How different ranking factors influence an academic institution's overall score

For many young adults there comes a time in life when the opportunity and decision to pursue a higher education becomes a reality. Students who chose to attend a school at the higher academic level know it is important to find the appropriate university that best suits their educational interests and personal needs. This milestone in each of these individual's lives, leads them to assess which college would provide them the best overall experience and quality of education in this important chapter of their lives. The process in finding a college usually takes place years before attending a selected school, and it is up to them to research personally appealing institutions. A question that arises during this period of exploration asks; what type of university information is being presented to these potential students, and how is this presented information of institutional ranks calculated? This leads to the topic of discussion involving college rankings, and the factors that play into deciding the order of these overall standings. In the past, university rankings have grown to become a critically influential resource for young students trying to decide where they should pursue a higher education. While reading an article from the *Economics of Education Review* examining university rankings, author *Julia Horstschraer* states " *From an economist's point of view, it is very important that prospective students apply for the universities that fit them best in order to maximize human capital production and to minimize drop-outs. Therefore, university rankings and indicators of excellence may provide valuable information for the decision of prospective students.* ". Although these standings are easily accessible if one were to search for a list of top tier universities, one would be overwhelmed with the number of accredited organizations claiming to have the best overall ranking over in their respective areas. The rankings representing the overall score of

a university have become the source of much controversy because of their tendency to involve politics, excessive self citation, and additional outstanding factors that can cause further inaccuracy amongst institutional ranks. The system of assessing universities and ranking them alongside their competitors is having an increasing effect on where students apply, and is helping to shape their perception of these institutions. This increased form of reference for deciding students shows the importance of having a ranking system that excludes vulnerable scoring strategies while limiting the amount of institutional bias present in all tested areas. Obtaining a high rank as an academic institution provides all schools great incentive to put forth effort to achieve such a position, but it also leaves opportunity for many to manipulate ranking evaluation methods for their benefit. The significant position changes in specific university ranks led many to become skeptical of ranking methods. The authors of an article regarding this issue, *Maxime Mussard* and *Alex Pappachen James,* state that " *The unreal growth of some universities in the Times Higher Education World University Rankings is questioned. We show that the universities and individuals can exploit a weakness of ranking tables by engaging in excessive self-citations.".* This discovery of unethical actions by certain colleges attempting to rank higher shows that error is present in rank assessment methods, but accounting enough for these errors allows testing to become more accurate when evaluating significance of the actual variables used in determining the overall score of an educational institution.

In this study, the properties that make up the ranking system and cause the distinct differences between schools will be analyzed to see just how each

ranking category impacts the final total perceived quality of the university in the year 2015, measured from zero to one hundred. This final institutional score will be described from now forth as our models sole dependent variable. Our data was obtained through the World Data Bank, an industry leader in providing reputable raw data from a large variety of topics and industries. The original data shows assorted rankings for exactly one-thousand different universities found in the United States and throughout the world. This data has been independently measured in a total of nine categories based on specifically established findings. These ranking categories include national rank, quality of education, alumni employment, quality of faculty, publications, influence, citations, and patents, as well as a description of the university's country of origin. These datasets are what we compiled to use in our model that will directly impact the overall score of the university, all else held constant, and will be henceforth considered our independent variables used in the overall model.

The variable labeled as national rank is used to measure the status of a university in comparison to other academic institutions located within its domestic market and the cumulative market alone. The list for the variable labeled quality education, represents a university's status as it pertains to its standard level of quality upheld amongst all of the institutions varying educational departments. Following this, the ranking for alumni employment shows the comparison to other universities for the rates at which their graduating seniors found full time employment opportunities within a year of their graduation. The next independent variable is that of the quality of faculty. This list is compiled by looking at an institution's faculty based on

student and peer reviews, as well as the amount of experience that the professors have throughout their educational careers. Additionally, there is a category that establishes order based on the accumulation of publications through creativity and research that has originated from either a faculty or an alumni of the university. Next, there is a list based on the grade of influence. This variable is established based off of the analysis of how a university is regarded and its level of impact throughout the community that it is rooted in. The following variables are categorized as citations. Similar to publications, this arranges a list based on the number of research papers or articles produced that are either funded by the university, produced by a faculty member, or contains university generated information that has been cited in a piece of literature. Our final independent variable is categorized as patents. This has been assorted based on an institutions number of patents which it has recently and historically filed for. This information tends to show the strength of a research university and therefore the lack of strength among universities that consider themselves to be based in a liberal arts foundation. These independent variables come from accredited results, but still have the chance to be rooted in bias simply based off of the system used to attain them. As previously stated, this topic of research can easily fall into a bias trap since individuals are highly prone to show preference for institutions that they share relations with, or value preconceived notions based off of historically incorrect rankings. Along with this, our articles state the conflicting incentive for institutions to produce data that is partial to them in hopes to promote their status' quality in the market. This could be particularly true in the data collected for the independent variables; quality of education, quality of faculty, and the institutions influence on its

community. These measurement errors should be greatly minimized due to our large set of data and our independent variables large basis on black and white statistics.

Now that our variables are defined and listed, a regression model can be designed to be further tested and analyzed in order to better understand the relationship between the variables. In creating this model we will implement a logarithm in order to see the marginal effect of a selected independent variable on the dependant variable. When using this function we can easily identify and relate the ranking variables in terms of percentage change while discussing their relationship to the dependent variable. We perform this function since we expect changes in our X value data to cause a constant percentage change in Y. This means that we can see the percentage change to our Y variable when our independent variables ranking either moves up by one ranking position or falls by one ranking position. Secondly, we will add a dummy variable “ dumUSA” from our country of origin variable data in order to see the effects of an institution on score if it is located within the United States. By using a dummy variable here, we are able to take categorical data and transform it into a dichotomous variable. This allows us the opportunity to incorporate this information into our model for testing. With the implementation of this dummy variable “ dumUSA” we are able to predict its coefficient to be positive inferring a positive increase on the dependant variable score. When creating this model, we predict that the variable coefficients will all be negative except for our dummy variable. We estimate this to hold true, since an increase in our independent variables, outside of dumUSA, means that they are falling in their respective rankings. This fall in

ranking would indicate a negative effect on the overall score of the institution, which is our dependent variable. Therefore our null hypothesis that will we will test will attempt to prove that our first eight independent variables will all cause our Y value to decrease as they increase. Meaning that our first eight variables will have a negative effect on our the academic institutions' score. Our initial regression model is as follows:

Log(score)= B0+B1natrank + B2qualeduc + B3alumempl + B4qualfac + B5publl+ B6influ

+ B7citation + B8patents + B9dumUSA + u

After initially testing our models regression we discovered that the coefficients of the independent variables turned out to hold negative values, as we expected. This provides an understanding of the effects these independent variables have on the dependent variable " score" which results in negative changes on the Y-axis as X increases positively, with the exception of " dumUSA" and its positive coefficient. This initial test proves to be a good indicator that our null hypothesis holds true. Next, we will run tests to determine if this model shows any signs of multicollinearity. If the model in fact does contain multicollinearity, this will produce results that indicate that two or more variables in our regression are highly linearly related. The result of multicollinearity being present in the model will cause OLS to remain biased, but it will also still be BLUE. Before performing this test, we suspect that there may be collinearity between the independent variables B4qualfac (quality of faculty) and B5publi (publication). The reason we suspect these variables to have a high linear relationship is due to the

fact that the quality level of the faculty employed should ultimately produce more publications considering that their work should be elite in their individual fields resulting in more recognition of their content. After running the correlation test we noticed that our predicted independent variables B4qualfac and B5publi didn’t hold an alarmingly high linear relationship in this model. Even though multicollinearity was not present between our suspected independent variables, after examining the model it showed that it was actually slightly present between the independent variables B5publi (publication) with B6influ (influence) and B7citation (citations). Although we did notice a correlation between these selected variables the linear relationship they held between one another didn’t reach a level of concern so our model did in fact pass the testing of multicollinearity.

For our next step, we will check to see if we can detect any forms of heteroskedasticity. When inquiring for heteroskedasticity, we will start by performing the Breusch-Pagen Test which checks for any linear heteroskedasticity. After performing this test, we realize that our model contains high levels of linear heteroskedasticity. Following this test, we will additionally use the White Test to also look for any non-linear forms of heteroskedasticity. Again, this test failed and proves not only that our variables contain linear heteroskedasticity, but also non-linear. To adjust for this error, we will look to create a quadratic variable or combine variables to offset this undesirable presence within our model. After creating and testing different adjusted forms of our model we came to the conclusion that this heteroskedasticity error could not be improved from our initial model. Along with this, we were unable to improve the misspecification that was indicated

in our model through the Ramsey RESET test. With this proven stubbornness for the model to change with our adamant adjustments, we will choose to not add any additional adjustments into our model. We will stick to the original model since it is the best form that we could create.

Understanding these variables and how they influence the overall outcome of the institutions score, can help educators focus on the academic qualities that most correlate with a high overall quality rating. This will allow schools to understand the importance of these measured statistics and can give them a path to improve areas that are lacking among their educational systems. By reviewing our model, institutions will learn that the quality of the faculty has the most effect on their overall score. Moreover, it provides a model that students and families can use in order to make more calculated and appropriate choices on where to pursue a higher education. After testing our hypothesis, we came to the conclusion that the variables selected will in fact have a negative effect on the overall score of a university as they increase. The testing of this hypothesis provided us with valuable results that can be shared to users referencing to these universities' scores understanding that manipulation of rankings and bias is in fact present and is recommended to take into consideration. When looking at our model we can state that the variable that increases a university's score the most, is the quality of the faculty. This developed model can give this specific audience additional resources that are hard to come by among other models, that tend to be developed and altered to meet the goals of a specific institutions or group of institutions.

Works Cited

- Article 1 – Mussard, Maxime, and Alex Pappachen James. “ Boosting the Ranking of a University Using Self-Citations.” *Current Science (00113891)* , vol. 113, no. 10, 25 Nov. 2017, pp. 1827–1827. *Academic Search Complete* , EBSCOhost, lib. tcu. edu/PURL/EZproxy_link. asp? url= http://search. ebscohost. com. ezproxy. tcu. edu/login. aspx? direct= true&AuthType= cookie, ip, uid&db= a9h&AN= 126435978&site= ehost-live. Accessed 5 Dec. 2017.
- Article 2 – Horstschräer, Julia. “ University Rankings in Action? The Importance of Rankings and an Excellence Competition for University Choice of High-Ability Students.” *Economics of Education Review* , vol. 31, no. 6, Dec. 2012, pp. 1162–1176. *Academic Search Complete* , EBSCOhost, doi: 10. 1016/j. econedurev. 2012. 07. 018. Accessed 5 Dec. 2017.