Data 101:
Guiding Principles for Faculty

February 2010

The use of data for making educational decisions and to assess educational outcomes has been legislated by political bodies and codified by accreditation. Faculty have always used data to inform the grading process – we gather data throughout the term to inform the letter grade assigned at the end. However, in today's educational environment faculty must also examine and use data to inform their practice and guide their work in order to improve student interactions, enhance pedagogical strategies, direct program development, and create evidence for program review. In addition, using evidence is essential as faculty address equity practices and advocate for educational opportunities that provide equitable access and outcomes for our diverse student populations.

Nevertheless, sometimes the word "data" sends people running in the opposite direction. Part of the reason for this negative reaction is the historical use (or misuse) of data as simple statistical ammunition for drawing inappropriate or misguided conclusions or even for more nefarious purposes. Occasionally data are used to assess a college's work or programs with a goal towards discontinuance rather than improvement. Another element of this reaction may be a general discomfort when conversing with those who gather, organize, and interpret the numerical data in purely statistical terms. To address issues reflected by this negative reaction, faculty must understand certain basic guidelines and concepts when dealing with data. Faculty need not become statisticians to skillfully find and use data, but they must learn to engage the data with a critical and analytical eye in order to interpret and apply it appropriately.

First and foremost, getting data, particularly the right data that can lead to positive changes, is not easy. Using the right data and refuting the wrong data are important learned skills. People often try to impose statistical rules on "data," creating a picture of sheer numbers as the only data with value, that move the decision making away from the human factors. As we all know, education is rife with human factors that trump statistical niceties. The term "data" implies facts, figures, narratives, eyewitness accounts and artifacts used to make a case. Guided by a few simple principles, accessing data and using it appropriately can be an important tool for faculty. Nothing is richer than knowing what you want to accomplish, identifying a good metric to assess outcomes, and moving forward as a result of good foundational data from those metrics. The data behave like a compass, keeping faculty on course. On the other hand, nothing is more damaging than making decisions based on the wrong data or over-reliance on metrics that are unreliable or missing the appropriate context. Consider the implications of a program review document that displays the descriptions and achievements of program participants disaggregated by ethnicity, over time. Using this data to determine appropriate budgeting and requests for faculty and staff provides power. Imagine enrollment management decisions based upon real student success, informed by course and program level data, in contrast to the typical high level institutional data that stops with overall success and retention injected into an allocation formula. Our decisions are only as good as our data, so we must be certain that the data we use is valid, relevant and actionable.

Statistically speaking, people often refer to data validity, reliability, and significance. While these matters are important considerations, no data will be perfect or statistically sound beyond all question. The real focus should be a learning-centered discussion about data and its implications, not a research-centered discussion. By this we mean that faculty should look to the outcome of the data discussions to improve learning as a primary focus rather than concentrating on technical terminology and constraints that overpower or paralyze dialog. Educational data are impacted by many variables such as socioeconomic influences, environmental opportunity, cultural impacts, personal skills, learning styles, teaching styles, language proficiency, etc. The variables are numerous. For this reason, the best strategy when dealing with data is to ask questions that can help to determine or clarify the validity, utility, and implications of the data being considered.

This paper suggests several principles faculty should acknowledge when looking at educational data and assessing its usefulness. It also demonstrates the application of these principles to a hypothetical scenario and indicates sources for getting data relevant to the California community colleges. The hypothetical scenario involves a Student Success Center that offers faculty mentoring, peer tutoring, textbook and computer access, drop in counseling, and Directed Learning Activities (DLA). The discussion is directed at discontinuing the Student Success Center and the data provided for this decision are as follows:

- Last semester 262 students used the facility; college headcount is 3,124
- Average GPA of those using the facility was 2.1; college average GPA of 2.5
- Only 5% of the total number of students transferring and completing Associate's degrees or certificates this year actually used the Student Success Center

General Data Quality Principles and Sample Questions to Ask

Begin with an overall question – What are the implications of this data on any decisions or policies? In order to describe the use of the guidelines to develop questions we will relate them to our hypothetical scenario at the Student Success Center.

Principle 1 – Use longitudinal data when possible.[1]   Data collected at one moment in time will frequently change, particularly in community colleges where the student population is so diverse. While data for a given population, class or program are helpful, these represent only a snapshot of the current conditions and variables which may or may not change significantly over time. Therefore, looking at data over time provides a better picture of the dynamics and many variables that influence whatever issue faculty may be examining. You might ask, "What number of students would we target as a meaningful number of students to use this facility? Were the number of students using the facility last semester's more or less than previous semesters? What factors may have influenced this last semester's student usage number? How do faculty measure success after students use this facility? Does use of the facility increase the students' success rate or GPA over time? How has this changed over time?"

---

1   Longitudinal data - A dataset is longitudinal when the same information is followed at various points in time with a goal of measuring change to inform improvement.

Principle 2 – Use data in context. Data without a rich and accurate context are meaningless at best and misleading in the worst case. Here, an example of data that a college should not use are student GPAs comparing the success center with the cumulative institution-wide student GPA (as provided in the scenario above) because the contexts are very different. Students directed to the center may be a particular population very different from the college as a whole, an inappropriate context in this situation. Rather faculty may want to examine student GPAs prior to using the facility and then examine this same set of students GPAs after using the facility. You may ask whether the very purpose of the facility is to attract students with low GPAs. You might ask, "Who are the students that access this facility? Are GPAs a good metric representing the function of the center or should it focus on student success in a single course? " If this GPA comparison does not address your key questions, don't hesitate to throw data out. Instead, you might begin with broader questions: "What type of success do faculty expect from students using the Student Success Center? What is the institutional context for the center?"

Principle 3 – Look for both direct and indirect data.[2]   Direct data actually measure the specific information faculty want. For example, measuring whether students who attended a particular activity actually learned the skills intended using a specific assessment measures direct data. Indirect data are those metrics that measure people's perception of the activity. You might ask, "Did students who accessed the facility report that the experience was helpful?" Direct data, often more accurate as to the actual activity, often do not provide guidance about how to make positive changes. Indirect data often provide insights into strategies to improve current practices. Good practice uses both direct and indirect data.

Principle 4 – Do not oversimplify cause and effect of data. In science, determining something does not cause a specific effect is often more important than concluding a specific cause did result in an effect. The reason for this is that until research has adequately tested all variables, one does not know that any specific cause was really the trigger, especially the only factor, causing a particular effect. Useful data never oversimplifies or over-values a single cause/effect relationship. Looking at our hypothetical scenario, one would not conclude that using the center caused a lower

---

2   Direct and indirect data- Direct data provide evidence of student knowledge, skills, or attitudes for the specific domain in question and actually measure student learning, not perceptions of learning or secondary evidence of learning, such as a degree or certificate. For instance, a math test directly measures a student's proficiency in math. In contrast, an employer's report about student abilities in math or a report on the number of math degrees awarded would be indirect data. Indirect data are sometimes called secondary data because they indirectly measure student performance. For instance, certificate or degree completion data provide indirect evidence of student learning but do not directly indicate what a student actually learned.

rate of transfer, degree, or certificate awards. One would also be hard pressed to say that the 5% of students that were successful for these outcomes were successful because of the center. There are many variables involved in any effect, particularly in education, and any one effect may have multiple potential causes.

Principle 5 – Use appropriate levels of data for appropriate levels of decisions. Make decisions based on the appropriate level of data. Do not use high-level institutional data (such as degree and certificate numbers) to make changes to homework assignments, number of sections, or the continuation for this success center. This high level data may indicate some further questions, but if the decision at hand involves altering sections, use data on sections. If faculty are addressing curriculum alignment, they should use data on specific curriculum. Using data about degree and certificate awards of those who accessed the Student Success Center may be useful to some extent, but it would be at too high a level to inform hours of operation or numbers of computer terminals within the center. Useful data have a close causal relationship to the appropriate level and authority of your target so the appropriate decisions can be implemented. In this scenario to inform hours of operation or number of computer terminals additional data is necessary. Indirect data such as student, staff and faculty perceptions about the hours of operation and direct data showing actual computer terminal use at various times throughout the day would be more appropriate to making these decisions.

Principle 6 – Perception is the reality within which people operate. When people perceive something in a particular way, one must deal with that perception. If a person is wed to a particular interpretation, data will not influence that perception unless used very carefully. Perceptions are most easily gathered through surveys or focus groups. Determine what the perception is and address that perception. For instance, in the case of the Student Success Center, asking students, faculty and administrators why students do or do not use the center provides information. The responses may or may not be factually accurate, but acknowledging what people perceive and asking how to improve the situation is important.

Principle 7 – Use of data should be transparent. Stakeholders should understand how to access data relevant to them. This is in contrast to using data as a powerful tool to marginalize or intimidate people, restricting access and limiting questions about the data and implications. Questioning data and prioritizing answers that data raise should be a collegial and transparent activity. Additional sources of data and subsequent questions to inform practice should be encouraged and valued. In addition to providing the Student Success Center data, faculty and staff should be asked what additional data should be examined.

Principle 8 – Consider carefully when to aggregate or disaggregate data.[3] Data is often aggregated (combined) to protect individual identities and confidentiality. Aggregated data examines student populations or course sections in contrast to revealing individual student results or faculty performance. For the Student Success Center, data clearly identified with a particular tutor or faculty member should not be contrasted with another individual's data. Data are often disaggregated to determine more discrete information and address components of the issue more effectively. In the Student Success Center scenario one might disaggregate data by students with basic skills needs, students with a specific ethnicity or first generation college students.

Principle 9 – Focus on data that is actionable.[4] It is always important to remember that data collection does not equate to action or improvement. Even the most valid and reliable data are not a substitute for action and will not by themselves motivate action. Some data provide information that leads to improved practice. Other data does not. Beware of data that are used to grade or rank issues, unless they are based on real values for the institution and the college, and do not provide information for improvement. US News and World Report ranks colleges and universities based on issues such as endowments, selectivity, and assessment by peer institution, alumni activity, etc. How would an institution improve these ranking criteria and do they actually represent criteria reflecting quality education? No Child Left Behind (NCLB) reveals another anomaly and danger in ranking. Once you have achieved the 90th percentile, a laudable achievement, a stable ranking at this 90th percentile is considered lack of improvement, yet the effort to go from the 90th to the 91st percentile is far greater than from the 40th to the 50th percentile. Instead faculty should select data that can be acted upon and used to change practice directly related to quality education. For the Student Success Center, don't just ask how many students used the Center; ask how students found the Center and what additional features would make it more useful, more inviting for students to use. Concentrate on data that leads to action and can inform change, intervention, or improvement.

---

3   Aggregated and disaggregated data - When data are combined together in a group this is aggregated data, e.g., all the sections of English 1A versus individual sections. When data are separated by certain variables, e.g., ethnicity, the data is disaggregated. There are benefits and problems with both.

4   Actionable data - is data that provides enough information to make changes. For instance, a school's ranking on SAT testing is not actionable. Disaggregating the school's data by average performance in math or English compared to a national average may provide information that can be addressed.

Principle 10 – Consider implications and the "What if?" Data need to be collected, examined, interpreted, and then appropriate decisions made to apply the data to make improvements. Once you examine the data ask, "If we do nothing with this data, what will happen? Where do the data project we are going? If these data are useful, what were the major factors that influenced them? If we act upon these data, how will we monitor the effect?"

Many educational researchers have described educational data as uneven, variable, lumpy, not precise, difficult to capture and describe, multifaceted, and a real challenge. But none of these difficulties make it something faculty should ignore. Ignoring data leaves us with only intuition, gut feeling, non-substantial arguments, and ineffective advocacy. Faculty owe students and our institutions more than that. Our task is to make the invisible nature of learning and education visible to others. Faculty can accomplish this task by learning how to analyze and interpret numerical data, surveys, artifacts, and other pieces of evidence and using this information to inform our practice.

"Education is not the piling on of learning, information, data, facts, skills, or abilities - that's training or instruction-but is rather making visible what is hidden as a seed."

Sir Thomas More

RESOURCES FOR LOCATING EXTERNAL DATA AND RESEARCH

Occasionally, senates and faculty have no input into creating institutional research agendas and need access to additional sources of data. Where can you get additional data to explore questions or look at limited evidence that has been provided? These excellent sources provide data at various levels and provide reliable information as it is officially reported by your college to external accountability reporting.

California Community College Chancellor's Office (CCCCO) Data Mart
http://www.cccco.edu/SystemOffice/Divisions/TechResearchInfo/MIS/DataMartandReports/tabid/282/Default.aspx

This interactive research site allows you to ask questions about student success, retention, awards, and student diversity by programs, colleges, or statewide. This is a powerful tool with information about your college and district.

CCCCO Accountability Report for California Community Colleges (ARCC) and ARCC Basic Skills Supplemental Report
http://www.cccco.edu/ChancellorsOffice/Divisions/TechResearchInfo/ResearchandPlanning/ARCC/tabid/292/Default.aspx

The Accountability Report are mandated reports to the Legislature which include important data about student success, retention, awards and progress statewide and by college. The basic skills supplemental ARCC contains the only real actionable data.

CalPASS
http://www.cal-pass.org/ (This requires a login that is very easy to apply for using the online request at the site. You will receive your personal login within about 48 hours and this is perhaps the most powerful database available for each individual college.)

"The only system that collects data about student success and transition from every segment of education, K-16. Informed by data, powered by inspiration and developed through collaboration—Cal-PASS partners identify problems, develop local solutions, and bring them to scale across regions and throughout California to achieve Success at Every Level."

CCCCO Fiscal Data resources:

http://www.cccco.edu/ChancellorsOffice/Divisions/FinanceFacilities/
FiscalServices/CCFS311PDFFiles/tabid/334/Default.aspx

"The CCFS-311 of a community college district is the vehicle for summarizing and communicating the results of budgetary decisions and transactions of all governmental, proprietary, and fiduciary funds for each annual fiscal period.

Fiscal Trend Analysis of the Unrestricted General Fund and Other Fiscal Data

http://www.cccco.edu/ChancellorsOffice/Divisions/FinanceFacilities/
FiscalServices/FiscalAccountability/DistrictFiscalTrendAnalysis04050809/
tabid/1564/Default.aspx

This analysis shows five-year trends for 50% law compliance, unrestricted reserves, number of FTES and percent of budget paid to payroll. These longitudinal data provide useful information about college trends and the impact of ongoing budget and planning decisions.

ACHIEVE

http://www.achieve.org/AboutAchieve

"Created in 1996 by the nation's governors and corporate leaders, Achieve is an independent, bipartisan, non-profit education reform organization based in Washington, D.C. that helps states raise academic standards and graduation requirements, improve assessments and strengthen accountability." Data relevant to individual states are particularly helpful and can be found in the state profile data link http://www.achieve.org/StateProfiles

(CPEC) California Postsecondary Education Commission

http://www.cpec.ca.gov/

"The Commission integrates policy, fiscal, planning, data, and programmatic analyses about issues concerning education beyond high school to the legislative and the executive branches of California government and to the general public."

NCES – National Center for Education Statistics

http://nces.ed.gov/

"The National Center for Education Statistics (NCES) is the primary federal entity for collecting and analyzing data related to education."

NCHEMS – National Center for Higher Education Management Systems
http://www.nchems.org/

The National Center for Higher Education Management Systems (NCHEMS) is a private nonprofit (501)(c)(3) organization whose mission is to improve strategic decision making in higher education for states and institutions in the United States and abroad. The NCHEMS Information Center for State Higher Education Policymaking and Analysis (The Information Center) provides state policymakers and analysts timely and accurate data and information that are useful in making sound higher education policy decisions. The Information Center is a comprehensive "one-stop-shop" for state-level higher education data and information, and a leader in coordinating the collection of missing data and information that are crucial for higher education policy analysis.