Sources of Information Students Use to Select Colleges – A Proposed Study using Multiple Response Analysis

Hossain, M. Shahadat
SUNY Potsdam

ABSTRACT

This study identifies the various sources of information students' use in selecting an institute of higher education. The literature does not provide any study that looks at all these various sources simultaneously to determine which is most preferred. This study proposes to provide an analysis that looks at complex data originating from multiple responses that is common in marketing research. We propose the use of multiple response analysis. The study will also investigate the level of importance students attach to each source of information. The study further proposes to apply the Rao- Scott correction factor to obtain the corrected chi-squared statistic. Thus, we propose a test of association in order to determine if there is evidence of a difference in information source selection between high school students and undergraduate students. The results can aid institutes of higher education in their marketing strategies.

INTRODUCTION

This study focuses on determining the most salient information delivery methods utilized by students when determining college choice. Six delivery methods are considered in depth, including college information pamphlets, open houses, newsmagazines that list college rankings, friends and family, teachers and guidance counselors, and the internet. Previous studies have investigated several of these sources independently while few have considered the aggregate of all six. Students may choose more than one source of information in their search of institutions of higher education. Studies to date, have primarily, employed anecdotal and qualitative analyses to determine these sources. Empirical evidence based on quantitative analyses is sparse in the literature. This may have resulted from the complexity of analyzing data stemming from the use of multiple sources in the decision making process. This study is thus motivated to analyze data involving more than one choice. Multiple response analysis enables researchers to analyze data resulting from more than one choice or "pick any" responses that is common in marketing research. This study proposes to employ multiple response analysis to determine if there is evidence of a difference in information source selection amongst high school and undergraduate students.

As described in Bergerson's study *College Choice and Access to College: Moving Policies, Research, and Practice to the 21st Century* (2009), the college choice process tends to be viewed as comprehensive, consisting of several steps before the final decision on a single college. The most widely-used model is Hossler and Gallagher's three-stage mode (1987), which includes predisposition, search, and choice.

The predisposition stage involves the development of students' college aspirations and expectations. Numerous factors including family socioeconomic status, parental involvement, peers, high school teachers and counselors, and the relative value these advisors place on attending college affect the student. The second stage, the search, normally occurs between tenth and twelfth grade. During this stage, students form choice sets and determine which institutional characteristics are most important. Students often take the SAT and/or ACT during this stage. The third and final stage is known as the choice stage. This is when students synthesize the information they have gained and use it to select an

institution and complete the enrollment process. This stage also tends to occur between tenth and twelfth grade.

SOURCES OF INFORMATION

Colleges mail brochures containing information relevant to graduating high school seniors. A study by Simões and Soares (2010) looked at the decision-making process of students at a university in Portugal. During registration for the college, the students were given a questionnaire since at this stage the decision-making process was still relevant. These students were grouped by their academic aptitude based on secondary school grades. The questionnaire asked the students to rank the most-used sources of information. In the category of marketer-controlled sources, college brochures were ranked in the top three most-used sources in 22.7% of respondents. Furthermore, 23.6% of respondents ranked university official guides, also a marketer-controlled source, in the top three most useful resources considered in their study.

In a 2006 study, Fischbach studied ways to attract future students in the increasingly competitive college recruitment environment at California State University, Northridge. In his study, he found that 79% of students indicated that they learned of an open house as a result of a mailed brochure. In another study, Thomas and Dawes' (1999) looked at factors that predict whom among recruited university applicants are most likely to attend. They found that participation in open house activities was one of four positively correlated variables. The Fischbach, a study undertaken to determine the long-term viability of open houses, also showed that the vast majority of participants perceived the open house positively and it satisfied their need to obtain information about the university (2006).

A study by McDonough, et al. (1998) looked into the effect that college rankings had on a student's decision to attend a college. 59.9% of students did not find rankings to be very important, 29.6% found them to be somewhat important, and 10.5% found them to be very important in their college search. In terms of ethnicity, Asian Americans were twice as likely to find rankings to be very important. Relative to students who rated rankings as unimportant, users who considered rankings to be important were frequently more likely to ask teachers for advice in high school. They tended to be higher-achieving students. Students who found the rankings as not important tended to be twice as likely to transfer to another college and also twice as likely to live with their parents. The study also found that 65% of first-generation college students find news magazines rankings unimportant compared to 58% of students with college-educated parents. Middle-income students tended to show the highest usage of the ranking system. In conclusion, high-achieving students who attended highly competitive post-secondary institutions and are focused on colleges that will give them good prospects for both graduate school and professional opportunities tended to use college rankings the most.

Pérez (2010) found that for U.S.-born Latino students, familial contacts and peer networks were critical influences on students in their search for a college. Similarly, in the Simõez and Soares (2010) study, it was found that 58.6% of respondents ranked former/current university students, an interpersonal source, as one of their top three most-used resources. Conklin and Dailey (1981) found that consistent parental assumptions of continued education increased the likeliness of a student attending college. A study by Joshi, et al. (2009) found several factors that were important to a student's decision on whether to attend a community or a four-year college. Students with parents having attained higher levels of education and income are more likely to attend four-year colleges. The study also found that parental encouragement of academic excellence resulted in a higher probability of attending a four-year university.

Studies have shown that guidance counselors and teachers also play a role as information-providers to potential college students. Pérez (2010) found that having contact with college professors prior to application was a critical influence in the college application process. The Simões and Soares study (2010) found that guidance counselors, considered a third-party independent source, were ranked within

the top three information sources by 11.3% of respondents. Further it was found that 30.3% of respondents believed their current teachers, an interpersonal source were a useful source of information for college. It is interesting to note that of the top three most-used resources in this study, two of them (former/current university students and current teachers) were interpersonal. This shows the heavy reliance high school students have on mentors and peers. Additionally, the study looked at the most-used sources relative to academic aptitude. There were significant differences between a student's overall high school grades and the sources he or she used to gain information about college. Students with the highest academic aptitude tended toward using interpersonal sources such as teachers while students with medium academic aptitude tended to favor marketer-controlled sources such as the university website.

The Simões and Soares (2010) further continued to describe that the university website, a marketer-controlled source, was the top most-used resource among respondents, with 80.9% of the respondents regarding this resource as one of the top three resources used to gain college-related information. A study by Dey Alexander (2005), strengthens Simões and Soares' 2010 findings, agreeing that usage of the internet in the search for colleges has become more widespread among prospective students. However, despite the popularity of college websites, Alexander's (2005) study showed that many websites are difficult to navigate, include incorrect or generic information, and generally are not designed for their target audience. Many university websites mimic the structure of the organizations that own them. Information such as fees and scholarships tend to be difficult to find. Many students do not know where to look for this information, so they give up the search assuming the information is not online. These websites failed to deliver the information that the student was looking for because the content was often not written for prospective students.

MULTIPLE RESPONSE ANALYSIS

Multiple response data (MRD) in research arise from questions that require the respondents to make multiple choices from among a restricted or unrestricted set of alternatives. Referred to as "pick any" data, they are commonly found in marketing (Decady and Thomas, 2000) and especially in the study of consumer behavior.

The use of specific products, brands and distribution channels, readership of magazines, contexts of product consumption, activities engaged in outside the marketplace (e.g. hobbies) and perceptions about the attributes of brands are examples of behavioral constructs that are commonly studied with the use of multiple response questions (Edwards and Allenby, 2001).
Loughin and Scherer (1998) gave an example in which farmers were asked to identify their primary sources of veterinary information. Five possible sources were listed and respondents were allowed to identify more than one source (Table 1).

Multiple response data are often collected as the first part of a multistage investigation to identify the most important items. The results are often the part of a descriptive study, which may lead the researcher to move from data to theory as opposed to theory to data.
The paper delves into the various applications available to researchers in analyzing multiple response data as identified by Decady (2000) in his Ph.D. thesis on the analysis of multiple response data The test of association developed by Decady and Thomas (2000) is applied to Table 1 as an example.
The analyses focus only on aggregated data.

Table 1: Loughin and Scherer (1998): Farmers' veterinary information sources by education.

	Information Source					Total	Total
Education	A	B	C	D	E	Respondents	Subjects

High School	19	38	29	47	40	173	88
Vocational	2	6	8	8	4	28	16
2-Yr. College	1	13	10	17	14	55	31
4-Yr. College	19	29	40	53	29	170	113
Other	3	4	8	6	6	27	14
Total	**44**	**90**	**95**	**131**	**93**	**453**	**262**

Average Response = 1.7 / Farmer

Possible Approaches to the Analysis of Multiple Response Data:

Pearson chi-square statistic assumption:

The classical Pearson test of association for tabulated data has a large sample chi-square distribution only when certain conditions are satisfied, namely:

a. the data consists of one or more simple random samples of *n* independent observations (or subjects) drawn from one or more large populations;
b. the observations (or subjects) can be classified into categories that are mutually exclusive and collectively exhaustive.

Under these conditions, the cell counts of the cross tabulation follow a multinomial distribution, (Decady, 2000).

With MRD variables each subject can respond to more than one count to each cell, ("pick any"). This will result in cells not being mutually exclusive and thus violate the above (b) assumption. Another problem that will arise will be the "within subject" correlation, inducing a correlation between the cells that is different from the correlation implied by the classical multinomial assumption. As a result the multinomial distribution will not be the correct distribution for the observed multiple counts aggregated over subjects, (Decady, 2000).

Rao and Scott (1981, 1984), Rao and Thomas (1988) suggest appropriate corrections of the Pearson statistic when dealing with data from statistical survey designs involving clustering. This is pointed out as there is a violation of assumptions upon which a valid Pearson chi-squared statistic is predicted when working with clusters. Decady (2000) in his dissertation observes that the clustering effect of MRD runs parallel to complex data and reviews the techniques used for analyzing complex data on MRD.

Statistical Methods for Categorical Data from Complex Data:

Three main families of procedures have been identified and applied in hypothesis testing on cross classified categorical data drawn from complex surveys, Rao and Thomas (1988) and Thomas (1993). These are:

(i) Tests based on the weighted least squares approach with tests provided by Wald statistics, (Koch, Freeman and Freeman, 1975; Binder, 1983).

(ii) Rao-Scott first and second order corrections to the classical χ^2 (Pearson) and G^2 (log-likelihood ratio tests), (Rao and Scott, 1981, 1984, 1987)

(iii) Based on the principles of resampling. Methods for jackknifing the classical χ^2 and G^2 tests developed by Fay (1975, 1982, 1985).

"One common thread among all these streams of research is that they all rely on some common method of variance and covariance estimation that accounts for the complexity of the data," (Decady, 2000).

Decady and Thomas, (2000) Analysis:

Decady and Thomas (2000) investigated procedures for testing for independence in a two way (RxC) table of aggregate counts where the column variable admits more than one response per subject (e.g. Table 1). As mentioned earlier, standard chi-squared tests of association do not apply to multiple response data. Decady and Thomas (2000) used the approximate large sample method of Rao and Scott (1981, 1984) to develop a simple first order corrected chi-squared test. This is represented below:

Let:

m_{ij} = The observed counts *{where the indices: i (for single-response variable) and j (for multiple-response variable) run from 1 to R and 1 to C, respectively}.*

m_{+j} = the column totals

n_i = number of subjects in each row *(where the n_i will be assumed to be fixed by design).*

τ_{ij} = the probability that a subject in the *i*th row will respond positively in the *j*th multiple response category.

τ_i = represent the Cx1 vector of probabilities in the *i*th row.

m_i = corresponding vectors of row counts.

For the above sampling model, the hypothesis of no association is equivalent to the hypothesis of homogeneity of response probabilities, i.e.,

$H_0: \tau_1 = \ldots, \quad \tau_R = \tau$ [1]

Where τ is the unknown vector of common response probabilities.

Thus under the null hypothesis, the row counts are samples from distributions having common probability vector τ, with estimate $\hat{\tau}$ given by:

$$\sum_i m_i / n_+ = m_+ / n_+ ,$$

where m_+ is the Rx1 vector of column totals m_{+j} and $n_+ = \sum_i n_i$ is the total number of subjects.

Estimates of the individual row probabilities τ_i are given by $\hat{\tau}_i = m_i / n_i$, i = 1,…,R. A statistic of chi-squared type for testing H_0 is given by:

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} n_i \frac{(\hat{\tau}_{ij} - \hat{\tau}_j)^2}{\hat{\tau}_j} \qquad [2]$$

where $\hat{\tau}_{ij} = m_{ij} / n_i$ represent the *j*th elements of the estimated probability vectors τ_i, i = 1,…,R and where $\hat{\tau}_j = m_{+j} / n_+$ represents the *j*th element of the estimated common probability vector $\hat{\tau}$. Statistic [2] is identical to Umesh's (1995) pseudo-chi-squared statistic and Loughin and Scherer's (1998) modified chi-squared statistic which Decady and Thomas (2000) expressed as in the following form:

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(m_{ij} - \hat{E}(m_{ij})\right)}{\hat{E}(m_{ij})} \qquad [3]$$

where the estimated expected counts $\hat{E}\ (m_{ij})$ in this multiple response situation are given by $n_i \hat{\tau}_j = n_i m_{+j} / n_{+}$.

Decady and Thomas (2000) observe that although the alternative expressions for the statistic X^2 given in equation [2] and [3] use appropriate estimates of the expected probabilities and counts, both Umesh (1995) and Loughin and Scherer (1998) noted that, for multiple response data X^2 does not exhibit the familiar large-sample chi-squared distribution under the null hypothesis. Thus, Decady and Thomas (2000) propose a remedy by using the first-order Rao-Scott procedure, which consists of dividing X^2 by a correction factor $\tilde{\delta}$ chosen to make the mean of the asymptotic distribution of the corrected statistic equal to the mean of a random variable on the appropriate number of degrees of freedom, namely (R-1)C .

The required correction factor has a very simple form:

$$\tilde{\delta} = 1 - \frac{m_{++}}{n_{+}c}, \qquad [4]$$

where $m_{++} = \sum_j m_{+j}$ is the sum of all the multiple response counts in the table.

It can be seen from equation [4] that the correction factor can be calculated directly from the aggregate table of multiple response data, provided the number of subjects per row is known; the expanded table is not required. The first-order Rao-Scott corrected chi-squared test of null hypothesis [1] then consists of referring the corrected statistic, given by;

$$X^2 = X^2 \Big/ \tilde{\delta}\ , \qquad [5]$$

to the upper $100\alpha\%$ critical point of a chi-squared distribution on (R-1)C d.f.

Loughin and Scherer (1998) data analyzed:

Decady and Thomas (2000) reanalyzed the Loughin and Scherer (1998) data reproduced in Table 1.For these data, Loughin and Schere's modified/pseudo-chi-squared statistic, similar to X^2 in equations [2] and [3] was calculated to be 20.85.

Decady and Thomas applies the correction factor, $\tilde{\delta}$, to obtain the corrected chi-squared statistic. $\tilde{\delta}$ value is calculated using equation [4] as follows:

$\tilde{\delta}$ = 1- (44+90+95+131+93)/ (262x5) = 0.654

Thus applying to equation [5]:

X^2 = 20.854/0.654 =**31.88** [d.f.= (R-1)C = (5-1)5 = 20]

Compared to the 5% critical value of a chi-squared distribution on 20d.f. =31.41

31.88 > 31.41 Thus: H_0 : Reject [*There is evidence of a difference in information source selection across educational level*.]

The value of the adjusted statistic is slightly greater than the 5% critical value, suggesting a test *p*-value slightly less than 0.05.

METHODOLOGY

This study proposes to survey high school students in the upstate New York region who plan to pursue a two year or a four year degree. These are high schools that attend Open Houses at SUNY Potsdam. We will also gather data from undergraduate students at SUNY Potsdam who plan to pursue a graduate degree. Students will be asked to select any of the six identified sources of information and have the option of selecting more than one. Multiple response analysis will be used to analyze this data. The level of importance of the six sources will be determined by using a five point likert scale with the verbal

anchors of 1= Least Important and 5= Most Important. Mean and standard deviation will be used to analyze the data for level of importance.

Survey instruments will be mailed to the high schools and teachers will be asked to administer them in the grade 12 classes. Responses will be mailed back to the author. The same surveys will also be administered amongst SUNY Potsdam seniors who plan to pursue a graduate program. Survey instruments will be provided to various faculties across campus and completed instruments will be returned to author via campus mail. Data will be gathered from January to March 2012 when students place the greatest emphasis in their selection of an institute of higher education.

DISCUSSION

While this is a proposed study, at this stage, it does provide us with a statistical method to analyze multiple response data. Once the data is collected and analyzed the results should aid admissions office to better strategize their marketing efforts to reach potential students. Decady and Thomas have conducted further work in the test of association with fine tuning in the statistic applied in this study. However, as this was the first of its kind we have limited this study to applying the very first published statistic with the correction factor they have proposed.

It is simple in approach and can easily be applied by practitioners and academics. The analysis of the Loughin and Scherer (1998) data goes to show the diverse range of multiple response data that researchers can come across and how simply one can analyze the data.
However, it is still difficult to analyze multiple response data any further than to test for association. For instance, if we were trying to find out the preferred source of information students use for selecting a program or college we would have to resort to descriptive statistics. This then can pose yet another problem as we would have to decide if we would deal with the total number of respondents or total number of subjects. The analysis of the importance of each of the sources will aid institutes of higher education to develop a more targeted approach in reaching prospective students.

REFERENCES

Alexander, Dey (2005). "How Usable Are University Websites? A Report on a Study of the Prospective Student Experience." Ausweb 05.

Bergerson, Amy Aldous (2009). "College Choice and Access to College: Moving Policies, Research, and Practice to the 21st Century." *ASHE Higher Education Report* 35.4 (2009): 1-141.

Bureau of Labor Statistics. "College Enrollment and Work Activity of 2010 high School Graduates". 8 Apr. 2011. 15 Nov. 2011 < http://www.bls.gov/news.release/ hsgec.nr0.htm>.

Conklin, Mary F. and Ann Ricks Dailey (1981). "Does Consistency of Parental Educational Encouragement Matter for Secondary School Students?" *Sociology of Education* 54.4 (1981):254-62.

Decady, Y. J. (2000). "The Analysis of Multiple Response Data in Marketing Research: A Rao Scott Adjusted Chi-Squared Approach," Ph.D. Thesis, Carleton University, Ottawa, Canada.

Decady, Y. J and Thomas D.R. (2000). " A Simple Test of Association for Contingency Tables with Multiple Responses," *Biometrics,* 56, (September)

Decady, Y. J and Thomas D.R. (2000). "Analyzing categorical Data with Multiple Responses per Subject," *2000 Proceedings of the Survey Methods Section: Statistical Society of Canada,* (June)

Dolinsky, Arthur L (2010). "The Adequacy of the information that Students Utilize when Choosing a College: An Attribute Importance and Information Sufficiency Approach." *College Student Journal* 44.3: 762-76.

Edwards, Y.D. and Allenby, G.M. (2001). " Multivariate Analysis of Multiple Response Data," *Fisher College of Business, Ohio State University,* (June).

Fischbach, Ronald (2006). "Assessing the Impact of University Open House Activities." *College Student Journal* 40.1 (2006): 227-34.

Joshi, Prathibha V., Kris A. Beck, and Christian Nsiah (2009) . "Student Characteristics Affecting the Decision to Enroll in a Community College: Economic Rationale and Empirical Evidence." *Community College Journal of Research and Practice* 33.10: 805-22.

Loughin T. M. and Scherer, P.N. (1998). "Testing association in Contingencty Tables with Multiple Responses," *Biometrics,* 54.

McDonough, Patricia M., Anthony Lising Antonio, MaryBeth Walpole, and Leonor Xóchitl Pérez (1998)**.** "College Rankings: Democratized College Knowledge for Whom?" *Research in Higher Education* 39.5 513-37.

Menon, Maria Eliophotoul, Anna Saiti, and Michalis Socratous (2007). "Rationality, Information Search, and Choice in Higher Education: Evidence from Greece." *Higher Education* 54.5: 705-21.

Padmannavar, Sunita S. and Milind Joshi (2011)**.** "Survey Analysis on the Web: Similarities and Differences in University Website Information." *International Journal of Engineering Science & Technology* 3.7: 6101-6113.

Pérez, Patricia A (2010). "College Choice Process of Latino Undocumented Students: Implications for Recruitment and Retention." *Journal of College Admission* 206: 21-25.

Rao, J.N.K. and Scott, A.J. (1981). "The Analysis of Categorical Data from Complex Surveys: Chi-squared Tests for Goodness of Fit and Independence in Two-way tables," *Journal of the American Statistical Association,* 76.

Rao, J.N.K. and Scott, A.J. (1984). "On Chi-squared Tests for Multi-way Tables for Cell Proportions Estimated from Survey Data," *Annals of Statistics,* 12.

Rao, J.N.K. and Thomas, D.R. (1988). " The Analysis of Cross-Classified Categorical Data from Complex Sample Surveys," *Sociological Methodology,* 18.
Thomas, D.R. (1993). "Inference using Complex Data from Surveys and Experiments," *Canadian Psychology,* 34, no. 4 (October)

Simões, Cláudia, and Ana Maria Soares (2010). "Applying to higher Education: Information Sources and Choice Factors." *Studies in higher Education* 35.4 (2010): 371-89.

Umesh, U.N. (1995). "Predicting Nominal Variable Relationships with Multiple Response," *Journal of Forecasting,* 14. In Decady, Y. J. (2000). "The Analysis of Multiple Response Data in Marketing Research: A Rao Scott Adjusted Chi-Squared Approach," Ph.D. Thesis, Carleton University, Ottawa, Canada.

U.S. Census Bureau. The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings. July 2002. 15 Nov. 2011 http://www.census.gov/prod/2002pubs/p23-210.pdf