AIR

Association for Institutional Research

PROFESSIONAL FILES | SUMMER 2017 VOLUME

Supporting quality data and decisions for higher education.

Letter from the Editor

Summer brings time to reflect and recharge. The Summer 2017 volume of AIR Professional Files presents four articles with intriguing ideas to consider as you plan for the next academic year.

Data governance is a pressing issue for many IR professionals, as sources of data proliferate and challenge our ability to control data integrity. In her article, *Institutional Data Quality and the Data Integrity Team*, McGuire synthesizes and interprets results from 172 respondents to an AIR-administered survey of postsecondary institutions on their data integrity efforts. She describes the current state of data governance and offers strategies to encourage institutional leaders to invest in data quality.

Those of us who work in assessment often take it for granted that assessment results will be used for learning improvement. Fulcher, Smith, Sanchez, and Sanders challenge this assumption by analyzing information from program assessment reports at their own institution. *Needle in a Haystack: Finding Learning Improvement in Assessment Reports* uncovers many possible reasons for the gap between obtaining evidence of student learning and using that evidence for improvement. The authors suggest ways to promote learning improvement initiatives, and share a handy rubric for evaluating assessment progress.

Institutional researchers are beset with requests to form peer groups, and it seems that no one is ever satisfied with the results. Two articles in this volume present very different methodologies for forming sets of comparison institutions. In her article, *A Case Study to Examine Three Peer Grouping Methodologies*, D'Allegro compares peer sets generated by different selection indices. She offers guidance for applying each index and encourages cautious interpretation of results. Rather than rummaging around for the perfect peer set, Chatman proposes creating a clone, or doppelganger university, one that is constructed from disaggregated components drawn from diverse data sources. In *Constructing a Peer Institution: A New Peer Methodology*, he walks us through the process of creating peers for faculty salaries, instructional costs, and faculty productivity. While the constructed peer approach has its challenges, the appeal of achieving a perfect fit peer is undeniable.

I hope your summer "reflection" inspires you to share your work with your IR colleagues through *AIR Professional Files*.

Sincerely,

Sharron L. Ronco

Sharron L. Ronco

IN THIS ISSUE...

EDITORS

Sharron Ronco
Coordinating Editor
Marquette University

Leah Ewing Ross
Managing Editor
Association for Institutional Research

Lisa Gwaltney
Editorial Assistant
Association for Institutional Research

ISSN 2155-7535

PROFESSIONAL FILE

ARTICLE 140

Association for Institutional Research

INSTITUTIONAL DATA QUALITY AND THE DATA INTEGRITY TEAM

Katherine A. McGuire

About the Author
Katherine A. McGuire is director of institutional research at Oxford College of Emory University.

Abstract
Data quality has become a pressing issue for many campuses in recent years, as colleges struggle to extract timely, accurate, and consistent information from ever-proliferating institutional data sources in order to meet strategic decision-making and accountability demands. In this mixed methods study, a survey and semi-structured interviews were used to examine data integrity teams, which are groups that try to improve the accuracy and usefulness of data in computing systems at institutions of higher education. A survey sent to a random sample of institutional researchers revealed that many campuses did not have data integrity teams. Where campuses had them, those teams frequently did not engage in activities like data auditing, creating or maintaining data standards documentation, or training staff on data standards issues. Interview participants from campuses with an established team reported that the greatest benefits were greater communication, collaboration, and awareness of data quality issues. Both survey respondents and interviewees reported that more data governance resources, including dedicated staff time, were needed to improve data quality. The implications of these findings for strategic data quality and best practices for institutions are discussed.

Keywords: Data quality, data governance

BACKGROUND

Computerized database systems have created a revolution in the capacity of organizations to store and rapidly retrieve information about their processes and people. The routine operations of colleges and universities have been profoundly affected by these broad-based changes in information management. All administrative and academic departments on a campus require access to information contained in institutional databases for their daily activities, whether it be directory information, student enrollment and academic records information, financial aid data, accounting and billing data, faculty and staff personnel data, donor records, grants management data, or facilities and scheduling information. In recent years, demand for information for accountability, institutional decision-making, and planning has placed increased scrutiny on data quality and data processes at postsecondary institutions.

Since early in the development of the field of institutional research, practitioners have expressed concern about the accuracy of data contained in student information systems. In a 1989 Association for Institutional Research (AIR) Professional File paper entitled "Data Integrity: Why Aren't the Data Accurate?," Gose described a number of major types of data errors, and noted that the human element was essential in maintaining data systems free from various types of "data corruption." By "human element," he presumably meant that improving communication between departments and individuals about data problems and data standards is crucial to improving data quality.

McGilvray points out that a persistent problem with data quality is that data management is one area where the trend toward greater integration and collaboration in organizations has lagged behind: "Our applications and business needs for information are integrated, but our behavior has not changed to work effectively in this world. For example, your company may need information to support end-to-end processes and enterprise decision-making, but the information is being created by an individual contributor from the business who has no visibility to other needs for the same information" (2006, p. 2).

Thus, data entry responsibilities frequently fall to the lowest-ranking and newest member of a department, someone who does not understand the needs of end users and in whom just enough training is invested to get the job done at a basic transactional level. Such employees tend to be rewarded for speed rather than accuracy, and often the department where data entry occurs is not directly impacted by data errors.

Colleges have adopted various strategies for improving campus data, all of which could be described by the umbrella term "data governance." Such strategies might include codifying data standards, creating standard operating procedures for data processes, developing master data sets for reporting, and assigning to specific personnel oversight of data in campus functional areas. All these strategies require that critical stakeholders regularly communicate and collaborate to identify problems, set standards and policy, oversee and review data and data processes, and help manage change that impacts data integrity. Some college campuses have instituted data integrity teams to serve this function. Data integrity teams are groups of stakeholders from diverse functional areas on campus that meet regularly to try to collaboratively address data problems as they arise, as well as to proactively implement improved data management policies and procedures.

Young and McConkey (2012) and McLaughlin, Howard, Cunningham, and Payne (2004) have described many of the activities that are appropriate for data integrity teams in higher education. Teams should first identify data stakeholders and their needs. They should institute consistent data definitions across the institution, such as by creating a data dictionary, and they should establish data use rules. They should draft data policies, communicate the importance of those policies, and monitor and report both the status of data quality efforts and compliance with standards. They should assign data stewards or custodians so that there is no ambiguity about who is responsible for data in a given area, and they should update such assignments when necessary. They should seek to understand external accountability and internal research and planning data needs, and should incorporate these needs into data standards decisions. Teams should be aware of data quality issues surrounding documentation, process gaps, and missing data. They should address issues of access, security, and integration of multiple data systems. Finally, data integrity teams should track how data decisions are made, as well as how conflicts between departments or members are resolved.

The present study examined the staffing, scope of activities, institutional environments, and effectiveness of data integrity teams on college campuses by means of a concurrent mixed methods research design, including an online survey and semi-structured interviews of postsecondary data users. Some of the research questions the study addressed were these:

1. What percentage of postsecondary institutions have formal data integrity teams? Can any institutional characteristics or organizational conditions be identified that seem to promote the development of data integrity teams?
2. Who typically serves on data integrity teams? Which institutional departments play leadership roles in data governance?
3. How well are data integrity teams supported by executive leadership and what authority do teams have to make and enforce data policy?
4. What are the typical tasks undertaken by a data integrity team? How effective do team members believe their teams are at solving various types of data quality problems?
5. What do team members perceive as the barriers to institutional data quality? How do they think these might be overcome? Are there any types of data problems that are insurmountable or unavoidable?

In the first phase of the study, randomly selected members of the higher education professional association, AIR, were invited to participate in a 20-minute online survey that asked questions about the demographic characteristics of their institution and whether it had a data integrity team. If the institution had such a team, questions followed as to who served on the team, core team activities, and team accomplishments and challenges. A second qualitative phase of the study interviewed individual data integrity team members at postsecondary institutions about their teams' activities and challenges. This study differs from previous data integrity research done by higher education information technology (IT) groups like EDUCAUSE (see Yanosky 2009) by focusing on the perceptions of professional institutional researchers rather than on IT leadership or staff, as well as in having a qualitative component.

Table 1. FTE Enrollment of Respondents' Institutions

FTE Enrollment	Frequency	Percent
Fewer than 1,000	18	11%
1,000–2,999	36	21%
3,000–9,999	55	32%
10,000–19,999	36	21%
20,000 or more	27	16%
Total	**172**	

Note: FTE = full-time equivalent.

METHODOLOGY

The quantitative phase of the study consisted of an online survey created and maintained in the online web survey tool SurveyMonkey (www.surveymonkey.com), and administrated by AIR. The survey contained item tracking so that AIR members whose institutions did not have data integrity teams or who were not members of their schools' data integrity teams answered a different set of questions than respondents who were on campuses and/or served on data integrity teams. A sample of 519 randomly selected members of AIR were sent an e-mail from AIR explaining the purpose of the survey and inviting them to participate by clicking on a hyperlink in the e-mail message. Descriptive data analysis was performed using Statistical Package for the Social Sciences (SPSS).

The qualitative phase of the study consisted of structured individual interviews. Each interview subject was a data integrity team member from a different postsecondary institution. Participants were recruited through the e-mail lists of two institutional research groups: the Georgia Association for Institutional Research, Planning, Assessment, and Quality (GAIRPAQ) and the Higher Education Data Sharing Consortium (HEDS). Additional potential subjects were located by a Google search of terms such as "university data integrity team," "college data governance," etc., and e-mail contact was made with relevant staff at institutions for whom data integrity team information was found online. Subjects were interviewed by phone using the online tool Skype, and interviews were recorded to MP3 files using the Skype recording tool Evaer. All semi-structured interviews were transcribed manually from the MP3 files, and the resultant data were coded and analyzed in QDA Miner Lite. Both thematic and content analyses were performed where appropriate.

RESULTS

Survey Results

A total of 205 AIR member respondents submitted the survey, for a 39% response rate. Of these, 197 responded to at least one item on data integrity and were included in the final analysis of survey results.

The majority (87%) of respondents were employed at postsecondary institutions. Of the 172 respondents employed on postsecondary campuses, by far the largest group was at institutions with both undergraduate and postgraduate programs (66%). Smaller percentages of respondents were from institutions with two-year (22%), four-year only (9%), and graduate-only (3%) programs. There were slightly more respondents from public (55%) than from private institutions; only four respondents (2%) were from private proprietary schools. The diversity of institutional student enrollment sizes represented in the sample can be seen in Table 1. Exactly half of the respondents were at multicampus systems, illustrating the potential complexity of data management at the institutions in the study.

Fewer than half (44%) of the 172 respondents from postsecondary institutions reported that their school had a data integrity team, and only 38 respondents (22%) reported leading or serving on a data integrity team. Table 2 shows the institutional characteristics of institutions that had data integrity teams.

Executive Advocacy of Data Quality Efforts

Respondents indicated they believed that campus executive leaders were overall supportive of efforts to improve data quality (see Table 3). With the exception of the chief business officer, whose rating decreased slightly when disaggregated, this confidence in leaders' support of data integrity was even more pronounced for respondents who were members of their institutions' data integrity teams.

Respondents' Ratings of Institutional Data Quality

Sixty-six percent of all institutional respondents said that they "Agreed" or "Strongly agreed" with the statement, "The overall quality of data in my institution's administrative computing system is high." There was virtually no difference in the percentage of respondents who rated institutional data quality highly who were on data integrity teams from those who were not. Respondents who reported that their campus did not have a data integrity team were asked why they thought it did not (see Table 4).

Table 2. Characteristics of Institutions with Data Integrity Teams

Institutional Characteristics	Number of Institutional Respondents with Data Integrity Team*	Percent of Institutional Respondents with Data Integrity Team*
Institutional Type		
Two year	18	47%
Four year only	5	46%
Four year plus graduate and/or professional	48	52%
Graduate and/or professional only	4	80%
Institutional Control		
Private for-profit	1	25%
Private not-for-profit	36	58%
Public	38	49%
Institutional FTE		
Fewer than 1,000	5	36%
1,000–2,999	19	59%
3,000–9,999	27	53%
10,000–19,999	15	54%
20,000 or more	9	50%

Note: FTE = full-time equivalent.
** "I don't know" and "No response" omitted from numerator and denominator.*

Table 3. Support of Campus Leaders for Data Integrity Efforts

The following campus leaders support efforts to address data integrity at my institution (Strongly agree or agree)	All Institutional Respondents (n=169)	Data Integrity Team Members Only (n=32)
President/Chief executive officer	56%	76%
Provost/Chief academic officer	69%	90%
Chief business officer/Chief financial officer	68%	61%
Chief student affairs officer	56%	68%
Chief Information officer	71%	84%

Table 4. Reasons Respondents' Institutions Do Not Have Data Integrity Teams

To the best of your knowledge, what are the reasons that your institution does not have a data integrity team? (check all that apply) (n=70)	Percent
Data quality is not a problem at my institution.	14%
Data quality issues are too contentious/political.	20%
Decision-makers are not aware of data quality issues.	27%
Decision-makers are not interested in data quality issues.	20%
Decision-makers do not have time to devote to data quality issues.	40%
Decision-makers do not have resources to devote to data quality issues.	43%

Composition and Leadership of Data Integrity Teams

Over 80% of survey respondents who were on data integrity teams worked in institutional research or assessment offices, as might be expected given the population sampled. As shown in Table 5, by far the most common functional area of team leaders was institutional research and related departments, followed by IT. Various other leader functional areas were mentioned in the open-ended comments for this survey item, including associate vice president and bursar, as well as cochairing arrangements.

Additional team members mentioned in the open-ended comments sections were online or e-learning coordinators, athletics, career services, the veterans' affairs office, and student life.

Data Integrity Team Characteristics

Over 80% of the respondents who served on their institution's data integrity team had been on the team for more than three years, and only about 15% had served for less than a year. The most common regular meeting schedules were monthly (24%) or quarterly (18%); a combined 32% said they met either irregularly or on an as-needed basis rather than keeping a regular schedule.

About 30% of the respondents said their data integrity team reported to the institutional research, institutional effectiveness, or assessment functional area. Another 16% reported to IT, 13% reported to academic affairs, and about 10% reported to the president or chief executive officer. A few other teams reported to executive cabinets or other entities. Several respondents said that their team either did not report to anyone or that they were not sure who their team reported to. Respondents indicated that the team reported to the individual or entity that oversaw it by face-to-face meetings or presentations (42%), memos or reports (13%), or both methods (40%). Most teams reported that they had only a limited range of data policy-making authority and that they referred data policy violators to another entity or person (see Table 6).

Team Activities and Effectiveness

Data integrity team members reported their team doing a variety of common data quality–related activities, as

Table 5. Team Leader's Department and Representation on Team

	Team Leader's Department (n=32)	Represented on Team (n=41)
Institutional research/Institutional effectiveness/Assessment	47%	100%
IT/Computing	24%	71%
Other (please specify)	16%	13%
Academic affairs/Faculty	3%	58%
Admissions/Enrollment management	3%	71%
Development/Advancement	3%	34%
Registrar	3%	79%
Business/Accounting	0%	66%
Financial aid	0%	74%
Human resources	0%	45%

Table 6. Team Authority to Make and Enforce Data-related Policy

Which best describes the team's authority to make data-related policy on your campus? (n=33)	Percent
We have a broad range of policy-making authority.	23%
We have a limited range of policy-making authority.	45%
We can make recommendations only.	29%
Which best describes the team's authority to enforce data-related policy on your campus? (n=32)	**Percent**
We have policy enforcement authority (e.g., can limit data systems access).	13%
We refer individuals who violate data policies to other entities (e.g., their supervisors).	53%
We have no authority to enforce policy.	27%

Note: "Other" responses not included.

summarized in Table 7. The activities that were most often cited as a focus of the team were identifying data gaps and inconsistencies, identifying data stewards, and considering institutional strategic reporting needs. The two items that respondents cited least often as being a focus of the team concerned data auditing and policy assessment.

Team members also reported on institutional and departmental environments and outcomes for data quality, as shown in Table 8. Although respondents indicated that advocacy and awareness of data quality issues existed on their campuses, only slightly over half agreed that having a data integrity team had improved institutional data quality. Many of the typical activities associated with data integrity teams, such as creating data documentation, training staff, documenting data steward responsibilities, and monitoring data quality, were occurring at a third or fewer of the institutions. Only a quarter of the respondents agreed that data users knew the procedure for reporting data problems.

Views of Non-Team Members on Data Integrity Practices

As noted previously, many of the AIR member respondents either did not serve on their campus data integrity team, were employed on a campus that did not have a data integrity team, or were not employed on a college campus. Respondents who reported that they were not currently on data integrity teams answered opinion questions about data quality issues on college campuses. Of these respondents, 85% agreed with the

Table 7. Frequency of Data Integrity Team Activities and Team Effectiveness

How often does the data integrity team focus on the following issues, and how effective is the team in each area?	Frequency of Team Activities (Percent "Sometimes" or "Often") (n=32)	Team Effectiveness (Percent "Effective" or "Highly effective") (n=31)
Identify data gaps and inconsistencies.	97%	66%
Identify data stewards (people responsible for maintaining data quality and reporting data issues).	97%	68%
Consider internal strategic data reporting needs.	93%	54%
Create new data policies.	90%	55%
Review current data policies.	87%	71%
Align data policies between departments.	87%	54%
Seek input from data stakeholders.	86%	57%
Address compliance or regulatory issues.	86%	61%
Establish needs, roles, and responsibilities of data stewards.	86%	58%
Determine who has or needs access to data.	79%	61%
Assess effectiveness of data policies.	79%	48%
Monitor data quality.	79%	57%

statement, "Every college or university should have a data integrity team." The majority of respondents (55%) believed that data integrity teams should report to the office of institutional research or institutional effectiveness; only 11% stated that the team should report to an IT function.

Respondents were also asked what they thought the activities of a data integrity team should be (see Table 9). The activities that respondents not on a data integrity team were likely to think most important differed somewhat from the activities that data integrity team members reported as teams' most frequently addressed issues, with data auditing and policy assessment assuming greater importance to the non-team-member respondents.

About a third of the respondents not currently on data integrity teams had served on one in the past; of these respondents, 65% rated their previous data integrity team to be highly or moderately effective.

Open-Ended Survey Comments

Around two dozen respondents gave additional reasons or commentary about why their institution did not have a data integrity team. About a third of the comments indicated that data quality issues were handled in an informal, ad hoc manner in response to specific problems or projects with whatever departments were impacted

Table 8. Institutional Environments and Activities for Data Quality Reported by Data Integrity Team Members

Indicate your level of agreement with the following statements about your institution:	Percent "Agree" or "Strongly Agree" (n=37)
My supervisor is aware of the importance of data quality.	90%
Data integrity team members serve as advocates for good data in their departments.	77%
Data quality is a strategic priority.	65%
Data stewards/managers exist in each functional unit that has data access and responsibilities.	58%
Having a data integrity team on my campus has improved data quality.	55%
Data quality is continuously monitored.	48%
Significant resources are devoted to data quality improvement efforts.	42%
The institution has a usable and complete data dictionary.	33%
All data users have easy access to data field documentation.	32%
Staff who work with data receive training about data standards.	32%
Data steward/manager responsibilities are clearly documented.	30%
There are regularly scheduled comprehensive data quality audits.	26%
Individuals who use data know how to report a problem or issue with data quality.	26%

by the particular issue. Similarly, several other respondents indicated that data quality issues were handled in a decentralized fashion within departments. Three participants said that they had previously had a data integrity team that had stopped meeting, and several others said that their institution was in the process of forming a data integrity team. Two respondents expressed the belief that data integrity teams were not useful because data quality issues were too complex to be solved by a single team.

Most of the respondents who served on data integrity teams commented on how data integrity could be improved at their own institution. Typical comments cited the need for more buy-in by both senior leadership and staff. More centralization of data quality efforts and user accountability for data quality were also mentioned by several respondents. Training for data users was one of the most frequently mentioned needs, as was creating or updating a data dictionary. The need for additional staff was a concern, and several respondents said that they believed their institution needed dedicated staff to oversee data integrity issues.

About 40% of the respondents not currently serving on a data integrity team answered the open-ended question, "How can data integrity be improved at institutions?" Twenty-five percent of the comments mentioned the need for greater executive buy-in and accountability, and nearly 20% of comments mentioned the need for some kind of accountability for data entry or data reporting staff. As Table 10 shows, team members and non-team members mentioned similar data quality solutions.

There were also several comments from both team members and non-team members about the need to understand the origins of information and filter out bad data before such data got into centralized data systems,

Table 9. Top Five Activities that Respondents Not on a Data Integrity Team Indicated Should Be Part of the Charge of a Data Integrity Team

What activities should be part of the charge of a data integrity team? Select all that apply. (n=139)	Percent of Respondents
Identify data gaps and inconsistencies.	94%
Review current data policies.	93%
Assess effectiveness of data policies.	88%
Monitor data quality.	87%
Seek input from data stakeholders.	87%

by technical validation or automation where appropriate: "Garbage in = garbage out. One of the most difficult challenges is controlling quality and consistency from point of entry."

SEMI-STRUCTURED INTERVIEW RESULTS

Demographics of Participants and Their Institutions

Interviewees were data integrity team members from seven institutions in the continental United States. Six participants were institutional research or institutional effectiveness administrators at the director level or higher; the other was an IT manager who specialized in data governance. Several different Carnegie types were represented among the institutions in the interview sample, including four baccalaureate colleges, one master's college, and two research universities. Regionally the South, Mid-Atlantic, Northeast, Pacific Northwest, and Midwest were represented. Six of the interviewees came from private not-for-profit institutions, and one was from a public institution. The total enrollments of the institutions ranged from just over 2,000 students to nearly 26,000 students.

Cross-case Analysis

As seen in Table 11, participants' institutions are compared side by side on a number of variables relevant to data integrity. These data were derived from the interview transcripts; in a small number of cases participants were not sure how to answer a question or became sidetracked to another issue when they were asked about it due to the loosely structured and organic nature of the interviews, so that the information could not be clearly ascertained from the transcripts.

Team Structure, Membership, and Leadership

There was a wide degree of variability in the structure of the data integrity teams represented in the sample. Some data integrity teams were effectively user groups for the main student information system on the campus, while others were outgrowths of the institution's business intelligence units. Sometimes there was only one team on a campus involved with data integrity, but at some institutions there were several teams with different specific functions. In some cases, this diversification of the data integrity function had to do with a working group of middle managers needing to rely on a higher-level executive committee to make policy; in other cases, it had to do with the size and complexity of the institution and the data issues encountered.

For some of the data integrity teams, particularly those that functioned as user groups for a specific data system (e.g., Datatel or Banner), membership was voluntary for those who had an interest in solving problems with institutional data. At other institutions, data integrity team membership was part of the job description for manager positions that involved working with data. Additionally, attendance might be expected at all meetings for some core members, while other staff attended only when there was a specific issue or problem being discussed that required their input.

Despite this variability in team structure between campuses, there was a relatively high degree of uniformity in the functional roles that were represented on campus teams. Typically, a single representative from each relevant department participated on the team. As might be gleaned from the demographic description of the study participants, institutional research and IT offices

Table 10. Topical Summary of Open-Ended Comments on How Institutions Can Improve Data Integrity

How can data integrity be improved at your institution/at institutions?	Data Integrity Team Member (Percent of Comments; n = 25)	Not on a Data Integrity Team (Percent of Comments; n = 52)
Increased accountability.	16%	19%
More/better training.	16%	12%
Greater executive buy-in.	12%	25%
Greater staff buy-in.	12%	10%
Centralization of data integrity efforts.	12%	6%
Dedicated staff.	12%	6%
Create/improve data dictionary.	12%	2%
Better communication or collaboration.	8%	14%
More staff overall.	8%	4%
Different unit in control of data integrity.	8%	Not mentioned
Automation of data entry or data validation.	4%	4%
More local unit autonomy in data quality decisions.	4%	Not mentioned
More time devoted to data quality.	4%	Not mentioned

were represented on such teams, and were frequently leaders or occasionally cochairs of the team. Additionally, staff from the registrar's office, financial aid, human resources, academic affairs, student affairs, and admissions office were members of nearly all the teams. Staff members from business and accounting, as well as development and alumni affairs, were represented at some but not all the institutions included in the interview sample. The differences in team membership and structure were often reported to be due to the existence of multiple different data systems on campuses, such as separate athletics, admissions, communications, or advancement databases, for example. Participants indicated that this multiplicity of data systems added an additional layer of complexity to data quality. Sometimes the data integrity team included users of a number of databases, and sometimes it included only users of the main student information system, which could be problematic when one database was used to populate another.

Activities and Processes of Teams

Different teams had different regular meeting schedules and agendas. Most typically, the main data integrity team met once a month. The frequency of team meetings seemed to vary with the structure of the data integrity function: the two teams with business intelligence or an analytics function were those meeting weekly. Typically, a meeting agenda was created at least in part from a call for topics, issues, or problems from team members.

Table 11. Cross-case Analysis of Interview Participant Data

Participant	Name of Team	Team Leader	Entity to Which Team Reports	Data Dictionary	Data Warehouse	Executive Sponsor
Participant A	Data management group	IT staff person	Administrative computing advisory group	Yes	Yes	None mentioned
Participant B	Data quality/data governance	Business intelligence manager	No formal reporting structure	Yes	Yes	None mentioned
Participant C	Data governance	None mentioned	Provost	Yes	Yes	VP for IE, provost
Participant D	Data standards group	IR	Steering committee composed of data stewards	Yes	Yes	Provost
Participant E	Data standards committee	Cochaired by IR and an academic dean	Executive-level cabinet	No	No	Academic affairs associate dean
Participant F	Data committee	CIO	Large ad hoc group of VPs	No	No	None
Participant G	Users group	Cochaired by IR and IT	Voluntary group, no formal reporting structure	No	No	CIO

Note: CIO = chief information officer; IE = institutional effectiveness; IR = institutional research; IT = information technology; VP = vice president.

Additionally, team meetings also usually spent time on updates of ongoing data quality projects. A few teams had regular reports from specific offices or groups, such as IT staff that were working on projects that might affect data and impact data users:

Participant: So we meet monthly. And we have split the meeting into several different things that happen. One thing that happens is that our project manager for our PeopleSoft implementation always gives an update because this is the only place where people who are not at very high levels can find out what's going on with our implementation. . . . For example, we're thinking of purchasing some BI [business intelligence] tools. The people who are going to have to work with these BI tools are the people at the data *standards committee meeting, not the cabinet. And so those are the people who need to know that this might be happening. . . . This is the only place where that . . . where they get that kind of update. So we always devote part of our meeting to that.*

Problem solving and change management were activities of the data integrity groups in the study that were frequently mentioned. Typically, problems or projects were submitted to the committees as an agenda item:

Participant: So once a month we put out a call for topics. We really just ask people, so OK, what's rubbing the wrong way? What's an issue now? And people bring these things up.

Additionally, changes in externally mandated compliance reporting or changes to institutional programs requiring adjustments to data collection and reporting strategies were often brought up in the data integrity teams. Examples of external policy changes that were mentioned were the change to the current federal Integrated Postsecondary Education Data System (IPEDS) race and ethnicity and human resources reporting standards. Technology changes, such as data system conversions or upgrades, might also typically be discussed in the data integrity group.

Frequently mentioned was the need for the formation of subgroups or subcommittees of team members with a particular interest or expertise in a specific data problem. Sometimes this was an issue of change management. These subcommittees would occasionally draw on personnel who were not regular members of the data integrity team if their expertise or input was needed. The typical protocol seemed to be for these subcommittee members to work on a problem outside the data integrity team, and then report back to and seek feedback from the team at its regular meetings until a data issue was resolved.

Data dictionaries were sometimes an activity of the data integrity team. Four of the participating institutions had data dictionaries and three did not. Both of the research institutions had data dictionaries and, perhaps not coincidentally, also had business intelligence models for reporting and analytics. Almost all the schools that had data dictionaries also had data warehouses, so it is probable that there is a relationship between the two outcomes. One of the research university participants belonged to a school that used the Data Cookbook, a commercially available data dictionary tool. This institution's participant described the tool as playing a positive role in developing consistent and accessible data standards and processes across campus, but also admitted that implementation and maintenance of the technology had been labor intensive.

Authority

Authority of the team to make and enforce data policy was handled in a number of different ways at the campuses in the study. Some teams had a clear charge from executive leadership while others were exclusively voluntary in nature. Teams seldom seemed to have broad authority to make data policy decisions. As indicated in the cross-case analysis, the usual arrangement was for a group of midlevel data managers to make data decisions at the field or project level, but to defer to an executive body on campus-wide policy decisions. Also noted in the cross-case analysis was that only about half of the participants reported having an executive advocate. Those that did spoke highly of the value of having an executive-level sponsor for data quality, particularly at the point of getting data integrity teams started:

Participant: And we have an advocate with my vice president, thank God, who used to be the CIO [chief information officer] here. . . . She's just that type of person that can just . . . that runs everything. But she's been a huge advocate for us. . . .

Interviewer: So she knows what the issues are.

Participant: Yeah. And you have to have an advocate, I would say. At least one.

There was some ambivalence from interviewees in response to questions about how much support data integrity teams and their efforts got from executive leadership. On the one hand, participants seemed to believe that leadership generally was supportive of the team itself. Where teams referred policy or strategic data decisions to an executive steering committee, participants reported that the steering committee respected their expertise and was willing to endorse their recommendations on most data policy matters.

Interviewer: So, do you feel like you get pretty good buy-in from executive leadership, then? You had said that, you know, recommendations go up to the cabinet level. I mean, are they pretty likely to approve things that the group, the data standards group, has recommended?

Participant: Yeah, I think as long as it's well-reasoned. I have to say, they're great about, what is you . . . I mean, what are you trying to do, why are you trying to do it, what's the benefit for the institution, what are the liabilities for the institution? And if you can present that, and they're all well-reasoned, they're like, "OK." . . . We have like 16 people on the data standards committee from across the institution. Everybody in that group buys into some things, and they've communicated back with their areas about it. We've probably picked up most of the rocks and seen what's underneath them. So when we go to the cabinet and try to make a recommendation, we've really, you know, we've really looked under a lot of rocks.

On the other hand, a number of the participants expressed the opinion that most leaders on their campus didn't have a very profound appreciation for the strategic importance of data quality or understand the kinds of data problems that existed on their campus. Additionally, some participants voiced frustration that data quality issues did not get the time, attention, or priority they needed:

Participant: I asked our interim provost—our provost is away briefly—so, I said to him, "Is it that people don't care? Because we had this one meeting, where everybody agreed we needed to meet, and we haven't met again. What's going on?" And he said, "I don't think it's that people don't care. It's that it doesn't seem "urgent." Something else usually . . . you know, that "urgent versus important" grid. It's very important, but not being seen as urgent."

Resource Issues

One of the greatest resource issues for teams was that of staffing and the related issue of staff time. With one exception, in which a data governance manager at a large research university oversaw the data quality processes at that institution, almost all participants mentioned team leaders as well as members who had other primary job responsibilities. Whereas there are clear benefits to having data integrity team members with deep understanding of the data needs of one or more specific functional areas, this arrangement can also mean that every person on the data integrity team has other, more-pressing responsibilities, making it difficult for team members to find time to dedicate to data integrity team projects. Several participants mentioned attendance problems at meetings. Workload was also given as a reason for not having data dictionaries or data warehouses. Of all the participants, only the two research universities had dedicated data governance staff or plans to add any.

Tools for communication between the team and data users were cited as a resource issue. Some participants mentioned that they placed data integrity group minutes or documentation like portable document formats (PDFs) of data dictionaries on an intranet site or used a tool like Moodle. Sometimes users accessed them but reportedly they often did not. Other interviewees said data policy decisions were sent to stakeholders by campus e-mail once, at the time they were implemented, which seemed to be problematic in terms of providing ongoing and readily accessible documentation to users. Only one of the institutions had implemented a "live" interactive metadata management tool. A perhaps related finding was that most participants reported that their team did not have a budget.

Benefits and Challenges

Almost every participant spoke about information-sharing and communication as key benefits of the data integrity team. The data integrity team was cited as a place where stakeholders were identified, impact of data decisions was explored, and users learned how data were created and used in other functional areas. Frequently the data integrity team was where users first became aware of compliance issues, technology changes, or program changes that might impact data collection or reporting needs.

Participant: So do we want to add that field? So we bring it to the table: Who-all does this affect? We think it affects me, institutional research, and the registrar's office. But who-all cares? It turns out financial aid. So it turns out, oh, this affects you, or maybe just confirming our instincts.

Interviewer: So it's a place for finding out who stakeholders are in decisions?

Participant: Yes. Yes, how does this affect, you know, other offices? That's a huge topic of the conversation, and that's been a huge benefit to this face-to-face meeting of folks.

Most of the participants also spoke of increased awareness of data integrity among data users as a benefit of the team, and several indicated that they thought that the team had raised the profile of data quality as a strategic issue on their campus.

In spite of this information-gathering function of the team, communication was also often cited as a challenge to working on data quality issues. Because members had different areas of domain expertise, they could not always easily explain to team members from another unit why a data element was problematic for them or how they knew a specific data point was incorrect. Members frequently used different technical vocabularies or conceptualized data or problems in varying ways. Even defining what constituted a data quality issue could be difficult:

Participant: Sometimes if you ask them, it's "No, we don't have a data quality problem," and then you go back and actually look and, "Well, yeah, actually, you do." "Oh, that's a data quality problem?" And then you talk about that. So it's getting people to kind of understand what their roles are and identify what it is they need to do.

A challenge that institutions seemed to struggle with was maintaining accessible documentation of not just data field standards, but also of procedures. One of the participants told how his school had recently "consolidated all of the handbooks—the students, the employees, the staff handbooks—into one college handbook, and that has reference to just about all the policies and guideline sets." However, this was not typical, as other institutions reported not having adequate documentation of policies, particularly those concerned with identifying and reporting data problems:

Interviewer: So in terms of the kinds of procedures you might have in place, you said you had a manual that has field-level kind of procedures. Are there also procedures for how you would report a problem? Like if you find a field that seems to have some discrepant or inaccurate data in it, and you think that maybe there's some sort of systematic issue, is there a written or formal procedure for how to initiate that?

Participant: Not really. What ends up happening is, either if it's an immediate problem they go to the IT helpdesk; if they think it's more of a systematic problem, it goes to the data standards group, which meets quarterly. It goes to them to reach a conclusion or a compromise on what should be done.

Training also seemed to be a challenge. None of the respondents reported that their institutions required any form of consistent training on data standards for all new personnel. In general, the standard seemed to be that departments within the college or university were in charge of training their own personnel, because of the difficulty in providing data systems training general enough to meet the differing, technically specific needs of users in diverse functional areas.

The participants mentioned several data areas as particularly problematic for users and teams. Parent names and contact information came up a number of times as an example of data that are of high importance to advancement offices but that are difficult to keep updated and challenging to use. Faculty data frequently were mentioned as a challenge, in part because two offices—human resources and academic affairs—are typically involved in creating and using these data, but also because those offices have different operational and reporting needs. Tracking student hiatuses (leaves) was cited as challenging. Also mentioned as problematic was integrating data from different campuses, or data from online and other special programs. In most of these cases it was clear that the complexity of the persons and activities represented by these data (online students in the military, faculty on sabbaticals, students whose parents were divorced and/or estranged) and not just technology limitations contributed to the difficulty of creating consistent and usable data.

Other Issues

A general observation was that the larger schools with a business intelligence and analytics orientation seemed to have more-advanced data quality processes. These institutions were more likely to have data dictionaries and data warehouses. Data governance tends to be a core component of a business intelligence and business analytics strategy. One of these respondents was careful to note, however, the integral role that a traditional institutional research orientation played in data quality.

> *Participant: The data needs to be in a way that people are confident in it, and you know how it's defined. . . . And I don't think anybody thinks about that like IR [institutional research] does. You don't have a research function in a typical corporate environment. You have a marketing or planning team or something like that, but not to the level that IR thinks about data governance. So it's been good for them to have us consulting on that. . . . Business intelligence, it won't work without good data. It won't. And you can't have good data coming in out of transactional systems that are not designed for reporting without some very formal sort of guidelines.*

Data system customization was also mentioned by some respondents as a factor in contributing to poor quality data. Although becoming less common as commercial enterprise resource systems replace legacy systems, users frequently have had the option to customize their data system and its fields to institutional needs. Frequently these customizations were poorly designed or documented, or documentation for the change has been lost over the years. In some cases, no current user knew the reasons for or specifics of the customization, which might no longer be necessary. Such customization can make finding and fixing data problems more difficult.

Finally, creative user methods of working around poor-quality data were mentioned as a barrier to improving data processes. Such strategies could include data silos like "shadow" spreadsheets kept by individual users, hasty "cleaning" of bad data to meet contingency needs, and insufficient documentation:

> *Participant: I think it's more that it's not being seen. The ways in which the system is broken are not immediately apparent, and the impact is not apparent. Because people have done an amazing job around here of work-around fixes.*

INTEGRATION OF QUANTITATIVE AND QUALITATIVE RESULTS

The most striking finding from the survey is the fact that only about half of the respondents said that their institutions had data integrity teams. This could explain why subject recruitment for the qualitative study was somewhat challenging. Both of the studies identified the same group of "usual suspects" among functional areas of team members, with institutional research and IT being the most common areas represented; staff from those areas frequently serve as team leaders. In addition, the studies identified broad representation by other campus departments. In both study phases, development or advancement was the most likely major function not to be included on the team, probably due to the development-specific data systems used at many schools as well as the unique types of data that advancement offices collect and use. Both methodologies found that IT and academic affairs were the most likely executive advocates for data integrity efforts.

Most data integrity team members in both the survey and the interviews reported that their teams were improving data quality on campus. Very few survey respondents from institutions that did not have data integrity teams believed that not having such a team indicated a lack of data quality problems on the campus; rather, it seemed to be related to a lack of resources, including time. This finding accords well with what interviewees said both about the difficulty of getting buy-in to data quality improvement efforts on campus and why their data improvement efforts were not as comprehensive as they would like them to be, and might explain why many institutions did not have a data dictionary or data warehouse even though they believed that having these resources would be beneficial. Data integrity work is by and large work that team members do in addition to their regular assignments, and respondents often reported difficulty maintaining momentum, particularly when organizational changes or crises demanded team members' attention. Several survey respondents from institutions without a data integrity team remarked in the open-ended comments that they had previously had a team but it could not be sustained. One of the interview participants reported having advised a department that she worked with that it needed to hire someone to attend to data governance issues, and several of the survey respondents stated in their open-ended comments that they believed dedicated staff were needed to oversee data integrity.

Both qualitative and quantitative study participants reported that their teams were participating in many of the same activities: identifying data issues, problems, and stakeholders; determining which offices did have or should have responsibility for which data; and evaluating current data policies and potential compliance or programmatic changes in data needs. Most of the participants in the qualitative study reported that their campus had identified data stewards, although their responsibilities were not always well-documented or official. Data dictionaries, a best practice recommendation in the data standards and data governance literature, were not found at most institutions in the survey sample, and were found in only half of the institutions in the interview sample. It would be reasonable to suppose that this absence is due to a resource issue. Most survey respondents reported that their teams were not performing data auditing and monitoring activities. Although mentioned by one or two of the interviewees, on the whole they did not discuss auditing when describing core team activities.

A subgroup of the survey respondents whose institutions did not have data integrity teams reported in the open-ended comments that their campuses preferred to deal with data issues in an ad hoc or decentralized fashion. Since many of the data integrity team interviewees cited communication and "getting everyone together at the table" as a benefit of the data integrity teams, this opportunity can be lost when data problems are dealt with in an ad hoc way. It is worth noting that interview respondents saw the value of having smaller groups working on specific problems that mainly impacted their respective units, as long as they reported back to the team. In the same vein, another interesting though divergent finding is that relatively few respondents in the open-ended items called for increased centralization of data integrity efforts, even though bringing diverse functions together was an often-mentioned strength of the team for interviewees.

Both parts of the study found that most teams had authority only at the data field level, and needed to defer to higher-level individuals or groups to make campus-wide policy decisions. Some interview participants believed the lack of policy-making authority of data integrity teams was a mechanism for keeping leaders in the loop about strategic data issues that might impact the institution as a whole. Although most respondents in both parts of the study believed that their campus leadership and their own supervisor supported data integrity efforts in a general way, they also believed that data quality issues were not very well understood by leaders. One of the ways that leaders support initiatives is by dedicating adequate resources to them, so it says something about executive buy-in that lack of resources was typically given as a reason that data quality efforts did not receive adequate attention.

Finally, both phases of the study identified similar benefits and challenges for data integrity teams. Better communication, awareness of data quality issues, and ability to collaboratively plan for organizational change impacting data systems were among the benefits mentioned by interviewees. Improving communication was also recommended by both survey respondents who were data integrity team members and those who were not as a way to improve data quality on campuses. Training was mentioned as a challenge by both survey and interview participants, as was maintaining readily accessible documentation about policies and procedures.

DISCUSSION

This study has shown that cross-functional data integrity teams on college campuses are identified with several positive outcomes by team members. Such teams provide a forum for communication about data gaps and problems, foster greater awareness about data systems quality issues, and can facilitate the creation of consistent campus-wide data standards as well as data user policies. However, the study also found that many campuses have not created or do not see the need for such teams. Additionally, teams often lack resources such as time or staffing to implement recommended best practices such as data dictionaries and data auditing.

McLaughlin et al. (2004) have put forth a number of data process models for postsecondary settings that could be applied to these results. For example, the evolution of information management is described as consisting of three stages: (1) decentralized data operations, (2) centralized data administration, and (3) distributed data management. The majority of respondents in this study reported struggling against decentralized data operations, where only internal reliability and immediate operational needs are considered. Data integrity teams were slowly moving campuses toward centralized data administration, focusing on how data will be used for reporting as well as on operational needs, and evaluating data in terms of internal validity as well as reliability. McLaughlin et al. argue that the increasing desire for integrated data by decision-makers necessitates that institutions must move toward distributed models, meaning models that account for data that are spread out over many different software systems. The challenge of good data increases as data are expected to serve ever-higher-level needs in the organization. One of the interview respondents articulated the importance of distributed systems:

> *Participant: An IR [institutional research] team to be effective really cannot manage it all by themselves. You have to have a distributed model, you know. Or you're going to die. Or you're not going to be successful. So that's what we're working on, is just getting it out into other people's hands. In a centralized data governance process, but distributed down the way that everyone feels confident pulling data, understands how it works.*

Finally, McLaughlin et al. (2004) have posited that there are three ways organizations can respond to data architecture failures: (1) masking or hiding problems, (2) coping and trying to circumvent data shortcomings, or, when these tactics inevitably fail to meet the need for enterprise analytics, (3) correcting deficiencies in the design of data systems and processes. The interview respondent quoted in the results section beautifully illustrated the strategy of coping with her description of "work-around fixes," as well as the role of these kinds of patches in concealing systemic data quality problems. Other interviewees and respondents to open-ended survey items described the creation of departmental or individual data silos as coping strategies. In the case of one research university in the interview sample that underwent reorganization, the critical need for strategic data was a driver in correcting existing data problems.

It is important to note that this study did not purport to directly measure campus data quality in any way, but only to measure participants' perceptions of data quality. However, for the purpose of this research such indirect measurement was deemed to be adequate because the term "quality data" is defined as data that are adequate to end users' needs. Since the respondents were business data end users, their subjective opinions about data quality were presumably based on professional experience and specialized knowledge or expertise.

A potential limitation of the survey is the small number of survey respondents who were serving on data integrity teams. There might also be a selection bias toward respondents who are very satisfied or very dissatisfied with their data integrity team and data quality on their campuses. Additionally, although AIR draws its membership from many different fields within higher education, AIR's member population might be weighted toward larger and/or more-affluent institutions with the budget resources to pay AIR's conference and membership fees, or toward larger institutional research offices whose staff are more easily able to get away from the office for professional development activities.

At least one group of researchers has identified a lack of connection in the data quality research literature between technological solutions and applied business information systems contexts (Sadiq, Yeganeh, & Indulska, 2011). In other words, teams need to be aware of and consider using technological solutions to the problems of data quality, whether this means implementing automated data validation and auditing systems, or electronic metadata management tools. If technological tools can help address resources limitations, the development of open-source data quality tools would be a promising applied research area.

To overcome the reluctance of campus leadership to invest in data quality efforts, better methodologies are needed to determine costs to higher education of poor data quality. Better research about the costs of poor data quality might be a necessary tool in moving data integrity front and center with institutional leaders who can set the data governance charge on their campuses. Another possible motivation for paying more attention to data quality could be the recent national press given to several high-profile cases of college and university data problems. The net effect might be to make stakeholders wonder if they are "minding the store" with respect to data quality on their own campuses. Ultimately, the case for data quality for colleges and universities is the business case of more-efficient and more-effective pursuit of educational mission in a time of resource constraints and high expectations.

REFERENCES

Gose, F. J. (1989). Data integrity: Why aren't the data accurate? *AIR Professional File, 33*.

McGilvray, D. (2006). Data governance: A necessity in an integrated information world. *Information management: How your business works.* http://www.information-management.com/issues/20061201/1069951-1.html

McLaughlin, G. R., Howard, R. D., Cunningham, L. B., & Payne, E. (2004). *People, processes and managing data* (2nd ed.). Tallahassee, FL: Association for Institutional Research.

Sadiq, S., Yeganeh, N. K. & Indulska, M. (2011). 20 years of data quality research: Themes, trends, and synergies. In H. T. Sheng & Y. Zhang (Eds.), *Proceedings of the Twenty-Second Australasian Database Conference* (Vol. 115, pp. 153–162). Darlinghurst, Australia: Australian Computer Society.

Yanosky, R. (2009). Institutional data management in higher education. White Paper. EDUCAUSE Center for Applied Research. http://net.educause.edu/ir/library/pdf/ers0908/rs/ers0908w.pdf

Young, A., & McConkey, K. (2012). Data governance and data quality: Is it on your agenda? *Journal of Institutional Research*, 17(1), 69–77.

PROFESSIONAL FILE

ARTICLE 141

Association for Institutional Research

NEEDLE IN A HAYSTACK

Finding Learning Improvement in Assessment Reports

Keston H. Fulcher
Kristen L. Smith
Elizabeth R. H. Sanchez
Courtney B. Sanders

About the Authors

Keston H. Fulcher is executive director of the Center for Assessment and Research Studies (CARS) at James Madison University, and associate professor of graduate psychology. Kristen L. Smith is doctoral graduate assistant for CARS. Elizabeth R. H. Sanchez is assistant for CARS. Courtney B. Sanders is doctoral graduate assistant for CARS.

Acknowledgments

Special acknowledgments and thanks to the following for their contributions: A. J. Good, Cathryn Richmond, and Alena Gordienko.

Keywords: Student learning improvement, use of results, student learning outcomes assessment, higher education assessment, assessment reporting

Abstract

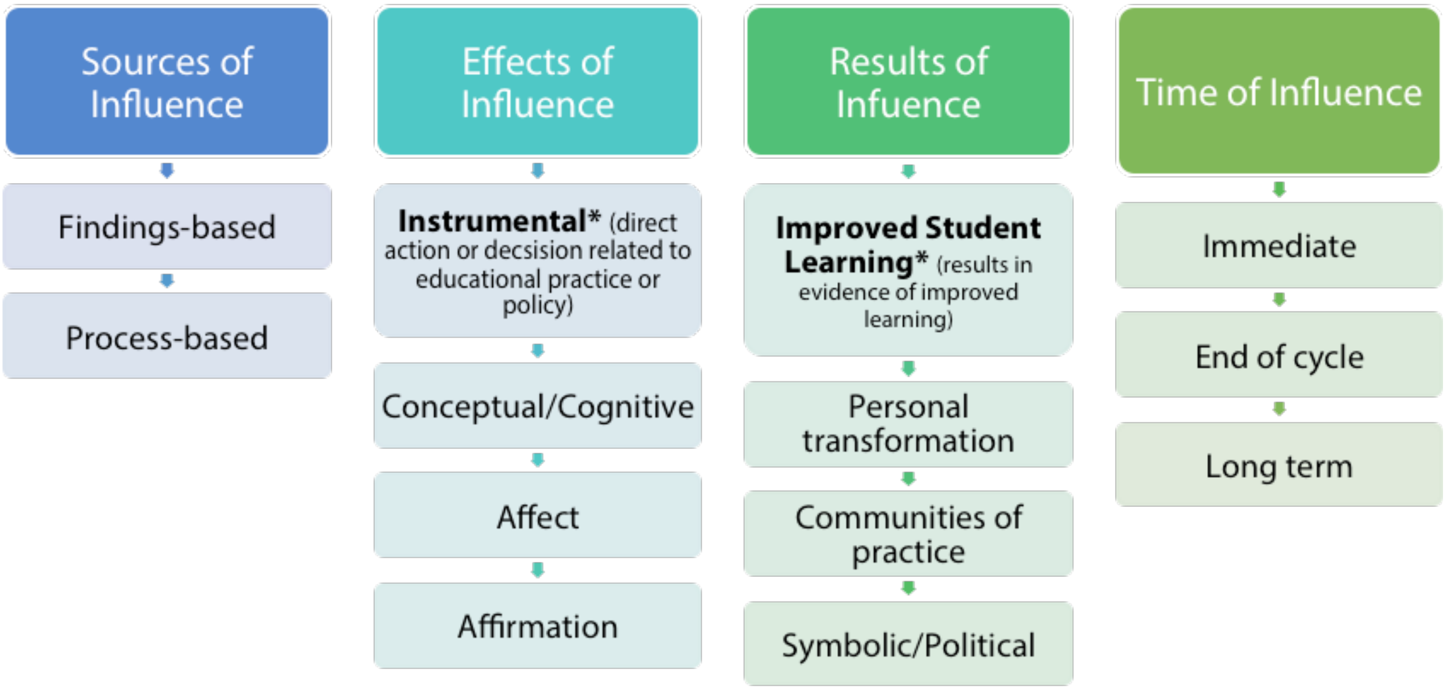
Higher education insiders trumpet the use of results for improvement as the most important part of the assessment cycle. Yet, at the same time, we acknowledge the rarity of improvement, especially at a program level. What are some reasons the most important phase of assessment occurs so infrequently? To seek answers, we investigated the "Use of Results" sections in 54 program-level assessment reports. In some respects, our findings were positive. On average, programs reported making approximately three curricular or pedagogical changes annually. A closer inspection, however, revealed concerns: (1) the curricular or pedagogical changes were not explicitly linked to learning outcomes, (2) programs rarely reported making changes that affect several classes, (3) many of the reported changes were unclear, (4) and few programs reassessed to determine if changes actually led to learning improvement. Our research concludes by providing suggestions for how programs can more effectively use results to inform changes, reassess students to determine if changes led to learning improvement, and report on improvement processes.

INTRODUCTION

For more than 30 years, higher education has refined assessment methodologies to meet accountability demands and demonstrate value (Ewell, 2009). Yet, as Suskie (2010, para. 8) observed, "Today we seem to be devoting more time, money, thought, and effort to assessment than to helping faculty help students learn as effectively as possible." Other researchers have come to a similar realization: although most institutions systematically collect assessment data, few use the data to improve student learning (Banta & Blaich, 2011; Blaich & Wise, 2011).

Why aren't assessment results used for learning improvement? There are several theories: It could be that institutions incorrectly assume that using results for improvement can emerge from only interesting research findings and well-crafted reports (Blaich & Wise, 2011). It could also be that inconsistent and vague communication surrounding the use of results for improvement confuses programs (Smith, Good, Sanchez, & Fulcher, 2015). Furthermore, accreditation requirements, rather than intrinsic interests, might be the main driver of assessment practices (Kuh & Ikenberry, 2009). Indeed, a myopic focus on assessment activities (e.g., identifying outcomes and gathering data) unintentionally neglects using results for student learning

Figure 1. Depiction of Current Study Within Jonson et al.'s (2014) Heuristic Model of Influence



improvement (Kinzie, Hutchings, & Jankowski, 2015).

Not using results to inform curricular and pedagogical changes remains a serious problem for higher education. To investigate the issue, we analyzed "Use of Results" sections in 54 assessment reports. While the current study emphasizes learning outcomes assessment at the academic degree program (e.g., bachelor's degree in biology), many concerns and findings likely generalize to other assessment and institutional effectiveness initiatives. Indeed, the inability to use results to make changes that promote improvement is an institutional concern.

Conceptualizing Use of Assessment Results

We are not the first assessment practitioners to examine why using results to improve student learning remains uncommon. For example, Jonson, Guetterman, and Thompson (2014) believed that higher education could benefit from a new, broader definition of use of results.

Instead of focusing on curricular and pedagogical changes intended to improve student learning, Jonson et al. (2014) created a model to describe various ways that discussing results can positively influence the culture of a university (Figure 1). For example, using assessment results for discussion can support taking direct action on educational practice or policy or changing people's ways of thinking about learning and assessment. Results for discussion can also alter people's emotions or attitudes regarding assessment practice and affirm the efficacy of an existing practice.

Jonson and colleagues (2014) further explained that each of the aforementioned influences could lead to the following outcomes:

- Evidence of improved student learning
- Transformation of stakeholders
- Building new communities of practice
- Generating support for policies and practice

The Jonson et al. (2014) framework sparks important conversations about how to define and measure using results for improvement, but we believe that a narrower, student-focused approach to using results would be of greater benefit to higher education. We define using assessment results for improvement as collecting and analyzing student learning data to support taking direct actions related to educational practice (i.e., making changes to curriculum and/or pedagogy) that lead to evidence of improved student learning (i.e., students' assessment scores show improvement after experiencing modified curriculum or pedagogy).

Adopting the more-narrow definition of use of results, one that centers on student learning improvement, allows us to keep in mind the overall intention of assessment and higher education. The current study is situated in the more-narrow definition, which might explain why we found so few examples of using assessment results in the 54 reports we examined.

Box 1. Hypothetical Example: 1980s Pop Culture Degree Program

At the conclusion of the 1980s Pop Culture degree program, students must be able to properly cite and reference a variety of sources in a research paper. In 2014–2015 the program used a rubric to evaluate all students' final research papers. Rubric scores revealed that students were not successful at citing or referencing sources. During a departmental discussion, program faculty confirmed that many students struggle to properly cite and reference sources.

After agreeing that the learning outcome of properly citing sources was both relevant and unmet, faculty agreed on curricular and pedagogical changes to address the issue. Before implementing new changes, faculty consulted with other instructors on campus and gathered information regarding what assignments could be effective at teaching such a specific skill set. Changes to the core courses of the 1980s Pop Culture program began in the fall of 2015. Specifically, the instructors of the two classes where writing is heavily emphasized—PCUL401 (1980s Politics and Culture) and PCUL404 (The 1980s and Today)—did the following:

1. Participated in a faculty development workshop during which the instructors found and agreed on examples of students' citing and referencing sources in their papers. Some examples were developing papers and others were advanced papers.
2. Shared the results of the past writing assessment with students, emphasizing that citing and referencing sources is a concern.
3. Provided modified examples of a developing and advanced paper to illustrate program expectations.
4. Created more in-class assignments to measure student progress, and encouraged students to rely on their own skills, instead of on online citation software, to create references.
5. Used the writing rubric to evaluate students' essays throughout the semester instead of using the rubric solely for the final research paper.

Results from curricular and pedagogical changes suggested that students' ability to cite and reference sources, as measured by the writing rubric, improved over time. Specifically, seniors' scores on the citing and sourcing element increased from 2.6 (between developing and competent) in 2015, the year before the curricular and pedagogical changes were implemented, to 3.2 in 2016 and 3.4 (between competent and advanced) in 2017, the years after the changes were implemented.

Understanding the Use of Results for Learning Improvement in Assessment Reports

Every year, academic programs at universities nationwide complete assessment reports that include a "Use of Results" section (Fulcher, Swain, & Orem, 2012). The current study examined the contents of these sections. More specifically, we investigated if changes to curricula or pedagogies were made based on assessment results and whether previous changes led to student learning improvement.

To evaluate the degree to which assessment reports conveyed using results for improvement, we first identified several ideal features of the "Use of Results" section in the assessment reports:

- Changes to curricula and pedagogies are made and reported.
- Changes to curricula and pedagogies are matched with an intended student learning outcome (i.e., what students should know, think, or be able to do).
- Changes to curricula and pedagogies are presented with a clear rationale (e.g., assessment data support changes).
- Reassessments demonstrate learning improvement (i.e., changes are at the program level and are effective).

To make the ideal assessment report more concrete, we provide an example from a hypothetical example: the 1980s Pop Culture degree program (Box 1).

The Current Research Study

Understanding how assessment reports could ideally connect assessment results to learning improvement efforts via curricular and pedagogical changes is important. We provided one simple example of a hypothetical program in an effort to clarify what the "Use of Results" section could, and should, include.

The current study focused on real programs attempting to use assessment results. We reviewed and qualitatively rated 54 program reports, comparing their features to our ideal assessment report. In doing so, we addressed five research questions (RQs).

Research Questions

RQ 1. How extensive in magnitude are the reported changes to curricula and pedagogies?

As we have explored, institutions and academic degree programs can use assessment results in different ways. Some use the results to inform changes to assessment instrumentation, while others use results to influence curricular and pedagogical changes. For those who used assessment results to change program curricula or pedagogies, we wanted to gauge the magnitude of the changes made, as described in assessment reports. That is, we wanted to see if the change was a course-level or a program-level change. If more students experience new curricula and pedagogy, we would expect to see more learning improvement at the program level.

We defined and evaluated magnitude of change in terms of minor, moderate, major, or extensive changes. An example of a change coded as minor in magnitude could include a new or modified course assignment based on previous assessment results. A change of moderate magnitude could be a new or modified unit or segment of the course curriculum. Major changes could entail a complete redesign of an entire course. Finally, extensive changes necessitate a restructuring of the curriculum or pedagogical approaches that involved several courses within a given academic program.

Again, we thought that perhaps programmatic changes of greater magnitude would be more likely to yield improved student learning. If faculty members are reporting that they only implemented changes of minor to moderate magnitude, this could help explain why no demonstrable student learning improvement exists. That is, using results to initiate only a minor or moderate change to curriculum, such as changing an assignment or unit in one course, might not be enough to move the needle at the program level.

RQ 2. To what extent are curricular and pedagogical changes linked to student learning outcomes?

To successfully improve student learning in a demonstrable way, faculty should focus assessment, pedagogical, and curricular efforts around specific student learning outcome(s) (Fulcher, Good, Coleman, & Smith, 2014). Once the learning outcome is identified, it should be clear how curricular and pedagogical modifications would enhance students' skills, knowledge, or abilities.

We defined and evaluated the match between changes and student learning outcomes, differentiating among four levels of connection in the "Use of Results" sections we evaluated:

1. It might be unclear how the change is linked to student learning.
2. It might be that the change is linked to student learning in general, but not directly to a specific student learning outcome of the program.
3. It might be that the change is linked to a specific, program learning outcome and yet lack specificity about why or how the change aligns with that particular learning outcome.
4. It might be that the change is clearly linked to a specific learning outcome in such a way that improvement seems likely.

Demonstrable program-level learning improvement can be achieved only through changes that match student learning outcomes. In other words, if we cannot determine what students should know, think, or be able to do as a result of the programmatic changes, how will we know if the changes were successful at improving student learning? Programs that can align changes with student learning outcomes in a clear and logical way should have greater success evidencing improvement.

RQ 3. What is the rationale behind curricular and pedagogical changes?
Often, there are numerous reasons that programs decide to implement changes to curricula or pedagogies; it is important to explain the rationale for making specific pedagogical and/or curricular changes (Fulcher et al., 2014). Ideally, the rationale provided in assessment reports is not only explicit, but also originates from different sources (e.g., direct assessment measures, accreditation recommendations, etc.). It is plausible that when changes lack robust supporting rationale, they are less likely to culminate in demonstrable student learning improvement. A lack of understanding or articulation of the rationale for curricular and pedagogical changes might contribute to why minimal learning improvements are found in assessment reports.

We defined and evaluated the rationale for curricular and pedagogical changes provided in assessment reports based on explicitness and type. For explicitness, we coded the report rationales as either stated, but not explained or stated with an explicit rationale. For type, we determined whether the source that contributed to the rationale was a direct measure, an indirect measure, anecdotal (e.g., conversations), accreditation or annual program review recommendations, or realignment of instruction with changes in programmatic learning objectives.

RQ 4. What is the typical stage of implementation for curricular and pedagogical changes?
Curricular and pedagogical changes take time to implement. For instance, Fulcher and colleagues (2014) suggested that it could take 3 to 5 years to make program-level adjustments and subsequently use assessment results to demonstrate improved student learning. In addition to time, change requires planning and foresight. In order to coordinate change efforts, programs should create an improvement timeline. Timelines articulate when baseline assessment data will be collected, when pedagogical or curricular changes will be implemented, and when students will be reassessed to determine whether their learning actually improved (Fulcher et al., 2014).

It could be the case that programs conceptualize processes of curricular and pedagogical changes 1 year at a time—correlative of the assessment reporting cycle. We encourage programs to look beyond an annual cycle. Creating a 3- or 5-year plan and timeline might help motivate programs to use assessment results, make changes, and reassess students to demonstrate improved learning.

For the current study, we defined and evaluated the stage of implementation of change in terms of five criteria. Change efforts could be in one of the following five stages:

1. Planning (a program is currently planning changes);
2. In process (a program is currently implementing changes; some

changes but not all have been made);
3. Completed but have not yet reassessed;
4. for efficacy (or effectiveness reassessed) but no demonstrable improvement evidenced; or
5. Completed and checked for efficacy (or effectiveness reassessed) and demonstrable improvement evidenced.

RQ 5. To what degree are programs able to close the assessment loop by using results to inform changes and subsequently demonstrate improved student learning?
The promise of quality assessment practice is to enhance learning for students and improve higher education. That is, if programs are typically unable to close the assessment loop by using results to inform changes and demonstrate learning improvement, then assessment practice is falling short of its promise.

We addressed RQ 5 via the fifth stage of implementation criteria discussed previously for RQ 4. More specifically, change efforts coded as being at Stage 5 of implementation represented instances of closing the assessment loop (i.e., change efforts coded as "Stage 5: Completed and checked for efficacy (or effectiveness reassessed) and demonstrable improvement evidenced" were used to address RQ 5).

METHOD

Our home institution is a mid-sized, 4-year, public university in Virginia. The State Council of Higher Education for Virginia (SCHEV) and our regional accreditor (Commission on Colleges of the Southern Association of Colleges and Schools, or SACSCOC) require colleges and universities to assess student learning. In compliance with their respective policies and guidelines, all academic degree-granting programs at our institution submit annual assessment reports for student learning outcomes. Each year graduate students, faculty members, and assessment specialists evaluate these assessment reports. Through feedback and consultation, several programs at our institution have demonstrated better assessment processes (Rodgers, Grays, Fulcher, & Jurich, 2013).

For this study, we examined all 54 exemplary assessment reports collected from the fall 2012–2013 reporting cycle. Fifty-four represents approximately half of our academic degree and certificate programs. Exemplary assessment reports received a score of 3.4 or higher out of 4, on a meta-assessment rubric (see Appendix A) (Fulcher & Bashkov, 2012; Fulcher & Orem, 2010). The 3.4 standard was set in 2011 by trained faculty using a modified Angoff procedure.

Our review included only exemplary assessment reports for practical reasons; we hypothesized that academic programs with established, high-quality assessment processes might be best poised to use assessment results to influence pedagogical and curricular changes (and subsequently demonstrate learning improvement). They also might be better equipped to reassess students' learning to determine if the implemented changes actually promoted learning improvement. Furthermore, programs in nascent stages of assessment (not close to exemplary) are likely focused on setting up assessment infrastructure. Such programs are typically establishing learning objectives, creating curriculum maps, and selecting assessment instruments. These programs, therefore, are less likely to have collected data and synthesized them into actionable results. Of course, use of results is a moot point to those programs that have not collected data. In essence, by focusing on exemplary reports we could rule out undeveloped assessment practices as an explanation for not using results to improve student learning. Within each exemplary assessment report, we identified specific descriptions of using results for improvement and then used an online Qualtrics survey to code each of the identified descriptions.

Procedures for Identifying and Coding Descriptions of Results

To locate specific descriptions of using results for improvement, a graduate student familiar with the meta-assessment rubric (see Appendix A) and the layout of assessment reports reviewed electronic copies of all 2012–2013 assessment reports in alphabetical order according to program name. The graduate student first read Section 6A, "Program Modification and Improvement Regarding Student Learning and Development," of the assessment report; this section asks program assessment practitioners to describe use of assessment results for student learning improvement. If there were no examples or evidence of use of results to improve student learning in Section 6A, the graduate student reviewed other sections in the reports. If the graduate student initially found no evidence of use of results, she set the report aside. Later, she rereviewed the report, reducing the chance of an overlooked example.

For each assessment report, the graduate student identified up to four examples that described use of results by electronically highlighting sections of the report in yellow. Note, of the 54 exemplary assessment reports, there were only two that had more than four examples. After the initial review and electronic highlighting, the graduate student randomized the order of the assessment reports and rereviewed them, converting the yellow highlighting to highlighting in red, yellow, green, or blue (i.e., example one was highlighted in red, example two in yellow, etc.).

Once the graduate student had reviewed all 2012–2013 exemplary assessment reports and highlighted all identified descriptions of use of results for improvement, three authors of this paper—along with three other graduate students—independently evaluated and coded the using-results descriptions via an online Qualtrics survey. Specifically, raters reviewed all highlighted descriptions—each representing an individual "use of results"—in their assigned assessment reports. The raters evaluated the following aspects of the descriptions:

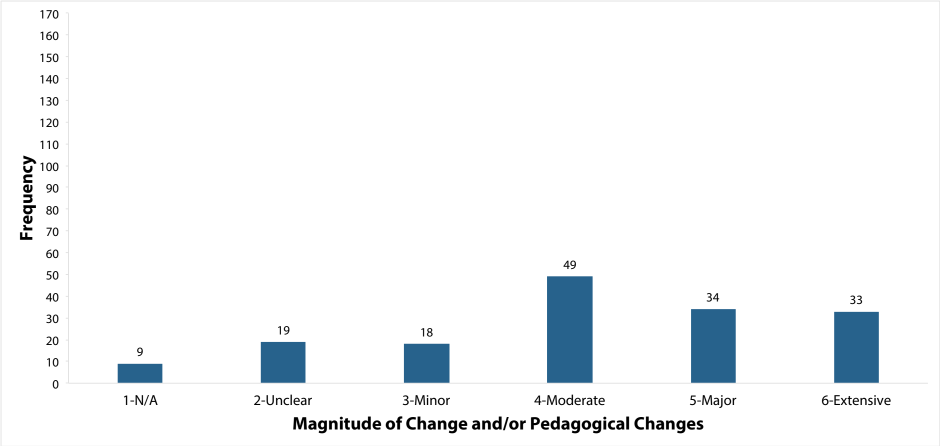
- Magnitude of change, defined by extent or magnitude of changes made to pedagogy, curricula, and so on (minor: changes to a small class assignment in one class; moderate: change to a unit within a class; major: major overhaul of a class; extensive: numerous changes that affect several classes)
- Extent to which faculty linked change to student learning objectives
- Rationale for needing change
- Reported stage of change implementation

The six raters were paired into three groups of two; each group was assigned a subset of the 54 exemplary assessment reports. Groups 1, 2, and 3 evaluated a total of 20, 21, and 13 different assessment reports, respectively. First, each rater independently coded the highlighted sections in every assigned assessment report, then each rater pair adjudicated to reach exact agreement on all coded sections. For instance, Raters 1 and 2 were paired and assigned 20 assessment reports to review; one of those assessment reports was from the Assessment & Measurement Ph.D. program. Each rater independently reviewed every highlighted description of using results within the Assessment & Measurement program assessment report. Then, using a Qualtrics survey, each rater coded the highlighted descriptions. Finally, they reviewed each other's ratings and adjudicated until they agreed on all ratings for the Assessment & Measurement report. Each rater pair repeated this process for every assigned assessment report.

RESULTS

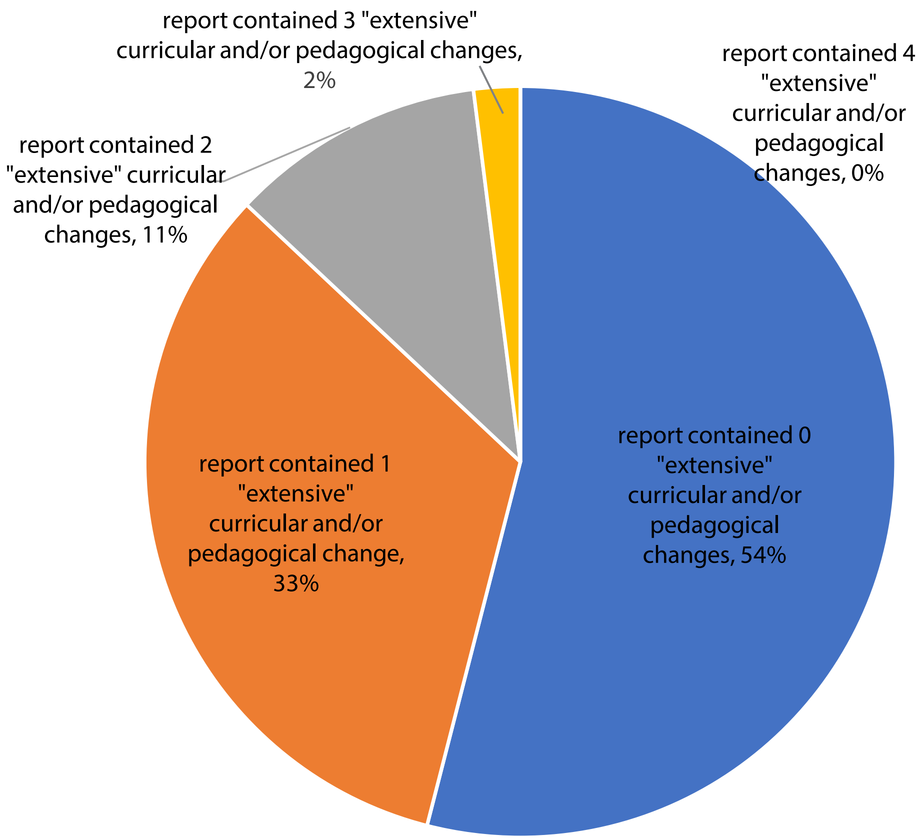
Across the 54 assessment reports, we identified and evaluated 162 different descriptions of using

Figure 2. Distribution of Magnitude of Curricular and/or Pedagogical Changes Across All Coded Program Assessment Reports



Association for Institutional Research

Figure 3. Percent of Program Assessment Reports That Contained Different Numbers of Curricular and/or Pedagogical Changes Coded as Being Extensive in Magnitude (e.g., Percent of Program Reports That Had Either 0, 1, 2, 3, or 4 Extensive Changes)



assessment results to make curricular or pedagogical changes. On average, we identified three descriptions per assessment report (M = 3.00, SD = 1.13). Clearly, reporting assessment data spurs talk of change. Nevertheless, only 8% of programs (among the 54 reports) could show that their pedagogical and curricular changes led to better learning outcomes. The following research questions (RQs) explore why so little learning improvement was reported despite so many changes within programs.

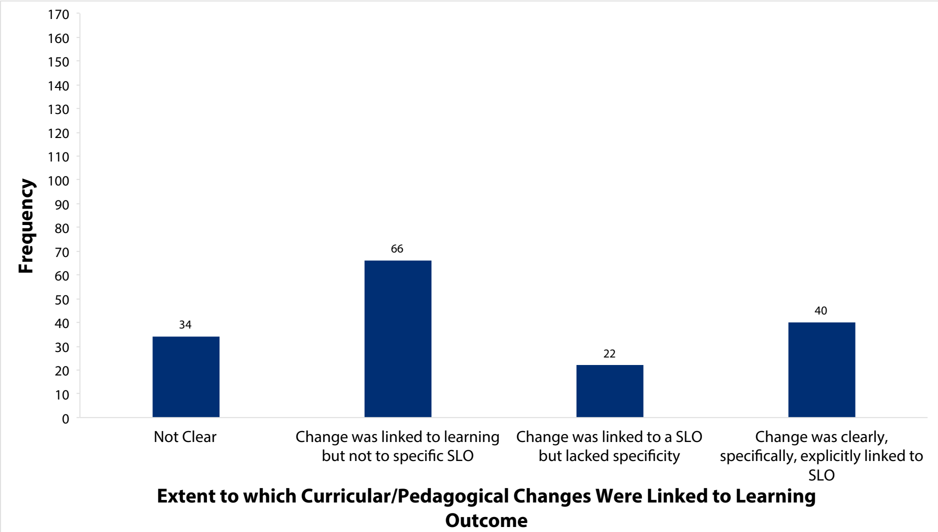
RQ 1. How extensive in magnitude are the reported changes to curricula and pedagogies?

Recall, researchers rated the magnitude of curricular and pedagogical changes described in programmatic assessment reports based on the reported magnitude of changes made to courses, curricula, pedagogies, and so on. For instance, a change that involved only minimal adjustments to one assignment in one course would be rated as minor. Such an adjustment would not be expected to have a demonstrable, positive effect on student learning, at the program or departmental level. Comparatively, a change that involved extensive modifications that affected multiple courses within the program or department would be expected to have a more demonstrable influence on program- or department-level student learning.

The magnitude of curricular and/or pedagogical changes was slightly negatively skewed (see Figure 2). In other words, the majority of the identified changes were coded as either moderate (a coded score of 4), major (a coded score of 5), or extensive (a coded score of 6) in magnitude. On average, the described changes were coded as moderate (M = 4.10, SD = 1.45). Nearly 20 of the identified changes were coded as unclear because, although faculty described a change, they did not provide enough information about the change to accurately identify its magnitude. For assessment reports in which faculty said they made a change, but then included no description of the change whatsoever, researchers applied the code "N/A."

Within each program assessment report, 54% had zero curricular and/or pedagogical changes coded as extensive in magnitude (see Figure 3). About 33% of the 54 assessment reports had one change coded as extensive in magnitude, 11% had two such extensive changes, and 2% had three. In addition, none of the 54 assessment reports contained four changes coded as extensive in magnitude. In essence, nearly half (46% of programs) reported the type of

Figure 4. Frequency of Identified Curricular and/or Pedagogical Changes That Were Linked or Aligned to Student Learning Outcomes



extensive pedagogical and curricular changes most often associated with learning improvement. However, these extensive changes equated to fewer examples of learning improvement than one might expect: only 8%. The results for RQs 2 to 4 provide more explanation to why these extensive changes led to so few examples of evidenced improvements.

RQ 2. To what extent are curricular and pedagogical changes linked to student learning outcomes?
Typically, curricular and pedagogical changes were linked to student learning generally (a coded score of 2), but were not matched to a specific, program-level student learning outcome (M = 2.40, SD = 1.08). As Figure 4 shows, approximately 34 out of the 162 identified curricular or pedagogical changes, or 21%, did not include enough details for raters to evaluate the alignment between the change and the program's student learning outcome(s). The lack of explicit alignment between changes and student learning outcomes might be contributing to the issue at hand: insufficient use of assessment results to evidence improved student learning.

For many, the link between curricular and pedagogical changes and specific student learning objectives might be implicit. However, documenting the use of assessment results to influence pedagogical or curricular changes that lead to improved student learning requires explicit connections between implemented changes and student learning outcomes. It seems that assessment practitioners and support services need to better conceptualize and articulate the importance of matching changes to student learning outcomes.

RQ 3. What is the rationale behind curricular and pedagogical changes?
About 80% of the identified descriptions of curricular or pedagogical changes provided a rationale that conveyed the need for change. But just over 50% of the descriptions of curricular or pedagogical changes provided a rationale and mentioned the source that supported the rationale (i.e., direct assessment measures, accreditation or program review recommendations, etc.). In addition, about 19% of the identified descriptions of curricular or pedagogical changes provided no rationale. The most frequently provided rationale behind the described curricular or pedagogical changes was data from direct assessment measures. In contrast, few cited accreditation/program review as a rationale for a given change; none mentioned curriculum realignment. Of the program assessment reports that provided a source explaining their intended curricular and/or pedagogical change, Figure 5 displays the percent of reports that cited various sources of rationales for changes.
Perhaps programs recognize the results of direct assessment measures, instead of feedback from accreditation/program reviews, as potential sources for change. In addition, some did not include any rationale to support changes to pedagogies and/or curricula. Perhaps the importance of understanding and describing the driving forces behind program-level changes is not recognized. Or, what might be a supportive rationale is not included because the report writer(s) believed the rationale was implied.

Figure 5. Of the Program Assessment Reports That Provided a Rationale and Source Explaining Their Intended Curricular and/or Pedagogical Change, Percent of Reports That Cited Various Sources of Rationales for Changes

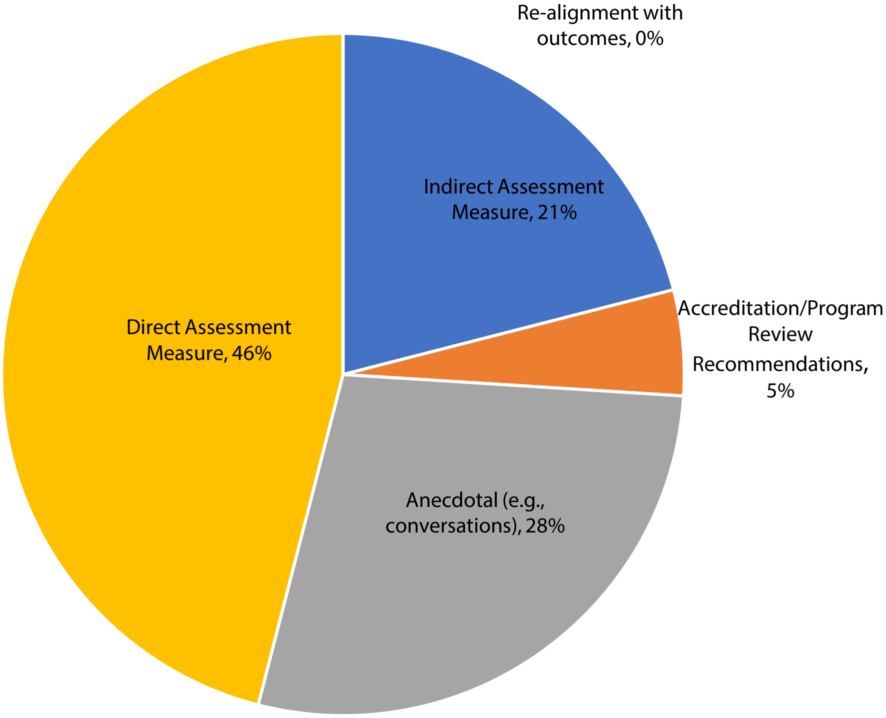
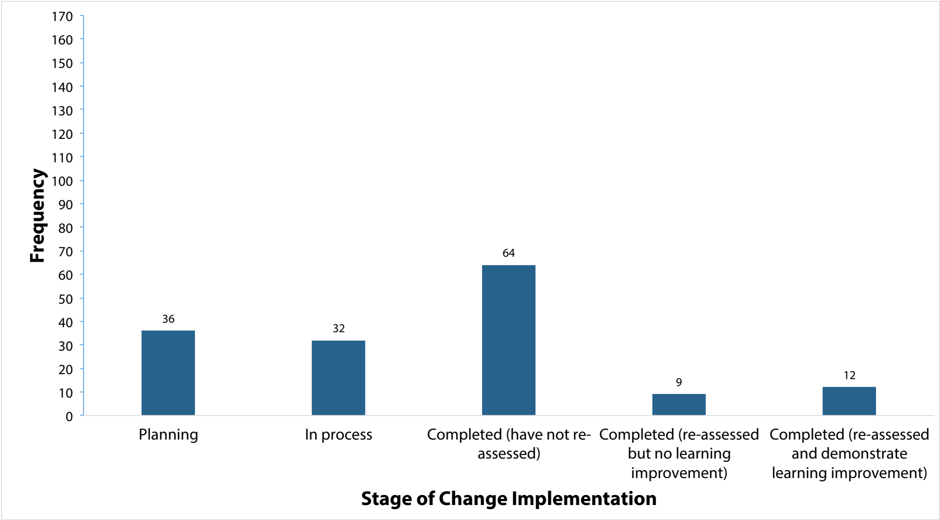


Figure 6. Distribution of Stage of Change Implementation Ratings Across All Coded Program Assessment Reports



Note, the meta-assessment rubric used at our institution in 2012–2013, the year of these reports (see Appendix A), does not require an explicit rationale to support curricular or pedagogical changes. Nonetheless, explicitly describing the rationale underlying change is an essential part of using results to demonstrably improve student learning (Fulcher et al., 2014). Given assessment measures were the most frequently cited rationale for curricular and pedagogical changes, intrinsic buy-in for change might be nonexistent. Alternatively, curricular and pedagogical changes that lack adequate rationale might not be well aligned with students' learning needs, program resources, faculty sentiments, or administrative agendas.

RQ 4. What is the typical stage of implementation for curricular and pedagogical changes?

Encouragingly, about 56% (85 out of 153) of the described curricular and pedagogical changes were complete. Yet, only 14% (21 out of 153) of all described curricular and pedagogical changes included follow-up reassessments (see Figure 6). Again, in 2012–2013 the crucial reassessment phase had not been explicitly stated in our institutional assessment cycle nor in our meta-assessment rubric (see Appendix A). Therefore, programs might not have been aware of the importance of reassessing. Alternatively, many might mistakenly believe that assessment work is done as soon as data are used for curricular and pedagogical change.

As Fulcher and colleagues (2014) explain, assessment practitioners, faculty members, and other stakeholders often confuse program changes with program improvements. A change is only an improvement when, upon reassessment, students demonstrate greater proficiency. Essentially, merely implementing curricular or pedagogical changes does not provide demonstrable proof of improved student learning, just as a pig never fattened because it was weighed. Assessment practitioners can do a better job of articulating and promoting the use of assessment results for improved student learning.

RQ 5. To what degree are programs able to close the assessment loop by using results to inform changes and subsequently demonstrate improved student learning?

As foreshadowed at the beginning of this section, only 8% of the evaluated curricular and pedagogical changes were implemented, reassessed, and demonstrated improved student learning. Our interpretation of this finding is that either programs are not closing the loop or our university programs do not know how to articulate such a process in an assessment report. Little integration of assessment processes with pedagogy and curricular design suggests a lack of clarity about learning improvement within our institution.

CONCLUSION

Even after more than 25 years of assessment practice at our university, finding evidence of student learning improvement in assessment reports is akin to finding a needle in a haystack. To understand more about this most important phase of the assessment cycle, we qualitatively reviewed and coded 54 exemplary assessment reports from academic programs across our campus. In these assessment reports, writers described changes to course scaffolding, use of different classroom pedagogies, course redesigns, and so on. Furthermore, the curricular and pedagogical changes described were typically coded as being moderate in magnitude and were primarily driven by data from direct assessment measures.

However, under scrutiny, the thread from the "Use of Results" section to demonstrable student learning was typically thin and loose. Few programs could demonstrate the positive impacts of the curricular and pedagogical changes they made. Based on descriptions in the assessment reports, programs rarely conducted follow-up reassessment research to determine whether curricular and pedagogical changes had a demonstrable impact on student learning outcomes. Perhaps this finding can help explain why use of assessment results has not contributed enough to improving student learning outcomes in higher education (Kuh, Jankowski, Ikenberry, & Kinzie, 2014).

The inability to empirically demonstrate improved student learning was not for lack of earnest efforts to improve. That is, some programs conceptualized curricular and pedagogical changes, provided some rationale to support these changes, and implemented the changes in their entirety. Yet, many of the program assessment reports lacked one or more critical elements, including

- Major or extensive pedagogical changes (i.e., changes at the program level);
- Tenable links between curricular and pedagogical changes and student learning outcomes;
- Convincing rationales to support curricular and pedagogical changes; and
- Adequate reassessment processes that can determine whether changes actually improved student learning.

Assessment Practitioners' Role in Bridging the Gap between Using Results and Demonstrating Student Learning Improvement

In general, higher education stakeholders have not successfully evidenced systematic improvements in student learning at the academic program level. While making some progress, our institution certainly struggles. From a policy perspective, being a good shepherd of resources suggests that institutions are making earnest efforts to improve. Academe's lack of demonstrating such improvement definitely contributes to the "Is college worth it?" conversation (Taylor et al., 2011).

To answer questions of worth and demonstrate the value of a college education, assessment results need to influence pedagogical and curricular changes at a program level. Ultimately, explicit gains in student learning should be clearly articulated via assessment reports, presentations, and other channels of dissemination. Assessment practitioners must do more to communicate the importance of student learning improvement initiatives.

Findings from the current study reflect Blaich and Wise's (2011) observation that excellent assessment—by itself—does not lead to learning improvement. In addition, our results suggested that practitioners could increase student learning improvement by helping programs

1. Develop and implement more widespread and multiyear curricular and pedagogical changes;
2. Situate improvement efforts within student learning outcomes;
3. Understand the important role of reassessment; and
4. Use a framework or step-by-step example to more effectively report and explain crucial information.

As Fulcher and Bashkov (2012) explain, we should not be surprised that assessment reports lack adequate descriptions of using results to demonstrate improved student learning. At our institution, we did not offer enough guidance with respect to how to report our learning improvement efforts. In addition, we realized that we have no assessment staff trained in pedagogy, curriculum, course redesign, course scaffolding, or organizational change. Lacking a holistic expertise within our own assessment office led us to engage in more-intentional partnership with our campus faculty development center. Doing so allows us to better serve faculty members as they create and implement curricular and pedagogical changes, and then reassess students' learning. For instance, the faculty development experts assist programs as they articulate student learning outcomes and align them with program theory.

We hope that the recommendations from the current study can assist institutions in better conceptualizing, articulating, implementing, reporting, and disseminating learning improvement success stories. We should also note that changes of greater magnitudes, alignment of actions, reassessing to determine effect of actions, and providing step-by-step examples for improvement can be extended beyond learning. The general principles could be applied to retention efforts, donor giving, or other important efforts. The more we discuss improvement, the better institutional decision makers we become.

Study Limitations and Future Directions

Thus far we have evaluated assessment reports from only one institution. Our findings might not reflect other institutions, especially those with different assessment practices and educational research initiatives. In addition, we have evaluated reports from only a single year's reporting cycle. Replicating our study across various reporting cycles, and across institutions, would reveal potential longitudinal trends and could provide external validity evidence for our findings.

In addition, future research should include interviews with faculty members who crafted the assessment reports. Through these qualitative data, institutional effectiveness researchers could further investigate faculty perceptions of the magnitude of their described changes to curriculum and pedagogy. A rigorous qualitative follow-up study could also provide crucial insights from faculty members to clarify why certain types of information and explanations were absent from the reviewed assessment reports.

REFERENCES

Banta, T. W., & Blaich, C. F. (2011, January). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 22–27.

Blaich, C. F., & Wise, K. S. (2011, January). *From gathering to using assessment results: Lessons from the Wabash National Study* (NILOA Occasional Paper No. 8). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension.* (NILOA Occasional Paper No. 1). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Fulcher, K. H. &, Bashkov, B. M. (2012, November). Do we practice what we preach? The accountability of an assessment office. *Assessment Update*, 24(6), 5–7, 14.

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig*. (NILOA Occasional Paper No. 23). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Fulcher, K. H. & Orem, C. D. (2010). Evolving from quantity to quality: A new yardstick for assessment. *Research and Practice in Assessment*, 5, 13–17.

Fulcher, K. H., Swain, M. S., & Orem, C. D. (2012). Expectations for assessment reports: A descriptive analysis. *Assessment Update*, 24(1), 1–2, 14–16.

Kinzie, J., Hutchings, P, & Jankowski, N.A. (2015). *Fostering greater use of assessment results: Principles for effective practice*. In G. D. Kuh, S. O. Ikenberry, N. A. Jankowski, T. R. Cain, P. T. Ewell, P. Hutchings, P., & J. Kinzie (Eds.), Using evidence of student learning to improve higher education (pp. 51–72). San Francisco, CA: Jossey-Bass.

Kuh, G. D., & Ikenberry, S. O. (2009). *More than you think, less than we need: Learning outcomes assessment in American higher education*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014). *Knowing what students know and can do: Current state of student learning outcomes assessment in U.S. colleges and universities.* Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Jonson, J. L., Guetterman, T., & Thompson, R. J. (2014). An integrated model of influence: Use of assessment data in higher education. *Research and Practice in Assessment*, 9, 18– 30.

Rodgers, M., Grays, M. P., Fulcher, K. H., &

Jurich, D. P. (2013) Improving academic program assessment: A mixed methods study. *Innovative Higher Education*, 38(5), 383–395.

Smith, K. L., Good, M. R., Sanchez, E. H., & Fulcher, K. H. (2015). Communication is key: Unpacking "use of assessment results to improve student learning." *Research and Practice in Assessment*, 10, 15–29.

Suskie, L. (2010). Why are we assessing? *Inside Higher Education*. Retrieved from: https://www.insidehighered.com/views/2010/10/26/suskie

Taylor, P., Parker, K., Fry, R., Cohn, D., Wang, W., Velasco, G., & Dockterman, D. (2011). Is college worth it? College presidents, public assess value, quality and mission of higher education. Pew Research Center, Washington, D.C. Retrieved from http://www.pewsocialtrends.org/2011/05/15/is-college-worth-it/

Association for Institutional Research

Appendix A. Assessment Progress Template (APT) Evaluation Rubric as Described in Fulcher & Orem (2010)

Assessment Progress Template (APT) Evaluation Rubric

1 – Beginning	2 – Developing	3 – Good	4 – Exemplary
1. Student-centered learning objectives			
A. Clarity and Specificity			
No objectives stated.	Objectives present, but with imprecise verbs (e.g., know, understand), vague description of content/skill/or attitudinal domain, and non-specificity of whom should be assessed (e.g., "students")	Objectives generally contain precise verbs, rich description of the content/skill/or attitudinal domain, and specification of whom should be assessed (e.g., "graduating seniors in the Biology B.A. program")	All objectives stated with clarity and specificity including precise verbs, rich description of the content/skill/or attitudinal domain, and specification of whom should be assessed (e.g., "graduating seniors in the Biology B.A. program")
B. Orientation			
No objectives stated in student-centered terms.	Some objectives stated in student-centered terms.	Most objectives stated in student-centered terms.	All objectives stated in student-centered terms (i.e., what a student should know, think, or do).
2. Course/learning experiences that are mapped to objectives			
No activities/ courses listed.	Activities/courses listed but link to objectives is absent.	Most objectives have classes and/or activities linked to them.	All objectives have classes and/or activities linked to them.
3. Systematic method for evaluating progress on objectives			
A. Relationship between measures and objectives			
Seemingly no relationship between objectives and measures.	At a superficial level, it appears the content assessed by the measures matches the objectives, but no explanation is provided.	General detail about how objectives relate to measures is provided. For example, the faculty wrote items to match the objectives, or the instrument was selected "because its general description appeared to match our objectives."	Detail is provided regarding objective-to-measure match. Specific items on the test are linked to objectives. The match is affirmed by faculty subject experts (e.g., through a backwards translation).
B. Types of Measures			
No measures indicated	Most objectives assessed primarily via indirect (e.g., surveys) measures.	Most objectives assessed primarily via direct measures.	All objectives assessed using at least one direct measure (e.g., tests, essays).
C. Specification of desired results for objectives			
No a priori desired results for objectives	Statement of desired result (e.g., student growth, comparison to previous year's data, comparison to faculty standards, performance vs. a criterion), but no specificity (e.g., students will grow; students will perform better than last year)	Desired result specified. (e.g., our students will gain ½ standard deviation from junior to senior year; our students will score above a faculty-determined standard). "Gathering baseline data" is acceptable for this rating.	Desired result specified and justified (e.g., Last year the typical student scored 20 points on measure x. The current cohort underwent more extensive coursework in the area, so we hope that the average student scores 22 points or better.)
D. Data collection & Research design integrity			
No information is provided about data collection process or data not collected.	Limited information is provided about data collection such as who and how many took the assessment, but not enough to judge the veracity of the process (e.g., thirty-five seniors took the test).	Enough information is provided to understand the data collection process, such as a description of the sample, testing protocol, testing conditions, and student motivation. Nevertheless, several methodological flaws are evident such as unrepresentative sampling, inappropriate testing conditions, one rater for ratings, or mismatch with specification of desired results.	The data collection process is clearly explained and is appropriate to the specification of desired results (e.g., representative sampling, adequate motivation, two or more trained raters for performance assessment, pre-post design to measure gain, cutoff defended for performance vs. a criterion)
E. Additional validity evidence			
No additional psychometric properties provided.	Reliability estimates (e.g., internal consistency, test-retest, inter-rater) provided for most scores, although reliability tends to be poor (<.60). Or, author states how efforts have been made to improve reliability (e.g., raters were trained on rubric).	Reliability estimates provided for most scores, most scores are marginal or better (>.60).	Reliability estimates provided, most scores are marginal or better (>.60). Plus, other evidence given such as relationship of scores to other variables and how such relationship strengthens or weakens argument for validity of test scores.
4. Results of program assessment			

Appendix A continued on next page

Appendix A continued

A. Presentation of results			
No results presented	Results are present, but it is unclear how they relate to the objectives or the desired results for the objectives.	Results are present, and they directly relate to the objectives and the desired results for objectives but presentation is sloppy or difficult to follow. Statistical analysis may or may not be present.	Results are present, and they directly relate to objectives and the desired results for objectives, are clearly presented, and were derived by appropriate statistical analyses.
B. History of results			
No results presented	Only current year's results provided.	Past iteration(s) of results (e.g., last year's) provided for some assessments in addition to current year's.	Past iteration(s) of results (e.g., last year's) provided for majority of assessments in addition to current year's.
C. Interpretation of Results			
No interpretation attempted	Interpretation attempted, but the interpretation does not refer back to the objectives or desired results of objectives. Or, the interpretations are clearly not supported by the methodology and/or results.	Interpretations of results seem to be reasonable inferences given the objectives, desired results of objectives, and methodology.	Interpretations of results seem to be reasonable given the objectives, desired results of objectives, and methodology. Plus, multiple faculty interpreted results (not just one person). And, interpretation includes how classes/activities might have affected results.
5. Documents how results are shared with faculty/stakeholders			
No evidence of communication	Information provided to limited number of faculty or communication process unclear.	Information provided to all faculty, mode and details of communication clear.	Information provided to all faculty, mode and details of communication clear. In addition, information shared with others such as advisory committees, other stakeholders, or to conference attendees.
6. Documents the use of results for improvement			
A. Improvement of programs regarding student learning and development			
No mention of any improvements.	Examples of improvements documented but the link between them and the assessment findings is not clear.	Examples of improvements (or plans to improve) documented and directly related to findings of assessment. However, the improvements lack specificity.	Examples of improvements (or plans to improve) documented and directly related to findings of assessment. These improvements are very specific (e.g., approximate dates of implementation and where in curriculum they will occur.)
B. Improvement of assessment process			
No mention of how this iteration of assessment is improved from past administrations.	Some critical evaluation of past and current assessment, including acknowledgement of flaws, but no evidence of improving upon past assessment or making plans to improve assessment in future iterations.	Critical evaluation of past and current assessment, including acknowledgement of flaws; Plus evidence of some moderate revision, or general plans for improvement of assessment process.	Critical evaluation of past and current assessment, including acknowledgement of flaws; both present improvements and intended improvements are provided; for both, specific details are given. Either present improvements or intended improvements must encompass a major revision.

This page left intentionally blank

PROFESSIONAL FILE

ARTICLE 142

Association for Institutional Research

A CASE STUDY TO EXAMINE THREE PEER GROUPING METHODOLOGIES

Mary Lou D'Allegro

About the Author
Mary Lou D'Allegro is associate provost at Paul Smith's College.

Acknowledgments
This paper is an update to a previous study published in *AIR Professional Files:* M. L. D'Allegro and K. Zhou, "A Case Study to Examine Peer Grouping and Aspirant Selection," *Professional Files* (Fall 2013), Association for Institutional Research. The following faculty inspired the author to develop additional novel peer selection indices not noted in previously published studies: Aaron Pacitti, Douglas T. Hickey Chair in Business and Associate Professor of Economics; and John Cummings, Dean of Science and Professor of Physics, both at Siena College. Thank you for intelligent and imaginative approaches in selecting institutional peers.

Abstract
This study considered three selection indices to choose institutional peers: (a) proximity, (b) percentile, and (c) normative. Although conceptually similar, only the proximity selection index had been previously studied. The purpose of this paper is threefold. First, the procedures used to generate the peer sets for each selection index are provided. Second, an empirical investigation was conducted to compare the institutional peers chosen by each selection index using those procedures. Third, the stability of peer selection over time was also ascertained from that enquiry.

Compiled separately from two data sets extracted three years apart, the three selection indices under investigation yielded remarkably different sets of peers. Fewer than half of the institutions used in this study were identified as peers at both points of time. Additional analyses revealed that the underlying distributions of the characteristics used to select peers might be just as influential as the characteristics themselves. The results did not produce sufficient evidence to endorse any one of the selection indices, but instead suggest that a combination of selection indices might be superior to any one selection index alone.

BACKGROUND

The continued increase in public scrutiny of higher education, the expanded demands of accountability, and the overall cynicism of the value of higher education have put colleges and universities on high alert. To counter this skepticism, colleges and schools have increased their efforts to evaluate their quality, efficiency, and effectiveness (Ruben, 2004). A growing and important segment of that evaluation is the comparison and benchmarking to like institutions (Qayoumi, 2012). Therefore, peer selection has become more prevalent. Moreover, higher education has seen the benefit of using peer comparisons and benchmarking to inform decision making and strategic planning.

This research builds on previous work that examined the methodology to choose a set of institutional peers. Specifically, that research investigated the usefulness of the proximity selection index and proposed standardized equation to foster ease of replication. In that work, the proximity selection index was deemed to be an appropriate methodology for the selection of a generic set of institutional peers (D'Allegro & Zhou, 2013). For this research, an institutional peer is defined as institutions that are similar with regard to certain delineating factors (Anderes, 1999; Trainer, 2008). A selection index is a numerical designation system to indicate the extent to which an institution is a potential peer.

Faculty proposed to the researcher two different approaches to peer selection indices that were not considered in the researcher's previously published work. The faculty's suggestions seemed rational because their methodologies might temper potential irregularities in the data. Particularly, their proposed selection indices either (a) relied on statistics that were less susceptible to the perils of non-normal distributions than the standard deviation used in the proximity selection index or (b) standardized the distribution so that imperfections in the data were minimized. As will be discussed in the "Methodology" and "Results" sections, non-normal distributions can acutely affect the set of peer institutions that are selected. This further confirmed that the process for determining peers seems to be arbitrary (Anderes, 1999). Accordingly, there is little or no evidence to the quality or adeptness of many processes to select a set of peers. Careful planning and investigation of the criteria used to select a set of institutional peers is still advised, but the researcher realized the frailty of even the most careful undertaking of selecting a set of institutional peers, including the conclusions of previous research.

At the heart of the paper is the description of three different selection indices and the ensuing peer sets created by each. Those selection indices were similar to the nearest neighbor rationale (McLaughlin, Howard, & McLaughlin, 2011). For all three selection indices, the distance between any given institution or comparison institution and the target institution on predetermined parameters was calculated. The divergence among selection indices is their underlying distributions. Correspondingly, the primary purposes of the study were to: (a) determine and document the differences, if any, in the institutional peer sets produced by each selection index; (b) conclude, from any differences, what index is best; and (c) ascertain the stability of peer selection over time.

METHODOLOGY

This study does not abandon previously applied principles and, as such, uses a variety of sources and methods to maintain a practical balance between stakeholder judgment and statistical analysis (Trainer, 2008). Credibility of the institutional peer sets relies on constituent input. Not only were faculty and staff consulted for this compilation, but in addition the concept for the alternative selection indices arose from the propositioned reasoning of two faculty members. Hence, selection methodologies were based primarily on constituent suggestions and on other documented peer selections.

In the original research, an attempt to find a quick, pragmatic method to choose a set of peers from two or three institutional characteristics was unsuccessful. Using different combinations of those institutional characteristics, it was discovered that the resulting peer sets were similar to the target institution with respect to some data elements but different with respect to others. Those differences were substantial enough to render the selection process ineffectual. This reinforces previous findings that institutional characteristics alone are not sufficient in choosing institutional peers (Shin, 2009).

Instead, a more-informed and more-comprehensive process was tested. The selection process entailed five steps outlined by D'Allegro and Zhou (2013): (a) identifying an initial set of peers, (b) choosing the preliminary set of variables, (c) transforming and standardizing variables, (d) determining the best set of variables to use, and (e) establishing the best selection strategy. This research is fundamentally undistinguishable from that research except for the last step. Therefore, a pithy summary of Steps 1–4 are provided, along with a comprehensive description of Step 5.

1. Identifying an Initial Set of Peers

The initial set of peers was selected a priori to this study. To recap, an initial set of institutional characteristics was identified to eliminate from further analysis institutions that would not realistically be considered a peer of the target institution. The initial set of institutions was chosen from an original list of private, nonprofit institutions that submitted data to the Integrated Postsecondary Education Data System (IPEDS) from the Data Compare Institutions website. The list was generated using the EZ group option (National Center for Education Statistics [NCES], 2012). Data for these institutions were collected for 2010 and 2011; these were the most recent data available at the time of the previous study. An updated data set was identically assembled using 2014 and 2015 information; these were the most recent data available at the time of this study. Note that for the target institution, the 2015 Basic Carnegie Classification did not change from 2010 (Carnegie Foundation, 2015). Furthermore, only the basic 2015 Basic Carnegie Classification was currently available on the EZ group option. Lists for both time periods were generated using the following criteria: (a) private not-for-profit institutions, 4-year or above; (b) highest degree awarded either a bachelor's degree, master's degree, or both; (c) baccalaureate college for arts and sciences, or baccalaureate college balanced arts and sciences, diverse fields; (d) enrolled full-time undergraduate students; (e) institution size between 1,000 and 9,999 students; (f) Title IV participant (federal financial aid eligibility); (h) located in the United States or designated as a U.S. Service School (e.g., U.S. Naval Academy), and (i) not a tribal college. These parameters align with the characteristics of the target institution. This is also on par with selection parameters recommended by previous studies (Anderes, 1999). As a result of applying these criteria, 285 institutions were selected for the previous study while the updated listed yielded 232 institutions.

2. Choosing the Preliminary Set of Variables

Other pertinent information was collected for each of these institutions. Relevance in the context of selecting peers are those data points that indicate the institution's priorities (Anderes, 1999; Cohodes & Goodman, 2012). For the most part, an institution's focus is on quality. As such, the target institution's own Key Performance Indicators (KPIs) were the starting point. KPIs are a mix of approximately 20 output or direct measures of quality and input or influencers of quality. Therefore, the initial set of variables chosen either had some influence on quality or included direct measures of institutional performance. Faculty and staff were also asked to rate the importance of each KPI, being mindful of the importance of using both input and output variables in the peer selection process.

The data also had to be easy to access for all or most institutions. Several sources were considered including: (a) National Survey of Student Engagement (NSSE) benchmarks, (b) American Association of University Professors (AAUP) Faculty Compensation Survey (2012), (c) Noel Levitz Student Satisfaction Inventory (NLSSI), and (d) *U.S. News & World Report* rankings (*U.S. News & World Report*, 2015). Nevertheless, not all institutions participate in the NSSE or NLSSI or administer these surveys within a reasonable time period to avail comparisons. Also, detailed AAUP faculty salary data are not available for many institutions. Consequently, data were obtained from IPEDS or the *U.S. News & World Report* rankings.

The preliminary set of 28 variables are shown in Appendix A, along with the institutional characteristics used to select the initial set of peers. Note that the KPIs have remained the same and, therefore, the faculty were not consulted again for this study. Therefore, no adjustments were needed for the updated data set.

3. Transforming and Standardizing Variables

There was a fair amount of variability in enrollment among the initial set of institutions. Moreover, the enrollment of the target institution was twice the size of most of the institutions in both data sets. Therefore, some of the data elements were standardized to mitigate differences due to institutional size (Gater, 2003; Huxley, 2009). This was accomplished by using the full-time equivalent (FTE) for enrollment as the divisor. Examples of data elements that were standardized by dividing by the FTE included the number of conferred bachelor's degrees, number of applicants, unduplicated annual enrollment, instructional expenses, and endowment.

Full-time and part-time faculty counts were combined into one data element. In effect, the proportion of full-time faculty was calculated by dividing the sum of full-time plus part-time faculty into the number of full-time faculty.

4. Determining the Best Set of Variables to Use

Of the 28 variables identified in Step 2, three were both output measures and among the target institution's KPIs: (a) ratio of conferred bachelor's degrees to FTE, (b) 1-year retention rate, and (c) 6-year graduation rate. These variables were also student centered—specifically student success focused—and aligned with the target institution's mission. To augment the data analysis and simplify its interpretation, the remaining variables were classified into

Table 1. Overall OLS Regression Models for the Three Performance Indicators: Ratio of Conferred Bachelor's Degree to FTE, 1-Year Retention Rates, and 6-Year Graduation Rates

		Standardized
Category	**Variable**	**Beta Coefficient**
Original Data Set		
Ratio of Conferred Bachelor's Degrees to FTE		
Admissions	25th Percentile Mathematics SAT	.348*
Faculty	Average Faculty Salary	–.142
Enrollment	Estimated Fall Enrollment per FTE	–.053
Institutional Characteristics	Selectivity	–.282**
Finance	Instructional Expenses per FTE	.166
1-Year Retention Rates		
Admissions	25th Percentile Mathematics SAT	.465***
Faculty	Average Faculty Salary	.135
Enrollment	FTE	.064
Institutional Characteristics	Selectivity	.301***
Finance	Instructional Expenses per FTE	.065
6-Year Graduation Rates		
Admissions	Percent of Students Receiving Federal Grant Aid	–.145**
Faculty	Average Faculty Salary	.211**
Enrollment	FTE	.090
Institutional Characteristics	Selectivity	.274**
	Proportion of Transfer Students	–.104**
Finance	Total Price of Attendance	.007
	Instructional Expenses per FTE	.224***
Updated Data Set		
Ratio of Conferred Bachelor's Degrees to FTE		
Admissions	Applicants per FTE	–.141*
Faculty	Percent of Faculty with Terminal Degree	.254**
Enrollment	12-Month Enrollment per FTE	.065
Institutional Characteristics	Selectivity	.038
Finance	Total Price of Attendance	.393***

1-Year Retention Rates		
Admissions	75th Percentile Mathematics SAT	.383***
Faculty	Average Faculty Salary	.086
	Percent of Faculty with Terminal Degree	.054
Enrollment	FTE	.131*
	12-Month Enrollment per FTE	–.050
Institutional Characteristics	Selectivity	.130
Finance	Total Price of Attendance	.089
	Alumni Giving Rate	2.229*
6-Year Graduation Rates		
Admissions	75th Percentile Mathematics SAT	.350***
Faculty	Average Faculty Salary	–.015
	Percent of Faculty with Terminal Degree	.127**
Enrollment	FTE	.158***
	12-Month Enrollment per FTE	–.040
Institutional Characteristics	Selectivity	.132*
Finance	Total Price of Attendance	.181**
	Alumni Giving Rate	.202***

*Note: * $p \le .05$, ** $p \le .01$, *** $p \le .001$.*

one of the following five groups: (a) admissions, (b) faculty, (c) enrollment, (d) institutional characteristics, and (e) finance.

As described in our previous research, several regression analyses, single-step ordinary least square (OLS), were used to identify the best variables to select a set of peers. In the first phase, regression models were compiled separately for the five variable categories for each of the three output measures, a total of 15 models. Because the analysis was still exploratory at this stage, the single-step enter method was preferred over other models. Distributing the variables into five groups allowed the inclusion of all variables into the model for that category (SPSS, 2012). Informed by previous research, the standardized beta weights were the determinants of what data elements would be used for peer selection (Hom, 2008).

In the second phase, an overall regression model for each output variable was computed using the best predictor(s) from each of the five regression models. The best predictor(s) had the smallest significance level associated with the standardized beta coefficient. The standardized beta weight's significance level indicates if a variable is, in fact, a predictor of the output variable (Cohen & Cohen, 1983). Although there were some exceptions, only one predictor from each category was chosen for the three overall models. This was deliberate because there were high correlations among predictors in any given category. In addition, the inclusion of only one or two predictors from each category forced a balance of institutional metrics for peer selection. The best predictors for each KPI regression model by category for the original and updated data sets are listed in Table 1.

5. Establishing the Best Selection Strategy

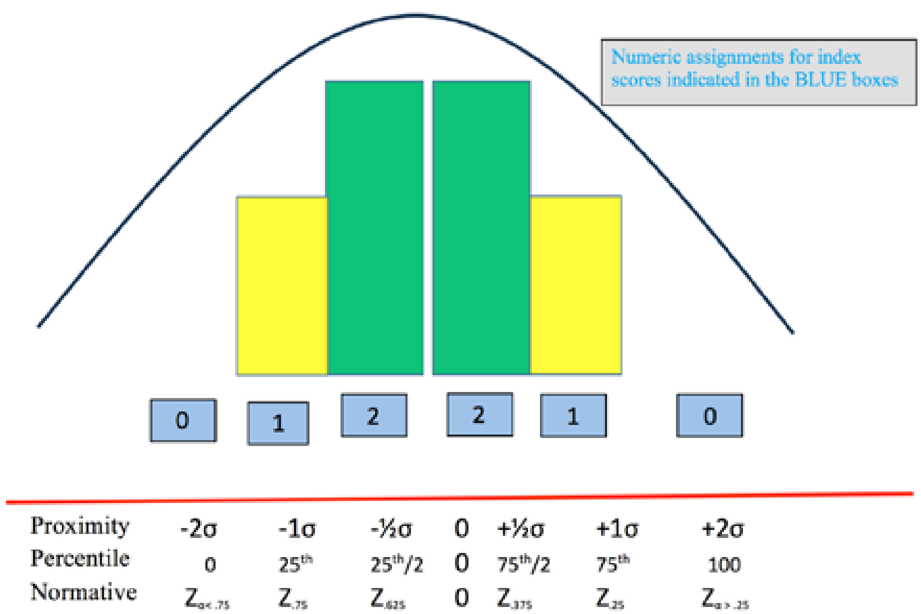
Peer institutions are determined by having metrics that are close to the target institution (McLaughlin et al., 2011). This is manifested in the computation of a selection index. Three selection indices were examined: (a) proximity, (b) percentile, and (c) normative.

The calculation of each selection index also involves several steps but the steps are basically the same for each: (a) identifying the most relevant parameters, (b) computing the numerical difference between the comparison and target institutions on each of those parameters, (c) averaging those differences across parameters, and (d) determining range cut-scores to delineate a peer from an almost-peer. The first step, identifying the most relevant parameters, has already been decided by the three overall OLS models mentioned in Step 4. Descriptions of Steps b–c are provided for each index below. The determination of range cut-scores are further described in the "Results" section.

Proximity selection index

As mentioned, the numeric differences between the target and each comparison institution were computed for each predictor. The mean of these differences determines an institution's propinquity to the target institution. For the proximity selection index, the unit of measurement is the standard deviation for each predictor. This is depicted in Figure 1, with the assumption for this illustration that the underlying data distribution for each predictor is normally distributed. For each predictor, a proximity index score of 1 was assigned to the comparison institution that was between one-half and one standard deviation of the target institution's metric, a score of 2 was given if the comparison institution was within one-half a standard deviation. Equally weighted, the average of the proximity index scores derives the proximity selection index. The two equations that compose the proximity selection index calculation are shown in Appendix B. An example on how to calculate the proximity selection index is provided in Appendix C.

Figure 1. Selection Index Numeric Assignments for Differences Between Target College and Each Institution in the Initial Data Sets



Percentile selection index

For the percentile selection index, differences between the target and each comparison institution were determined for each predictor as it was for the proximity selection index. Moreover, the logic is the same and is shown in Figure 1. However, the boundaries for each percentile index score is determined by the first and third quartile cut-scores, and not by the data distribution's standard deviation as it was for the proximity selection index. In effect, the percentile selection index ensures an equal number of comparison institutions in each partition.

A slight diversion is in order. Normal distributions are not assumed and skewed variables can still produce accurate results (Smith, 2012). Yet, extreme values or outliers on the low end or high end of the distribution can affect or skew the distribution and drag the mean away from a true measure of central tendency. Outliers on both ends might also affect the distribution's kurtosis. Kurtosis refers to the width of the peak of the distribution around the measure of central tendency (Hembree, 2013). In turn, this exaggerated dispersion could unduly increase the standard deviation and, thus, stretch the distribution segments. Consequently, a disproportional number of comparison institutions would receive larger index scores than they deserve because they would be more likely to fall in a subdivision closer to the mean. This might not be a problem per se, but could compromise the ability of the selection index to distinguish a peer from a non-peer.

On the other hand, the percentile selection index distribution is partitioned with an equal number of comparison institutions in each section. Unlike the proximity selection index, outliers are less likely to affect the percentile selection index because the percentile selection index relies on the median as the center of the distribution and not a potentially displaced mean. Therefore, the percentile selection index could be advantageous to the proximity selection index, especially for skewed data distributions.

For each predictor, a percentile index score of 1 was assigned to the comparison institution that was within 25 percentile points of the target institution metric, and a score of 2 was given if the comparison institution was within 12.5 percentile points of the target institution. This is a smaller partition than the proximity selection index, given a percentile index score greater than 0 is awarded if the comparison institution is within 50 percentile points or half the percentile selection index distribution versus approximately 68% of the proximity selection index distribution. Equally weighted, the average of the percentile index scores derive the percentile selection index. The two equations used for computing the percentile selection index are shown in Appendix B. An example of how to calculate a percentile selection index is provided in Appendix C.

Normative selection index

Before the boundaries for each normative selection index were established, values for each predictor were converted to z-scores. Each predictor was standardized with the resulting distribution having a mean of 0 and standard deviation of 1 (SPSS, 2012). That said, the standard normal distributions were derived from using the original distribution's mean and standard deviation. Therefore, effects of the outliers and resulting asymmetrical distributions were not completely eradicated. However, the advantage of these transformed distributions is the fact that the new distributions were symmetrical. In essence, the normative selection index is a hybrid of both the proximity and percentile selection indices. As with the proximity selection index, the mean and standard deviation determine distance or probability. However, as with the percentile selection index, the use of the standard normal distribution, ensures that the distribution is sectioned into equal parts.

Another benefit of transforming the original distribution to the standard normal distribution is that the cut-points are easier to compute and conceptualize. As mentioned, the curve created by the z-scores represented by the x-axis and resulting probabilities plotted on the y-axis, in a standard normal distribution is symmetrical (Weiss, 2015). The difference in the proportion of the total area under the curve that is to the right of the z-score between the comparison institution and target institution was used to determine distance from the target institution.

For each predictor, a normative index score of 1 was assigned to a comparison institution that was within one-fourth the distance of the total standard normal distribution's area from the target institution. As with the percentile selection index, a score of 2 was given if the comparison institution was within one-eighth of the area or distance from the target institution's probability corresponding to the z-score. Equally weighted, the average of the normative index scores derives the normative selection index. The equations used to compute the normative selection index are shown in Appendix B. An example on how to calculate a normative selection index is provided in Appendix C.

RESULTS

For the original data set, there were 58 peers and 47 almost-peers across the three peer selection indices. There were fewer peers in the updated data set, 34. There were 55 almost-peers. Across data sets, the normative selection index in the original data set produced the largest number of peers, 51. The percentile selection index in the

Table 2. Index Score Peer and Almost-Peer Classifications for the Three Selection Indices

Selection Index	Peer		Almost-Peer	
	N*	Percent**	N*	Percent**
Original Data Set				
Proximity	813	65.6%	426	34.4%
Percentile	638	53.8%	547	46.2%
Normative	756	60.7%	487	39.3%
Updated Data Set				
Proximity	750	60.1%	498	39.9%
Percentile	595	64.5%	327	35.5%
Normative	606	60.2%	400	39.8%

*Note: * Count of index scores for each predictor for each peer and almost-peer.*
*** Percent of index scores that were 1 (Almost-Peer) or 2 (Peer).*

updated data set produced the fewest number of peers, 26, just slightly more than half the size of the largest set of peers or set of almost-peers.

Selection Index Ranges

Proximity selection index

For the original data, the range of the resulting proximity selection index was 1.33 to 1.78 for the peers and almost-peers. The updated data set posted a range that was slightly more compressed, ProxI Range = 1.44 to 1.78, for the peers and almost-peers. The cutoffs for the peer set was the 95th percentile, while the almost-peers were institutions between the 90th and 95th percentiles.

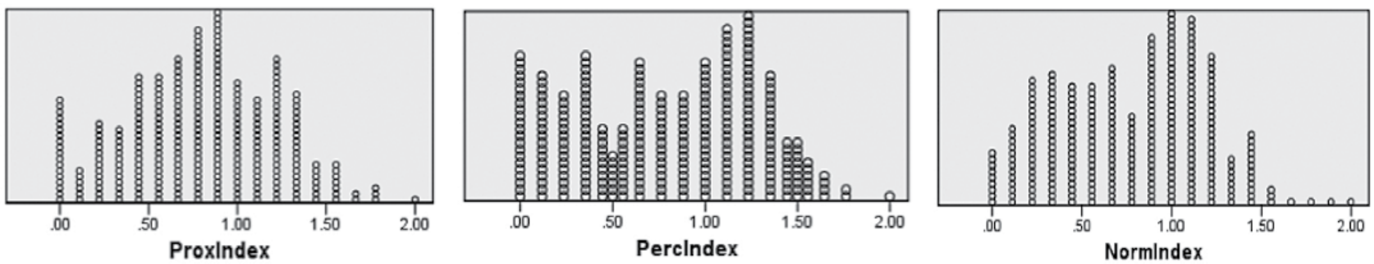
The set of proximity peers and proximity almost-peers changed between the original data set and the updated data set. In part this was due to the smaller set of initial peers in 2016 compared to 2013 (N = 285, N = 232, respectively). The smaller number of initial peers in the updated data set was the result of several circumstances. For 46 of the original data initial set of institutions, the Basic Carnegie classification level changed in 2015 to a master's level. The enrollment of six of these original data set initial institutions dropped below 1,000, and one institution closed.

Examining the individual proximity index scores for each predictor in the original data set, the proximity index scores were more likely to classify a comparison institution as a peer than an almost-peer (65.6%), although the number of peers and almost-peers were the same. This is seen in Table 2. The updated data set was similar in that 60.1% of the proximity index scores categorized a comparison institution as a peer although peers make up only two-fifths (42.3%) of both sets (11 vs. 15, respectively).

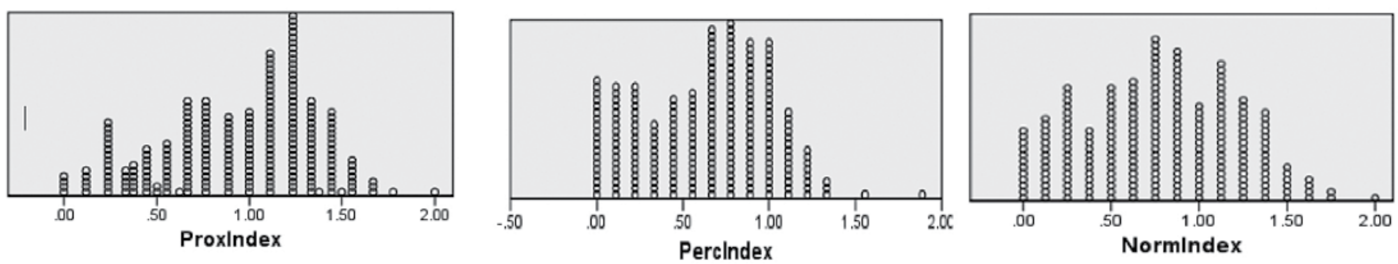
Percentile selection index

For the original data set, the percentile selection index range used to determine the peer institutions and almost-peer institutions was the same as the proximity selection index for the updated data set (PercI Range = 1.44 to 1.78) but more compressed than the percentile selection index for the updated data set (PercI Range = 1.11 to 1.56). For comparative purposes, the same cutoffs used for the proximity selection index were also applied to the percentile selection index, 95th percentile or higher for peers and

Figure 2. Selection Index Distributions



UPDATED DATA SET



between the 90th and 95th percentiles for almost-peers.

As seen in Table 2, the original data set, the percentile selection index methodology had a more equitable split between the two peer groups with almost 54% (53.8%) of the percentile index scores classifying comparison institution as peers. However, the peer set is more than twice the size of the almost-peer set (21 vs. 9, respectively) and, therefore, the percentile index scores do not follow the individual percentile index score classification proportions. For the updated data set, about one-third (35.5%) of the percentile index scores categorized a comparison institution as an almost-peer, but the number of percentile selection index peers and almost-peers is similar (12 vs. 14, respectively).

Normative selection index

For the original data set, the normative selection index range (NormI = 1.22 to 1.89) was larger than the other selection indices. Therefore, the range for the updated data set was more compressed (NormI = 1.38 to 1.75) than the range for the normative selection index in the original data set. Again, the same cutoffs used for the other two selection indices were also applied to the normative selection index: 95th percentile or higher and between the 90th and 95th percentiles for peers and almost-peers, respectively.

In the original data set, the proportion of individual normative index scores that classified a comparison institution as a peer (60.7%) is the inverse of the actual proportion of peers (36.7%) to almost-peers. For the updated data set, the proportion of index scores that classified a comparison institution as a peer (60.2%) is more analogous to the actual proportion of peers and almost-peers, with more than half (55.6%) of the institutions at or above the 95th percentile.

Selection index distributions

An examination of each selection index distribution is shown in Figure 2. As seen, the distributions are shaped differently from what was expected. For example, in the original data set the proximity selection index should be very susceptible to outliers, but in fact it was more normally distributed than the percentile selection index that had a more pronounced right skewness. The probabilities associated with the percentile selection index were moderately uniform, yet

Table 3. Skewness and Kurtosis for the Predictors and Each Selection Index

Selection Index	Skewness	Kurtosis
Original Data Set		
25th Percentile Mathematics SAT	.56	–.15
Percent of Students Receiving Federal Grant Aid	.83	.28
Average Faculty Salary	.74	.99
FTE	.95	.72
Total Price of Attendance	–.01	–.22
Instructional Expenses Per FTE	1.79	4.10
Alumni Giving Rate	.70	.37
Proximity Selection Index	.00	–.51
Percentile Selection Index	–.03	–1.12
Normative Selection Index	.02	–.69
Updated Data Set		
Applicants per FTE	.80	.50
75th Percentile Mathematics SAT	–1.44	–.33
Percent of Faculty with Terminal Degree	–1.44	1.97
Average Faculty Salary	.54	.38
FTE	1.08	1.26
12-Month Enrollment Per FTE	4.01	21.06
Total Price of Attendance	–.11	–.79
Alumni Giving Rate	–.82	.63
Proximity Selection Index	–.31	–.71
Percentile Selection Index	–.01	–.57
Normative Selection Index	.06	–.71

multimodal or with more than one peak. For the updated data set, the proximity selection index distribution was more left skewed than both the percentile and normative selection index distributions and all distributions generated by the original data set.

All but the normative selection index distribution is misshapen. Again, the percentile index distribution appears to be multimodal.

To further investigate these incongruities, the skewness and kurtosis for each selection index were also computed. Results are shown in Table 3. In brief, the differences in asymmetry of the selection index distributions affect peer selection.

Paradoxically, only the proximity selection index distribution for the original data set was not left skewed. This is shown in Table 3 and Figure 2. Although slight, the percentile selection index was the most skewed (SE = –.03) of the original data set selection indices. For the updated data set, the proximity selection index posted the largest skew (SE = –.31) and the percentile selection index had the smallest skew (SE = –.01). Overall, the distributions gleaned from the updated data set seem to be more normally shaped than the original data set distributions.

Delving deeper into the data, it was discovered that the predictors used in the selection indices were also skewed. Skewness and kurtosis for the continuously scaled predictors are also shown in Table 3. Except for the Total Price of Attendance (SE = –.01) predictor, all were positively or right skewed in the original data set. The Instructional Expenses Per FTE predictor was the most skewed (SE = 1.79). For the updated data set, one-half (4) of the predictors were left skewed and one-half were right skewed. The 12-Month Enrollment Per FTE predictor (SE = 4.01) was the most skewed. Yet, even with the more pronounced skewness of the predictors in the updated data set compared to the original data set, the equitable proportion of left skew to right skew predictor distributions in the updated data set seemed to balance all the selection index distributions.

Examining the selection index distributions' kurtosis can also be informative. As seen in Table 3, all the selection index distributions for both the original and updated data sets had negative kurtoses. Negative kurtosis is associated with distributions with a flatter distribution compared to a normal distribution. A normal distribution would have a kurtosis of zero (DeCarlo, 1997). Not surprisingly, the percentile selection indices had the most negative kurtosis, indicating that it is less peaked than the other distributions. This is decipherable in Figure 2. In this regard, the selection index is working as intended. On the other hand, for the updated data set the kurtosis was similar across selection indices. The percentile selection index was the most peaked, albeit negative.

Table 4. Peer Overlap Across Peer Selection Indices

Selection Index/ices	Percent Overlap	
	Peer	Almost-Peer
Original Data Set		
Proximity/ Percentile	21.4%	4.3%
Proximity/ Normative	33.3%	21.3%
Percentile/ Normative	23.8%	6.4%
All 3	19.1%	2.1%
Updated Data Set		
Proximity/ Percentile	5.9%	12.0%
Proximity/ Normative	2.9%	20.0%
Percentile/ Normative	5.9%	4.0%
All 3	0.0%	0.0%

Selection index combinations

Comparison institutions were seldom chosen for membership in more than one selection index peer group. This is shown in Table 4. The original data set peer groups have the most overlap with one-third (33.3%) of the comparison institutions either a proximity or normative selection index peer.. This might be because the distribution for the proximity peer selection index is normally shaped and, therefore, the percentile and normative peer selection index transformations did not make much of a difference. For the updated data set, no comparison institution was a member of all three peer selection indices and only four comparison institutions (5.9% each for proximity/ percentile and percentile/ normative peer selection index combinations) were chosen for two peer selection index groups. Again, the peer selection index distributions—or, more precisely, the differences among the distributions—could have contributed to the uniqueness of each peer selection index membership. The proximity selection index is left skewed, the percentile selection index is multimodal, and the normative selection index is the most symmetric but slightly right skewed. Note that, unlike the original data set, overlap among peers was more prevalent for the three almost-peer selection index groups compared to the peer selection index groups.

Feasibly, symmetry could be achieved by combining selection indices, as was the case for the updated data set. Comparison institutions that were (a) only a normative selection index peer (NORMATIVE ONLY), (b) both a percentile and a normative selection index peer (BOTH), or (c) neither a percentile nor a normal selection index peer (NEITHER), were further investigated. As a starting point, the difference or distance between the average of each of these selection index peer sets and the target institution were examined for each continuously scaled predictor. This is seen in Figure 3. For the original data set, target institution was closest to the means derived from BOTH institutions for almost half (42.9%) of the seven predictors. For the updated data set,

Figure 3. Target Institution Comparisons to the Normative Selection Index Peers Only, Both Normative and Percentile Selection Index Peers, and Neither a Normative or Percentile Selection Index Peer

	Original Data Set		
	NORMATIVE	**BOTH**	**NEITHER**
25th Percentile Mathematics SAT	X		
Total Price of Attendance		X	
Average Faculty Salary		X	
FTE		X	
Total Price of Attendance	X		
Instructional Expenses Per FTE			X
Alumni Giving Rate			X
COUNT	2	3	2

	Updated Data Set		
	NORMATIVE	**BOTH**	**NEITHER**
Applicants/ FTE	X		
75th Percentile Mathematics SAT		X	
Percent Faculty with Terminal Degree	X		
Average Full-time Faculty Salary		X	
FTE	X		
12 Month Enrollment/ FTE		X	
Total Price of Attendance		X	
Alumni Giving Rate		X	
COUNT	3	5	0

	Both Data Sets		
	NORMATIVE	**BOTH**	**NEITHER**
TOTAL	*5*	*8*	*2*
PERCENT	*33.3%*	*53.3%*	*13.3%*

NORMATIVE: Normative Selection Index Peers Only
BOTH: Peers that are both Percentile and Normative Selection Index Peers
NEITHER: Peers that are neither Percentile nor Normative Selection Index Peers

BOTH institutions posted predictor means closest to the target institution for over half (62.5%) of the eight predictors. Combined across data sets, the target institution was closer to the institutions that were BOTH peers more frequently (53.3%) than the other two groups. Next was the NORMATIVE ONLY institutions, with one-third (33.3%) of the predictor means being nearest to the target institution compared to the other two groups. The NEITHER peer institutions fared the worst, with the distance between the target institution and the peer institution being the closest for only two predictors (13.3%) across data sets. In sum, institutions that are both percentile and normative selection index institutions tended to be the nearest to the target institution compared to the normative selection index–only institutions or those institutions that were in neither the percentile nor normative selection index peer groups.

A closer examination of the target institution's position on each continuously scaled predictor's distribution corroborates these findings. In Figure 3 green indicates the position of the target institution at the high end (right) of that predictor's distribution, yellow indicates the target institution in the middle of the predictor's distribution, and red indicates the target institution at the low end (left) of the distribution. As seen, there was no clear pattern. That is, the target institution's position on the distribution did not seem to influence peer selection index membership. This might be good news, in that the selection indices were somewhat unaffected by the target institution performance compared to other institutions.

CONCLUSION

This study investigated the use of three peer selection indices: proximity, percentile, and normative. These selection indices were applied to a predetermined set of institutions using institutional characteristics based on constituent feedback and institutional priorities. To select a set of peers that was well-informed and aligned with those priorities, the following steps were executed: (a) determination of what data to use, (b) data element standardization, (c) regression modeling to identify the predictors that were best correlated with key institutional attributes, (d) computation of index scores and corresponding selection indices, and (e) ascertaining the appropriateness of the selection indices. The last step was accomplished by comparing peer sets that were identified for each selection index to each other as well as considering the impact of the distributions of the predictors that make up each selection index.

As mentioned, the crux of the paper was to describe each selection index and to explore the differences among the three selection indices' peers. This research is innovative in that this was the first time that two of the selection indices, percentile and normative, were formally introduced and investigated. Moreover, the three selection indices were investigated simultaneously. As with our previous research, no selection index is endorsed outright but rather the plausibility and limitations of each was discussed. That said, selection index methodology holds promise as a robust and legitimate peer selection tool.

Because of the number of institutions in the initial data sets (N = 285 and N = 232 for the original and updated data sets, respectively), the 95th percentile of the selection index was established as the cutoff for choosing peers. Another set of almost-peers was also identified from institutions that were between the 90th and 95th percentiles. The two-tiered system to classify the immediacy of the institutions to the target institution is practical because of the relatively small range of index scores for all three selection indices. In turn, there might not be any meaningful differences regarding nearness to the target institution between institutions in the 90th to 95th percentile range and those in the 95th percentile to maximum range.

The results are not conclusive, but nonetheless indicate that using selection index composites—in particular a combination of the percentile and normative selection indices—can be useful. Although not a factor for the data sets used in this research, being mindful of the target institution's distribution position could be important and warrants further investigation.

Comparisons between the original data set and the updated data set reinforce the importance of regularly verifying an institution's list of peers. The peer institutions that were chosen changed over time, regardless of the selection index used. In fact, less than one-half (44.7%) of the proximity selection index peers or almost-peers identified in the updated data set were part of the original data set of proximity selection index peer or almost-peer list. The percentile selection index was somewhat less stable across data sets. Only one-third (33.3%) of the original data set percentile selection index peer or almost-peer institutions made the updated data set percentile selection index peer or almost-peer list. Likewise, only one-third (33.3%) of the normative selection index peer or almost-peers in the original data set were also peers in the updated data set. That said, the original data set of normative selection index almost-peers was very large compared to the other normative selection index peer sets, essentially ensuring some correspondence.

Admittedly, the Carnegie Classifications were modified between the extraction of the original and updated data sets and those modifications affected the selection of the initial set of peers and, ultimately, the final selection of peers. It is expected that Carnegie Classification will be updated more frequently and, therefore, the time of extraction of the two data sets used in this study was apropos. The results of this study can be taken as a warning that peer lists can become outdated and unsuitable. As this research demonstrates, it is reasonable to expect that institution characteristics and priorities change over time.

Finally, peer list differences among the selection indices demonstrate the importance of due diligence before, during, and arguably even after the selection process. Feedback from faculty and staff are key to this thoroughness. Beforehand, engaging constituent input not only helps to identify institutional priorities but, afterward, also reinforces their importance. Additionally, participation of constituents increases acceptance and use of the chosen set of peers.

Examining the set of institutions gleaned by each selection index affords both a comparison of the appropriateness of each institution as a peer and the set of institutions as a reasonable peer group (D'Allegro & Zhou, 2013). To that end, the choice of initial set of institutions is crucial. These institutions should be approximate to the target institution by proxy of both institutional characteristics and the predictors that will ultimately be used to choose a set of peers.

RECOMMENDATIONS

This study validates that peer selection based on a multistaged approach is necessary but not sufficient. Careful vetting of the appropriateness of the actual statistical steps and methodology are needed. As an example, several OLS regression models were generated to determine the best predictors of institutional quality, the mainstay of the target's priorities. However, other methodologies could be employed, including discriminant analysis, factor analysis, and variable match (Anderes, 1999).

Preliminary scrutiny of the variables to choose peers should be undertaken. To ensure the best mix of institutional characteristics to choose peers, this research engaged a two-stage regression modeling process. In the first stage, the best predictors were chosen from five different institutional characteristic categories. The second stage confirmed the correlation of the predictors to three institutional quality measures. Additionally, the location of the target institution on each potential predictor distribution and other anomalies should be identified and considered a priori to the actual determination of peers.

The examination of the selection indices is also in order. As the comparison of the distributions for each selection index revealed, resulting non-normal distributions had a profound impact on the selection of peers.

Both the type of institution as well as the purpose of the peer selection are key in determining the most appropriate information to collect (Shin, 2009). The use of historical information and data trends are posited as options but might not be fitting. As was the case in this study, historical information gleaned a different set of peers than more-current data.

Following the logic of the use of a two-tier taxonomy, two sets of peers were identified: peers and almost-peers. This affords the flexibility of choosing peers for different purposes and audiences. In addition, it somewhat mutes the imperfections of the peer selection methodologies.

The purpose of the study was to provide reasonable peer selection options. As stated, peer comparisons have many applications, such as determining quality, benchmarking salaries, evaluating programs, informing policy, and setting strategic direction. Coupled with the wide variety of institutional types and missions and the inconclusiveness of the results, no single selection index can be upheld to be better than the other selection indices. Accordingly, care should be taken to determine the best selection index or combination of selection indices. As for the latter, selection index combinations should be further investigated. A set of institutions determined to be a peer by two or more selection indices might prove to be more trustworthy and steadfast than the selection of peers from only one selection index. This seemed to be the case in this study, in which there was less distance from the target institution for most of the predictors for the combined selection index peers compared to the normative selection index–only peers, or, for comparison, institutions not selected by either the percentile or normative selection indices.

As of this study, there are few publications on peer selection methodologies. Evidence that is more conclusive is needed about peer selection models and the effect that target institution type might have on those models. As mentioned, the impact of peer comparisons on institutional quality and improvement has not been studied. Evaluation that invokes the use of peers seems to be in vogue but the question remains: Are peer comparisons or benchmarking superior to other types of comparative assessments or non-comparative evaluation? Further research should be able to address.

REFERENCES

American Association of University Professors (AAUP). (2012). *AAUP Faculty Salary Survey*. Retrieved from http://data.chronicle.com/category/ccbasic/21/faculty-salaries/

Anderes, T. (1999). Using peer institutions in financial and budgetary analyses. *New Directions for Higher Education*, 108, 117–123.

Carnegie Foundation. (2015). *The Carnegie classification of institutions of higher education.* Retrieved from http://carnegieclassifications.iu.edu/

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavior sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohodes, S., & Goodman, J. (2012). *First degree earns: The impact of college quality on college completion rates.* Retrieved from https://dash.harvard.edu/bitstream/handle/1/9396433/RWP12-033_Goodman.pdf?sequence=1

D'Allegro, M. L., & Zhou, K. (2013). A case study to examine peer grouping and aspirant selection. *AIR Professional Files*, 13, 1–16.

DeCarlo, L. T. (1997). On the meaning of the use of kurtosis. *Psychological Methods*, 2(3), 292–307.

Gater, D. S. (2003). *Using national data in university rankings and comparisons.* Paper presented at the Association for Institutional Research International Forum, Tampa, FL.

Hembree, B. (Ed.). (2013). Production process characterization. In C. Croarkin & P. Tobias (Eds.), *Engineering Statistics Handbook* (pp. 1064–1200). Gaithersburg, MD: National Institute for Standards and Technology.

Hom, W. (2008). Peer grouping: The refinement of performance indicators. *Journal of Applied Research in the Community College,* 16(1), 45–51.

Huxley, R. G. (2009). Formation of community college clusters using the CCSSE benchmarks for extra-large community colleges. *Journal of Applied Research in Community College,* 16(2), 103–108.

McLaughlin, G., Howard, R., & McLaughlin, J. (2011). *Forming and using peer groups.* Paper presented at the Association for Institutional Research International Forum, Toronto, ON.

National Center for Education Statistics (NCES). (2012). *Compare individual institutions: Select institutions.* Retrieved from http://nces.ed.gov/ipeds/datacenter/InstitutionByGroup.aspx

Qayoumi, M. H. (2012). *Benchmarking & organizational change.* Alexandria, VA: Association of Physical Plant Administrators of Universities and Colleges.

Ruben, B. D. (2004). *Pursuing excellence in higher education: Eight fundamental challenges.* San Francisco, CA: Jossey-Bass.

Shin, J. C. (2009). Classifying higher education institutions in Korea: A performance-based approach. *Higher Education,* 57(2), 247–266.

Smith, M. K. (2012). *Summary statistics for skewed distributions*. Retrieved from https://www.ma.utexas.edu/users/mks/statmistakes/skeweddistributions.html

SPSS. (2012). PASW Statistics 18 command syntax reference. Chicago, IL: Author.

Trainer, J. F. (2008). The role of institutional research in conducting comparative analysis of peers. *New Directions for Higher Education*, 2008(141), 21–30. Doi:10.1002/he.290

U.S. News & World Report. (2015). *College rankings and lists*. Retrieved from http://premium.usnews.com/best-colleges

Weiss, N. A. (2015). *Introductory statistics* (10th ed.). Boston, MA: Pearson.

Association for Institutional Research

Appendix A. Data Elements Used for Peer and Aspirant Selection: Time Frame, Indicator Type, and Source

Variable	Time Frame	Indicator Type	Indicator Source
Admit Yield	2011–2012, 2014–2015	Admissions	IPEDS
Number of Applicants, Total	2011–2012, 2014–2015	Admissions	IPEDS
Percent of Applicants Admitted	2011–2012, 2014–2015	Admissions	IPEDS
SAT Critical Reading 25th Percentile Score	2010–2011, 2014–2015	Admissions	IPEDS
SAT Critical Reading 75th Percentile Score	2010–2011, 2014–2015	Admissions	IPEDS
SAT Math 25th Percentile Score	2010–2011, 2014–2015	Admissions	IPEDS
SAT Math 75th Percentile Score	2010–2011, 2014–2015	Admissions	IPEDS
Percent of Full-Time Undergraduates Receiving Federal Grant Aid	2010–2011, 2013–2014	Admissions	IPEDS
Average Salary Equated to 9-Month Contracts of Full-Time Instructional Staff: All Ranks	2011–2012, 2014–2015	Faculty	IPEDS
Full-Time Primary Instruction Head Count	Fall 2011, Fall 2015	Faculty	IPEDS
Part-Time Primary Instruction Head Count	Fall 2011, Fall 2015	Faculty	IPEDS
Percentage of Faculty Holding Terminal Degrees	2011–2012, 2015–2016	Faculty	U.S. News & World Report
Estimated Fall Enrollment	Fall 2010, Fall 2015	Enrollment	IPEDS
Full-Time Equivalent (FTE)	Fall 2010, Fall 2015	Enrollment	IPEDS
Total Enrollment, Unduplicated	2010–2011, 2014–2015	Enrollment	IPEDS
Percentage of Classes Enrolling Fewer Than 20 Students	2011–2012, 2015–2016	Enrollment	U.S. News & World Report
Carnegie Classification: Basic (Arts & Sciences or Diverse Fields)	2010, 2015	Institutional Characteristic	IPEDS
Carnegie Classification: Enrollment Size & Setting	2010, 2015	Institutional Characteristic	IPEDS
Carnegie Classification: Undergraduate Profile (Transfer & Full-Time Proportions)	2010, 2015	Institutional Characteristic	IPEDS
Geographic Region	2011–2012, 2014–2015	Institutional Characteristic	IPEDS
Religious Affiliation	2011–2012, 2014–2015	Institutional Characteristic	IPEDS
Endowment (FASB)	2009–2010, 2013–2014	Financial	IPEDS
Instructional Expenses Per FTE (FASB)	2009–2010, 2013–2014	Financial	IPEDS
Tuition Total Price for In-District Students Living on Campus	2011–2012, 2014–2015	Financial	IPEDS
Alumni Giving Rate	2011–2012, 2015–2016	Financial	U.S. News & World Report
Bachelor's Degrees Conferred	2010–2011, 2014–2015	Student Success	IPEDS
Graduation Rates, Total Cohort (6 Years)	As of Aug. 31, 2010, As of Aug. 31, 2014	Student Success	IPEDS
Retention Rates, Total Cohort (1 Year)	Fall 2010, Fall 2014	Student Success	IPEDS

Appendix B. Equations Used to Compute Each Selection Index

PROXIMITY SELECTION INDEX EQUATIONS

ProxSvar1= (TIvarx – CIvarx)/ SDvarx
varx ϵ{1,…x}

ProxInstitution= average (ProxSvar1…ProxSvarx)
institution ϵ{1,…n}

Where:
ProxS = Proximity Score
ProxI = Proximity Selection Index
TI= Target Institution
CI= Comparison Institution
Var1-Varx= Predictors

0 assigned to ProxS when: ProxS < -1 or ProxS > 1
1 assigned to ProxS when: -1< ProxS < -.5 or .5 < ProxS < 1
2 assigned to ProxS when: -.5 < ProxS < .5

PERCENTILE SELECTION INDEX EQUATIONS

PercSvar1= (TIvarx – CIvarx)
varx ϵ{1,…x}

PercInstitution= average (PercSvar1…PercSvarx)
institution ϵ{1,…n}

Where:
PercS = Percentile Score
PercI = Percentile Selection Index
TI= Target Institution
CI= Comparison Institution
Var1-Varx= Predictors

0 assigned to PercS when: PercS < -.25 or PercS > .25
1 assigned to PercS when: -.25 < PercS < -.125 or .125 < PercS < .25
2 assigned to PercS when: -.125 < PercS < .125

NORMATIVE SELECTION INDEX EQUATIONS

NormSvar1= (TIvarx – CIvarx)
varx ϵ{1,…x}

NormInstitution= average (NormSvar1…NormSvarx)
institution ϵ{1,…n}

Where:
NormS = Normative Score
NormI = Normative Selection Index
TI= Target Institution
CI= Comparison Institution
Var1-Varx= Predictors

0 assigned to NormS when: NormS < -.25 or NormS > .25
1 assigned to NormS when: -.25< NormS < -.125 or .125 < NormS < .25
2 assigned to NormS when: -.125 < NormS < .125

Appendix C. Examples on How to Calculate Each Selection Index

PROXIMITY SELECTION INDEX

Proximity Index

Notation Var_x	Predictors	Target Institution Value	Comparison Institution Value	Standard Deviation	$ProxS=(TI_{varx} - CI_{varx})/ SD_{varx}$			
					Equation	Result (Prox S)	Assigned*	
Var_1	FTE	1,810	1,400	200	(1,810-1,400)/200	2.05	0	Because ProxS > 1
Var_2	Average Faculty Salary	$ 45,000	$ 50,000	$ 15,000	(45,000-50,000)/15,000	0.33	2	Because ProxS> -.5 and ProxS < .5
Var_3	Total Cost of Attendance	$ 42,000	$ 37,500	$ 6,000	(42,000-37,500)/6,000	-0.75	1	Because ProxS > -.5 and ProxS < -1
					AVE of $ProxS_1$, $ProxS_2$, $ProxS_3$		1.00	

* 0 Assigned to ProxS when: ProxS < -1 or ProxS > 1
1 Assigned to ProxS when: -1< ProxS < -.5 or .5 < ProxS < 1
2 Assigned to ProxS when: -.5 < ProxS < .5

PERCENTILE SELECTION INDEX

Percentile Index

Notation Var_x	Predictors	Target Institution Value	Comparison Institution Value
Var_1	FTE	1,810	1400
Var_2	Average Faculty Salary	$ 45,000	$ 50,000
Var_3	Total Cost of Attendance	$ 42,000	$ 37,500

Notation Var_x	Predictors	Target Institution Value	Comparison Institution Value	$PercS=(TI_{varx} - CI_{varx})$			
				Equation	Result (PercS)	Assigned*	
Var_1	FTE	55.00%	40.00%	.55-.40	0.15	1	Because PercS > .125 and PercC < .25
Var_2	Average Faculty Salary	50.00%	62.00%	.50-.62	-0.12	2	Because PercS> -.125 and PercS < .125
Var_3	Total Cost of Attendance	62.00%	32.00%	.62-.32	0.30	0	Because PercS > .25
				Ave of $PercS_1$, $PercS_2$, $PercS_3$			1.00

* 0 Assigned to PercS when: PercS < -.25 or PercS > .25
1 Assigned to PercS when: -.25< PercS < -.125 or .125 < PercS < .25
2 Assigned to PercS when: -.125 < PercS < .125

Appendix C. Examples on How to Calculate Each Selection Index

NORMATIVE SELECTION INDEX

Normative Index

Notation Var_x	Predictors	Target Institution Value	Comparison Institution Value
Var_1	FTE	1,810	1,400
Var_2	Average Faculty Salary	$ 45,000	$ 50,000
Var_3	Total Cost of Attendance	$ 42,000	$ 37,500

Converted to the Area Corresponding to the Standardized Normal Scores*

Notation Var_x	Predictors	Target Institution Value	Comparison Institution Value	Equation	Result	Assigned**	
				NormS=(TI_{varx} – CI_{varx})			
Var_1	FTE	0.35	0.45	.35-.45	-0.10	2	Because NormS > -.125 and NormS < .125
Var_2	Average Faculty Salary	0.75	0.88	.75-.88	-0.13	1	Because NormS > -.25 and NormS < -.125
Var_3	Total Cost of Attendance	0.88	0.48	.88-.48	0.40	0	Because NormS > .25
				Ave of $NormS_1$, $NormS_2$, $NormS_3$		1.00	

* To compute the area:
1. Calculate the z-score= ($Value_{Varx}$ - $Mean_{Varx}$)/$StdDev_{Varx}$
2. Find the area under the curve to the right of the z-score (NORM.S.INV function in Excel)

** 0 Assigned to NormS when: NormS < -.25 or NormS > .25
1 Assigned to NormS when: -.25< NormS < -.125 or .125 < NormS < .25
2 Assigned to NormS when: -.125 < NormS < .125

This page left intentionally blank

PROFESSIONAL FILE

ARTICLE 143

Association for Institutional Research

CONSTRUCTING A PEER INSTITUTION: A NEW PEER METHODOLOGY

Steve Chatman

About the Author
Steve Chatman is an institutional research analyst in the Office of Institutional Research and Decision Support at the University of California, Merced.

Abstract
Whatever your method of selecting institutions for comparison and benchmarking, you can both increase the validity and accuracy of those comparisons and extend the value of comparisons to department and college levels by constructing a peer institution from disaggregated components. This paper will demonstrate the methodology using the National Study of Instructional Costs and Productivity (Delaware Cost Study), the Faculty Salary Survey by Discipline (Oklahoma State University [OSU]), and Academic Analytics, LLC, to construct better peer institutions with comparative statistics at campus, college, and department levels for faculty salaries, instructional costs, and research activity. The methodology can also be used to fine-tune traditional peer methodologies and should be added to the institutional research arsenal of cluster-, threshold-, hybrid-, and panel-based peers.

NARRATIVE

In the most influential institutional research document describing peer institution selection, Paul Brinkman and Deb Teeter (1987, p. 7) wrote, "In developing peer groups, it is unrealistic to expect to find perfect matches, 'clones' as it were, for the home institution." In fact, practitioners soon discover that the use of even a handful of narrowly described thresholds (same schools and colleges of same relative sizes) will eliminate all other universities, and the researcher is left with an off-the-rack fit instead of a tailored fit. This paper asserts that Brinkman and Teeter were wrong about finding perfect matches. There is an alternative that will produce a near-perfect match: that is, a clone or doppelganger university. It just will not be a brick-and-mortar university. In fact, it won't exist except on spreadsheets or in computer code.

Traditional methods of peer group selection can be classified into developed or predetermined types. These types are not mutually exclusive and most peer selection processes incorporate elements of multiple types. Predetermined types are easily communicated publicly and include the following:

1. Natural peers are based on geography, athletics conferences, consortiums, or similar factors. These peers are particularly useful when communicating with legislators or the public in general.
2. Traditional peers are based on long-term associations or rivalries (e.g., Ivy League, State versus University of).
3. Jurisdictional peers are based on political, legal, and administrative systems (e.g., state, regional, campuses of the university system, accreditation regions).
4. Classification-based peers are most often based on Carnegie Basic Classification or a subset thereof.

Developed peers rely on measured characteristics and can vary from simple (e.g., disciplinary composition clusters, public Research 1 and 2 [R1 and R2]), to complex (e.g., student characteristics, funding levels, composition by student levels, professional programs), and include the following:

1. Cluster analysis is more statistically complex. It sorts institutions into groups based on composition dimensions.

Association for Institutional Research

Table 1. Home U Instruction by Department and College Expenditures Compared to Expenditures at National Research Universities (Data Are Fictitious)

Home U Degree Programs / Majors	CIP	Delaware Discipline if Different	Home U FTE Students (Ugrad SCH / 15 + Grad SCH / 12)	Home U Instruction Expenditure	Home U Instruction $ / FTE Student
Anthropology	45.02		127	$888,679	$6,975
Cognitive Sciences	30.25	42.00 Psychology	208	$1,508,545	$7,269
Economics	45.06		229	$1,100,499	$4,810
History	54.01		114	$1,035,698	$9,078
Literatures and Cultures	16.01		458	$3,719,811	$8,125
Management	52.02		115	$565,035	$4,928
Political Science	45.10		173	$1,721,097	$9,968
Psychology	42.01		827	$3,734,230	$4,517
Sociology	45.11		273	$1,236,805	$4,529
School of Social Sciences, Arts, and Humanities			2523	$15,510,399	$6,148
Applied Mathematics	27.03	27.00 Mathematics and Statistics	782	$3,300,100	$4,218
Bioengineering	14.05		40	$805,709	$19,943
Biological Sciences	26.01		605	$3,392,147	$5,611
Chemistry	40.05		492	$2,905,605	$5,905
Earth Systems Sciences	40.06		104	$1,607,946	$15,506
Physics	40.08		219	$1,941,943	$8,863
School of Natural Sciences			2242	$13,953,450	$6,223
Computer Science and Engineering	14.09	11.07 Computer Science	223	$2,474,021	$11,083
Environmental Engineering	14.14	14.08 Civil Engineering	113	$1,632,681	$14,498
Materials Science and Engineering	14.18		77	$844,570	$10,921
Mechanical Engineering	14.19		124	$2,047,071	$16,529
School of Engineering			537.0	$6,998,343	$13,032
Writing Program	23.13		725.2	$4,340,547	$5,985
Home U Overall			**6,027.3**	**$40,802,739**	**$6,770**

Delware Cost Study Instruction $ Per FTE Student	Home U Instruction $ Per Student / National Research Univ $ Per Student	Home U - Delaware Instruction $ Per Student	Weighting National Instruction Expenditure by Home U FTES	$ Difference Times Home U FTE Students
$5,865	119%	$1,110	747,299	141,380
$5,632	129%	$1,637	1,168,828	339,717
$5,930	81%	-$1,120	1,356,784	-256,285
$6,157	147%	$2,921	702,411	333,287
$5,762	141%	$2,363	2,638,036	1,081,775
$6,948	71%	-$2,020	796,704	-231,669
$6,809	146%	$3,159	1,175,687	545,410
$5,632	80%	-$1,115	4,656,162	-921,932
$5,111	89%	-$582	1,395,644	-158,839
$5,802	106%	$346	$14,637,554	872,845
$5,172	82%	-$954	4,046,918	-746,818
$15,849	126%	$4,094	640,300	165,409
$6,824	82%	-$1,213	4,125,677	-733,530
$7,254	81%	-$1,349	3,569,331	-663,726
$9,531	163%	$5,975	988,365	619,581
$8,417	105%	$446	1,844,165	97,778
$6,785	92%	-$563	$15,214,754	-1,261,304
$10,175	109%	$908	2,271,230	202,791
$11,181	130%	$3,317	1,259,167	373,514
$15,508	70%	-$4,587	1,199,285	-354,715
$10,748	154%	$5,781	1,331,140	715,931
$11,286	115%	$1,746	6,060,822	937,521
$4,942	121%	$1,043	3,583,938	756,609
$6,553	**103%**	**$217**	**39,497,068**	**1,305,671**

For example, institutions can be sorted based on relative mix of disciplinary degrees awarded.
2. Threshold analysis is straightforward and easily communicated. The characteristics of potential peers have to fall within a range above and below the measured characteristic of the home institution. For example, if headcount enrollment at the home institution is 20,000, then peers would have enrollments between 17,500 and 22,500. Thresholds can be similarly applied to full-time equivalent (FTE) enrollment, admissions scores, in-state enrollment percentage, or almost anything commonly measured.
3. Panel analysis relies on the expertise of professionals, typically institutional executives, who either nominate potential peers or eliminate potential peers identified by other methods.
4. It is more common for the methodology to be a hybrid of other types in various sequences (e.g., cluster analysis followed by threshold analysis and then submission to a panel).

In contrast with developed or predetermined institutional peers, the constructed peer methodology described in this paper is typically built from departmental or disciplinary components. Unlike institutional peers, the constructed peer methodology can use disciplinary components that vary from one department or school to another. Psychology might select Psychology peers and Biology might select a different set of Biology peers. But even when the home institution is constrained to compare with a given institutional set, the constructed peer methodology can be based on the elemental characteristics of those institutions. Because the result is constructed from disciplinary components, the result will be useful at the level of the department and will be more accurate when aggregated to college and institutional levels.

In spite of the availability of data to support a constructed peer methodology by department, especially for faculty salaries and disciplinary expenditures, the methodology has not contributed to the discussions of peer institution groups that were popular in the 1980s and that continue to dominate institutional research practice: various cluster analysis techniques and some measure of judgment (panel, hybrid, threshold, panel) about institutional key or performance statistics (Brinkman & Teeter, 1987; Terenzini, Hartmark, Lorang, & Shirley, 1980; Trainer, 2008; Xu, 2008). There are two very good reasons to revisit peer methodology. First, good disaggregated data are available for critically important institutional research elements including faculty salaries (e.g., OSU since 1974), instructional costs and productivity (Delaware since 1992), and faculty research activities (Academic Analytics, LLC). Second, disciplinary composition should always be an institutional research consideration because it dramatically affects every aspect of teaching, research, and service; and every aspect of the student experience. There is less variance among universities by program than among programs within a university (Chatman, 2010).

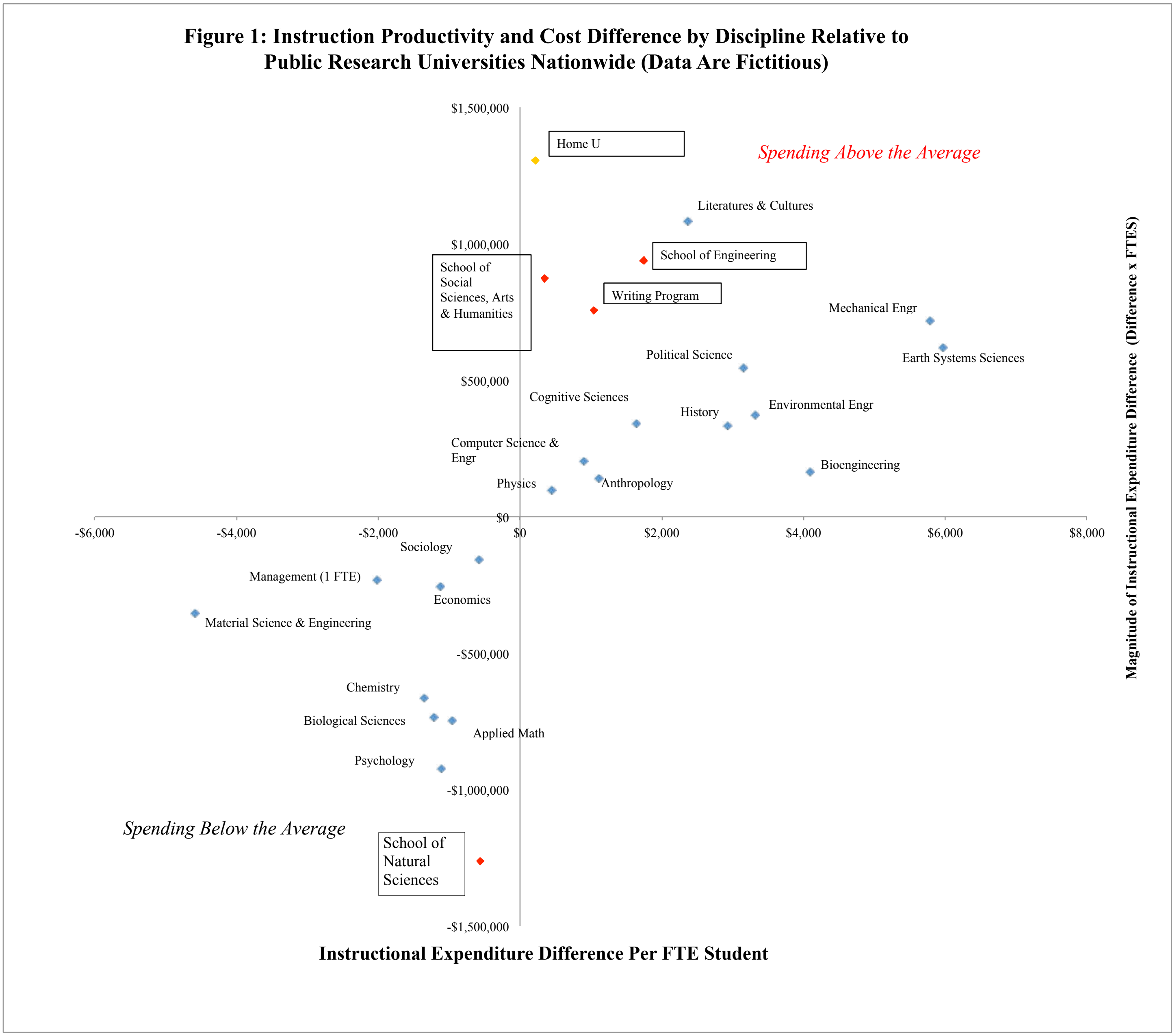
METHODOLOGY

Information from the Delaware Cost Study, the OSU Faculty Salary Survey by Discipline, and Academic Analytics, LLC, will be used to construct doppelganger universities with comparative statistics at campus, college, and department levels for faculty salaries (OSU), instructional cost (Delaware), and faculty research and scholarly activity (Academic Analytics, LLC). The central feature of the methodology is constructing a peer by weighting comparative per capita or mean values to reflect the home institution composition. The methodology will be introduced using per capita instructional costs from the Delaware Cost Study. The other applications are similar in that they find a comparator per capita figures at the lowest available level of aggregation and weight that per capita figures using home institution amounts to create constructed or doppelganger departments. The resulting departments can be combined with others to produce a constructed peer or doppelganger university. The data shown are fictitious but generally reflect the characteristics of the University of California, Merced, a university that grew from farmland to research university in 10 years and continues to grow at a very rapid rate. The nearly 7,000 undergraduates in 2016 had Hispanic, Pell Grant recipient, and first-generation majorities.

Table 2. Department and College Level Faculty Salary Comparisons Using School of Natural Sciences at Home U and OSU Research Universities Average Salaries (2012–2013)

Home U (Actual)								
Ladder Rank	**Content Area**	**Four-Digit CIP Code**	**Salary**	**Head-count**	**OSU R1 & R2**	**Home U Expenditure**	**Comparator-Based Expenditure**	**Home U / OSU R1&R2**
Professor	Applied Mathematics	27.03			122,866			
Assoc. Prof.	Applied Mathematics	27.03	82,000	4	83,941	328,000	335,764	98%
Asst. Prof.	Applied Mathematics	27.03	77,200	4	73,884	308,800	295,536	104%
Professor	Biology, General	26.01	142,400	3	126,463	427,200	379,389	113%
Assoc. Prof.	Biology, General	26.01	83,717	6	84,375	502,302	506,250	99%
Asst. Prof.	Biology, General	26.01	74,040	10	72,848	740,400	728,480	102%
Professor	Biomedical/Medical Engineering	14.05	149,400	1	155,250	149,400	155,250	96%
Assoc. Prof.	Biomedical/Medical Engineering	14.05	99,300	1	104,157	99,300	104,157	95%
Asst. Prof.	Biomedical/Medical Engineering	14.05	89,400	2	83,843	178,800	167,686	107%
Professor	Chemistry	40.05	117,667	3	135,046	353,001	405,138	87%
Assoc. Prof.	Chemistry	40.05	88,650	2	84,958	177,300	169,916	104%
Asst. Prof.	Chemistry	40.05	74,667	6	74,369	448,002	446,214	100%
Professor	Ecology, Evolution, Systematics, and Population Biology	26.13	109,350	2	128,697	218,700	257,394	85%
Assoc. Prof.	Ecology, Evolution, Systematics, and Population Biology	26.13	82,500	1	91,106	82,500	91,106	91%
Asst. Prof.	Ecology, Evolution, Systematics, and Population Biology	26.13	78,750	4	77,694	315,000	310,776	101%
Professor	Physics	40.08	151,700	1	122,345	151,700	122,345	124%
Assoc. Prof.	Physics	40.08	85,425	4	84,901	341,700	339,604	101%
Asst. Prof.	Physics	40.08	78,960	5	75,386	394,800	376,930	105%
School of Natural Sciences								
Professor	Overall		130,000	10	131,952	1,300,001	1,319,516	99%
Assoc. Prof.	Overall		85,061	18	85,933	1,531,102	1,546,797	99%
Asst. Prof.	Overall		76,961	31	75,020	2,385,802	2,325,622	103%
				59		5,216,905		
					Mean Overall	**88,422**	**99,708**	**89%**

Figure 1. Instruction Productivity and Cost Difference by Discipline Relative to Public Research Universities Nationwide (Data Are Fictitious)



Comparing Instructional Costs at the Constructed Peer Institution

Please note that the data here and elsewhere in the report are fictitious and are offered to illustrate the methodology. Steps 1 through 4 describe the methodology for one department, Sociology. The same steps apply to other disciplines/departments and the results can be aggregated to colleges or to the university total.

1. The home institution instructional expenditure in Sociology was $1.2 million.
2. The expenditure per FTE student (based on Sociology student credit hours [SCHs] by level) was $4,529 at the home campus.
3. The per student expenditure in

sociology for research universities (R1 and R2) from the Delaware Cost Study was $5,111, compared to $4,529 at the home campus. The home institution therefore spent 89% of the "expected" amount, or $582 less per student.

4. The home institution had 273 FTE students in Sociology and therefore spent about $159,000 less to deliver sociology instruction than expected.

Steps 1 through 4 were repeated for the other departments and then aggregated to the college and university levels in Table 1. For the School of Social Sciences, Arts, and Humanities, the instructional expenditure was 106% of the constructed peer; Engineering was 115%; and Natural Sciences was 92%. Overall, the home institution instructional expenditure was 103% of the constructed research university peer. The difference per FTE student overall was $217, or $1.3 million in total.

In this example, all public research universities were used for comparison but Delaware supports analysis by selected peers and the peer set could even vary based on the department or college, especially if the home institution participates in a data-sharing consortium (e.g., Association of American Universities Data Exchange [AAUDE]). It is easy to imagine that an Engineering peer set could differ from a Natural Sciences peer set, etc.

Table 1 shows the detail behind computing comparisons and the difference between the local university and the comparative figures per FTE student by department, college, and campus. Figure 1 arrays expenditure differences along two axes. The x-axis is the difference per FTE student and the y-axis is the difference for all FTE students (difference per student times number of FTE students). The two axes of Figure 1 are used because a big difference per FTE student in a small department can have less institutional impact than a small difference in a large department.

In examining the scatterplot in Figure 1, it is clear that the per student institutional composite was close to that for the constructed peer, but that there was a great deal of variation by department and school. If the analysis was limited to institution-level measures, the school and departmental differences would have been obscured. That is a danger of institution-level measures. The composite can be at the mean peer value, suggesting normative performance, but be made up of values showing wide variation. Funding at the institutional level without consideration of disciplinary patterns makes that misleading outcome more likely. The results by school show that one school, Natural Sciences, is spending less than expected for natural science disciplines and is helping to offset the other schools that are spending more than expected for their disciplines. Both schools (Natural Sciences; and Social Sciences, Arts, and Humanities) are actually spending very similar amounts per FTE student. However, the expected expenditure for natural sciences is $563 more per FTE student in this example. It is reasonable to expect the dean of Natural Sciences to make these differences known in the next budget cycle. Please recall that these are not actual amounts and are used to illustrate the methodology; even if accurate, however, the results are not intended to be prescriptive. They do not show programs to be cut or where investments are needed, but they do identify areas of greater or lesser spending than is average and raise the question of whether those spending differences were intentional or a historical artifact.

Other Examples

The technique is generally applicable. Any comparative measure from an outside source that is available at a low level of aggregation can be weighted to reflect local composition and thereby create more-accurate, more-valid, and more-useful statistics for the department, school, and university. The following will illustrate the methodology using faculty salaries and faculty professional research activity but it could be extended to almost any measure.

Faculty Salary Comparison

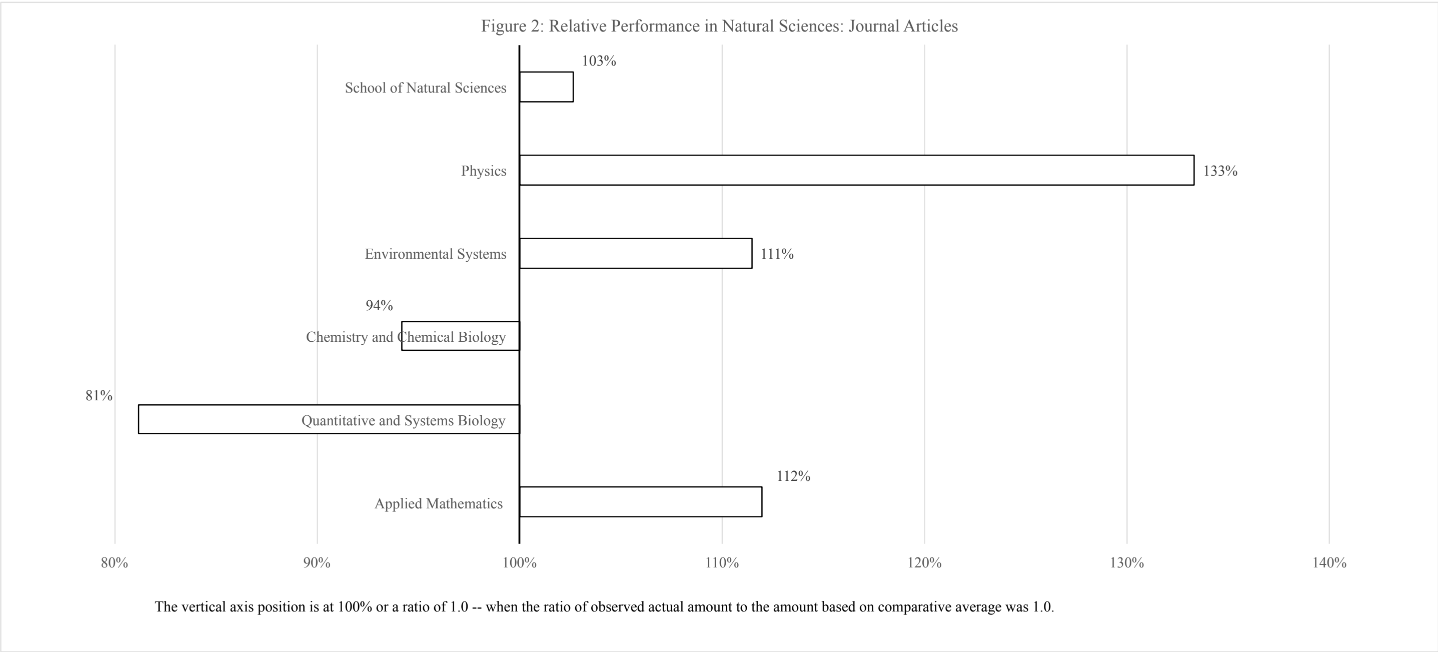
The predominant factors associated with variance in faculty salaries are discipline and rank. Unless the comparator peer set has the same faculty composition by rank and discipline, there will be error that might be masked at the campus level. That error can be controlled by constructing a peer that does have the same disciplines and ranks in the same amounts. The following example illustrates the methodology using OSU Faculty Salary Survey by Discipline averages for public R1 and R2 institutions. As was the case

Table 3. Home U Data Compared to Public and Private University Faculty Academic Analytics for Natural Sciences (Data Are Fictitious)

	Academic Analytics (Per Capita)							
Natural Sciences Disciplines	**Academic Program from Academic Analytics**	**Home U Tenured and Tenure Track Count from Academic Analytics Records**	**Books (2005-2014)**	**Journal Articles (2011-2014)**	**Citations (2010-2014)**	**Grants (2010-2014)**	**"Grant Dollars (2010-2014)"**	**Honors and Awards (Lifetime)**
Applied Mathematics	Home U	11	0.3	10.3	55.6	1.1	78,497	0.3
	Academic Analytics		0.2	9.2	101.6	1.5	180,000	0.9
Actual Output as Percent of Comparative-Average Based Output in Discipline								
Quantitative and Systems Biology	Home U	40	0.1	10.2	153.1	1.0	181,365	0.3
	Academic Analytics		0.2	12.5	180.4	1.3	340,000	0.4
Actual Output as Percent of Comparative-Average Based Output in Discipline								
Chemistry and Chemical Biology	Home U	16	0.3	14.6	288.3	1.1	206,538	0.3
	Academic Analytics		0.2	15.5	330.2	1.8	330,000	1.1
Actual Output as Percent of Comparative-Average Based Output in Discipline								
Environmental Systems	Home U	27	0.1	12.2	152.6	1.5	235,405	0.4
	Academic Analytics		0.2	10.9	142.5	1.4	190,000	0.5
Actual Output as Percent of Comparative-Average Based Output in Discipline								
Physics	Home U	18	0.2	22.4	125.1	0.9	110,077	0.6
	Academic Analytics		0.3	16.8	200.0	1.2	150,000	0.7
Actual Output as Percent of Comparative-Average Based Output in Discipline								
School of Natural Sciences	Home U	107						
	Academic Analytics							
Actual Output as Percent of Comparative-Average Based Output in Discipline								

Academic Analytics (Weighted)					
Books	Journal Articles	Citations	Grants	Grant Dollars	Honors and Awards
3.3	113.3	611	12.0	863,464	3.0
2.2	101.2	1,118	16.5	1,980,000	9.9
150%	112%	55%	73%	44%	30%
4.0	406.0	6,123	41.2	7,254,594	10.0
8.0	500.0	7,216	52.0	13,600,000	16.0
50%	81%	85%	79%	53%	63%
4.8	233.6	4,612	18.1	3,304,601	5.0
3.2	248.0	5,283	28.8	5,280,000	17.6
150%	94%	87%	63%	63%	28%
2.7	328.1	4,120	40.0	6,355,939	11.1
5.4	294.3	3,848	37.8	5,130,000	13.5
50%	111%	107%	106%	124%	82%
3.2	403.2	2,252	16.9	1,981,379	10.1
5.4	302.4	3,600	21.6	2,700,000	12.6
60%	133%	63%	78%	73%	80%
18.0	1,484.2	17,718	128.2	19,759,977	39.1
24.2	1,445.9	21,064	156.7	28,690,000	69.6
75%	103%	84%	82%	69%	56%

Figure 2. Relative Performance in Natural Sciences: Journal Articles



for instructional expenditures, the mean salaries for the comparators by discipline and rank are weighted by the local university composition and the total expenditures are used to create college and institutional comparisons. For this example, the methodology will be applied to the School of Natural Sciences and illustrated using Chemistry. As shown in Table 2, Chemistry professors are paid $135,046, on average, at R1 and R2 schools. The home institution had three professors. If the home department paid the three professors exactly the national mean, the home department would have spent $405,138. The home department actually paid professors $353,001, or 87% of the average. For all departments in the School of Natural Sciences, the home school spent $1,300,001 on professor salaries. If every department in the school had paid the national public R1 and R2 average to each professor, the school would have spent 99% of the aggregated $1,319,516 amount.

The constructed peer methodology is especially useful at Home University (Home U), an 11-year-old public research university, because its mix by rank and discipline is atypical. Because it is a new university, Home U has a much higher proportion of assistant professors and a much lower proportion of professors than is typical. It also has more STEM faculty than is typical of a public university. The unweighted campus mean, not adjusted for the higher proportion of assistant professors and lower proportion of professors, would be well below a simple institutional-level comparator even though both the comparisons by rank and the weighed institutional mean were above the comparator averages. This is illustrated for Natural Sciences in Table 2. By rank, faculty salaries were at or close to the national average: professor salaries were 99% of average, associate professors were 99% of average, and assistant professors were 103% of average. However, the overall average for the home institution was 89% of the overall national average. A result based on component comparisons that is different from the overall comparison is an example of the Yule–Simpson effect, defined as a trend appearing in different groups of data that disappears or reverses when the data are aggregated. In this case, means were close to the average by rank but substantially lower overall. As was the case for instructional costs, large differences for a few faculty should not be cause for alarm, but substantially different patterns by discipline might be cause for discussion, or there might be a strategic plan to recruit substantially more-competitive faculty in one area or another. The results are not prescriptive but should be illuminating.

Faculty Professional Activity

The third example relies on data from Academic Analytics, LLC, a service that gathers federal grants, books, honorific awards, journal and conference publications, and citations for individual faculty and makes those data available to subscribing institutions. The data values shown here are fictitious but the measures shown are available from Academic Analytics and are used with permission. Because faculty are identified by disciplinary area and institution type by Academic Analytics, the mean values for all faculty in a disciplinary area can be used as a comparative standard (Table 3). To make the explanation less complicated, analysis will again be limited to the School of Natural Sciences.

For example, and using the comparative subset of these pseudo value statistics in Physics, the comparative average values per faculty member in Physics were about 0.3 books (2005–2014), 16.8 journal articles (2011–2014), 200 citations (2010–2014), 1.2 grants (2010–2014), $150,000 grant dollars (2010–2014), and 0.7 honors and awards (lifetime). Because the home department had 18 faculty members, the comparative average–based outcome for the 18 faculty members in Physics was 5.4 books, 302 journal articles, 3,600 citations, 21.6 grants, $2,700,000 grant dollars, and 12.6 honors and awards. Actual counts were compared to the comparative average–based outcomes and expressed as percentages (60% to 140% for this Physics example). The comparative average–based outcomes and observed amounts can be aggregated to school and campus levels and can be used to identify areas of relative strengths. Those relative amounts can be expressed graphically. For this example, the relative percentages for journal articles in Natural Sciences disciplines are shown as Figure 2. Again, comparison at the school level (103%) obscures a substantial range by department (133% in Physics to 81% in Quantitative and Systems Biology). For the School of Natural Sciences, journal articles, citations, and number of grants were stronger. Books, grant dollars, and number of honors and awards were lower. That would be a reasonable pattern for a very young university with a disproportionately small number of full professors. As was true for other comparisons, the results are not prescriptive and, especially in this case, should not be used to establish a rigid individual faculty norm for evaluation. The norms are more meaningful at discipline and school levels.

SUMMARY

There are remarkably few published productivity standards in higher education (Chatman, 2016). Instead, analysis is typically parochial, treating history as a comparative standard, or, at the institutional level, treating a cluster of other universities as a comparative standard. The process of selecting peer institutions uses any of a variety of methods or combinations of predetermined or developed peer methods that have been well described elsewhere (Brinkman & Teeter, 1987) and continue to dominate higher education (the National Center for Education Statistics' Executive Peer Tool, or ExPT). This is true even though much better data sources are available that support comparative analysis at the department level or at even smaller aggregates. This paper offers a constructed peer methodology as yielding a better, more-accurate, and more-valid peer because it accurately reflects the disciplinary composition of the home institution and isolates the comparison to the variable being considered.

A constructed peer institution for comparison has important advantages to peers from traditional institutional methodologies. First, the process of constructing a peer produces comparative values at all levels of academic aggregation (e.g., department, school or college, and university). Second, the normative or standard values used to construct the peer can be tailored by department, school, or college so that each level can be based on its own tailored set of institutions. Perhaps the social sciences college and the engineering college of an engineering-focused university should have different peer sets. Third, the methodology is generalizable. The same steps used to construct a faculty salary peer can be used to produce a student satisfaction peer, an alumni engagement peer, a facility utilization peer, a development peer, etc. If a comparative measure can be expressed at the level of a department and at a per capita rate common across institutions (e.g., faculty or FTE students) then the per capita rate can be inflated to reflect the home institution and support a direct comparison. For example, the mean level of satisfaction by disciplinary area for a comparable set of institutions can be weighted by local number of students by major and then compared at the college or institutional level. Fourth, in every case the constructed peer fits the home institution accurately. It has the same programs in the same relative and absolute amounts. For example, it has exactly the same number of faculty overall and by rank and discipline. It is a clone or doppelganger. Given that disciplinary differences are ubiquitous, institutional values used in comparison that ignore those differences might reflect disciplinary composition more than real differences. In other words, the home institution might appear to spend less on instruction per student because it is primarily a social sciences institution comprised of disciplines associated with less-expensive instruction. Likewise, student satisfaction and engagement varies by area of major (Chatman, 2010) and institutional comparisons of satisfaction or engagement will reflect disciplinary composition differences. Institutional measures that ignore differences in disciplinary composition (e.g., Voluntary System of Accountability™) can obscure real differences. Fifth, a variety of relative performance measures can be combined to yield a consistent dashboard or performance profile for departments, colleges, and the institution. For example, the measures described in this paper produce an academic summary that includes cost per credit hour, faculty salaries, and faculty professional activities for a constructed peer that mirrors the home institution.

A constructed peer also has two substantial disadvantages. First, it is more difficult to make transparent; also, in many cases, policies about sharing and reporting information among institutions prevent making the detail available. Second, it requires more effort on the part of the user to understand and the provider to describe because it is less familiar. It is more difficult to explain to higher education constituencies. A university president or chancellor will likely choose to report comparison to the average faculty salary for Pac-12 institutions over the average faculty salary for a peer constructed from the bottom up using various combinations of Association of American Universities (AAU) public institutions. And, while it is less accurate and less valid, comparisons at the institutional level are often very similar to the constructed institutional average. Using an older sister university—for example, the overall faculty salary comparison to OSU's Faculty Salary Survey by Discipline—showed the sister university faculty salary average to be 9% higher. The comparison based on analysis using the constructed peer methodology by rank and discipline was 7% higher. If the only purpose of the peer comparison is to compare institutional-level values, then this method of peer construction is probably not worth the additional effort and loss of transparency. However, if the value of comparisons is extended to school and department levels, then constructed peers are preferable. If the methodology were to become more common, then its reporting would be less of a problem. We regularly use many summary measures and indices as if the meaning were simple and straightforward when they are actually remarkably complex. Some examples include the consumer price index, unemployment rate, Dow Jones industrial average, and even wind chill.

REFERENCES

Brinkman, P. T., & Teeter, D. J. (1987). Methods for selecting comparison groups. *New Directions for Institutional Research*, 53, 5–23.

Chatman, S. P. (2010). *Institutional versus academic discipline measures of student experience: A matter of relative validity*. Association for Institutional Research Professional File Series, 114, 1–20.

Chatman, S. P. (2016). *Instructional productivity standards by discipline and level, finally*. Paper presented at the Annual Forum of the Association for Institutional Research, New Orleans.

Terenzini, P. T., Hartmark, L., Lorang, W. G., & Shirley, R. C. (1980). A conceptual and methodological approach to the identification of peer institutions. *Research in Higher Education*, 12(4), 347–364.

Trainer, J. (2008). The role of institutional research in conducting comparative analysis of peers. *New Directions for Higher Education*, 141, 21–30.

Xu, J. (2008). *Using the IPEDS peer analysis system in peer group selection*. Association for Institutional Research Professional File Series, 110, pp. 1–16.

Thank You!

AIR expresses sincere appreciation for the members who serve as peer reviewers. Authors who submit materials to AIR Publications receive feedback from AIR members. The following individuals reviewed manuscripts submitted to this volume of *Professional Files*.

Matt Anson

Dave Becher

Anne Marie Brady

René Cintrón

Amanda Clark

Linda Clark

Liying Cui

Nora Galambos

Jennifer Kobrin

Ethan Kolek

Teresa Isbell

Jessica Lillegaard

Sun No

Gee Lockhart Sigman

Pa Thao

Benita Thompson

Ying Vuthipadadon

Annie Weber

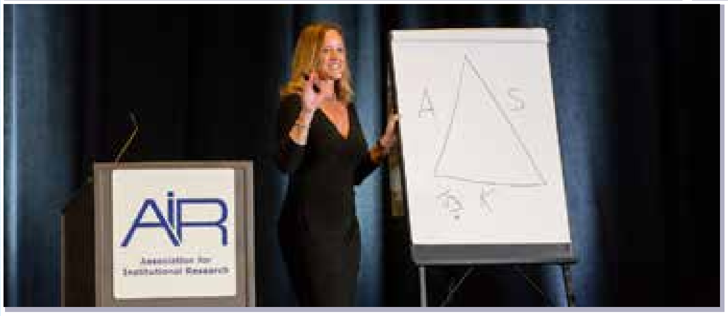
Mike Williford

About AIR *Professional File*

Professional File volumes are journal-length publications grounded in relevant literature that synthesize current issues, present new processes or models, or share practical applications related to institutional research. All submissions are peer-reviewed. For more information about AIR Publications, including *Professional File* and instructions for authors, visit www.airweb.org/publications.

Association for Institutional Research
Christine M. Keller, Executive Director and CEO
www.airweb.org

ISSN 2155-7535



AIR provides higher education professionals with learning opportunities that better equip them to meet the needs of data-informed decision making at their institutions. Whether online or face-to-face, in the office or at our annual conference, AIR's professional development opportunities provide outstanding return on investment for those working in IR and related fields, and the institutions they represent.

A Holistic Approach to IR
This course provides a foundation for participants to meet and navigate the ever-growing demands for data and information in the current higher education landscape.

IPEDS Keyholder Essentials: A Beginner's Guide
Created for data providers with less than 9 months of experience as keyholders, this course covers basic concepts and definitions.

IPEDS Keyholder Efficiencies: Reducing the Reporting Burden
Created for data providers with 10-24 months of experience, this course is designed to enhance your knowledge as an IPEDS Keyholder.

Data and Decisions Academy®
The Data and Decisions Academy is a series of online, self-paced courses designed for community college IR professionals.

AIR Forum
Learn, connect, and share at the largest gathering of IR professionals in the world, and take advantage of recorded sessions, presentation slides, handouts, and scholarly papers available after the conference.

ASSOCIATION FOR INSTITUTIONAL RESEARCH
Data and Decisions for Higher Education

LEARN MORE AT WWW.AIRWEB.ORG