and financial aid information. For this reason, there is no need for publications to dispute which aspects are more or less important in determining the quality of a college or university.

A school would no longer be classified as "better" or "worse" than another; it simply has different attributes. Using this clustering model, schools would be less likely to engage in practices that will alter certain statistics—such as test scores or research fund amounts—in the attempt to influence their position in a ranking. With a clustering approach, competition between schools would remain primarily within clusters, rather than between them. That is, any given school is more likely to compete with another school that is similar to itself, rather than one with an entirely different profile. Finally, perhaps the most obvious application of clustering schools according to specific features is that prospective students will be better equipped with correct information about different schools. This information would lower the likelihood that a prospective student would choose a based solely on prestige or other reputational factors.

Managerial Implications

Decision makers and managers in higher education, such as university administration and admissions representatives, would benefit from a detailed segmentation of colleges and universities. Officials in higher education can use the five generated segments to (a) improve student prospecting and targeting, (b) market specific programs to students who could be a good fit with the institution, and (c) promote their respective school's differing attributes.

Targeting is perhaps the most logical next step in applying the segmented clusters. Schools would find that they could better streamline their marketing efforts if they have a well-defined target student. That is, a college—more specifically, its admissions department—could better identify potential students that would fit the college's profile. Focusing on engaging students that would be a good fit for the school, versus trying to appeal to any and all incoming college students, would mean that any outreach efforts would be better received.

Similarly, specific departments at various colleges could focus more seriously on engaging interested students. If it can be determined that a student is a good fit for *and* is interested in a particular school, communicating frequently with such a student can influence their ultimate decision. For example, a student wanting to obtain an advanced medical degree and have opportunities for research would consider schools in Cluster 2 of the five-cluster model. Responding to this expressed interest, schools in this group should illustrate how their program offerings could best suit this particular student. Knowing that at least the school profile aligns with the potential student means that individual departments (in this example, a department of medicine) can work on building relationships with interested students that are more likely to join the program.

With a clustering model, colleges lose the ability to "prove" their superiority over other schools, as they may have been able to with traditional rankings. This is especially important for schools that typically would not be able to compete with "top" schools according to many rankings. Factors that would not generally be associated with top schools, such as a small campus or an offering of only undergraduate programs, might in fact be a leading consideration for many students.

Implications for Prospective Students & Parents

The clustering model presented in this paper is also useful for prospective college students and their parents. Students can benefit from the model by (a) seeing all available options based on a set of attributes, (b) identifying the group or groups that include school types closet to their preferences, and (c) better understand colleges' missions.

If a prospective student uses only college rankings to inform themselves about different schools, they're likely to overlook crucial information regarding types of schools that are excluded from such lists. For example, schools that do not have much funding available for research (a measure indicative of prestige) habitually rank very low or are excluded altogether from rankings. In reality a student might be unfazed by the fact that a school offers few research opportunities. If this is the case, this statistic makes no difference to the student and they should be aware that they have other options.

The qualities used to describe the different segments are areas students tend to examine when considering colleges. Therefore, each student can easily determine a list of comparable schools based on their individual interests. For students, narrowing all possible college choices down to a manageable list can be a daunting task that would be enormously simplified by a cluster model. For parents of prospective students, one top priority is commonly evaluating which schools are most realistic for their student in terms of finances and other logistical respects.

Being able to see the specific ways in which colleges differ from one another gives insight into their overall objectives and top priorities. A school's "personality"—the environment, culture, and overall feel—is often just as important to a student as other qualities like major offerings or class sizes. Having an understanding of a college's values and goals further enables students to determine their true best fit.

Future Research

This paper has presented the possibility of organizing colleges into distinct groups based on their specific characteristics. Further research of this topic could address whether cluster membership can be predicted for any given school. A form of logit regression would further support an argument for grouping universities as a preferred approach compared to rankings. Continued analysis could also explore whether the original five-cluster model would be applicable to institution types that were not included in this study. For example, two-year and for-profit institutions were not included. Additional research could explore a clustering solution for those institutions. Also, public and private four-year universities could be separated with cluster analysis.

References

Altbach, P. G. (2012). The globalization of college and university rankings. *Change: The Magazine of Higher Learning, 44:1*, 26-31, DOI: 10.1080/00091383.2012.636001

Anderson, N. (2013). Five colleges misreported data to U.S. News, raising concerns about rankings, reputation. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/education/five-colleges-misreported-data-to-us-news-raising-concerns-about-rankings-reputation/2013/02/06/cb437876-6b17-11e2-af53-7b2b2a7510a8_story.html

Anderson, N. (2015). The US News and World Report college rankings; A public vs. private dilemma. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/grade-point/wp/2015/09/09/the-u-s-news-and-world-report-college-rankings-a-public-vs-private-dilemma/

Asymmetric Information. (2016). In Investopedia. Retrieved from http://www.investopedia.com/terms/a/asymmetricinformation.asp

Chamberlin, E. H. (1933). The theory of monopolistic competition: A re-orientation of the theory of value. Cambridge, MA: Harvard University Press.

Chamberlin's monopolistic competition. (2012). In Policonomics. Retrieved from http://www.policonomics.com/chamberlins-monopolistic-competition/

Cochi Ficano, C. K., & Cunningham, B. M. (2002). The determinants of donative revenue flows from alumni of higher education: An empirical inquiry. Retrieved from https://www.ilr.cornell.edu/sites/ilr.cornell.edu/files/The%20Determinants%20of%2

0Donative%20Revenue%20Flows%20from%20Alumni%20of%20Higher%20Education_0.pdf

Edward Chamberlin. (2013). In *New World Encyclopedia online*. Retrieved from http://www.newworldencyclopedia.org/entry/Edward_Chamberlin

Hair, J. F. Jr., Black, W., Babin, B. J., Anderson, R. E. (2010). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice Hall.

Hejwosz, D. (2010). Students as consumers: The commercialization of higher education in the United States of America. *Liberté! World Edition.* http://liberteworld.com/2010/05/31/students-as-consumers-the-commercialisation-of-higher-education-in-the-united-states-of-america/

Hunt, S. D. (2000). A general theory of competition. Thousand Oaks, CA: Sage Publications, Inc.

Lawson, R. A., & Zerkle, A. (2006). Price discrimination in college tuition: An empirical case study. *Journal of Economics and Finance Education, 5*. Retrieved from http://www.economics-finance.org/jefe/econ/Lawsonpaper.pdf

Lehmann, D. R., Gupta, S., Steckel, J. H. (1998). *Marketing research*. Addison-Wesley Educational Publishers Inc.

Lilien, G. L., Rangaswamy, A., De Bruyn, A. (2013). *Principles of marketing engineering* (2nd ed.) State College, PA: DecisionPro, Inc.

Peterson, R. A. (2000). A meta-analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters, 11*, pp. 261-275. DOI: 10.1023/A:1008191211004

"Ranking affects the financial resources of public colleges." (2016). *The National Bureau of Economic Research*. Retrieved from http://www.nber.org/digest/septoct07/w12941.html

Tierney, J. (2013). "Your annual reminder to ignore the U.S. News & World Report college rankings." The Atlantic. Retrieved from http://www.theatlantic.com/education/archive/2013/09/your-annual-reminder-to-ignore-the-em-us-news-world-report-em-college-rankings/279103/

"U.S. News breaks online traffic record." (2013). *U.S. News & World Report*. Retrieved from http://www.usnews.com/info/blogs/press-room/2013/09/12/us-news-breaks-online-traffic-record

"What's wrong with college rankings?" (2015). *College Confidential*. Retrieved from http://www.collegeconfidential.com/bad-rankings/

Appendix A

Initial Sources & Attributes Measured in Each

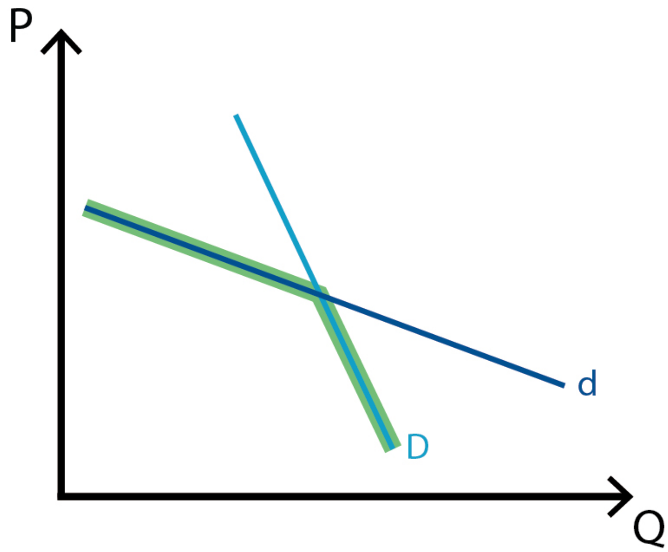
Source	**Year**	**Attributes Included**
Bloomberg Businessweek	2014	Ranks compiled according to: employer, alumni, student surveys; job placement rates; starting salaries
Brookings Institution	2015	Value-added score computed based on curriculum value, alumni skills, STEM orientation, completion rates, student aid
Business Insider	2014	Ranks based on evaluations of alumni potential, as reported by hiring managers
Center for Measuring University Performance (CMUP)	2013	Total research, faculty awards, National Academy members, doctorates awarded, endowments, annual giving
College Factual	2013	Statistics on test scores, educational resources, degree completion, post-grad earnings
CollegeNET/Payscale SMI	2014	Tuition, economic background, graduation rate, early career salary, endowment
Daily Beast	2014	Future earnings, quality of education, affordability, on-time graduation, campus quality, activities & clubs, nightlife, diversity, sports
Fiske	2016	Ratings calculated using measures of: academics, campus setting, student body, financial aid, housing, food, social life, extracurricular activities
Forbes	2015	Cost, enrollment
Global Language Monitor (TrendTopper MediaBuzz)	2016	Ranks determined according to strength of brand equity (mentions in social media, blogs, web pages, global print, other electronic media)
Kiplinger	2014	Measures of competitiveness, graduation rates, academic support, cost & financial aid, student indebtedness
LinkedIn	2015	Rankings calculated by job category based on graduates in relevant professions, desirable jobs for each profession
Money	2015	Equal measures of quality of education, affordability, outcomes
New York Times	2014	College Access Index calculated by Pell grants awarded, net price for low- and middle-income families, endowment
Niche	2016	Grades (A-D-) assigned for: academics, value, professors, student surveys, diversity, student life, athletics, campus quality, local area, safety
Parchment-MyChances.net	2016	Rankings calculated considering college matchups based

Source	Year	Attributes Included
		on expectations
Payscale-College ROI	2015	20-year net ROI, total 4-year cost, graduation rate, typical years to graduate, average loan amount
Princeton Review	2016	Ratings, based on quality of life, fire safety, green rating, academic, professors interesting, professors accessible, admissions selectivity, financial aid
Reuters	2015	Rankings calculated based on measures of innovation, including academic papers, research, patent filings
StartClass	2015	Acceptance rate, students offered admission, number of applications
Times Higher Ed	2015	Measures based on: teaching, international outlook, research, citations, industry income
U.S. News & World Report	2016	ACT/SAT scores, graduation rates, major offerings, location, other measures of academic excellence
Washington Monthly	2015	Social Mobility (recruiting and graduating low-income students), Research (producing cutting-edge scholarship and PhDs), and Service (encouraging students to give something back to their country)

Source: Internal documents

Appendix B

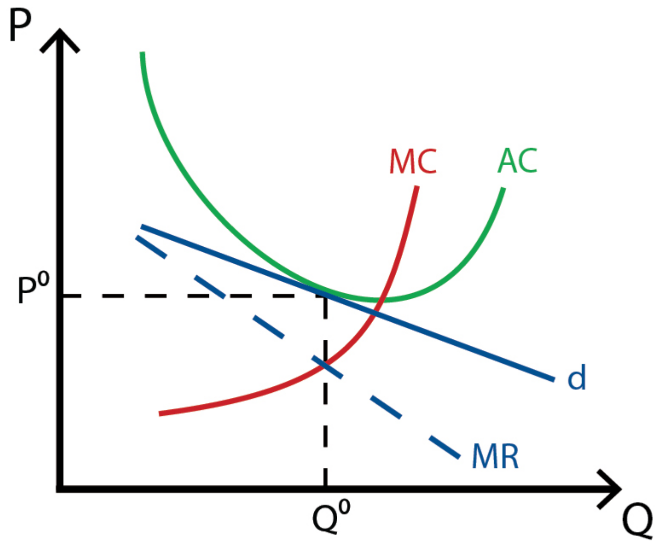
Demand for Products in Monopolistic Competition



Source: Perceived vs. Actual Demand. (2012). [Graph illustration Chamberlin's monopolistic competition]. *Policonomics*. Retrieved from http://www.policonomics.com/chamberlins-monopolistic-competition/

Appendix C

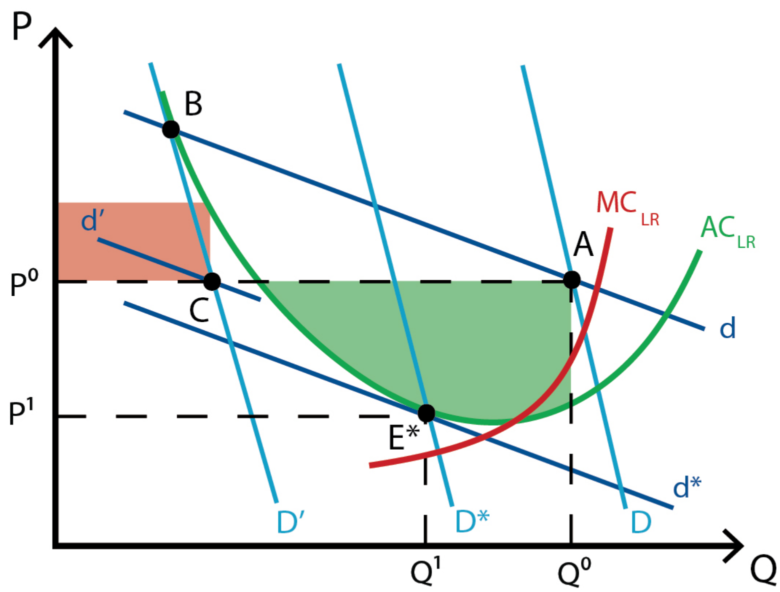
Short-run Equilibrium in Monopolistic Competition



Source: Equilibrium in the short-run. (2012). [Graph illustration Chamberlin's monopolistic competition]. *Policonomics*. Retrieved from http://www.policonomics.com/chamberlins-monopolistic-competition/

Appendix D

Long-run Equilibrium in Monopolistic Competition



Source: Equilibrium in the long-run. (2012). [Graph illustration Chamberlin’s monopolistic competition]. *Policonomics*. Retrieved from http://www.policonomics.com/chamberlins-monopolistic-competition/

Appendix E

Sources & Variables Removed Due to Unsubstantial Information

Source	Variable(s) Removed
Fiske	All variables removed
Niche	All variables removed
Forbes	All variables removed
Business Insider	All variables removed
Wall Street Journal	All variables removed
New York Times	All variables removed
Gallup Poll	All variables removed
LinkedIn	All variables removed
StartClass	All variables removed
Psychology Today	All variables removed
Bloomberg Businessweek	All variables removed
Daily Beast	All variables removed
Kiplinger	All variables removed
PayScale-College ROI	All variables removed
Center for Measuring University Performance (CMUP)	Institution type; Measures in Top 25 Control; Measures in Top 26-50 Control; 2011 Federal Research; 2012 Endowment Assests; 2012 Faculty Awards; 2012 Doctorates Awarded; 2011 Postdocs; 2011 Median SAT
Times Higher Ed	All variables removed

Source: Internal documents

Appendix F

Final Dataset Including Dummy & Continuous Variables

Name	Description	Ranking Source	Original Source	Variable Type
Federal Pell grant aid per student	average amount of Pell grant (needs-based) aid awarded to each student	Brookings Institution	IPEDS	Continuous
Mean amount of student financial aid from institution	financial aid funded by the college itself, rather than federal or other sources	Brookings Institution	IPEDS	Continuous
Curriculum value	the labor market value of the college's mix of majors; calculated by determining the national median earnings for all bachelor's degree holders in the labor force by major, using the Census Bureau's 2013 American Community Survey (ACS), made available by the Integrated Public Use Microdata Series (IPUMS); A weighted average for each school is then calculated using the actual number of graduates in each major, with data from IPEDS	Brookings Institution	Brookings Institution	Continuous
Graduation rate, twice normal time	the percentage of enrolled students who graduate from the college in eight years for four-year programs and four years for two-year programs	Brookings Institution	Brookings Institution	Continuous

Name	**Description**	**Ranking Source**	**Original Source**	**Variable Type**
Retention rate	the share of students from the full-time and part-time adjusted fall 2012 cohorts still enrolled in fall 2013	Brookings Institution	Brookings Institution	Continuous
Percent of awards at bachelor's or higher level	share of degrees awarded at at least Bachelor's level	Brookings Institution	Brookings Institution	Continuous
Percent of gradautes in STEM fields	the percentage of graduates who complete a degree in a field of study that prepares them for an occupation demanding high levels of science, technology, engineering, or math knowledge	Brookings Institution	IPEDS	Continuous
Mid-career salary, median (all alumni)	for schools not reported, input the average [$77,685] of the values from schools that did report a mid-career salary	Brookings Institution	PayScale	Continuous
CMUP 2012 Annual Giving per capita*1000	Annual giving amount per 1000 students; 0=not reported	CMUP	CMUP	Continuous
CMUP 2011 Total Research per capita*1000	Research amount per 1000 students; 0=not reported	CMUP	CMUP	Continuous
CMUP 2012 National Academy Members per capita*1000	Number of National Academy members per 1000 students; 0=not reported	CMUP	CMUP	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
CMUP 2012 Doctorates Awarded per capita*1000	Number of doctorates awarded per 1000 students; 0=not reported	CMUP	CMUP	Continuous
CMUP 2012 Natl Merit Scholars per capita*1000	Number of National Merit Scholars per 1000 students; 0=not reported	CMUP	CMUP	Continuous
Faculty receiving significant awards	number of faculty reciving prestigious awards, relative to the number of full-time faculty; 0 for missing values	CMUP	CMUP	Continuous
Faculty in national academies	number of factulty in the National Academies, relative the number of full-time faculty; 0 for missing values	CMUP	CMUP	Continuous
Three-year cohort default rate	Three-year cohort default rate	College Scorecard Government Data	FSA	Continuous
1-New England region	(CT, ME, MA, NH, RI, VT) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
2-Mid East region	(DE, DC, MD, NJ, NY, PA) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
3-Great Lakes region	(IL, IN, MI, OH, WI) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
4-Plains region	(IA, KS, MN, MO, NE, ND, SD) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
5-Southeast region	(AL, AR, FL, GA, KY, LA, MS, NC, SC, TN, VA, WV) recoded into	College Scorecard Government	IPEDS	Categorical-Dummy variable

Name	**Description**	**Ranking Source**	**Original Source**	**Variable Type**
	dummy variables	Data		
6-Southwest region	(AZ, NM, OK, TX) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
7-Rocky Mountains region	(CO, ID, MT, UT, WY) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
8-Far West region	(AK, CA, HI, NV, OR, WA) recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
1-Large city/suburb	combination of Locale categories: City: Large (population of 250,000 or more) and Suburb: Large (outside principal city, in urbanized area with population of 250,000 or more); recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
2-Midsize city/suburb	combination of Locale categories: City: Midsize (population of at least 100,000 but less than 250,000) and Suburb: Midsize (outside principal city, in urbanized area with population of at least 100,000 but less than 250,000); recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable

Name	Description	Ranking Source	Original Source	Variable Type
3-Small city/suburb	combination of Locale categories: City: Small (population less than 100,000) and Suburb: Small (outside principal city, in urbanized area with population less than 100,000); recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
4-Town	combination of Locale categories: Town: Fringe (in urban cluster up to 10 miles from an urbanized area), Town: Distant (in urban cluster more than 10 miles and up to 35 miles from an urbanized area), Town: Remote (in urban cluster more than 35 miles from an urbanized area); recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
5-Rural	combination of Locale categories: Rural: Fringe (rural territory up to 5 miles from an urbanized area or up to 2.5 miles from an urban cluster), Rural: Distant (rural territory more than 5 miles but up to 25 miles from an urbanized area or more than 2.5 and up to 10 miles from an urban cluster), Rural: Remote (rural territory more than 25 miles from an urbanized area and more than 10 miles from an urban cluster);	College Scorecard Government Data	IPEDS	Categorical-Dummy variable

Name	Description	Ranking Source	Original Source	Variable Type
	recoded into dummy variables			
1-Small/very small, primarily nonresidential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
2-Small/very small, primarily residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
3-Small/very small, highly residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
4-Medium size, primarily nonresidential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
5-Medium size, primarily residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
6-Medium size, highly residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government	IPEDS	Categorical-Dummy variable

Name	Description	Ranking Source	Original Source	Variable Type
		Data		
7-Large, primarily nonresidential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
8-Large, primarily residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
9-Large, highly residential	Carnegie Classification-Size & Setting; recoded into dummy variables	College Scorecard Government Data	IPEDS	Categorical-Dummy variable
No. of Title IV low income students (public & private)	sum of number of Title IV students, $0-$30,000 family income from both public and private institutions; 0 for schools not reported	College Scorecard Government Data	IPEDS	Continuous
Percentage of undergraduates who receive a Pell Grant	Percentage of undergraduates who receive a Pell Grant	College Scorecard Government Data	IPEDS	Continuous
Percent of all federal undergraduate students receiving a federal student loan	Percent of all federal undergraduate students receiving a federal student loan	College Scorecard Government Data	IPEDS	Continuous
Avg. Net Price for Title IV institutions	Average net price for Title IV institutions (public and private institutions)	College Scorecard Government Data	IPEDS	Continuous

Name	**Description**	**Ranking Source**	**Original Source**	**Variable Type**
Net tuition revenue per full-time equivalent student	Out-of-state tuition and fees	College Scorecard Government Data	IPEDS	Continuous
Instructional expenditures per full-time equivalent student	Instructional expenditures per full-time equivalent student	College Scorecard Government Data	IPEDS	Continuous
Average faculty salary	Average faculty salary	College Scorecard Government Data	IPEDS	Continuous
Enrollment of undergraduate degree-seeking students	Enrollment of undergraduate degree-seeking students	College Scorecard Government Data	IPEDS	Continuous
Admission rate	Admission rate; not reported=999	College Scorecard Government Data	IPEDS	Continuous
Completion rate for first-time, full-time students at four-year institutions	Completion rate for first-time, full-time students at four-year institutions (150% of expected time to completion/6 years)	College Scorecard Government Data	IPEDS	Continuous
Share of undergraduate students who are first-time, full-time degree-/certificate-seeking undergraduate students	Share of undergraduate students who are first-time, full-time degree-/certificate-seeking undergraduate students	College Scorecard Government Data	IPEDS	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
Adjusted cohort count for completion rate at four-year institutions	Adjusted cohort count for completion rate at four-year institutions (denominator of completion rate)	College Scorecard Government Data	IPEDS	Continuous
Adjusted cohort count for completion rate at four-year institutions	Adjusted cohort count for completion rate at four-year institutions (denominator of completion rate), pooled for two-year rolling averages	College Scorecard Government Data	IPEDS	Continuous
First-time, full-time student retention rate at four-year institutions	First-time, full-time student retention rate at four-year institutions	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-white	Total share of enrollment of undergraduate degree-seeking students who are white	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-black	Total share of enrollment of undergraduate degree-seeking students who are black	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-Hispanic	Total share of enrollment of undergraduate degree-seeking students who are Hispanic	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-Asian	Total share of enrollment of undergraduate degree-seeking students who are Asian	College Scorecard Government Data	IPEDS	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
Undergrad enrollment-American Indian/Alaska Native	Total share of enrollment of undergraduate degree-seeking students who are American Indian/Alaska Native	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-Native Hawaiin/Pacific Islander	Total share of enrollment of undergraduate degree-seeking students who are Native Hawaiian/Pacific Islander	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-2/more races	Total share of enrollment of undergraduate degree-seeking students who are two or more races	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-non-resident aliens	Total share of enrollment of undergraduate degree-seeking students who are non-resident aliens	College Scorecard Government Data	IPEDS	Continuous
Undergrad enrollment-unknown race	Total share of enrollment of undergraduate degree-seeking students whose race is unknown	College Scorecard Government Data	IPEDS	Continuous
Part-time undergrad, degree-/certificate-seeking students	Share of undergraduate, degree-/certificate-seeking students who are part-time	College Scorecard Government Data	IPEDS	Continuous
Percentage of undergraduates aged 25 and above	Percentage of undergraduates aged 25 and above	College Scorecard Government Data	IPEDS	Continuous
Proportion of faculty that is full-time	Proportion of faculty that is full-time	College Scorecard Government Data	IPEDS	Continuous

Name	**Description**	**Ranking Source**	**Original Source**	**Variable Type**
No. of branch campuses	Number of branch campuses	College Scorecard Government Data	IPEDS	Continuous
Midpoint of the ACT cumulative score	Midpoint of the ACT cumulative score; 999=not reported	College Scorecard Government Data	IPEDS	Continuous
Average SAT equivalent score of students admitted	Average SAT equivalent score of students admitted; 000=not reported	College Scorecard Government Data	IPEDS	Continuous
Original amount of the loan principal upon entering repayment	The original amount of the loan principal upon entering repayment	College Scorecard Government Data	NSLDS	Continuous
Median debt for students who have completed	The median debt for students who have completed	College Scorecard Government Data	NSLDS	Continuous
Median debt for students who have not completed	The median debt for students who have not completed	College Scorecard Government Data	NSLDS	Continuous
Endowment per capita*1000	endowment amount per 1000 students; caluclated from total endowment reported by NACUBO and undergraduate enrollment reported by govt; 0=not reported	NACUBO	NACUBO	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
Parchment Score (Elo points)	starting amount of points (1500), minus points for being the not-chosen school and points added for being the chosen school; input base score of 1500 for missing values	Parchment	Parchment	Continuous
Princeton Review Total Points	sum of 8 factors rated by Princeton Review (Quality of Life, Fire Safety, Green, Academic, Professors Interesting, Professors Accessible, Admissions Selectivity, Financial Aid)	Princeton Review	Princeton Review	Continuous
TTMB1-private colleges	dummy variable indicating whether a school is present on TTMB's Private Colleges list; 1=included on list, 0=not included on list	TrendTopper MediaBuzz	Global Language Monitor (GLM)	Categorical-Dummy variable
TTMB2-public universities	dummy variable indicating whether a school is present on TTMB's Public Universities list; 1=included on list, 0=not included on list	TrendTopper MediaBuzz	Global Language Monitor (GLM)	Categorical-Dummy variable
TTMB3-private universities	dummy variable indicating whether a school is present on TTMB's Private Universities list; 1=included on list, 0=not included on list	TrendTopper MediaBuzz	Global Language Monitor (GLM)	Categorical-Dummy variable
TTMB4-not mentioned by TTMB	dummy variable indicating which schools are not included in any of the 4 TTMB lists; 1=is not included on any of 4 lists 0=is	TrendTopper MediaBuzz	Global Language Monitor (GLM)	Categorical-Dummy variable

Name	Description	Ranking Source	Original Source	Variable Type
	included on a TTMB list			
TTMB5-public colleges	dummy variable indicating whether a school is present on TTMB's Public Colleges list; 1=included on list, 0=not included on list	TrendTopper MediaBuzz	Global Language Monitor (GLM)	Categorical-Dummy variable
US News rank 1-36	indicates whether a school is ranked between 1-36 according to US News, regardless of which list the school appears on; recoded into dummy variables	U.S. News & World Report	U.S. News & World Report	Categorical-Dummy variable
US News rank 37-80	indicates whether a school is ranked between 37-80 according to US News, regardless of which list the school appears on; recoded into dummy variables	U.S. News & World Report	U.S. News & World Report	Categorical-Dummy variable
US News rank 81-199	indicates whether a school is ranked between 81-199 according to US News, regardless of which list the school appears on; recoded into dummy variables	U.S. News & World Report	U.S. News & World Report	Categorical-Dummy variable
Missing from US News	indicates whether a school was not mentioned on any US News ranking list; 1=not mentioned, 0=mentioned; recoded into dummy variables	U.S. News & World Report	U.S. News & World Report	Categorical-Dummy variable

Name	Description	Ranking Source	Original Source	Variable Type
Classes with fewer than 20 students (%)	share of undergraduate classes with 20 or fewer students; a value of 0 was assigned for those schools that did not report this value	U.S. News & World Report	U.S. News & World Report	Continuous
1-WA Monthly National Universities	indicates whether a school is included on Washington's Monthly's National Universities ranking list; recoded into dummy variables	Washington Monthly	Washington Monthly	Categorical-Dummy variable
2-WA Monthly Liberal Arts Colleges	indicates whether a school is included on Washington's Monthly's Liberal Arts Colleges ranking list; recoded into dummy variables	Washington Monthly	Washington Monthly	Categorical-Dummy variable
3-WA Monthly Master's Universities	indicates whether a school is included on Washington's Monthly's Master's Universities ranking list; recoded into dummy variables	Washington Monthly	Washington Monthly	Categorical-Dummy variable
4-WA Monthly Baccalaureate colleges	indicates whether a school is included on Washington's Monthly's Baccalaureate Colleges ranking list; recoded into dummy variables	Washington Monthly	Washington Monthly	Categorical-Dummy variable
Federal work-study funds spent on service	percentage of federal work-study grant money spent on community service projects	Washington Monthly	Washington Monthly	Continuous
Science & engineering PhDs awarded	number of science and engineering PhDs awarded by the university; 0 for missing values	Washington Monthly	Washington Monthly	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
Predicted grad rate	The graduation rate prediction formula uses data from 2010 to 2012 and includes the percentage of Pell Grant recipients and students receiving student loans, the average ACT/SAT score, the admit rate, the racial/ethnic and gender makeup of the student body, the number of students (overall and full-time), and institutional characteristics such as whether a college is primarily residential.	Washington Monthly	Washington Monthly	Continuous
Research expenditures	total amount of an institution's research spending	Washington Monthly	Washington Monthly	Continuous
Professions & Applied Sciences count	number of majors available that are considered Professions & Applied Sciences by Wikipedia's Outline of Academic Disciplines	Wikipedia	Wikipedia	Continuous
Natural Sciences count	number of majors available that are considered Natural Sciences by Wikipedia's Outline of Academic Disciplines	Wikipedia	Wikipedia	Continuous
Social Sciences count	number of majors available that are considered Social Sciences by Wikipedia's Outline of Academic Disciplines	Wikipedia	Wikipedia	Continuous
Formal Sciences count	number of majors available that are considered Formal Sciences by	Wikipedia	Wikipedia	Continuous

Name	Description	Ranking Source	Original Source	Variable Type
	Wikipedia's Outline of Academic Disciplines			
Humanities count	number of majors available that are considered Humanities by Wikipedia's Outline of Academic Disciplines	Wikipedia	Wikipedia	Continuous
Bus Acc Y/N	Denotes whether or not an institution has a business program accreditation, regardless of which accreditation; 1=has business accreditation, 0=does not have business accreditation	N/A	Wikipedia	Categorical
Football Y/N	Denotes whether or not an institution has a football team, regardless of Division; 1=has football, 0=does not have football	N/A	Wikipedia	Categorical
ROI	mid-career median salary (Brookings)/4 year cost (calculated from Govt Data)	N/A	Brookings Institution/PayScale; CollegeScorecard Data	Continuous

Source: Internal documents

Appendix G

Factor Loadings, Final EFA Model

	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Factor 7**
Federal Pell grant aid per student	**0.877**	-0.201	-0.091	0.250	-0.097	0.019	0.164
Percentage of undergraduates who receive a Pell Grant	**0.853**	-0.262	-0.142	0.261	-0.105	0.057	0.169
Three-year cohort default rate	**0.741**	-0.261	-0.143	0.275	0.002	0.209	-0.097
Mid-career salary, median (all alumni)	-0.181	**0.793**	-0.048	-0.102	0.055	-0.003	-0.058
ROI	0.152	0.025	-0.082	**0.837**	0.333	0.130	-0.006
Avg. Net Price for Title IV institutions	-0.268	0.147	-0.032	**-0.866**	-0.120	-0.019	-0.032
Net tuition revenue per full-time equivalent student	-0.239	0.439	0.035	**-0.762**	-0.151	0.006	0.013
Instructional expenditures per full-time equivalent student	-0.123	**0.788**	0.131	-0.201	-0.028	-0.010	-0.128
Average faculty salary	-0.251	**0.762**	0.185	-0.188	0.329	-0.017	0.050
Enrollment of undergraduate degree-seeking students	-0.080	0.200	0.116	0.242	**0.876**	-0.039	0.098
Admission rate	0.155	-0.046	-0.071	0.143	-0.017	**0.847**	-0.056
Adjusted cohort count for completion rate at four-year institutions	-0.092	0.254	0.131	0.137	**0.881**	-0.090	-0.009
Professions & Applied Sciences count	0.000	-0.105	0.288	0.137	**0.775**	0.031	-0.009
Natural Sciences count	-0.034	0.066	**0.835**	0.052	0.087	-0.108	-0.127
Social Sciences count	-0.143	0.196	**0.683**	-0.093	0.276	-0.022	0.116
Formal Sciences count	0.019	0.036	**0.839**	0.060	0.046	-0.075	-0.129
Humanities count	-0.241	-0.014	**0.768**	-0.142	0.179	-0.088	0.137
Midpoint of the ACT cumulative score	0.067	0.000	-0.155	-0.038	-0.049	**0.847**	0.104
Proportion of faculty that is full-time	-0.029	0.164	0.153	0.199	-0.078	-0.041	**-0.725**
Undergrad enrollment-black	**0.893**	-0.037	-0.043	-0.018	-0.016	0.087	-0.182
Undergrad enrollment-Hispanic	-0.021	0.277	0.102	0.257	-0.027	0.016	**0.736**
Undergrad enrollment-Asian	-0.093	**0.751**	0.060	0.095	0.151	-0.056	0.335

Source: Internal documents

Appendix H

Variables Used for Segmentation & Discrimination in Cluster Analysis

Segmentation Variables	**Discrimination Variables**
Federal Pell grant aid per student	East region
Three-year cohort default rate	North region
Mid-career salary, median (all alumni)	South region
ROI	West region
Avg. Net Price for Title IV institutions	Small campus
Net tuition revenue per full-time equivalent student	Medium size campus
Instructional expenditures per full-time equivalent student	Large campus
Average faculty salary	Bus Acc Y/N
Enrollment of undergraduate degree-seeking students	Div 1
Admission rate	Div II
Professions & Applied Sciences count	Div III
Humanities count	Specialty-focused institution
Midpoint of the ACT cumulative score	Advanced Professional School
Proportion of faculty that is full-time	1-WA Monthly National Universities
Undergrad enrollment-black	2-WA Monthly Liberal Arts Colleges
Undergrad enrollment-Hispanic	3-WA Monthly Master's Universities
Undergrad enrollment-Asian	4-WA Monthly Baccalaureate colleges
	Missing from US News
	TTMB1-private colleges
	TTMB2-public universities
	TTMB3-private universities
	TTMB4-not mentioned by TTMB
	TTMB5-public colleges
	CMUP Top Research Universities
	Football Y/N

Source: Internal documents

Appendix I

Eigenvalues & Variance Explained by EFA Model

	Initial Eigenvalues			Rotation Sums of Squared Loadings		
Factor:	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	5.911	26.869	26.869	5.911	14.727	14.727
2	3.758	17.081	43.95	3.758	13.88	28.607
3	2.257	10.26	54.21	2.257	12.505	41.112
4	1.629	7.404	61.613	1.629	11.777	52.889
5	1.321	6.005	67.619	1.321	11.71	64.599
6	1.27	5.775	73.393	1.27	7.043	71.642
7	1.006	4.572	77.965	1.006	6.323	77.965

Source: Internal documents

Appendix J

KMO & Bartlett's Test Values

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.811
Bartlett's Test of Sphericity	
Approx. Chi-Square	18926.362
df	231
Sig.	0.00

Source: Internal documents

Appendix K

Sizes of Each Cluster in Final Cluster Model

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Number of observations	185	90	505	65	330
Proportion	15.7%	7.7%	43.0%	5.5%	28.1%

Source: Internal documents

Appendix L

Confusion Matrix from Discriminant Analysis

Actual / Predicted cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Cluster 1	**79.50%**	02.20%	09.20%	01.10%	08.10%
Cluster 2	07.80%	**72.20%**	00.00%	01.10%	18.90%
Cluster 3	17.00%	00.60%	**58.40%**	11.10%	12.90%
Cluster 4	03.10%	01.50%	09.20%	**72.30%**	13.80%
Cluster 5	02.10%	05.50%	11.50%	14.50%	**66.40%**
Hit Rate (percent of total cases correctly classified):					**65.79%**

Source: Internal documents