Peer Grouping: The Refinement of Performance Indicators

Willard Hom
Chancellor's Office, California Community Colleges.

Community colleges operate under much scrutiny these days, and these institutions have experienced a growing emphasis on performance indicators as paths to institutional accountability. California's system of 109 community colleges recently developed and implemented an innovative accountability program that used peer group comparison as one of its elements. This article describes California's use of peer grouping in terms of its development, mechanics, and implications for the future.

Introduction

For purposes of evaluation and planning, higher education institutions have often tried to compare themselves to other institutions on selected performance indicators. This common practice gives college officials some idea of "where their institutions stand" with respect to a certain performance indicator. But college officials generally recognize that such a comparison is only valid and fair if it involves colleges with similar characteristics. In fact, the Integrated Postsecondary Education Data System (IPEDS) tries to address this historic need by integrating a "peer selection" procedure on its website (US Department of Education, 2007). The Carnegie Foundation has provided a framework for identifying peer institutions since 1970 (Carnegie, 2007; McCormick & Zhao, 2005). A study from the National Center for Educational Statistics (NCES) actually identified groups of similar colleges through a cluster analysis of various postsecondary institutions in the U.S. (Phipps, et al., 2001) and a new analysis extends that work (Goan & Cunningham, 2007). The National Survey of Student Engagement (NSSE) system has attempted to help an institution to find its institutional peers in its data system (National Survey of Student Engagement, 2007). Finally, one publisher even devoted an entire volume of its series to the topic of peer identification (McCormick & Cox, 2003). Hurley (2002) documents the various efforts that have focused upon finding peer institutions for community colleges.

Postsecondary educational institutions have company when it comes to this need to compare themselves for evaluative purposes. California's secondary education system essentially administers its accountability program, with its Academic Performance Index (API), in such a way that each school's performance has a set of similar schools (its "peers") to facilitate a fair comparison (California Department of Education, 2007).

In any case, to determine which institution is a "similar school" in the postsecondary environment, an analyst can undertake an extensive analysis to categorize or classify a specific population into groups so that each group contains institutions that closely resemble one another. Researchers often label such an analysis as peer grouping. Researchers use peer grouping to "level the playing field" or to avoid the problem of "comparing apples to oranges." In essence, peer grouping is a method of arranging data so that people can compare institutions while "controlling" for specific, measured factors (often referred to as "confounders") that would otherwise lead to unfair and invalid comparisons of performance.

How California Constructed Its Peer Grouping

California used peer grouping to satisfy an assortment of legislative, political, and technical criteria. California's central office for the state's 109 community colleges (henceforth known as the System Office) had responsibility for implementing state-level mandates that affect these institutions. The System Office therefore had authority to plan and implement a peer grouping plan that the state legislature had passed in 2004 (Chancellor's Office, 2007).

Historically, state oversight bodies (such as the California Department of Finance and the Legislative Analyst's Office) have preferred straightforward forms of evaluation such as institutional rankings. Rankings can give analysts quick, if simple, insights into relative performances of institutions. When the rankings involve a rate or proportion for each institution, this form of performance reporting provides busy oversight bodies with the most fa-
miliar and useful type of policy-making information. On
the other hand, college officials have argued that their
missions are so complex that any single numeric perfor-
mance indicator will fail to give a valid and fair portrayal
of their institutional performance. These officials hold that
performance indicators must account for the different en-
vvironments of each institution (i.e., the kind of students
enrolled and their proximity to receiving transfer institu-
tions, et al.), the limited accuracy and coverage of admin-
istrative data, and unique historical events (i.e., natural
disasters and relevant policy changes, et al.) before policy
makers and local voters can take these indicators seriously.
At the same time, most college officials face a dilemma
with two major elements. They have scarce resources with
which to develop the data and analysis systems to fulfill
the aforesaid requirements of performance indicators, and
they tend to lack the background in evaluation research to
envision a methodological solution that can satisfy the
public’s demand for performance rankings. In this situa-
tion, the System Office for the community colleges devel-
oped a peer grouping approach that largely satisfied the
needs and interests of these two political groups: over-
sight bodies and college officials. To a large extent, the
peer grouping approach also served the mission of the
System Office. It found a “middle ground” that allowed
the oversight agencies to understand community college
performance in new and objective ways while capturing
much of the diversity in institutional environments and
institutional missions. This compromise (a) promoted the
chances of improved funding from the Governor’s Office
and the Legislature and (b) fortified the System Office’s
role as a problem-solving go-between in the community
college scene—two strategic goals for the state agency re-
sponsible for helping to manage the state’s higher educa-
tion system.

The state’s current effort at accountability, known as
Accountability Reporting for the Community Colleges
(ARCC), emerged from the demise of a program known as the Partnership
for Excellence (PFE). Beginning in 1998, the PFE program acted as a first
step toward accountability in the state system, but oversight bodies
showed dissatisfaction with its limited analyses (Legislative Analyst’s
Office, 2007). Although the state leadership chose to abandon PFE, much
of the basic data collection mecha-
nisms and concepts for performance
indicators eventually served as foun-
dations for the ARCC performance
indicators.

To identify the members of a
particular peer group, the researcher
can apply a variety of classification methods but a statisti-
cal method known as cluster analysis has often been the
tool of choice (Everitt, et al., 2001; Lorr, 1987). In disci-
plines such as business, medicine/public health, psychol-
ogy, and political science, researchers have historically
applied cluster analysis to obtain peer groupings. Hurley
(2002) discusses how cluster analysis, among other meth-
ods, has been employed in various community college peer
grouping efforts.

Cluster analysis has a number of advantages and
disadvantages that the System Office considered before it
chose this method as its tool for classifying the colleges
into groups for comparison. One critical advantage that
this quantitative tool had over subjective tools (such as a
jury of expert opinion for a reputation-based classifica-
tion) is the capacity of cluster analysis to prevent politi-
cally biased peer grouping that would favor certain col-
leges. Table 1 displays the advantages and disadvantages
of this method for the System Office. The advantages and
disadvantages in Table 1 are really general considerations.
Cluster analysis has a decisive technical advantage over
other quantitative classification methods or data mining
tools in that the user can proceed despite the lack of any
information about the “true group membership” for each
entity in the population under study (Han & Kamber,
2001). In the case of California’s community colleges, there
is no way to identify the true group membership of each
institution.

The most notable technical disadvantage of cluster
analysis is that “it is difficult to tell whether a cluster
analysis has been successful. Contrast this with predic-
tive modeling, in which we can take a test data set and see
how accurately the value of the target variable is predicted
in this set. Generally speaking, the validity of a clustering
is often in the eye of the beholder; for example, if a cluster
produces an interesting scientific insight, we can judge it
to be useful. Quantifying this precisely is difficult, if not

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Familiarity among the institutional researchers at the colleges and districts</td>
<td>Complex and difficult to understand for the lay audience</td>
</tr>
<tr>
<td>2 Wide availability in standard statistical program, promoting opportunities for replication</td>
<td>Potential for method bias</td>
</tr>
<tr>
<td>3 Acceptance as an effective tool in many disciplines of research</td>
<td>Inherently fuzzy in nature while seeming precise</td>
</tr>
<tr>
<td>4 Flexible in handling user requirements and data</td>
<td>Potential for counter-intuitive results</td>
</tr>
<tr>
<td>5 Relatively easy to learn for most analysts who have not already had exposure to the method</td>
<td>Some difficulty in demonstrating the validity of its results</td>
</tr>
<tr>
<td>6 Sufficient capacity to define clusters with any number of user-specified variables</td>
<td>Somewhat esoteric graphic displays for results</td>
</tr>
</tbody>
</table>
impossible, since the interpretation of how interesting a clustering is will inevitably be application-dependent and subjective to some degree…” (Hand, Mannila, & Smyth, 2001).

The recent experience with peer grouping in ARCC may offer the analyst in other states some guidance. One major development in California’s peer grouping project was the decision to give each community college a separate peer group for each performance indicator. As a result, the 2007 ARCC report produced a total of six peer groupings (one for each of the six performance indicators that had sufficient data in 2007). This policy recognized the reality of the multiple missions that the public expects its colleges to perform. Research has shown that the system’s community colleges do tend to specialize to some degree in the missions that they strive to achieve, making it misleading and unfair to evaluate such institutions on the basis of a single performance indicator (Gill & Leigh, 2004). Bailey (2003) pointed out, “Community college administrators and faculty articulate several central college missions, including granting degrees, transfer, workforce development, worker upgrading, and remediation…No strong consensus has emerged that identifies one or two central purposes of the community college…”

Development of the peer groups for each indicator went through specific steps (variable selection, regression modeling, and cluster analysis). Figure 1 shows the work flow that the ARCC project used for peer grouping through cluster analysis. The steps of literature review, field input, and review of COMIS (the System Office’s database of community college data) functioned as a variable search-definition process. These steps occurred simultaneously in order to meet the schedule of the oversight agencies. The System Office planned for these steps to conclude before its analysis moved to the next step (the search for correlation), and each step could have used a different analyst operating independently of other analysts if needed.

In the search for the correlation step, the System Office executed a number of sub-steps. In the first sub-step, there was an exploratory data analysis or EDA (Mosteller & Tukey 1977). The EDA helped the System Office to decide which potential peer grouping variables had inadequate data quality to contribute to the prediction of a performance indicator. The prevalence of missing values and unrealistic distributions of values helped to establish the usability of specific variables. The extent of missing values had a decisive weight in the usability of a variable. If there were too many missing values in a variable, the peer group analysis either had to exclude some colleges from the peer grouping effort or use imputed values in place of the missing values. Omission of a college from the peer grouping would amount to excusing that college from the project’s form of accountability for a particular performance indicator. On the other hand, imputation of missing values would have created a debate about the validity of a college’s peer group membership. The EDA sub-step also helped to determine the need for possible data transformations of variables before testing for linear correlation (the next step).

The second sub-step, the actual search for correlation, tested for a bivariate relationship between each peer grouping variable and each performance indicator. This sub-step used scatter plots and correlation coefficients to judge association. For a robust test for bivariate correlation, the analysis used Spearman rank correlation coefficients, as well as the standard Pearson correlation coefficients. These tools helped staff to decide which data elements or variables to drop from further analysis. This screening of variables enabled the peer grouping effort to concentrate on a parsimonious set of variables. The bivariate correlations also helped the next step, the regression model, by helping to identify potential instances of multicollinearity.

The next step in Figure 1 used ordinary least squares regression to find the best set of variables that predicted a performance indicator. This step aimed at identifying those variables that had a statistical relationship to a performance indicator and a theoretical connection from a causal perspective. Consequently, staff identified the fewest number of variables and the most “valid” variables for the ensuing step of cluster analysis. This is important for two reasons. (1) The analyst should only attempt to group colleges according to environmental factors that have an empirical and a theoretical relationship with a performance indicator, and (2) the cluster analysis will have more clarity (or transparency) if it uses the fewest number of variables possible. In the regression modeling, staff

![Figure 1. Steps in Formulation of Peer Groups](image-url)
employed a standard regression analysis procedure that has been detailed elsewhere (Chancellor’s Office, 2007). When staff completed this step, it had a small number of variables for use in the next step, cluster analysis.

In the cluster analysis step, staff needed to weigh the specific criteria that applied to California’s accountability program. First of all, the System Office designed the peer grouping so that a college could belong to a different peer group for a different performance indicator as a means to achieve valid comparisons. So, the number of cluster analyses that the System Office conducted equaled the number of performance indicators that it was mandated to analyze. The state’s oversight agencies sought comparison colleges for each college in the system. If a college had no comparison colleges, that is, if it had a peer group containing only itself, then that college essentially escaped accountability for its result on a specific performance indicator. So, the cluster analysis needed to minimize the occurrence of one-member peer groups. (To a lesser extent, the cluster analysis also needed to have few, if any, peer groups with only two or three members in them.) In some respects, the criterion for a minimum of small peer groups forced the System Office to create some peer groups that joined some colleges that would have been classified as different (i.e., relatively dissimilar) enough to appear as “outliers” in the system. The need to avoid too many instances of one-college peer groups related to another criterion for the peer grouping. The peer groups needed to divide the population of colleges so that the size of the groups did not become too large. That is, the System Office needed to avoid the creation of a peer group that contained a disproportionate number of the colleges. This criterion implements the notion that equal-sized groups would provide more equitable comparison situations than peer groups that had dramatically different numbers of colleges within them. In order to achieve these criteria, the System Office applied the Ward’s method of clustering along with a target number of six clusters (so that approximately six groups of eighteen colleges per group would result). The System Office clustering also used two other options in the cluster analysis for each peer grouping. Staff standardized each clustering variable (i.e., each predictor from the regression model for a performance indicator) so that different scales of measurement would have no effect on the clustering. Secondly, staff chose the hierarchical clustering method so that the process of the group formation could be examined for sensitivity. Standard references on cluster analysis explain these clustering options. (Everitt, et al., 2001; Hair & Black, 2000; Lorr, 1987; Aldenderfer and Blashfield, 1984).

How Peer Groups Were Used

The preceding section covered the process that led to the construction of the peer groups. In this section the applications of the peer groupings are summarized. These applications largely embody the political and administrative issues that precipitated the mandate for the ARCC. Figure 2 below outlines the major elements involved in the application of the peer groups. The diagram basically displays the time sequence of the elements. After the System Office completed its peer groupings, colleges received a draft report containing these results as well as a wide array of related performance data. Each college used this draft information for its different performance indicators (including how these results compared to the results of its peer groups) to write a “self-assessment.” This self-assessment provided the college’s perspective on the reported results—essentially creating a formal way for colleges to offer explanations or responses to findings in the ARCC report.

Each college then submitted its self-assessment to the System Office, which incorporated these brief narratives into the final ARCC report. The final ARCC report then became available to the public through the System Office website, satisfying the legislation that mandated the ARCC project. However, release of the public report did not signify the last step in the ARCC process (and application of the peer grouping information).

The bottom three boxes of Figure 2 denote the three major actors who used the peer grouping information in the public report. The elected Board of Trustees for each community college district had to “interact” with the ARCC report and the peer grouping information. Within one year of the report’s release, each district had to submit docu-
ment of board interaction to the System Office. Through this process, local elected officials were expected to hold local college officials accountable for their performance. The peer groupings aided in board evaluations of college performance because they provided trustees a direct comparison of the local college to other colleges with similar challenges for a particular performance indicator—a tool that they hitherto had lacked. State oversight agencies applied the peer grouping information in the public report to their policy formulation and budget decisions. Although state oversight agencies did not use the ARCC report to attempt specific interventions within one or more districts, they did employ the report’s peer group information as part of a general diagnosis of the system’s performance. Last, but not least, the public media and press used the peer group information in the public report to alert the community to the shortcomings or strengths of their local institutions. The legislation that required the peer grouping does not mention the public media and press, but there is obvious value in helping communities to receive information that may help them to initiate local change (such as change in college policy or college personnel or local bond measures). In a sense, the public media/press acted here as part of a local accountability mechanism without direction (or funding) from the state government. Although Figure 2 omits a block at the bottom level for college officials (i.e., college presidents, vice-chancellors, and deans), we note that this group of actors began their application of the peer grouping in an earlier stage, the self-assessment stage, and this group will probably continue to apply that information throughout the year as college-based planning evolves.

Discussion

California’s use of peer grouping had innovative qualities in both its methodological dimension and its administrative dimension. The use of a regression model to identify environmental factors for each performance indicator, in concert with a cluster analysis that used those environmental factors to define a peer group for each performance indicator, marks a milestone in community college accountability analysis. The integration of state-level indicators with college-level indicators, the use of four different perspectives of overall community college performance, and the breadth of the indicators also helped to distinguish California’s implementation of peer grouping (Chancellor’s Office, 2007). The integration of a technical advisory group, a college self-assessment, a web-based public report, the dissemination of peer grouping data and methods, and the requirement for local board interaction all set examples for future peer group applications in other states. The dissemination of the peer group data and methods, along with opportunities to submit corrected data, reinforced the System Office’s efforts to ensure a valid and equitable accountability process. A number of colleges actually used the System Office-supplied “peer group toolbox” to replicate the peer group results that appeared in the public report.

California’s early experience with peer grouping demonstrates the dynamics of accountability that can evolve from the inherently political nature of performance measurement and reporting. Peer grouping also creates a dilemma for college officials who must interact with different types of stakeholders. For an audience of accountability officials who focus upon the relative performance within a peer group, the names of the peer group colleges have fairly low relevance. For this audience, a college would probably prefer to have peer group members who have relatively weak performance records. On the other hand, a college will tend to prefer high-performing (i.e., high reputation) colleges as peer group partners if it needs to impress its local community especially the voting public, employers of its alumni, and potential students.

Analysts who face a “high stakes” accountability situation, inadequate data, insufficient staffing, or strong opposition to performance reporting may expect more difficulties in trying to use California’s approach to peer grouping. The peer grouping implementation in California enjoyed a conducive environment. The legislation for ARCC did not tie the results to funding or direct repercussions such as heightened restrictions and authorized special staffing to implement the new accountability project. California has progressed far in its development of a comprehensive data system. Furthermore, the initial legislation allowed for time and a panel of experts to plan a system of accountability. Although various officials at the colleges voiced opposition to the accountability initiative (and to the peer grouping), this opposition did not expand into an organized effort of non-compliance or for counteractive legislation. Probably the most frequent hurdle in the peer grouping process was the difficulty that some college officials had in accepting peer groupings that associated their colleges with institutions that they had not traditionally viewed as peers. This difficulty appeared to occur most often with respect to peer colleges that had different geographic characteristics, different overall enrollment characteristics, or different public images.

In the planning of the ARCC peer grouping, the System Office had concerns about the ability of college administrators (i.e., the chief executive officers, presidents, et al.) to understand the peer grouping idea and the cluster analysis method. The idea and statistics behind it could conceivably have alienated executives because of its “black box” quality. As a remedy for this potential problem, the System Office worked with institutional researchers at the colleges to emphasize and facilitate their role in explaining these technical points to executives, hopefully allay-
ing their suspicion, discomfort and opposition regarding the peer grouping approach. Because a large proportion of the institutional researchers at the colleges had some exposure beforehand (i.e., graduate school coursework) with cluster analysis, their technical assistance (counseling) at the local level apparently diminished the potential problems that the esoteric nature of this approach could have generated. In fact, institutional researchers often played the role of presenters or co-presenters of the ARCC data before local trustees, and they often had to explain the cluster analysis to these elected officials.

It is noteworthy that the ARCC project started in 2004 with a proposal for accountability that a national panel of experts helped to produce (Chancellor’s Office, 2007). Although the panel of experts had no involvement in the development of the peer grouping, the set of performance indicators in the ARCC clearly benefited from their input and support at the outset. Their comments and endorsement of the basic plan lent critical credibility to the accountability plan that eventually included a peer grouping component. In addition, the System Office worked to put everyone’s focus on all of the report’s data rather than on the unique comparisons created with the peer grouping. As one would expect, the news media tended to focus upon the peer group results because of its comparative aspects. Because of a unique set of circumstances, college officials actually publicly promoted peer group results and the ARCC rather than trying to let the ARCC occur with minimal public awareness. These elements of expert input from around the nation and the unique press/media environment that occurred in 2007 may make the California experience somewhat special.

Conclusion

California’s use of peer grouping may serve as a useful “test case” in community college accountability across the nation. When state legislatures seek to closely examine the performance of their community college systems, they may refer to California’s experience with peer grouping as a possible model for their own version of college accountability. Nevertheless, as time unfolds, the long-term benefits (and costs) of California’s peer grouping, along with its own unique political and social climate, will determine how, and if, peer grouping will continue as a tool in California’s program for accountability, ARCC. If peer grouping in California were to fall into disfavor to the extent that an impasse over its use occurs, then some form of ranking could possibly replace peer grouping. Naturally, the likelihood that any ranking system could supplant the peer grouping approach would depend upon how legislators and state officials view the feasibility and public appeal of a ranking system (or any other alternative method for performance evaluation for that matter). An abandonment of peer grouping may not occur at all, of course, if the stakeholders in the system (local voters, state oversight bodies, and college officials) can see a net benefit from the peer grouping and both validity and fairness in this approach.

It seems unlikely that college officials will advocate an end to peer grouping. Peer grouping is essentially a substitute for hierarchical ranking of colleges. The latter option has little support among college officials (Phillipe and Boggs, 2003; McClenney, 2007), while classification systems (like peer grouping) do have support. Technically speaking, the peer grouping system in California cannot transform into a ranking system either. Because the cluster analysis produces “clusters” of colleges that constitute a so-called categorical (or nominal) variable, no one can sort the clusters to create a state-wide ranking of the colleges. Any one of the six clusters in an ARCC peer grouping has no logical “position” relative to the other five clusters; the clusters (and thus the colleges) cannot be sorted into a meaningful sequence outside of their peer group. Therefore, the use of cluster analysis for peer grouping does not create a “path” to a hierarchical ranking of the colleges, a quality that would tend to maintain the support of college officials.

In terms of parties that want to consider the adoption of a peer grouping system, we note the following point from Bailey (2003): “…Although each of the classification systems presented in this volume can be used to generate interesting and important insights, classification systems need to be much more systematically defined, and the link between those goals and the measures proposed need to be clearly articulated…. But since different goals will in all likelihood be best served by different schemes, it makes sense either to propose a variety of systems or to provide easily available data that can be used to generate many systems. If we are looking for a system that can be used to benchmark practices, it will be different from one used to distribute money or to help select a sample of community colleges for a research project…. The California peer grouping effort implemented the concept of peer groupings for specific purposes, but the mechanism that it developed for a general accountability plan (i.e., reporting comparisons of performance indicators) may not work for a state that has a different purpose for its peer grouping (i.e., funding or research).

A final thought relates to the capacity of community colleges and their state (central) offices. The use of cluster analysis linked to multiple regression for an official peer grouping effort mandated by the legislature can succeed despite the complexity of this concept. That is, community college systems can succeed at implementing a relatively sophisticated performance measurement system even though senior-level officials may have little technical expertise in the selected methodology. Moreover, a critical
element of success for a sophisticated performance measurement system is the participation of technically proficient staff in the community college system (at the local college and central office levels). If California’s community colleges had lacked a cadre of qualified researchers, the peer grouping system would probably have faltered in the discussion stage – far short of implementation.

References


Willard Hom is the director of the Research and Planning Unit at the System Office, California Community Colleges. Before joining the System Office in 1999, he conducted social research for the California Department of Health Services and the California Employment Development Department. He received a bachelor of arts in political science from the University of California, Davis, and a MBA from California State University, Sacramento.

His recent work has focused on accountability models; his past work has dealt with survey methods, market research and organizational analysis. Willard is also president of the National Community College Council for Research and Planning. He can be reached at 916/327-5887.