Title: The relationship between data quality and mixed results: Case studies from the American-affiliated Pacific region.

Authors and Affiliations:

Wendy Kekahio  
Senior Evaluator  
McREL’s Pacific Center for Changing the Odds  
1003 Bishop St., Honolulu, HI 96813

Akiemi Glenn, PhD  
Senior Researcher  
McREL’s Pacific Center for Changing the Odds  
1003 Bishop St., Honolulu, HI 96813

Sena Pierce  
Research Associate  
McREL’s Pacific Center for Changing the Odds  
1003 Bishop St., Honolulu, HI 96813
Abstract Body

Background / Context:


The Institute for Educational Sciences supports rigorous and applied educational research coordinated through Regional Education Lab (REL) programs which serve ten United States (U.S.) regions (http://ies.ed.gov/ncee/edlabs). The REL Pacific program is unique as it supports educational research across the Pacific including the State of Hawai‘i; the U.S. territories of Guam, the Commonwealth of the Northern Marianas Islands, and American Samoa; and three independent countries in financial and political compact with the U.S. – the Federated States of Micronesia, the Republic of the Marshall Islands (RMI), and the Republic of Palau, known collectively as the freely associated states (FAS). In addition to differences in language, culture, and governance across these entities, there are great divides in the technical infrastructure, resources, and human capacity that support baseline educational data collection, maintenance, and use. For example, while educators in Hawai‘i explore new ways to utilize the wealth of student level data available within the Hawai‘i state longitudinal data system, a high school vice principal in RMI enters student attendance data by hand for approximately 1,000 students stored in a stand-alone computer.

The exploration of causal relationships within natural educational settings is a complex enterprise that requires sophisticated experimental design and a high level of control. Studies can include the manipulation of a treatment or variable and the observation of its effect, or vice versa, the observation of an effect and a search for its cause, which has been likened to a detective searching for the culprit of a crime (Scriven, 1976 cited in Shadish, Cook and Campbell, 2002). Educational research conducted in the U.S.-affiliated Pacific Islands has largely followed the path of a detective looking for the source of specific issues, e.g., poor exam results, poor academic performance, or high rates of college remediation. A baseline study of extant data is the first step in identifying variables for control or manipulation in later experimental designs. What happens, however, when the basic building blocks of educational data (e.g., student identifiers, demographic, and enrollment information) are compromised, i.e., weakened by less-than-desirable data quality standards, missing, or not reported or not collected in the first place? How can researchers respond when challenges in data quality ensure mixed results of the best laid plans?

The panel will explore unexpected challenges with educational data in the Pacific as case studies from REL Pacific work to address the larger question: How can researchers respond when challenges in data quality ensure mixed results?

Panel Description / Outline:

The purpose of the panel is: (a) to present case studies of the impact of compromised, missing, or unreported or non-collected baseline educational data in the development and implementation of robust research; and (b) to generate discussion on how to resolve complications in foundational data. Each panelist will present a study that had to respond to or account for compromised,
missing, or unreported or non-collected baseline data. Though each study varies in scope, design, and purpose, all have been conducted in the U.S.-affiliated Pacific Region in areas of emerging data solvency (i.e., the ability to meet local needs to access and use reliable data to inform policy or practice).

Each panelist will present the background, methodology, and expected results of a study that was conducted in the region. Panelists will then discuss the shift in actual or expected results as a direct consequence of challenges with the baseline educational data available in the study setting. Each panelist will present their study followed by five minutes for clarification questions from the audience lead by the moderator. At the conclusion of each presentation the moderator will present a predetermined question related to the specific data challenge of the panelist. The audience will be instructed to consider the question and discuss with fellow audience members. The moderator will then solicit responses from the audience and conclude with responses from the panel. Potential discussion questions follow:

- What do you believe is the role of the researcher in the face of data that has been misinterpreted or misrepresented due to compromised data?
- Based on your experience, what considerations or extra steps can be taken when designing a study based on unknown extant data limitedly or not explored in prior research?
- In your experience what are effective ways for researchers to address the absence of key student data in existing educational data management systems? Is that the researcher’s role?
- In your experience, what actions can researchers take upon discovering key data is unavailable after a study has commenced?

The studies presented during the panel will be used to contextualize the challenge of implementing sound research in the face of unreliable or unavailable baseline data. The studies are briefly outlined in the following sections, organized by data challenge.

**Purpose / Research Question**

**Case One – Compromised Data in a U.S.-Affiliated Country**

Fifteen years ago a state education agency (SEA) in a U.S.-affiliated country in the North Pacific developed and implemented statewide testing procedures for 8th-grade students. Subsequent tests were created for 6th- and 4th-graders. The SEA began to notice a steady decline in the test scores of its students and contacted the REL Pacific for assistance in analyzing and interpreting the assessment data. The study was conducted as the first step to explore the decline in performance for further investigation.

**Case Two – Missing Data in a U.S. Pacific State**

A Research Alliance working with REL Pacific in a U.S. Pacific state requested a study to explore the distribution of students from the freely associated states (FAS) and their academic performance in relation to other student groups in the state. Two research questions guided the study: (a) What is the total number of students from the FAS in the Pacific U.S. state; and (b) what is the academic performance of FAS and non-FAS students in the Pacific U.S. state by
grade-level groups, as measured by norm-referenced and criterion-referenced standardized assessments?

**Case Three – Unreported Data in Two U.S. territories**
Educational leaders in two U.S. territories expressed concern about the college-readiness of graduating high school students. In response, a study was proposed to explore the condition of college readiness in the two U.S. territories. Two research questions guided the study: (a) What is the current state of academic readiness of high school students in the two U.S. territories for successful transition to postsecondary education; and (b) what are the sources and quality of available data on college entrance for Pacific Region students?

**Population / Participants / Subjects**

**Case One – Compromised Data**
The study included test data from approximately 2,000 4th-, 6th-, and 8th-grade students in a U.S.-affiliated country in the North Pacific from 1997 to 2011. Students were largely from similar cultural backgrounds and spoke English as a second language.

**Case Two – Missing Data**
The study included data from students attending pre-kindergarten through Grade 12 at public schools within a U.S. Pacific state from 2010 to 2012.

**Case Three – Unreported Data**
The study included graduating high school seniors who attended public high school in two U.S. territories in the 2009-2010 academic year.

**Research Design**
All three studies presented in the panels explored extant data.

**Data Collection and Analysis**

**Case One – Compromised Data**
Data from the U.S.-affiliated North Pacific country included 10,000 separate files in over 500 folders. Data were transferred from the SEA to REL Pacific after a data-sharing agreement was established. Data were analyzed using descriptive statistics.

**Case Two – Missing Data**
The REL Pacific received student-level demographic data and performance data on a U.S. nationally norm-referenced assessment and a criterion-referenced assessment from 2010 to 2012 from the U.S. Pacific state Department of Education (DOE) following the establishment of a data sharing agreement. FAS student-level data was aggregated by grade level (elementary – grades 3 to 5, intermediate – grades 6 to 8, and high school – grade 10) to increase the likelihood that at least 40 students would be disaggregated into target student groups (U.S. Pacific state Department of Education, 2011). Data was analyzed using descriptive statistics.
Case Three – Unreported Data
Data sharing agreements were reached with the Departments of Education in two U.S. territories. Student-level data for graduating seniors in the 2009-2010 academic year were received by REL Pacific. In addition, the cohort data was sent to the National Student Clearinghouse (NSC) Student Tracker in an effort to determine if students were enrolled in any college subscribing to the NSC database. Data were analyzed using descriptive and inferential statistics.

Findings / Results

Case One – Compromised Data
Analysis of the data revealed that the prior results of test performance were generated without matching student data. In other words, mean values were produced for different administrations of the tests, but did not reflect the performance of the same students across tests over time.

Case Two – Missing Data
Anticipated findings will include FAS demographic information such as home language, ELL status, place of birth, and age of entry in relation to students’ academic performance. It is expected that findings will demonstrate FAS students perform lower than other student groups on standardized tests. However, a limitation of the study is the lack of a single variable to clearly disaggregate first-generation and later-generation FAS students. The result is that potential differences in the two populations, for example English language proficiency and familiarity with the U.S. education system, may skew the overall academic profile of the FAS student population and lead to an incomplete baseline study upon which to base future research.

Case Three – Unreported Data
The anticipated findings from the study are expected to be limited due to the absence of student enrollment data in local institutions of higher education. Although the study utilized college enrollment data from the NSC Student Tracker, the institutions of higher learning within the two U.S. territories did not report enrollment data to the NSC. Therefore, though findings about student performance in high school will be used in later research, the lack of college enrollment data prevented an exploration of high school academic readiness in relation to college enrollment in the territories.

Conclusions

Extant educational data was included within the design of each study presented in the panel. In each case, data provided baseline information that could be built upon for future empirical studies. However, system-level challenges with compromised, missing, or unreported/uncollected data limited and/or redirected expected findings. Although the process of conducting empirically based research in areas of emerging data readiness is akin to taking one step forward and two steps back, the challenges are not unique to the Pacific. The studies presented in the panel serve to highlight the need for increased scholarship and discussion around the role of the researcher when faced with compromised, missing or unreported educational baseline data.
Appendices

Appendix A. References
