Forming and Using Peer Groups Based on Nearest Neighbors with IPEDS Data

Gerald McLaughlin (DePaul University)

Richard Howard (Consultant)

Josetta McLaughlin (Roosevelt University)

ABSTRACT

Institutional performance benchmarking requires identifying a set of reference or comparator institutions. This paper describes a method by which an institution can identify other institutions that are most similar to itself using a methodology that identifies the nearest institutional neighbors based on a balanced set of metrics accessed from IPEDS data. The Nearest Neighbor methodology is robust and flexible; it is easy to understand and to explain to others; and it is a hybrid method integrating judgment and analytical techniques. Use of the method is discussed, and it is compared to other methodologies such as Cluster Analysis.

INTRODUCTION

The use of reference or comparator groups in higher education has become common practice. There are various types of groupings, among them peer groups, aspiration groups, natural groups, and competitor groups. For this paper, the term reference group is used as a general term that refers broadly to peer groups that are constructed on the basis of similar key characteristics. The paper is organized around the seven steps that are required to identify peer groups and a case study that demonstrates the application of these steps. The nearest institutional neighbors are identified using a balanced set of metrics available through data from the Integrated Postsecondary Education Data System (IPEDS).

Exploration of various statistical methodologies for forming reference groups in US higher education began more than 20 years ago (Terenzini, et al., 1980; Teeter & Brinkman, 1987; McLaughlin & McLaughlin, 2007). The primary objective was then, as it is now, to find an appropriate method for benchmarking the performance of one institution relative to a group of institutions. The overall goal of this effort was thus to identify an appropriate means for making judgments about the relative performance of institutions. The development of reference groups paralleled the development of performance benchmarking as a common feature for many of our institutions, especially those that are funded by various states and public monies. Benchmarking has in fact become a requirement of various accrediting agencies who are interested in how institutions perform when compared to other similar institutions. In addition, institutions that operate within the financial markets now have a means for providing information specific to the higher education sector that is required by the various bond agencies and

other financial organizations who evaluate the financial stability of institutions (Townsley, 2002; Gaither, Nedweek, & Neal, 1994).

Two basic statistical procedures are commonly used to form groups – Cluster Analysis based on a cluster algorithm that identifies relatively homogenous groups and a Nearest Neighbors statistical methodology based on a distance score between a target institution and other institutions which are similar (McLaughlin & McLaughlin, 2007). The advantage to using statistical methods is that such procedures are relatively objective. The disadvantage is that the outcomes are sometimes complicated to explain to the end user of the analysis. Critics have also suggested that problems potentially surface with respect to comparability, substitutability, and the additive attributes of some procedures (McLaughlin & McLaughlin, 2007; Hom, 2005).

This paper focuses on the second of the two statistical procedures – Nearest Neighbor statistical procedures. Though the procedure itself is relatively objective, the context in which the analysis is done requires that various judgments be made concerning the overall process. Decisions associated with forming peer groups are heavily nuanced by both the political and analytical context in which the analysis takes place. The methodologies chosen for forming peer groups ultimately depends on the answer to questions concerning the appropriate variables for selecting reference institutions and the appropriate methodology for use in analyzing these variables.

Seven steps will be used to describe how a Nearest Neighbor methodology is used in identifying peer groups. We will first provide a general discussion of these steps and then we will show the key characteristics of the methodology as it fits within the steps. While we describe this methodology by identifying a purpose for developing the

reference group and conclude by presenting results, in reality the steps rarely represent a linear sequence. The completion of one step will frequently result in an iterative, but hopefully heuristic, cycle of revisiting previous steps while simultaneously moving to the next step in the sequence. The sequence of steps is:

1. Clarify the purpose for developing and using the reference group(s).
2. Determine the composition of the comparison -- what type, what size, and how many reference groups to form.
3. Select a methodology for forming the reference group(s).
4. Identify measures of interest and targets for outcomes.
5. Determine how much difference makes a difference.
6. Collect and analyze the data.
7. Present results and adjust the process

Differentiating Between Cluster Analysis versus Nearest Neighbor Methodology

The choice of methodology for this study on forming peer groups is best understood through comparison with Cluster Analysis. The conceptual difference lies in differentiating between techniques that begin with a set of data points from which clusters are formed (i.e., Cluster Analysis), and those that begin with a single institution as the data point and identify other institutions that are close to the reference institution based on a distance measure (Nearest Neighbor methods). This paper describes the latter, i.e. a Nearest Neighbor method.

Cluster Analysis is a generic name for methods that identify objects that are similar on some attribute(s) (Romesburg, 2004). It is used widely in many professions

(see for example, Punj & Stewart) and, in the case of higher education, is frequently used to develop classifications of institutions. Similarly, it is used by administrators to inform decisions in planning and management.

There are many forms of Cluster Analysis from which to choose (Hartigan, 1975; Gordon, 1981; Fraley & Raftery, 1998) with methods ranging from heuristic to formal based on statistical models. Many methods follow a hierarchical strategy (Fraley & Raftery, 1998), Four generic steps are generally followed for hierarchical Cluster Analysis – data collection for creating of a data matrix, standardization of the data matrix, computation of values to measure similarities among all pairs of data objects, and use of a clustering method to show the hierarchy of similarities among these pairs (Romesburg, 2004).

By contrast, Nearest Neighbor methodologies used in this study, though closely associated with the study of whether a data set is clustered (Cherni, n.d.), focus on the distances that occur from a data point to its Nearest Neighbor(s) (Clark & Evans, 1954). Like Cluster Analysis, the Nearest Neighbor method is a widely used generic application -- ranging from ecology and psychiatry to archeology.— that can be applied to multiple models (Cherni, 2005; Diggle, 2003; Clark & Evans, 1954; Skellam, 1952). Much of the work on Nearest Neighbor methods employs R-trees due to their efficiency and popularity (Tao, Papadias, & Shen, 2002).

The Nearest Neighbor methodology employed in this study is discussed in the following sections. Due to the non-linear nature of the process, the information is contained in multiple sections.

DISCUSSION AND APPLICATION OF STEPS AND METHODOLOGY

> *In this age of accountability, transparency, and accreditation, colleges and universities increasingly conduct comparative analyses and engage in benchmarking activities. Meant to inform institutional planning and decision making, comparative analyses and benchmarking are employed to let stakeholders know how an institution stacks up against its peers and, more likely, a set of aspirant institutions—those that organizational leaders seek to emulate. (James F. Trainer, 2008)*

The following section begins with a discussion of the context within which the Nearest Neighbor methodology is used to form a peer or reference group. Each of the seven steps is discussed in sequence as a means to address the on-going complexities of this context. The Nearest Neighbor methodology is further described in the case study that follows this section. The case describes peer group formation at a southeastern university..

1. Clarify the purpose for developing and using the reference group(s).[1]

Institutions traditionally list a number of reasons for establishing a peer group. Often these reasons include requirements for accountability by various state and public agencies along with requirements from various accrediting agencies that the institution demonstrate an acceptable level of effectiveness and efficiency in its operations. In other situations, the initiative for performance benchmarking can come from internal

[1] There are several uses for reference groups. For our purposes, we are considering a group of institutions that can be used to compare an institution against a group of institutions that are similar on specified attributes. The terms peer, reference and comparison will be used interchangeably in the description.

concerns. Typically these concerns are brought forth by various advocacy processes that likely involve resource issues. For example: Are there sufficient numbers of faculty members given the number of students and programs? Are faculty salaries competitive with salaries of faculty at peer institutions? Is the development office raising sufficient advancement funds? Do we have the appropriate degree programs given the institution's size and programmatic characteristics?

As can be seen by these common and frequently asked questions that reflect both internal and external pressures, any initiative that requires formation of peer groups for performance benchmarking has the potential for significant political, social, and economic impacts. The impacts can affect (positively or negatively) the potential professional status of faculty, administrators, staff, and ultimately students. As such, it must be recognized up front that any activity leading to peer group formation will in reality be influenced by political agendas from across the campus and often beyond.

Given the political context, the peer group formation process should start with a statement of the purpose for which the comparisons will be used. This purpose can be extremely broad, e.g., comparing overall institutional effectiveness with other "peer" institutions. It can also be an extremely focused purpose, e.g., comparing the adequacy of faculty salaries or setting goals for faculty research funding. Traditional foci of comparisons include salaries, staffing, adequacy of funding, expenditures, assessing outcomes such as graduation, debt and debt repayment, and numerous institutional characteristics, an example being those found in ranking publications such as *US News and World Report*. Specific attention can be focused on primary areas of concern to include finance, enrollment, staffing, and facilities. Undergirding all of these discussions

is the basic mission of the institution. A useful template for this audit can be a PEST assessment of the Political, Economic, Social and Demographic, and Technical contexts of the referenced institution (McLaughlin & McLaughlin, 2007).

One decision that frequently needs to be made at this point of the process involves the desirability of creating one general set of "peers" versus creating different sets of "comparator" institutions to be used in different comparative analyses, for example, one for salary comparisons, one for retention and graduation comparisons, one for financial comparisons, etc. Another decision that should be made at this point involves the intended use of the comparator group's metrics. For example, if the use is in planning, it may be desirable to select a peer group where the institution is at the median and set a goal at a higher or lower quartile. In this case, it will be desirable to construct a relatively large comparator group of 20 to 30 institutions and to establish performance benchmarking goals for specific areas at different percentiles of the group. On the other hand, it may be desirable to have multiple smaller groups for student and faculty outcomes. This is discussed in the following step.

2. Determine the composition of the comparison --what type, what size, and how many reference groups to form.

Two factors should be considered in discussions about the purposes for developing and using the reference group(s): (1) what type of reference groups to form and. (2) the size of the(se) group(s). If the purpose is single, general and focused, a single general reference group will likely be sufficient for comparative purposes. One caution is that in most cases where institutions are employing a general aggregate group, there is sometimes a subconscious sense underneath the surface that there are

in reality two groups -- those whom we are like and those whom we would like to be like (i.e., the aspiration group mentioned in the earlier Trainer quote (2008). In general, the larger and more complex an institution, the more likely it will be that multiple comparison groups will be necessary. The smaller liberal arts colleges will usually need a single group. Again, however, the need for one or multiple comparison groups will be determined by the intended purpose for the creation of the group(s).

In addition, once the purpose of the comparisons has been defined, there are a number of types of comparison groups that can be created. The most common comparison group, and the one pursued later in this discussion, is the identification of a ***Peer*** *group*. These institutions are *similar* to the reference university, i.e., your university, on most primary or key attributes. Traditionally, they will be approximately the same size, have a similar general mission, have somewhat similar student bodies and curricula, and have similar resources.

The second reference type is the ***Aspirational*** group. This group is comprised of institutions that have one or more attribute(s) or characteristics that the home institution desires to attain but has not yet attained. Frequently these attributes are perceived to lead to a higher status (Carnegie), greater resources, and a higher level of performance on indicators (graduation rates, research grants, etc.). Otherwise, the institutions have similar characteristics. It is not uncommon for institutions to identify such groups based on one of the popular rankings such as *US News and World Report*. These institutions are sometimes considered to be “preferred peers”.

A third type of reference group is the ***Competitor*** *group* which is comprised of institutions that compete with the home institution for some resource. For example, a

frequent competitor set of interest would be those institutions where students go when they do not enroll after your institution accepts them. In this situation, there are several organizations that will help you identify where "your" students go (one form of competition) after they are accepted by your institution. These primary organizations include the National Student Clearinghouse (www.studentclearinghouse.org) and ACT (www.act.org). Competitor groups can also be established in terms of faculty; if faculty are offered a position at your institution but do not accept it, where do they go? One key reality about competitor groups is that they do not always have to be higher education institutions. For example a primary competitor for students can frequently be a military service or a local business or industry.

The fourth grouping, ***Predetermined*** groups**,** are those institutional groupings that already exist for other purposes. *Predetermined groups* include traditional groups such as faith-based institutions, natural groups such as athletic affiliation, and jurisdictional groups comprised of institutions that are part of a legal or geographical jurisdiction. Similar to predetermined groups are classification groups such as those formed by the Carnegie Classification process. These particular classifications are used extensively in national studies, i.e., *US News and World* rankings and AAUP salary studies. [2]

There are several basic strategies for determining the appropriate size of a reference group. As noted earlier, one strategy is to identify a larger group of institutions

[2] For a more extensive discussion of types of groups see D. J. Teeter and P. T. Brinkman, "Peer Institutional Studies//institutional Comparisons," in Primer for Institutional Research, J. Muffo and G. McLaughlin, eds., (Tallahassee: Association for Institutional Research, 1987, 89 – 100, D. J. Teeter and P. T. Brinkman, "Peer Institutions," in Primer for Institutional Research, MA Whiteley, JD Porter, and RH Fenske, eds., (Tallahassee: Association for Institutional Research, 1992), 63-72), and G.W. Mclaughlin and J.S. McLaughlin, The Information Mosaic, AGB, 2007, Washington DC, Chapter 7

-- 25 or 35 institutions. These can be used as a norm group from which goals and objectives can be developed. For example, it can be used as a norm group in the formation of an aspiration group using one or more characteristics of institutions selected from the predetermined group. On the other hand, similar attributes from the group of 25 or 35 institutions can be used to set standards at a point other than the mean or median of the group. For example, from a set of 25 or 35 comparable regional institutions, one might identify a retention and graduation rate as the median for the group and to be used as the peer comparison and then set retention and graduation rates at the 75th percentile as an aspiration or "stretch" goal. A major advantage to having a larger group is that many of the data exchanges such as CUPA-HR, CSRDE, and NSSE may only contain a subset of institutions that are in the reference group. Starting with a larger norm group makes it more likely that there will be a sufficient number of comparison institutions that are participating in the data exchanges noted above. The alternative to this strategy would be to identify smaller focused specific groups. For planning purposes, one might form a group of four or five very similar institutions as current peers and a second group of four or five institutions that represent an aspiration group. However, the smaller the group(s), the more risk there is that there will be political opposition to their appropriateness.

Another decision point in determining size is whether to use different groups for different purposes. For example, strong arguments for different groups can be made on the basis of the resource base and student characteristics. With respect to resources, it may be desirable to look only at other institutions within a similar sector such as public or private not-for-profit, private for-profit, urban/rural, etc. However, when looking at

student characteristics, it may be more desirable to consider only other institutions with similar curriculum profiles and a similar balance of residential/commuter, graduate/undergraduate/professional,minority/ethnic student characteristics, or socioeconomic status. When one is looking at competitors, the reference group will very likely differ substantially from any general peer or aspirational groupings.

When forming peer groups, there is a strong relationship between the required similarity attributes and the number of similar institutions. A good approach to better understanding this relationship is to access the website provided by Carnegie Foundation where you can conduct an initial assessment of similar institutions based on broad classification categories. The following question can be explored: In terms of broad characteristics, how many institutions are similar to mine? Do not be surprised if you find that even with a limited number of characteristics, there are very few -- if any -- institutions that are similar to your institution. As every institution has always argued, "(w)e are different".

3 . Select a methodology for forming the reference group(s)

There are several primary methodologies used to form a reference group. One option that is always available uses reference predetermined group such as mentioned in Step 2.. Teeter and Brinkman (1987) point out that the major reference groups tend to come from predetermined groups such as institutions in an athletic conference, institutions in a jurisdiction such as a state, and/or traditional groups. If these are viable alternatives, it may be appropriate to identify goals or objectives for performance indicators relative to these groups. For example, an institution may set as a

performance the median of its athletic conference. Work done by the Big 10 is an example of an athletic conference where comparative analysis is viable (Secor, 2002).

Where a predetermined group does not seem to exist, the general procedure often involves judgment, analytics, or some combination. Typically, the judgment builds on the expert opinion of the institution's key stakeholders. This methodology tends to be fairly simple but extremely politically sensitive. If there are different factions involved in the decision process, it can also be quite contentious.

With respect to analytics, one common approach used in higher education is Cluster Analysis, or some form of Cluster Analysis such as Q Factor Analysis. In this methodology, a large group of institutions are defined in terms of a multidimensional space formed from the variables selected in Step 4 - measures which are typically related to the areas of interest determined in Step 1. When using these metrics, they are traditionally converted to some type of standardized measure, after which a composite measure is formed. (For example, at SUNY Albany, Terenzini, Hartmark, Lorang, and Shirley (1980) standardized the variables, created factors, and conducted a Cluster Analysis using the factors. Following Cluster Analysis, they used the factors in conducting a Discriminate Analysis to examine the location of clusters in a multidimensional space.

Obviously, numerous quantitative decisions must be made that have a bearing on the results of this type of analysis. Questions to be answered include: How does one standardize the variables before doing the Cluster Analysis? Does a dollar in salary count more than a dollar in tuition or fees? Should variables reflect magnitude or should they reflect relative magnitude? In other words, are variables based on size such as

number of faculty and number of students, or are variables based on ratios such as students per faculty and average salary per faculty?

Elayne Reiss, Sandra Archer, Robert Armacost, Ying Sun, & Yun (Helen) Fu,(2010) used a methodology employing Cluster Analysis in a systematic sequence to identify comparable institutions. Interestingly, they used data sources beyond the core IPEDS to include Web of Science, (http://apps.isiknowledge.com), NSF (http://www.nsf.gov), the Carnegie Foundation (http://www.carnegiefoundation.org/ classifications/), and *US News and World Report* (http://www.usnews.com/). As noted earlier, this methodology for forming homogenous groups has also found acceptance outside of higher education (Kerschbaum, 2008; Blankmeyer, LeSage,Stutzman, Knox, & Pace, 2010).

In terms of the cluster methodology, there are multiple procedures and these can be based on multiple criteria. In general however there are no definitive rules for the number of clusters that are appropriate.[3]. There is also a discussion of Cluster Analysis options on the SPSS website at http://support.spss.com/productsext/statistics/ documentation/19/clientindex.html . The major advantage of the cluster methodology is that it tends to be more objective than some of the other methods. The disadvantage is that it is rather complicated to explain. With respect to higher education institutions, a conceptual issue is that the institution of interest can be on the outer boundary of a cluster and actually be more similar to those institutions in another cluster.

The form of analysis chosen for this study differs in that it uses the higher education institution of interest as the centroid in the space defined by the variables and

[3] See http://www.statsoft.com/textbook/cluster-analysis/?button=1 for a good discussion of Cluster Analysis

then looks at the distance of other institutions in that to the target institution. As noted earlier, this methodology, sometimes referred to as Nearest Neighbor, has several variations. For example the distance can be measured with various metrics that are typically first standardized and that can be weighted (Weeks & Daron, 2000).

One variation of this form involves selecting institutions for the analysis that have a certain set of characteristics and excluding those without those characteristics. The advantage to this methodology is that it ensures that the institution of interest is at the center of the most similar possible institutions given the variables selected for the analysis. The difficulty with this analysis is that there is no clear number of institutions that should be used in the analysis. Determining an appropriate number of institutions thus requires application of judgment and a continued discussion on the purpose of forming the reference group.

4 Identify measures of interest and targets for outcomes.

The selection of measures and standards should be a function of the purpose for which the institution is benchmarking. This selection is likely the most critical step -- if not the most critical step -- in the process for creating the reference group, especially given that groups are created using metrics considered to be “key attributes”. There needs to be at least general agreement among decision makers about these attributes. If not, there will be strong arguments that the resulting comparison group is not appropriate; as a result, the initiative has a high probability of being stopped before it begins. This is a reflection of the political nature of creating useable reference groups, and the political issues need attention from the beginning of the process.

For purposes of the case study, we will define measures as those aspects of the institution which allow us to identify comparable peer institutions for the purpose of our activity. We will consider standards to be those outputs that are used to actually benchmark performance. In reality the distinction between measures and standards is much less clear than would be defined for this discussion. Some of the measures which are inputs and processes for institutional operations may also be considered as performance measures. For example, if an institution's retention and graduation rates are important in defining institutional context, this does not preclude these same inputs from being used as performance indicators to measure the outcomes of the institution.

A number of tools are available to support the key concept behind identifying the measures for selecting peer institutions. It is necessary to use measures which will ensure sufficient comparability while making it feasible for the institution to achieve standards. For example, if the basic nature of an institution is its urban nature coupled with a focus on graduate education, then these measures – urban and graduate education -- would be essential as part of the institution's description. There are a large number of sources for alternative measures that are appropriate for describing an institution of higher education.

A tool widely used by corporations as a more classic starting points for choosing measures is the balanced scorecard. The traditional balance scorecard is made up of four primary complements that evaluate the institution from different functional perspectives: the customer perspective, the financial perspective, the internal business perspective, and the innovation and learning perspective (Kaplan and Norton, 1996). In translating this tool for use in higher education, it will be necessary to create categories

such as enrollment, finance, academics, and mission-based activities. If this sounds familiar, it is a tool that is already used by many institutions in their most recent strategic plan. Another good source for identifying primary measures is the recent work done by the Carnegie Classification System in developing their new classification system (2005).

One of the aspects of any set of measures is obviously that they must be available across the range of institutions to which comparisons are being developed. While this seems to be intuitive, there are some conditions under which it is not a given that all data be available. For example, recent work done to benchmark institutions in Canada used the US IPEDS data as a core data set, and measures of Canadian institutions were used to estimate the responses they would have made to the IPEDS data set (Xu, 2008). In another example, Pike and Kuh (2005) combine statistical methodology with institutional averages of student engagement They use Q Factor Analysis to derive groups of institutions based on the amount of engagement and types of engagement reported by their students on the NSSE surveys.

Institutions can also decide to collaborate to develop their own data exchange with internal data. This can be a rather limited set of data such as the Consortium for Student Retention Data Exchange (CSRDE) that collects data to describe student retention and graduation. It can also be a large multipurpose initiative such as the Association of American Universities Data Exchange (AAUDE) which is an ongoing initiative, or the NACUBO benchmarking data exchange which is focused on a broad set of operational activities. They provide the benchmarking service for members while referring those wanting to form peer or comparison groups to the various NCES/IPEDS tools.

5. Determine how much "difference" makes a difference.

In the preceding discussions, there have been several discussions about determining the amount of homogeneity, or similarity, which is appropriate. It is also important to come to some agreement about how much "difference" is consistent with the purposes of forming groups. The first discussion should involve identification of those factors that are sufficiently significant such that institutions will not be considered if they do or don't have the characteristic. For example, if an institution has a hospital as part of its organizational structure, it may decide that it only wants to look at itself relative to other institutions that also have a hospital. Another example is doctoral programs. If an institution is primarily focused on undergraduate and Master's level instruction, it may decide to exclude all institutions that are in the Carnegie basic category of Doctoral from its consideration.

A second consideration in determining if the importance of a "difference" is related to the importance of the variable or attribute. Placing weights on factors that are considered more important can be done in most quantitative methodologies. For example, this can be done in Cluster Analysis by including a variable multiple times. If a standardization procedure is used, a variable can be standardized to increase the functional weight. If basing the analysis on the Nearest Neighbor methodology, weights can be used in a similar manner. When one is using a simple metric such as "Same," "Similar," and " Different", a determination decision needs to be made concerning how much "difference" is important and represented by the scale.

An example of weighting to reflect "difference" is the work done by several Canadian institutions (Lang, 2000). These institutions started with basic categories of

enrollment, financial, library, demographic context, and degrees awarded. Within these factors they identified 23 individual aspects of their institutions. Since they were looking at a range of aspects for differing institutions, they developed for different sets of weights for these 23 measures. One set of weights provided a General Slate perspective which is the general perpective.. One set of weights provided a Research Slate perspective. One set of weights comprised a Compensation Slate and one set of weights provided a Government Ability to Pay Slate. For example, FTE enrollment was given a weight of 5% in the Base Slate and in the Government Ability to Pay Slate. In contrast, it was given a weight of 2% in the Research Slate and 0% in the Compensation Slate. These percentages were then multiplied times the standardized differences between various target institutions and the other institutions in the set of institutions under consideration.

At the aggregate level, determining how much “difference” makes a difference requires determining the appropriate homogeneity of the clusters. While there are some standards on the amount of information lost from combining institutions into a group, there does not seem to be any hard and fast rule as to when a cluster is appropriate. This is true both for methods where the overall group is being divided into subgroups -- such as the use of analysis trees -- and for the method where institutions are being added to existing groups. It should be noted that cluster analysis is becoming one of the key topics in Data Mining. (For example, see Han and Kamber, 2006.). Data Mining often suggests computing one of the Maximum Likelihood Information criteria and then running multiple samples of the data looking for a consistent pattern with a Scree Test. (For example see http://www.statsoft.com/textbook/cluster-analysis/ n.d.)

6. Collect and analyze the data.

This step has been mentioned several times throughout the preceding sections. It is obvious that for institutions in the United States, the IPEDS data are the main source of data for forming reference groups, which in turn makes performance benchmarking feasible. It should also be noted that several organizations such as the National Center for Education Statistics (NCES), the Association of Governing Board (AGB), and the National Association of College and University Business Officers (NACUBO) provide various tools for forming reference groups. In addition, many of the websites designed to help students select an institution will have options for selecting characteristics or attributes of institutions that can also be used to select a reference set of institutions. This is also true of the Carnegie Foundations website where one can select institutional characteristics and then view what institutions have those characteristics.

7. Present results and adjust the process.

The results of forming reference groups is an iterative process. It is this iteration that will typically bring judgment to bear on the analysis at all steps. This is a hybrid methodology that merges together analysis and judgment. In fact, if reference groups are being formed for applied purposes such as performance benchmarking, it is highly unlikely that the process will be fully quantitative or linear.

APPLICATION: THE CASE STUDY

The target institution is a southeastern land-grant university with very high research and numerous doctoral programs. Institutions are thus selected which confer

Bachelor, Master, and Doctorate degrees. The goal was to use this university as the reference institution and then to identify based on a distance measure a group of institutions that were similar.

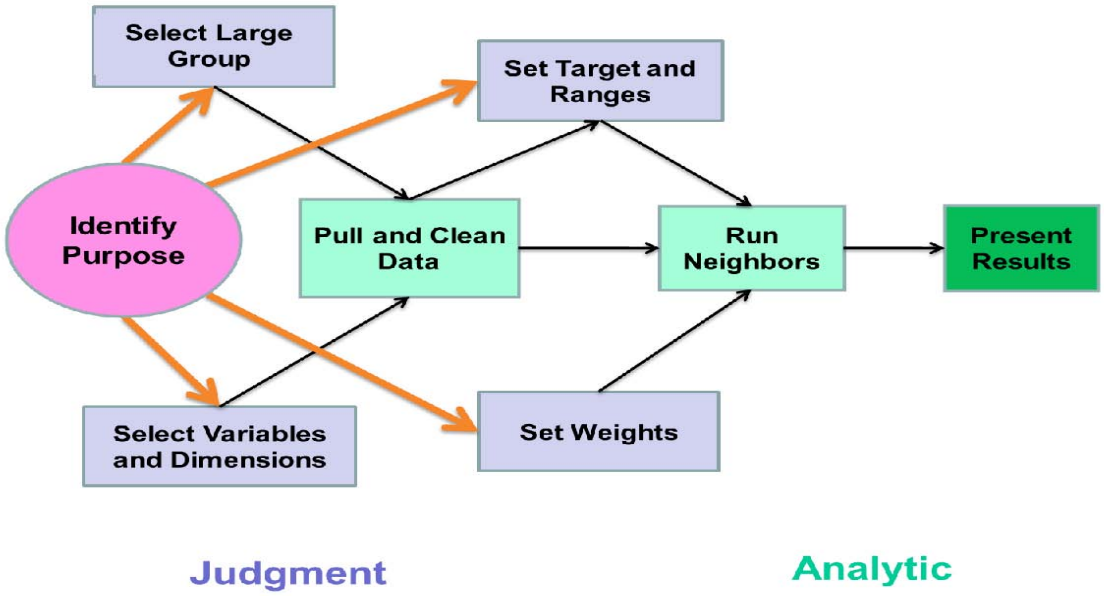
Step one was to clarify the purpose for developing and using the reference group. In this case, there was no specific focused agenda item that required developing a specialized peer, or reference group. However, there are historically multiple initiatives common to this type of institution where having a reference group would be of value. Since the intent was to identify institutions most similar to a target institution, it was desirable to develop a methodology that was general in nature and capability and that was flexible and transparent to potential users. The decision was to develop a reference group that could be used for goal setting through multiple performance benchmarking type activities.

Step two was to determine how many reference groups to form and to determine the size of each group. The methodology that was chosen is flexible and can create multiple reference groups or can identify a single reference group. The size of the group can range from very small to several hundred. Because the intent in this case was to create a group for multiplel uses, the decision was to create relatively large group of manageable size -- in the neighborhood of 25 or 30.

The methodology chosen in step 3 of the process is shown in Figure 1. As can be seen, this methodology represents the Nearest Neighbor methodology discussed earlier where a specific institution is identified as the target institution. In this case a large group of institutions is selected that represents a primary reference group. As noted, the target institution is a southeastern land-grant University with very high

research and numerous doctoral programs. Institutions are thus selected which confer Bachelor, Master, and Doctorate degrees. Note that this excluded institutions that did not offer Doctoral degrees since it was considered highly unlikely that any institution that did not offer Doctoral programs would be accepted as a comparable institution to a major research university. In addition private for-profit institutions were excluded as were institutions outside of the United States and the District of Columbia. Finally institutions were required to be Title IV eligible. This resulted in an initial group of 559 institutions. Eleven institutions were then removed because of excessive missing data.

Figure 1. Methodology for forming reference groups



Measures of interest and targets for outcomes were identified in step 4 through . several iterations of discussions based on two questions concerning: 1) the key areas in

the operation of an institution and 2) the items that are available for these areas. These areas and the items chosen are shown in Table 1

Table 1: Areas and items used to identify neighbors *

1)	**Institutional Characteristics:**
	a) Population Density,
	b) Region,
	c) Carnegie Basic,
	d) Carnegie UG Profile,
	e) Carnegie Enrolment Profile,
	f) Carnegie Size and Setting,
	g) Control,
	h) Hospital.
2)	**UG Market Characteristics:**
	a) FTE Students,
	b) UG Freshmen Applicants/UG HC,
	c) UG (IS) Tuition and Fees,
	d) % Discount Rate (Fees),
	e) % FT-FT DS Accepted,
	f) Yield of FT-FT DS,
	g) Freshman Retention Rates,
	h) 6 Yr Graduation Rates.
3)	**Student Characteristics:**
	a) % White Students,
	b) % UG as Female,
	c) Dorm Capacity as %FT UG,
	d) % UG as Full Time,
	e) %UG Entering in First-Time Full-Time Degree Seeking Cohort,
	f) % FTFTDS Cohort with Pell Grants,
	g) Student Services $/FTE Student,
	h) % UG 25 Years and Older.
4)	**Academic Characteristics:**
	a) IPEDS Student/Faculty Ratio,
	b) % FTE Staff as Faculty,
	c) Research & Service $/FTE Faculty,
	d) % Full Time Faculty as White,

	e) % FT Faculty as Female,
	f) Average Faculty Salary,
	g) % FTE Faculty as Tenure Track,
	h) Instruction and Academic Support $/FTE Student
5)	**Curriculum Characteristics:**
	a) First Prof and PhD's as % Degrees,
	b) Engineering as % Bachelors,
	c) Educ/Leisure/Family Science as % Bachelors,
	d) Other STEM as % Bachelors,
	e) Bus/Pub Admin/Legal/ Communications as % Bachelors,
	f) Applied PhD's as % (First Prof + Doctoral),
	g) Educ/Leisure/Family Science as % Graduate,
	h) Technology and Health Science as % Degrees
6)	**Financial Characteristics**
	a) Net Tuition + State Dependency/Core Revenues,
	b) Tuition and Fee and State Revenue/FTE Student,
	c) Endowment $/FTE Student,
	d) Net Income Ratio,
	e) Financial Viability,
	f) Primary Reserve Ratio,
	g) Return on Net Assets,
	h) % Change in Endowment

* Underscored variables weighted 2 as being more important

As in the Canadian study described by Lang (2000), variables were assigned different weights. While Lang assigned percentages to sum to 100, the importance of individual items were multiplied by differing amounts. Outcome measures were not uniquely differentiated from the variables used in the analysis since many of the variables -- such as Retention Rate -- are both variables of interest and also outcome measures.

Determining how much difference makes a difference was determined in step 5. The following steps were used in computing the proximity of each institution to the target institution based on the 48 items:

1) All items were given a different score of “0”,,”1”, or “2”. Zero indicated that the institution was the **Same** as the target institution on the item. A score of 1 indicated that the institution was **Similar** to the target institution on the item. A score of 2 indicated that the institution was **Different** from the target institution on the item.
2) For Categorical variables, judgment was used to determine the degree to which an institution was the Same, Similar, or Different. Categorical variables included all institutional type variables. For example, in the case of the major research land-grant University the Basic Carnegie category of Very high research/doctoral was considered to be the same, High research/doctoral was considered to be similar, and all other institutional categories were considered to be different.
3) For Continuous items, basic differences were established using the standard deviation of the item. The following definitions were used:

Let Δ = |Target Institution minus Other Institution|, then

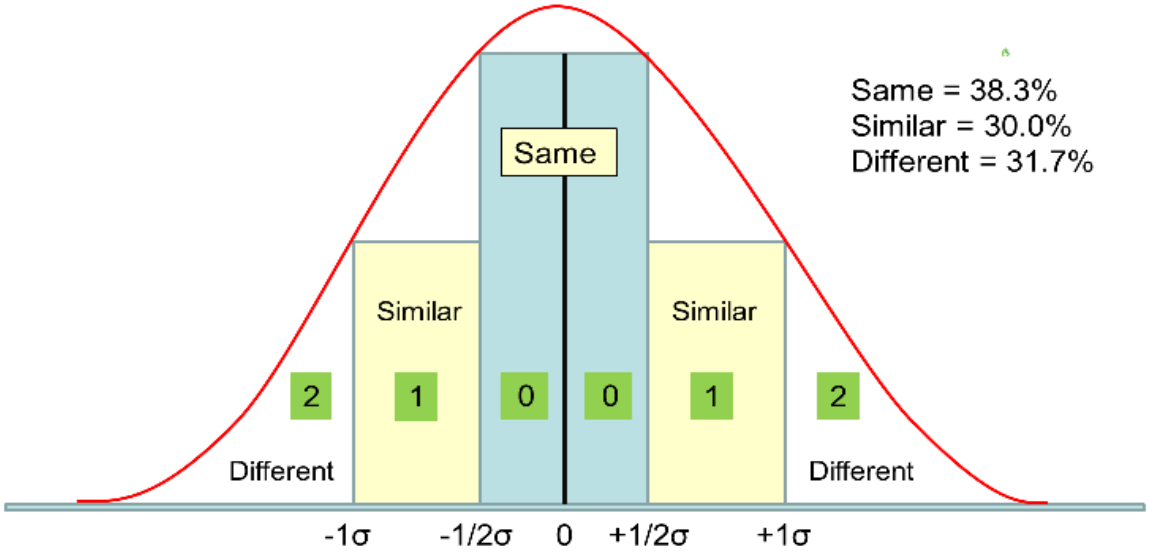
Same = If $\Delta \leq$ ½ Standard Deviation then $X_i = 0$;

Similar = If ½ SD < Δ < 1SD then $X_i = 1$;

Different = if $\Delta \geq$ Standard Deviation then $X_i = 2$.

Some adjustments were made for high levels of skewness where the gaps for difference were reduced. The methodology allows for adjusting either the upper or lower boundaries for similarity. The result of using the standard deviation results in the distribution of scores shown in Figure 2.

Figure 2: Distribution of difference scores for normally distributed items



4) For each institution, the similarity score was weighted and summed across the 48 items. This sum is then divided by the sum of the weights so that each institution gets a Proximity Index which is an aggregate score between zero and two.

Data for Collection and analyse were obtained from the IPEDS Data Center. (http://nces.ed.gov/ipeds/datacenter/). The appropriate .uid and .mvl files were developed and used as appropriate to extract the data. Financial data, programmatic data based on degrees conferred, and general institutional and staffing data were extracted as three different datasets and converted to Excel spreadsheets. After the spreadsheets were sorted in terms of UNITID, they were copied – pasted into a master Excel spreadsheet. This master Excel workbook used formula from various worksheets to create the balanced scorecard where the indices were computed. It then connected these indices to a worksheet one which computed the weighted differences and the

Proximity Indices. This worksheet in turn was connected to one where sorts could be made based on proximities.

In the final step, results are presented and the process is adjusted. In this case analysis, institutions are similar based on their overall proximity and also based on their proximity in terms of the six specific measures used to compute the overall proximity. The following figures show results that were found to be of interest. (See Figures 3-6.)

Figure 3: Proximity of 50 institutions to a Southeastern Land Grant University

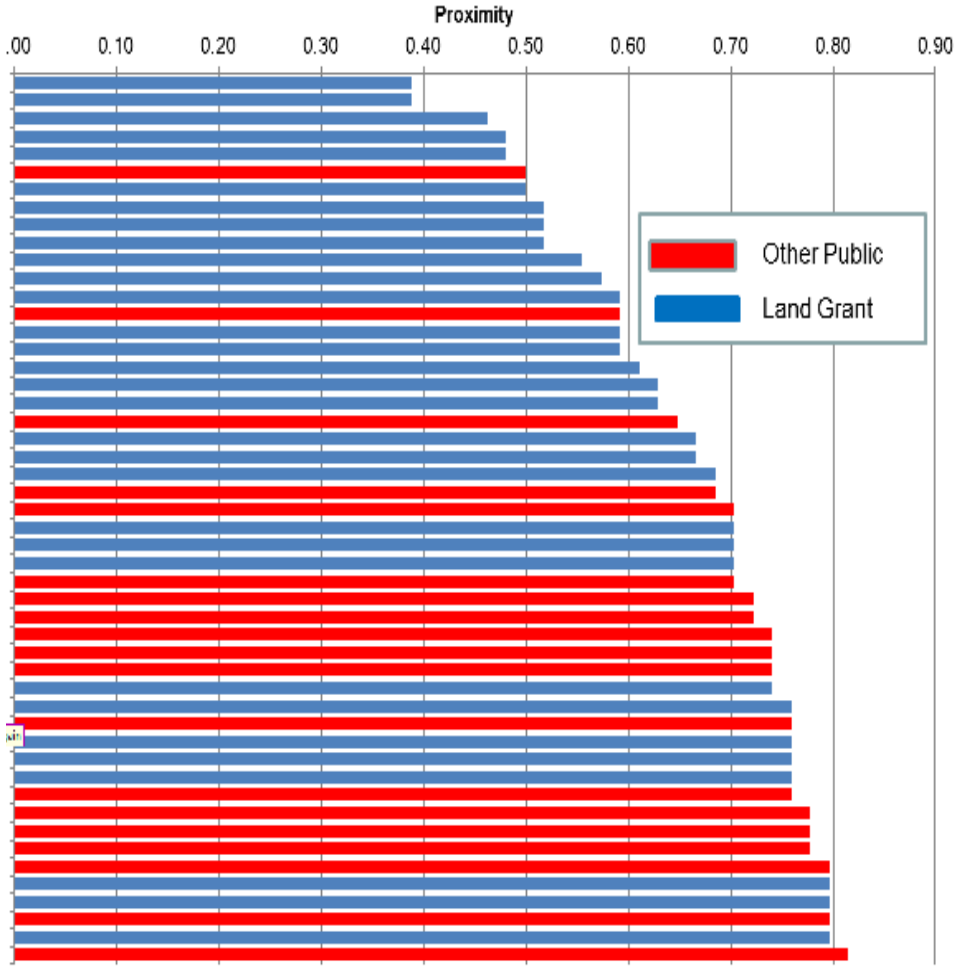


Figure 4: Similarity of three types of institutions to South-eastern Land Grant Research University on 6 Measures

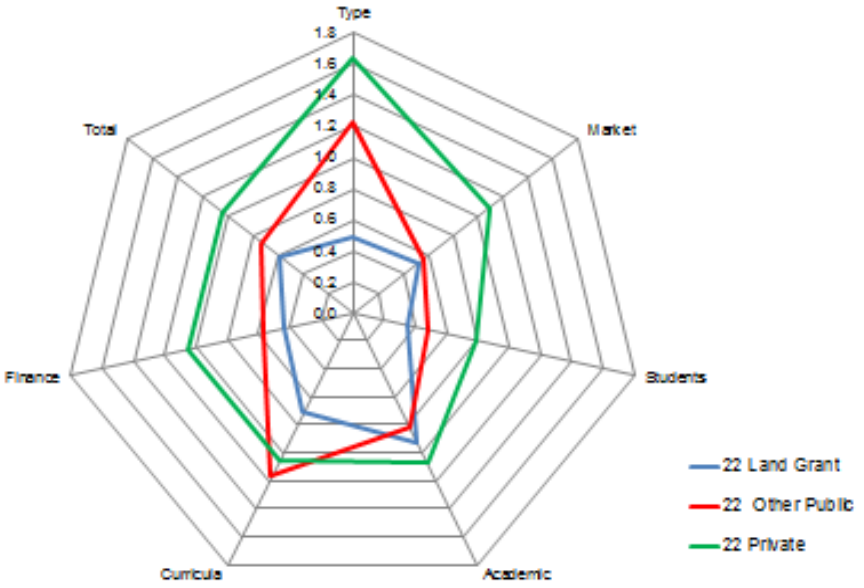


Figure 5: Similarity of three types of institutions to South-eastern Land Grant Research University on Curricula Characteristics

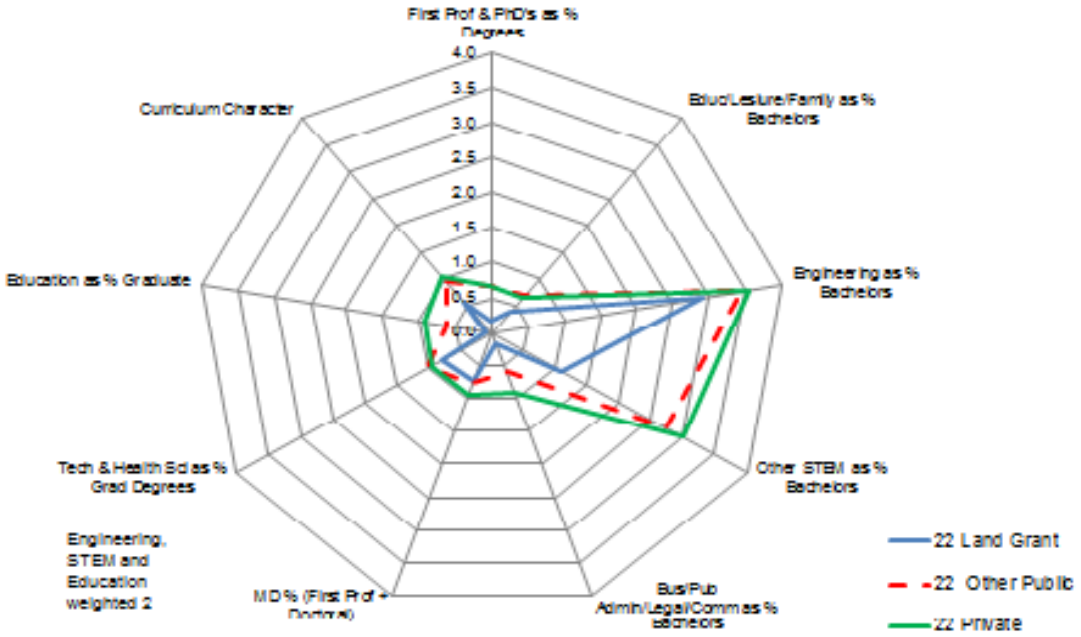
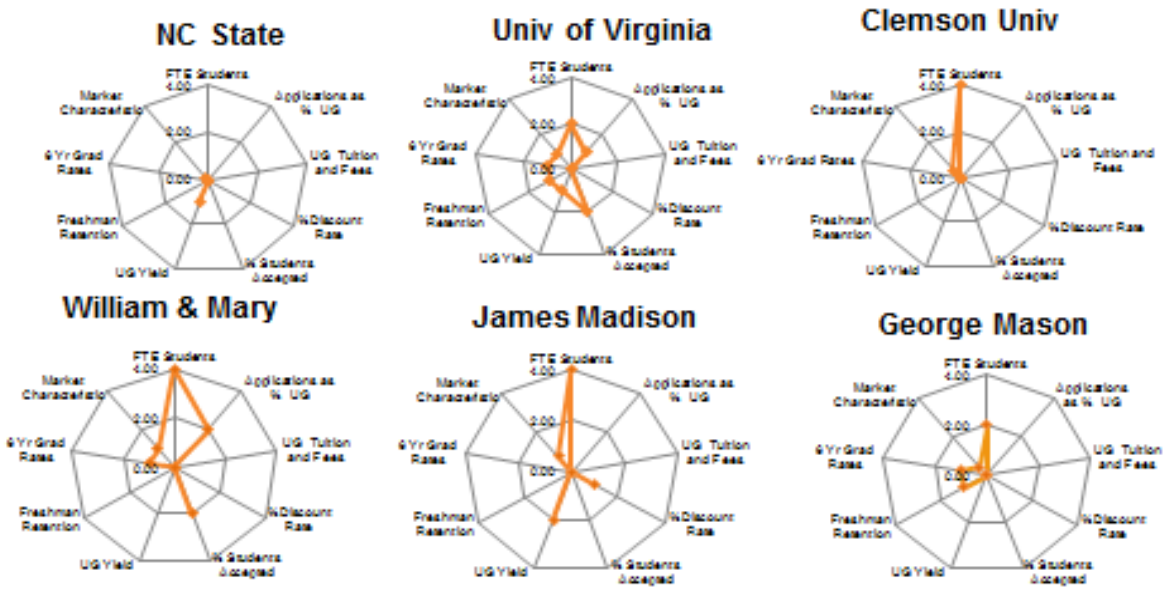


Figure 6: Competitors to South-Eastern Land Grant Research University



Figures 4 through 6 are only a sample of the graphs that can be used to describe the relationship of the focus institution to other institutions. They are based on the proximity measures and show the focus institution as the center of the comparisons. In addition to these, and after a comparison group is identified, it is also helpful to plot the distribution of institutional scores on key metrics relative to the scores of the comparison group.

LESSONS LEARNED

During the past several years, the Nearest Neighbor methodology has been used to create comparison groups for a number of institutions. In general, these institutions were smaller, private liberal arts colleges and were requesting comparison groups that would allow them to do institution-wide assessment and evaluations. Institutional concerns varied but in general revolved around curricular issues, endowment and tuition

rates and faculty salary questions. The following reflect important lessons that were learned or reinforced from this case study:

The Role of Stakeholders. Conversations with the senior stakeholder to set the parameters of the process were important. From these conversations, the stakeholder understood that they had an active role in the selection of the comparison institutions and that it was not simply an analytic process, but one that required their judgment and input to be successful.

Data Consistence. IPEDS data format, structure and definition tend to change from year to year. This requires that once the data are down loaded, (particularly if you are using programs that were used in the past) they be reviewed to make sure that you got what you thought you were getting.

Spreadsheet Complexity. The spreadsheets which are the output of the analytic process are large and complex. Reviewing the outcomes with the stakeholder was much easier if the stakeholder had a working knowledge of Excel. When this was not the case, someone from the campus with that knowledge needed to be present for the conversation. In the ensuing conversation. the flexibility of the model is demonstrated by asking the stakeholder to do the manipulations. We found this to be a critical step as it gave them an understanding of how the model worked and how to customize it for their particular institution by setting the weights for each of the measures. Going through this process also gave them a greater appreciation for how the model could be used and more confidence in the appropriateness of the resulting comparison group. Giving the stakeholder the capacity to test different scenarios and to work with other campus leaders enhanced the model's use.

SUMMARY

The preceding discussion desribes a basic Nearest Neighbor methodology for forming reference groups and building comparisons for institutional benchmarking. It has also demonstrated this methodology based on using publicly-available IPEDS data for a major land-grant southeastern research University.

There are two important points that follow from this case study. First, in today's higher education environment, institutions are not faced with the choice of having reference groups but are faced with the choice of how they want to develop their reference groups. Institutional reference groups are being provided to the public through numerous mechanisms. Vanity ratings and the popular press use various criteria to group institutions with each other. The federal government is also grouping institutions through its college navigator. (http://nces.ed.gov/collegenavigator/). Education Trust is comparing institutions (http://www.collegeresults.org/), as are -- in the broader sense -- the Association of Governing Boards (http://agb.org/benchmarking-service), The Institute for College Access and Success (http://ticas.org/), NACUBO (http://www.nacubo.org/Research/NACUBO_Benchmarking_Tool.html), IPEDS, and *the Chronicle of Education* (http://chronicle.com/article/2011-Salary-Explorer/126972/). All arel providing mechanisms to facilitate comparisons between the target institution and other institutions. Therefore the question is not: *Will you be compared?* But: *To whom and how will you be compared?*

Does it make sense to build a reference group consistent with your decision making needs? Does it make sense to use a methodology such as the one described above that provides both the quantitative objectivity of national databases and also the

judgmental expertise of key stakeholders? If the answer to these last two questions is yes, we seriously encourage that the methodology discussed above be considered for use.

Second, based on a qualitative rather than a quantitative insight, there is no better way to conclude this discussion then to share the insights that came from similar initiatives in Oregon (Weeks, Puckett, and Daron, 2000, p 20)

> In a dynamic political environment, the analysis applied to as sensitive and issue as peer comparisons must necessarily reflect the adjustments and compromises that are part of the political process. In return, decision making that draws from sound analysis is more likely to avoid the manipulations of the purely political process. Building a relationship of centralized analysis and decentralized decision-making requires trust compromise on both sides, but the result is more likely to be long-lasting.

REFERENCES

Blankmeyer, E., LeSage, J. P., Stutzman, J. R., Knox, K. J., and Pace, R. K.. "Peer-Group Dependence in Salary Benchmarking: A Statistical Model", *Managerial and Decision Economics, 32*: 91–104 (2011) OR 2010.

Brinkman, Paul T. "Effective Institutional Comparisons." *New Directions for Institutional Research, 53* (1987).

Brinkman, P. T. and Teeter, D. J.. "Methods for Selecting Comparison Groups." *New Directions for Institutional Research, 53* (1987).

Cherni, Sofiya, "Nearest Neighbor Method." Department of Mathematics and Computer Science, South Dakota School of Mines and Technology. Rapid, City, SD. May 9, 2005. Retrieved 5/15/2005, from http://www.mcs.sdsmt.edu/rwjohnso/html/sofiya.pdf

Chronicle of Higher Education, Salary Explorer, http://chronicle.com/article/2011-Salary-Explorer/126972/ n.d.

Clark, P. and Evans, F., "Distance to nearest neighbor as a measure of spatial relationships in populations," *Ecology, 35*: 445-453, 1954. In Cherni, S. 2005.

Diggle, Peter J., *Statistical Analysis of Spatial Point Patterns,* New York: Oxford University Press Inc., 2003.

Education Trust, College Results On-Line, http://www.collegeresults.org/ n.d.

Fraley, Chris and Raftery, Adrian E., "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," The Computer Journal, 41-8, 1998.

Gaither, Gerald, Brian P. Nedweek, and John E. Neal. (1994), "Measuring Up: The Promises and Pitfalls of Performance Indicators in Higher Education," ASHEERIC Higher Education Report Series, 95-5(23-5).

Han, Jiawei and Micheline Kamber. (2006), *Data Mining*, (2nd Edition), Elsevier, San Francisco CA.

Kaplan, Robert S. and David P Norton. *The Balanced Scorecard*. Boston: Harvard Business School Press, 1996.

Kerschbaum, Florian, "Building A Privacy-Preserving Benchmarking Enterprise System ," *Enterprise Information Systems , 00,(00)*, February 2008, 1–15,

Lang, D. W. (2000). Similarities and Differences: Measuring Diversity and Selecting Peers in Higher Education. *Higher Education, 39*: 93–129.

McLaughlin, G.W. and McLaughlin, J.S. *The Information Mosaic: Strategic Decision Making for Universities and Colleges*, AGB, 2007, Washington DC, Chapters 6 and 7

NCES, College Navigator, http://nces.ed.gov/collegenavigator/, n.d.

Pike, G. R. and Kuh, G. D., "A Typology of Student Engagement for American Colleges and Universities," *Research in Higher Education, . 46*,(2), March 2005, pp. 185-209

Punj, Girish, and Stewart, David W., "Cluster Analysis in Marketing Research: Review and Suggestions for Application," *Journal of Marketing Research, XX*, 134-48, May 1983.

Reiss, E. , Archer, S., Armacost, R. , Sun, Ying, and Fu, Yun (Helen), "Using SAS® PROC CLUSTER to Determine University Benchmarking Peers," SESUG,

Savannah GA, September, 2010. (http://uaps.ucf.edu/doc/SESUG_Benchmarking_2010.pdf)

Rodríguez-Déniz, Héctor, and Voltes-Dorta , Augusto, "Hierarchical Clustering For Airport Benchmarking," (http://www.sauder.ubc.ca/Faculty/Research_Centres/Centre_for_Transportation_Studies/~/media/Files/Faculty%20Research/OPLOG%20Division/OPLOG%20Publications/Rodriguez-Deniz_Voltes-Dorta_Cluster_WP.ashx), n.d.

Romesburg, H. Charles, *Cluster Analysis for Researchers*, Lulu Press. 2004.

Secor, R, "Penn State Joins the Big Ten and Learns to Benchmark," *New Directions for Higher Education 118*, San Francisco: Jossey Bass, 2002) 65 – 77.

Skellam, J. G. "Studies in Statistical Ecology. I. Spatial Pattern," *Biometrica*, 39, 346-362, 1952.

Teeter, D. J. and Christal, M. E., "Establishing Peer Groups: A Comparison of Methodologies." *Planning for Higher Education, 15* (2) (1987).

Teeter, D. J. and Brinkman, P. T., "Peer Institutional Studies//institutional Comparisons," *Primer for Institutional Research*, J. Muffo and G. McLaughlin, eds., (Tallahassee: Association for Institutional Research, 1987, 89 – 100.

Teeter D. J. and Brinkman, P. T. "Peer Institutions," *Primer for Institutional Research*, MA Whiteley, JD Porter, and RH Fenske, eds., (Tallahassee: Association for Institutional Research, 1992), 63-72).

Townsley, M. T. (2002). *The small college guide to financial health: Beating the odds*. National Association of College and University Business Officers.

Trainer, J. F. The Role of Institutional Research in Conducting Comparative Analysis of Peers, Chapter 2, *New Directions For Higher Education, no. 141*, Spring 2008.

Weeks, S.F. Puckett, D.and Daron, R,."Developing Peer Groups for the Oregon University System: From Politics to Analysis (and Back)," *Research in Higher Education, 41,* (1), 2000.1-20.

Xu, J. , "Using the IPEDS Peer Analysis System in Peer Group Selection," *The AIR Professional File,* Number 110, Winter 2008.