Technical Paper:

Data Infrastructure for Partnership Research: Structures and Processes Used in the Stanford-SFUSD Partnership

Moonhawk Kim
Jim Shen
Laura Wentworth

May 16, 2019

Executive Summary

There has been growing interest in the education sector in developing research practice partnerships. These partnerships are long-term, mutualistic, strategic relationships between researchers and practitioners that produced research related to the challenges experienced by practitioners and generalizable to the broader field (Coburn, Penuel, and Giel, 2013; Coburn and Penuel, 2016). There are a number of steps researchers and practitioners take to develop these partnerships. One of the more essential steps is the development of infrastructure for exchanging administrative data.

There has been little documentation of how partnerships develop the necessary infrastructure for exchanging, accessing, and using data. Some research practice partnerships have discussed the need for this data infrastructure (Lopez-Turley & Stevens, 2015; Conaway, et al., 2015; Nelson, London, & Strobel, 2015). Other studies on research use (Farley-Ripple, 2015; Honig, et al., 2017) and data use (Coburn and Talbert, 2006; Farrell, 2015) by practitioner partners suggests they need enhanced data infrastructure to support research use, data use, and generally evidence use by practitioners.

To help the field understand what the necessary data infrastructure looks like and how it gets operationalized, this technical paper describes the development and structure of the data infrastructure for the Stanford-SFUSD Partnership. Established in 2009, the Stanford-SFUSD Partnership supports 25-30 active projects at any given time, most of which are research projects requiring administrative data as part of their studies.

In this technical report, we describe the a) history context, and rationale b) the institutional structures and personnel and c) the systems for organizing and exchanging the data within the Stanford-SFUSD Partnership. We also discuss d) the benefits and challenges with this curriculum set of structures and systems, and e) discuss our future plans for improving the data infrastructure.

Key findings from this technical paper are:

1. **Allow ample time and resources the management of agreements**: It is a task in and of itself to develop and maintain an agreement for data exchange and use.
2. **Partnerships need personnel in all organizations to manage data**: To create the structures for data exchange and use of data within research across two or more institutions, there needs to be personnel attending to data from within each of the organizations involved.
3. **Personnel and structures need on-going development**: Even with personnel and structures managing the data, it is important to continuously build and strengthen internal skills, knowledge, and structures for data management within the practice institution (in this case a school district) and the research institution (in this case a university).
4. **Anticipate deviation from the standards for data exchange and use:** Partnership leaders must anticipate that other personnel will deviate from the standards of the data exchange and use agreement because of pre-existing relationships and the desire to expedite the data exchange.

Table of Contents

I. The Data Infrastructure Needed for a Research Practice Partnership **5**
History and Context 5
Developing and Streamlining Data Use Agreements 5
Institutional Structures and Personnel 6
University Research Staff: Stanford CEPA Data Manager 7
District Leader: SFUSD's Supervisor of Analytics 9
Third Party Broker: Ed Partners' Partnership Director 10
II. Organizing and Exchanging Data for a Research Practice Partnership **12**
Gathering Data from Around the District 12
Data available through the district's central data sources 12
Assessment Data 13
State Assessments 13
District Assessments 13
Separate Systems 14
CALPADS 14
Documentation 15
How to transfer data from SFUSD to warehouse 15
How the warehouse stores the data 16
Providing Data to Research Teams 17
III. Lessons Learned from the current configuration **18**
The benefits of this process 18
The challenges of this process 19
The Future of Data Infrastructure for Partnership Research 20
IV. Findings and Discussion **22**
Do not underestimate the management of agreements 22
Partnerships need personnel in all organizations to manage data 22
Personnel and structures need on-going development 22
Anticipate deviation from the standards for data exchange and use 23
References **24**
Appendix A: Sample language in agreement to house SFUSD data at Stanford **25**
Appendix B: Sample language in amendments to data agreements **33**

I. The Data Infrastructure Needed for a Research Practice Partnership

History and Context

Established in 2009, leaders from Stanford University Graduate School of Education (Stanford GSE), San Francisco Unified School District (SFUSD), and California Education Partners (Ed Partners) launched the Stanford-SFUSD Partnership. The purpose of the partnership was to move from the relationship with Stanford GSE and SFUSD from a happenstance relationship to a strategic collaboration. The general goal of the partnership is to support generalizable and actionable research that could be used by SFUSD leaders in local decision-making and beyond. The Dean of Stanford GSE, the Superintendent, Deputy Superintendent, and Chief of Research of SFUSD, as well as senior staff at Ed Partners have been involved in designing and leading the partnership from the beginning until the present day. Ed Partners, a third party non-profit, staffs a Partnership Director that supports the Stanford-SFUSD Partnership with facilitation, coaching, and infrastructure development, operations, and maintenance to support the approximately thirty active projects at any given time.

One of the most important things the SFUSD research department, Stanford GSE Dean's office, and the Partnership Director accomplished was signing a data warehousing agreement authorizing the establishment of a data warehouse at Stanford to facilitate the delivery of the SFUSD data for research to Stanford researchers. For the first two years of the partnership, Stanford researchers and the SFUSD research department administrators felt challenged by the volume of data exchanges and the time and resources it took to prepare those data extracts. Also, it took time to develop the data use agreements for each project, which had to be reviewed by the Stanford and SFUSD legal departments.While SFUSD worked with many researchers, 20% of the research projects on average were coming from Stanford University, which also motivated the school district to take steps like agreeing to warehouse SFUSD administrative data at Stanford.

Developing and Streamlining Data Use Agreements

To address these challenges with data exchange, in 2011, the Partnership Director worked with the SFUSD research department, Stanford GSE Dean's office, and the leadership of Stanford University's Center for Education Policy Analysis (CEPA) to streamline the agreements that need to be in place for data use and access. To do this, the leaders created a data use agreement template which Stanford researchers could easily fill out when applying to SFUSD to conduct research and request data. If Stanford researchers used the data use agreement template, the Stanford and SFUSD legal teams would not need to review the agreement as the template maintained consistent legal language that both Stanford and SFUSD legal counsel agreed upon. Also, an even larger commitment by SFUSD and Stanford to this partnership was when the members of the Partnership established an "umbrella" warehousing agreement between Stanford University GSE, CEPA and SFUSD to house SFUSD data and distribute the necessary data for their research projects to all Stanford researchers with a data use agreement. Different than a data use agreement template, which simplified legal contracts, the warehouse would require personnel

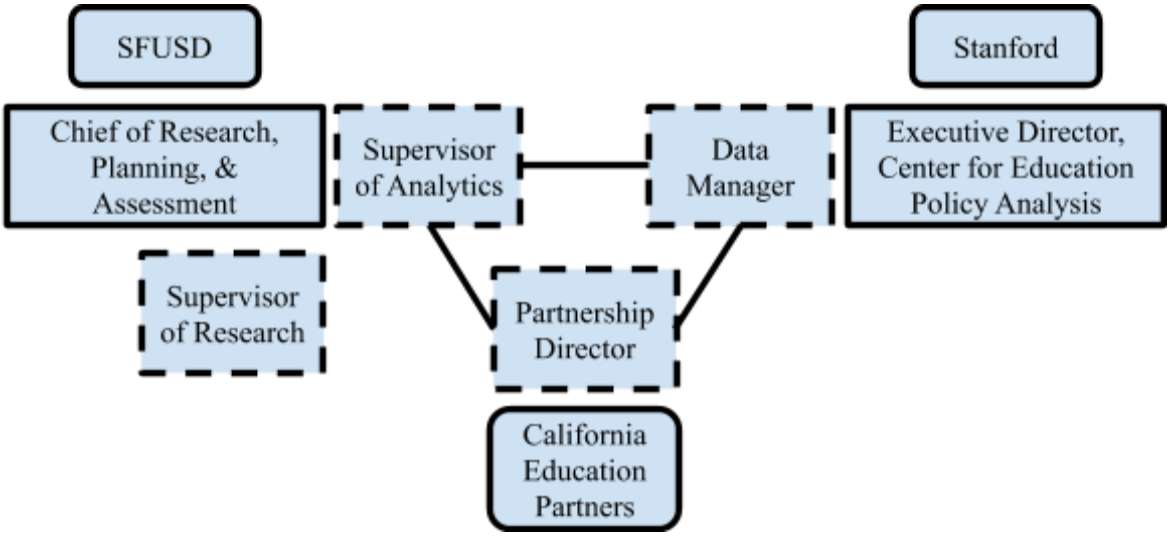
and management beyond hiring a partnership director. This undertaking key in moving the Stanford-SFUSD Partnership from a federation of projects to an actual partnership.

The elements included in the original data use agreement which warehoused the SFUSD data at Stanford, referred to as the "umbrella" agreement, were compiled according to the data needed for existing research projects between Stanford researchers and SFUSD administrators. Based on the original negotiated projects, this included K-12 student data, early education data, staff data (teachers and principals), and other types of student and school level data allowable by FERPA. One variable that is commonly in other data use agreements for research, which SFUSD does not share with any researchers is the indicator of whether students are eligible to participate in the Free and Reduced Price Lunch program. Federal regulations limit the use of this data solely to administer the lunch program, and SFUSD's legal department considers the use of the data as an indicator for income level in analysis to be outside what is legally permitted.

To keep the "umbrella" agreement updated, the Stanford research department, the Partnership Director, and the Stanford and SFUSD legal departments needed to amend the agreement when Stanford research projects required additional elements. (See Appendix A for an example of the language in the Stanford-SFUSD Partnership's most recent agreements between Stanford and SFUSD) From 2011-2014, the umbrella agreement was amended three times; from 2014-2017 it was amended once. (See Appendix B for an example of one of the amendment templates.) We also found that data use agreements for each project often needed to be amended to add new data elements or to extend the time frame of the research project. Sometimes researchers forget to request a variable of interest, while other times projects expand in scope and require additional data elements not previously requested or new data that the district previously did not collect. Therefore, the SFUSD research department and Stanford and SFUSD legal teams developed an amendment template which project leaders could fill out and sign.

Institutional Structures and Personnel

Figure 1:



As seen in Figure 1, the data infrastructure involves the two partnering organizations, SFUSD and Stanford University, and a facilitating third party non-profit, Ed Partners. Each of the organizations have internal structures that play a role in developing and maintaining the data infrastructure.

For SFUSD, the Supervisor of Analytics of the Research, Planning, and Assessments (RPA) department is the point person for managing and housing SFUSD data for research, and about 20% of his time is supported to data management for research, including supplying data to Stanford as well as other research institutions. For Stanford, the data manager at CEPA is the point person for managing and housing SFUSD data for research within Stanford University, and about 50% of his time is focused on maintain data infrastructure for the Stanford-SFUSD Partnership. For California Education Partners, the Partnership Director supports the Supervisor of Analytics and Data Manager in the data compilation and exchange. Another important role that the Partnership Director plays is to work with SFUSD's Supervisor of Analytics and the Executive Director of CEPA to negotiate the agreement which allows Stanford to house SFUSD data. The Partnership Director devotes about 10-20% of her time to supporting data infrastructure depending upon the whether the larger agreements are being negotiated. The roles of the supervisor of analytics at SFUSD, the data manager at CEPA, the partnership director are described in more detail here.

University Research Staff: Stanford CEPA Data Manager

The Stanford research data manager is in charge of maintaining the SFUSD-CEPA Data Warehouse. The data manager receives the raw data from SFUSD, organizes and cleans the data for storage in the data warehouse, and provides the approved data samples to research teams operating at Stanford University.

Fulfilling these responsibilities requires both technical and non-technical skills. On the technical side, the most basic requirement for the research data manager is proficiency in data management and cleaning in a statistical software package. Most researchers at Stanford utilize Stata for their research, so producing data products in Stata is most useful for the role at Stanford. An important part of this process is having strong organizational skills to track and document the data elements being received from the district and provided to the research teams.

Additionally, the research data manager needs to have some knowledge of statistical methods and research design principles. While the data manager does not conduct the actual statistical analysis at Stanford, understanding the research design and methods enables the data manager to ensure the data received is suitable for intended use for the data. For example, one project planned to study the effects of tracking based on the math validation test using a regression discontinuity design, which requires student test scores and the tracking assignment. However, the district initially provided a data file with only the tracking assignment, which was insufficient for the study. By understanding the data required for the intended design, the data manager was able to communicate with the district to procure the necessary data.

The data manager also needs to be able to read and interpret data use agreements and memorandums of understanding written by legal departments as well as communicate data issues with district and university staff. The data manager has to be able to operationalize the requirements set in data use agreements, such as what methods to use for data transfers and how to store the data in the warehouse. It also requires understanding the relevant data security and legal expectations and how those impact the transfer and storage of data. Most importantly, the data manager must be able to interpret the data use agreements that govern the data elements that each research team is approved to receive.

Often times the legal language does not correspond with the names of the specific data elements extracted from district information systems, requiring the data manager to work with district staff to work together to resolve these differences. For example, when a data use agreement lists "social emotional survey items" as one of the variables, the data manager has to determine which of the possible variables in the universe of SFUSD data correspond to this request. He consequently needs to deduce from the description of the project and use his understanding of the construct and intended use of desired measurements listed in the agreement.

A particularly nuanced component of the data manager's role is managing the expectations of faculty and student researchers who are not familiar with the inner workings of how the partnership acquires data. For researchers the district and the warehouse may initially appear to be a monolithic black box from which they receive data, not understanding that the RPA office must first collect the data from all of the data generating departments within the district before sending the data to Stanford. Simultaneously, researchers have their own deadlines to meet, leading to them pressuring the district and the data warehouse to provide the data as quickly as possible. The data manager helps explain the process to researchers and sets realistic expectations for how quickly the district and warehouse can provide the requested data.

With the need to resolve technical questions, legal language issues, and explaining the data acquisition and transfer process, the research data manager has to be adept at communicating with both the district supervisor of analytics and the Stanford researchers he is supplying with data. Given the importance of maintaining documentation and keeping multiple parties informed of problems encountered and their solutions, much of the communication in the data transfer process takes place in written form via emails. The data manager also frequently meets with researchers at Stanford in person to discuss issues with the data. In both cases, the data manager needs to understand and have solutions for the issues at hand but also to be able to communicate it to people who may not be familiar with the data or the data transfer process.

Managing a data warehouse in an RPP is a time intensive process during the initial set up phase where the infrastructure and procedures have yet to be built up and codified. However, once the initial groundwork is in place, maintenance of the data warehouse takes less time and is often a routine process. The data manager at Stanford currently spends approximately 50% of his time working on the SFUSD-CEPA data warehouse. Other responsibilities include providing data management and cleaning for other projects, managing CEPA's restricted use federal data

license, and working with the Stanford GSE administrative and IT staff to perform functions specific to CEPA.

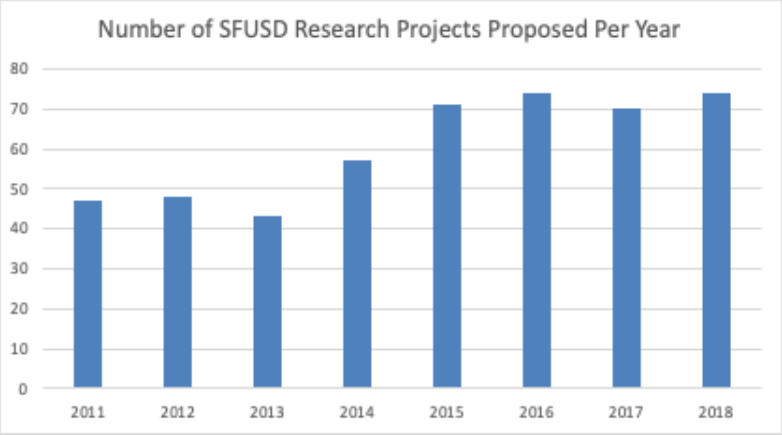
District Leader: SFUSD's Supervisor of Analytics

While SFUSD's supervisor of analytics is the counterpart to Stanford's data manager, the context is very different. Although managing data, data quality, and data exchanges with external partners would require a full-time staff dedicated to the work, the district does not have the necessary resources to devote to the work. As a result, the supervisor of analytics has multiple domains of responsibility, only one among which is the Stanford partnership.

The analytics supervisor's primary responsibility is to carry out quantitative analysis and to create more complex and analytical reports that go beyond the standard descriptive summary reports by RPA for the district. Projects in this domain range from extensive statistical evaluation of district's programs and initiatives to innovative summaries of student outcomes that offer novel insights to school leaders.

The position's secondary domain of responsibility is to improve the data analysis pipeline. On one end, the analytics supervisor collaborates closely with the Department of Technology to improve data systems and data governance processes. On the other end, the supervisor works with counterparts in other departments to develop and implement ongoing professional development of data analysts throughout the district. Although indirectly, this body of work does contribute to strengthening the long-term sustainability and development of the partnership with Stanford.

Figure 2:



The third domain pertains to overseeing and managing the process of exchanging data with external researchers and analysts. This involves reviewing and signing data use agreements (DUAs), working with district staff to collect and prepare appropriate data elements, and transferring the data securely to external partners. To give readers an understanding of the volume of research projects handled by the SFUSD research department, since 2011, the amount of research projects with SFUSD requesting data from the Stanford Data Manager has tripled, going from three projects in 2011 to nine projects in 2018. Also, as seen in Figure 2, the amount of research proposals received from SFUSD between 2011 and 2018 has grown from 47 proposals in 2011 to 74 proposals in 2018. Interestingly, between 2011-2018, of the 504 research proposals

submitted to SFUSD, almost 20% of the research applications submitted to SFUSD were from Stanford University researchers.

This last domain includes the partnership with Stanford University. The analytics supervisor reviews DUAs submitted by Stanford researchers to ensure that the requested elements are available and appropriate. In addition, the position manages the process of one-off transfers for data outside the umbrella agreement as well as the regular annual transfers for those elements covered by the umbrella agreement.

Over time, the technical aspects of data exchanges between SFUSD and Stanford have become streamlined and more efficient. However, the most challenging aspects on the district's side, for which the analytics supervisor needs to be highly skilled, are the following: investigate and obtain where in the district relevant data elements reside and trace data definition and quality issues resulting from alterations to the technical and organizational structures surrounding data in the district.

First, given the high degree of decentralization and of diversity in data sources, data are scattered throughout the district's central office departments, offices and teams. Sometimes, the official data about aspects of the district are collected and maintained by entities outside the district, e.g., Early Education's Quality Rating and Improvement System (QRIS) and Early Childhood Environment Rating Scale (ECERS). The analytics supervisor needs to be curious, tenacious, and patient in tracking down data.

Second, due to the fact that the district's data systems have gone through significant transformations and that thorough documentation of data definitions is lacking, the analytics supervisor ultimately spends a great deal of time investigating data quality and definition issues. For this reason, although data exchanges is the tertiary responsibility, supporting it—for non-Stanford projects as well as Stanford ones—ultimately requires the majority of the supervisor's time. Given the volume of research projects, with about 75 research projects proposed to the school district each year, most of which require complex data extractions and exchanges, the district would ideally have a staff person who would be devoted one-hundred percent to being the data manager for external research and evaluation partnerships.

Third Party Broker: Ed Partners' Partnership Director

The partnership director is an employee of a third party, non-profit organization, California Education Partners, whose leaders have ties to Stanford GSE, SFUSD, and a San Francisco-based foundation that funds Stanford GSE and SFUSD. In the early days of the partnership, the partnership director worked closely with the SFUSD research department to review Stanford research proposals and also improve access, agreements, operations, and engagement between Stanford faculty and SFUSD leaders. The partnership director also supports the facilitation of relationships between Stanford faculty with the desire to work in partnership with SFUSD administrators with district leaders that want to work on research with Stanford

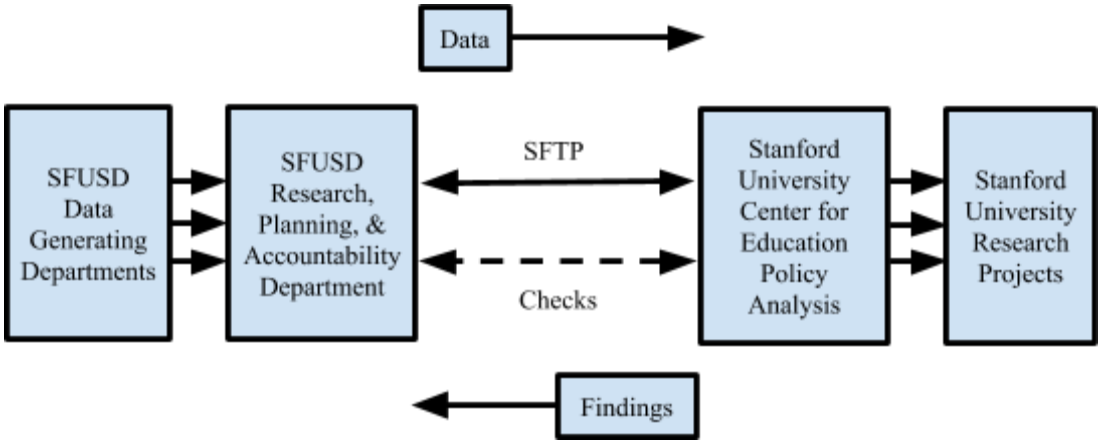
faculty. Finally, the partnership director maintains and implements the overall strategies for the partnership.

In the more recent history of the partnership, after SFUSD hired a supervisor of analytics, the partnership director relinquished most of the execution of SFUSD data exchange between the Stanford data manager and all other departments across the school district. Now, the partnership director supports the development, maintenance and renewal of the data use agreement which gives Stanford permission to house, exchange, and distribute SFUSD data. Maintaining the agreements about research, including data use agreements, takes less than 5% of the partnership directors' time, although during intense periods of agreement renewals, this percentage of time might increase.

II. Organizing and Exchanging Data for a Research Practice Partnership

There are specific steps the SFUSD supervisor of analysis and Stanford GSE data manager take to organizing and exchanging the data for the work of the research practice partnership. As seen in Figure 3, departments give data to SFUSD research department's Supervisor of Analytics, data is sent over the SFTP to Stanford's data manager. The Stanford data manager and SFUSD supervisor of analytics often go back and forth with these data checks to explain data sources, definitions, and receive specific details of needed data sets based on agreements. Ultimately, Stanford researchers then sending back findings from the analysis of the SFUSD data they receive. The process in Figure 2 shows a visual of this process and the subsequent sections provide in the sections titled: gathering data from around the district, how to transfer data from SFUSD to the data warehouse, how the warehouse stores data, and providing data to research teams.

Figure 3: The process for organizing and exchanging Data



Gathering Data from Around the District

SFUSD has launched a major project to enhance governance and management of data throughout the district. Currently, however, data reside in numerous physical and organizational places with unclear ownership. Accordingly, a major component of the SFUSD analytics supervisor's responsibility is locating and obtaining the data to transfer to Stanford.

Data available through the district's central data sources

Most of the student level data is available through what is known as "RPA Tables," which is a data mart that refreshes nightly from the district's student information system (SIS). Elements such as student demographics, school enrollment, dates of entry and exit are available through this source. The district maintains two historical snapshots of this data mart every school year---one from the California school census day (first Wednesday of October) and another from the end of the year.

While most of the data in RPA tables are automatically extracted and processed from the SIS, some data require staff intervention to validate and correct before the district can make them

available for research and analysis. Attendance is the main example for data requiring manual processing by district staff, to ensure accuracy for compliance and funding purposes.

Frequently, other data may require further validation and correction depending on any disjuncture that appears between practices implemented at schools and how data systems are structured. For example, changes at school sites in how grading periods are structured throughout the year can lead to errors in grades and credits data. Correcting errors in the data can sometimes lead to delays in the annual data transfer cycle.

Assessment Data

Students' assessment data constitute a large and important part of data for research and analysis. These data fall into one of two categories: California state assessments and district-wide formative assessments. The sources of data differ for these two categories.

State Assessments

The district receives data files from the state for California Assessment of Student Performance and Progress (CAASPP) and the English Language Proficiency Assessments for California (ELPAC), previously the California English Language Development Test (CELDT). The state's Department of Education (CDE) controls the compilation and the timeline of the public availability of these assessment data. The district is not permitted to share them with Stanford until the CDE lifts the embargo after an annual assessment cycle.

A challenge for longitudinal analysis is that state assessments—the main instrument for tracking student achievement—can significantly change over time. The current state assessment for California—the Smarter Balanced Assessment—is now in its fifth year, prior to which the state relied on a different and incomparable assessment. Even for internal analyses at the district, this break in the data series has provided an insurmountable obstacle, and internal evaluations simply limit the temporal scope of longitudinal analyses to that which can be covered by a single assessment.

District Assessments

The district administers numerous assessments throughout each school year. Many of these focus on literacy at the elementary level, such as Fountas and Pinnell (F&P). F&P is administered multiple times each year, in Spanish as well as English for students receiving instruction in Spanish. Other assessments include Reading Inventory (RI), Math Benchmark, Math Milestone Tasks, Integrated Writing Assessment (IWA), and Interim Assessment Blocks (IABs).

At the pre-kindergarten (PK) level, assessment data are different and challenging. The two principal assessments at this level are the Desired Results for Children and Families (DRDP) and Phonological Awareness Literacy Screening (PALS). Ages and Stages Questionnaire (ASQ) assess children's developmental milestones. Lastly, a set of assessments—Early Childhood Environment Rating System (ECERS), Classroom Assessment Scoring System (CLASS), and

Quality Rating and Improvement System (QRIS)—track the quality of classrooms and programs rather than students. In PK assessments, the involvement of 1) city and county governments as well as state and federal ones and 2) various third-party entities for technical assistance (e.g., WestEd) inject an additional layer of complexity in accessing and maintaining consistent data.

Separate Systems

SFUSD departments maintain separate data systems with limited interoperability among them. The three main systems are the student information system (SIS), the employee information system (EIS), and the financial information system (FIS). While RPA processes and transfers the SIS data (as discussed above), Human Resources independently transfers all the staff data extracted from the EIS. SFUSD currently does not regularly share data from the FIS with Stanford researchers.

In addition to these three main systems, district's data are currently maintained in numerous disparate systems and processes. For example, Special Education Department uses the Special Education Information System (SEIS). BASIS is used for students' behavior, attendance, and intervention data. Students' assignment to schools is done through yet another system maintained by the Educational Placement Center (EPC). Data on student homelessness is maintained separately as a "program" with complex entry and exit dates rather than as a simple flag on student status. On the contrary, students' involvement in diverse and complex Section 504 plans is currently reduced to a single indicator for the whole school year.

Such decentralization and dispersion of data and data systems make it challenging to systematically collect the metadata as well as the data to transfer to Stanford. Supervisor of Analytics expends a great deal of time tracking down the correct owner of data elements listed in the umbrella DUA and working with appropriate analysts inside and outside the district to obtain data in the correct structure.

CALPADS

California Longitudinal Pupil Achievement Data System (CALPADS) is the state's accountability data system. It collects and compiles data from local educational agencies (LEAs) and reports back authoritative metrics, such as graduation rates and students qualified to receive special funding, such as the Local Control Funding Formula (LCFF).

Researchers—including the Stanford partnership—sometimes request CALPADS data extracts as the most reliable and vetted versions of district's data files. This poses three challenges. First, the state does not independently collect and verify the data. Ultimately, the data in CALPADS is what the district has collected, processed, and submitted. Second, although the district submits raw data to CALPADS, the state has specific definitions and procedures for processing and calculating accountability metrics. As CALPADS matures, CDE is slowly becoming more transparent with these, but they have remained inaccessible by the district for many years.

Lastly, the district does not have a well-established process for working with CALPADS data. This results in failures to incorporate into the district's data systems any corrections that return from CALPADS or creation of new and inaccurate data records when authoritative CALPADS records of the same data element exist.

Documentation

In the process of gathering the data, the district data staff also assemble the documentation for the data. The external documentation lists the variables included in each data file, the description of the field, and their possible values. For the data that originate within the district's central data sources, particularly the RPA tables, the district maintains internal codebooks and documentation that it can share with the Stanford data manager and researchers. Similarly, the EIS data has its own set of documentation that comes from the HR department. Data that come from separate systems within SFUSD has documentation that is not directly integrated in the documentation for the RPA tables that the administrators of the system in question provide to the supervisor of analytics to forward to Stanford. Data that originate from external sources such as the state or third party testing services have their own documentation prepared by the respective source.

The internal documentation inherent within the data can potentially contain variables such as the date of extraction, date of data collection, and value descriptions for variables. This depends on the data files in question; the EIS data includes the date of extraction as one of the variables in the data, but the specific date of extraction is not available for the RPA tables. In some cases, the documentation is simply missing.

How to transfer data from SFUSD to warehouse

Once SFUSD RPA or HR has collected the data at the district, the next step is to transfer the data to the CEPA warehouse. Under previous versions of the data use agreement between Stanford and SFUSD, district staff would write the data files to CDs and hand deliver the discs to Stanford. There are obvious inefficiencies and security concerns in using such a method. In 2014 the data use agreement that governs the data warehouse was amended to allow for electronic transfers of the data between SFUSD and Stanford.

Currently CEPA maintains a secure FTP server for file transfers between SFUSD and Stanford. CEPA provisions and maintains user accounts on the server with access controlled folders for district staff to upload the data files to be transferred to Stanford. Using an FTP server enables quick turnarounds for file transfers from the district; there have been multiple instances where projects had tight schedules and deadlines where the ability for the district to transfer data quickly has been crucial. When SFUSD needs to receive individually identifiable data from research staff, such as analytical results or datasets with student identities that need to be shared with district staff, the CEPA research data manager can use the same server to send data in reverse. The FTP also also allows for tighter and more individually-customizable security by allowing access only to specific individuals and rescinding access when it is no longer needed.

SFUSD can provide data in several different formats. Data that are extracted from SFUSD systems are typically shared in some type of spreadsheet (e.g. Excel or CSV) format. When sharing data extracted from SFUSD administrative data systems, district staff also provides the SFUSD documentation files on the FTP as well. SFUSD also shares data that comes from external sources, such as standardized testing results that the district receives from the California Department of Education. With these sorts of data, the documentation is provided to the district from the same source or is publicly available.

Transfers of student data are driven by the schedule that SFUSD operates under. The warehouse receives one major extract of the SFUSD student information system from SFUSD RPA once a year, with the data typically becoming available by the October after the end of a given school year. SFUSD HR provides two major staff data extracts per year in April and October. Other data files, typically files needed for a specific project or files being newly integrated into the regular data transfers from SFUSD, are provided throughout the year as needed or available.

Once SFUSD uploads the data to the FTP server, the CEPA data manager downloads the data from the FTP server. To help ensure the security of the data, the data are left on the FTP for as little time as possible, although the data use agreement between SFUSD and CEPA allows the data to remain on the FTP for up to 14 days. While the FTP server is password protected, keeping data off the FTP is the most secure way of preventing unauthorized electronic access to the data.

How the warehouse stores the data

One of the key things that the warehouse does is organize the data that it receives from SFUSD. Currently the warehouse uses a self-documenting folder naming system that tracks the year and month of receiving the raw data as well as the type of data received. For example, staff data files received in April 2018 would be placed in a folder named “201804_Staff.” This system is self-documenting by making it apparent which folder contains what set of raw data and when the data was received. The warehouse also places documentation files that accompanied the raw data into the same raw data folder.

This file naming system was chosen because of the timing and frequency of the data transfers. The warehouse does not receive data sufficiently frequently where tracking the data at more than a monthly interval is necessary; the file naming system for the raw data files is sufficient for accurate tracking of the data for the purposes of the data warehouse.

The warehouse stores the data received from SFUSD on an encrypted desktop computer with two hard drives configured in a mirrored array to limit the possibility of data loss in the event of a hard drive failure. An additional backup of the data is kept on an encrypted external storage drive that is kept in a locked cabinet. Keeping the raw data off of servers outside of the direct control of the data manager reduces the likelihood that unauthorized users can gain electronic access to the data. Encrypting all of the computers and drives that the data is stored on reduces the possibility that an adversary can access the data even if they gain direct physical access to the computer. While no security arrangement is foolproof, taking steps to safeguard the data is

important not only because of its contractual obligation to do so but also as part of its ethical responsibility to the students and staff of SFUSD as well as to ensure that Stanford is a trustworthy partner in the RPP.

To make use of the raw data, the warehouse writes a set of syntax files to process the raw data into "master" data sets as an intermediary step before providing the data to research teams. The master data files are cleaned versions of the yearly raw data organized into longitudinal data files. This is done for the data files most commonly used by the research teams at Stanford, such as the yearly student data extracts and the biannual staff data extracts. By pre-processing the raw data into the master data files, the warehouse is generally able to significantly reduce the turnaround time between receiving a data request from a research team and providing a research team with the data that they have requested. The only instance in which this does not hold true is if a research project requests data that the warehouse does not already hold and requires SFUSD to transfer the files to Stanford specifically for that research project. Like the raw data, the master data are stored locally on the encrypted warehouse computer. However, the syntax files are backed up onto Stanford servers, as the syntax files themselves do not contain any sensitive information.

Providing Data to Research Teams

When SFUSD approves a research team's data use agreement, the data manager provides the appropriate sample of data as outlined in the data use agreement. Typically research teams are restricted in the number of years of data and variables out of each data file that they are allowed to receive. For data that exists in the master data files, this is as simple as extracting the appropriate subset of data as defined in the data use agreement. For data that is not in the master data or already in the data warehouse, the data manager additionally cleans the data before ensuring that the correct subset is being provided to the research team. This is more time consuming but only done in instances where projects have requested data that is not commonly used by other research teams.

In both instances, the data manager provides the data with a set of anonymized unique student identifiers instead of using official SFUSD student ID numbers. This is another step to protect the confidentiality of individually identifiable information. Only in rare cases, with explicit approval in the data use agreement, do research teams receive data such as names or official ID numbers that would enable researchers to individually identify students. The unique student identifier, whether the scrambled or official version, is the primary key used to link observations between the separate data files provided to the research teams.

Once the data has been prepared, the data files for each project are uploaded to an access controlled folder on the FTP server. Research teams have fourteen days to download the data, after which the data is removed from the server. It is the responsibility of each research team, which have their own data use agreements with SFUSD, to securely store the data that they receive from the data warehouse.

III. Lessons Learned from the current configuration

As the Stanford and SFUSD personnel develop and execute this infrastructure, a few benefits and challenges about the process emerges. Additionally, these insights also suggest ideas for the future of the partnership and the functioning of the data warehouse.

The benefits of this process

When the process works as intended, there are numerous advantages to the process. One of the main benefits of the current framework is having a single point of contact at both SFUSD and Stanford for the data transfer process. This reduces the opportunities for miscommunication between the district, data warehouse, and researchers when setting up research projects. The sets and types of data required for each project, what data a project has already received, and known problems with the data are all project-specific issues that are more easily tracked and resolved through a centralized warehouse.

While the warehouse could be centralized at the district end instead of the university, the institutional structure of the university allows it to better serve the research needs of the partnership. When researchers at Stanford have questions about or issues with the data, they are able to go directly to the data manager to resolve these issues rather than having to different departments within the district. This also serves to filter out issues that can be resolved just within the data warehouse from those that actually need involvement from the supervisor of analytics or the departments at the district that generated the data.

By having the same staff handling the data transfers to the warehouse and providing data to the research teams, the warehouse is able to build up a large body of institutional knowledge. This encompasses both the data generation and transfer process as well as aspects of the data itself. Knowing which departments are responsible for generating each data file, how variables within datasets are produced and how reliable they are, and the availability of certain datasets and variables in specific years, are all part of the body of institutional knowledge built up in formal documentation as well as interpersonal connections over the years. This has helped reduce the time that researchers spent waiting for known answers to data questions, made it easier to find the correct information to answer new questions, and reduced the amount of time that district staff spend answering questions.

Another advantage of the data warehouse is managing the transfers of longitudinal data. By standardizing and organizing the data provided by SFUSD, the warehouse ensures that the data provided to the research teams is consistent across years as well as providing updates as available in the same format. The Stanford Data Manager organizes the data and performs some data cleaning functions, but is not making coding decisions or extensively preparing the data for analysis. Frequently research projects have data use agreements that allow them to have not only historical longitudinal data but data that has yet to be generated from subsequent school years. The data warehouse enables both timely access to new data as well as ensuring that the data matches with the previous years' of data.

The challenges of this process

The data warehouse does face some challenges even with its almost ten year history. The largest challenges stem from a few sources which cause pressure on the data warehouse to deviate from the standard process.While the procedures outlined above generally work well and serve the needs of facilitating research while protecting the security the data, there have been instances where departments within SFUSD and researchers at Stanford have bypassed the official data transfer pipeline. Various departments within SFUSD often collaborate with Stanford researchers to answer specific research questions. Prior to establishing our data use agreements and strict procedures for data exchange, some departments sent data directly to the researcher at Stanford. When this happened the data was frequently transferred through unsecure means, such as district staff emailing unencrypted/unscrambled data spreadsheets to researchers.

While this was easy and efficient for the individual departments and researchers, it sidesteps the institutional safeguards built into the research practice partnership and the data warehouse pipeline. One obvious issue is that data was not properly sanitized and transferred in a way that does not properly protect the personal information of the people within the data. It also makes it more difficult for the data warehouse to support the researchers in the future if and when they come to the warehouse for follow up data; if the supervisor of analytics and the data manager are not informed about the rationale for the data requests and what sets of data the project needs, it is difficult to provide the data that the research team needs.

This issue typically arises out of the lack of knowledge about the standards for data exchange among individual departments and researchers. For instance, both researchers and district staff may not understand that all projects need centralized approval not only from SFUSD but also from Stanford's Research Compliance Office/Institutional Review Board (IRB). While the data provided to research projects through the warehouse is anonymized and may not end up being classified as human subjects research due to the inability to individually identify subjects, Stanford research projects need the IRB to make that determination even if SFUSD has approved the project.

While not specific to the partnership, the IRB process itself can be a source of confusion for both Stanford researchers and SFUSD staff. The IRB process is specific to the university and does not involve district staff, who only need the end result of having the approval letter from the IRB. At Stanford, the IRB office typically asks researchers to obtain formal approval via its online submission process for human subjects research. Most research conducted using only using administrative data is in the exempt from ongoing IRB approval given researchers never make contact with the human subjects as well as the research itself utilizing secondary data. However, IRB has also occasionally told researchers that they do not need to go through the formal process, which deviates from the usual process of submitting the formal approval letter to the district.

In other cases, SFUSD staff or Stanford researchers have wanted to expedite the data transfer for reasons such as the district wanting results for time-sensitive policy decisions or Stanford researchers needing to make deadlines for publications or conference submissions. This can be a

source of friction as they may attempt end runs around the formal data transfer process or otherwise pressure the partnership to get the data as quickly as possible. Another logistical issue that has arisen given the longevity of the partnership is researchers with ongoing projects who graduate or otherwise leave Stanford university and still expect to receive data from the warehouse. Because the data warehouse only serves researchers working within the Stanford-SFUSD partnership, researchers who have left must be sponsored by faculty remaining at Stanford so that their research technically remains a Stanford project.

The efforts to eliminate these unsupervised data transfers comes from both sides of the research practice partnership. The partnership director, supervisor of analytics, and data manager work to inform SFUSD departments and Stanford researchers of the proper procedures for transferring data between the district and Stanford.

The Future of Data Infrastructure for Partnership Research

While the Stanford-SFUSD Partnership has made great strides in strengthening its data infrastructure, there is a need for having a data warehouse within the school district to support multiple partnerships. SFUSD has developed internal capacity to manage data exchange and improved the compilation of data for use in research by having a full time person to manage a data warehouse for research purposes. Stanford University has been providing a substantial proportion of the funding for this position, and SFUSD now wants to think of a way that they can rely less on the resources from Stanford and also make the data for research available to all partnerships. To make an adjustment for this drive for more access to the data for research beyond Stanford is by requesting fees from researchers requesting data on a project-by-project basis or an annual fee for institutional level partnerships with other partnering universities where more than one professor or researcher is accessing data.

On the SFUSD side, there is also a move to create a data archive specifically for research within the district. Currently, data are pulled by the research department personnel project-by-project and there is not a single data warehouse maintained for research. To build the internal data warehouse within SFUSD, the research department needs to work closely with the Department of Technology, the department that owns the hardware infrastructure for storing data at the district. Part of the challenge of developing this internal warehouse is having the staff resources to develop and maintain the databases.

The Unified Data Systems (UDS) project that launched in the Spring of 2018, however, may significantly improve the systems at SFUSD and obviate a separate data warehouse within the research department. In particular, the vision of UDS is to rigorously and centrally archive all the data in the district. The data will be validated and warehoused in the cloud. Custom-generated data marts will provide access to real-time and accurate data for all reporting and analysis needs. One such data mart can be dedicated to supplying the Stanford data warehouse. Realization of this vision, anticipated in approximately 2020-2021, will eliminate the need for a data manager on the SFUSD side to track down and supply data extracts to Stanford.

On the Stanford side, the warehouse could be expanded and improved on in various ways. Most of the changes would be from a technical perspective and would require an amendment to the data use agreement with the district. One of the changes that would lead to a more robust warehouse would be improving the way that the warehouse stores the data by implementing an off site backup. Currently the warehouse has an encrypted computer that stores the data locally and an encrypted external backup drive that also exists on site. However, the data currently cannot be stored off site, leaving the warehouse vulnerable to data loss in the event of physical destruction or theft of the current storage devices. While the data could be reconstructed from the district archives and the code for processing the data is stored in multiple locations, this would be a more time intensive process than having a remote backup.

Another change that the warehouse could implement for scalability is introducing an intermediary step of data organization and storage between the raw data receipts from the district and the master data files that are used to provide data to research teams. Currently, the warehouse maintains obsolete data files (whether due to errors discovered in the data, superseded by more complete files, or other reasons) in the raw data repository but rewrites the code for generating the master data files to exclude those and only use the updated versions. While the code is self-documenting in that it is clear which is the most current set of raw data files from which files that the code references, introducing an intermediary storage location of the most updated files for each type would potentially reduce the chances for confusion in the event that multiple data managers had to reference the code or if the type of data files stored in the warehouse increased.

Similarly, the warehouse could expand the practice of producing master data files from only using them for data that multiple projects would potentially need to encompass all of the types of data stored at the warehouse. While there currently there are few potential users of one-off data files, if the RPP scaled up in size and scope of the research projects that it conducts, the types of data that researchers would need on a regular basis would increase and justify the creation and maintenance of more master data files.

IV. Findings and Discussion

As a whole, this technical paper provides a broad overview of the current status of the data infrastructure supporting the Stanford-SFUSD Partnership. The patterns across the larger paper suggest a set of five insights about develop data infrastructure for a research practice partnership. These insights outline a set of broader implications which may be useful for other leaders developing data infrastructure within research practice partnerships.

Do not underestimate the management of agreements

It is a task in and of itself to develop and maintain an agreement for data exchange and use. The beginning history of the development of this data infrastructure was very much tied to the development of this data use agreement to help facilitate one of the largest challenges in the partnership - access to data. The data use agreement needs to be revisited every three years, based on the recommendation of the Stanford and SFUSD legal departments, and takes about a year to negotiate as all organizations need their legal counsel to review any updates to the agreement language.

Partnerships need personnel in all organizations to manage data

To create the structures for data exchange and use of data within research across two or more institutions, there needs to be personnel attending to data from within each of the organizations involved. The partnership needs a data manager within SFUSD to manage the relationship with the multiple departments that hold the data and the Department of Technology that manages the raw data. For example, from 2010-2016, SFUSD received 325 application for research in the district, about 50 research application a year on average, and a substantial amount of projects needing administrative data. With about 18% of the research projects from 2010-2016 coming from Stanford University, it also helps to have a data manager within the Stanford as well as the school district to help researchers access and understand data more easily. This allows Stanford researchers to have a point person for questions and requests, and it reduces the amount of time the district leaders spend responding to researchers and increases the time district leaders can spend on data quality. When we started warehousing the SFUSD data at Stanford, the data manager received three requests for administrative data in 2013, and by 2018 was receiving nine requests for administrative data in 2019, which is triple the amount of data requests in five years. With the strengthening of the partnership and consequently a growing amount of Stanford researchers requesting data, the Stanford data manager provides needed capacity to manage the growing data infrastructure requirements needed for partnership research.

Personnel and structures need on-going development

Even with personnel and structures managing the data, it is important to build and strengthen internal skills, knowledge, and structures in an on-going fashion with the practice institution (in this case a school district) and the research institution (in this case a university). For example, SFUSD needs to create its own warehouse of data for use in research. It is a herculean effort to collect data from across the school district departments and organize data for research within the

school district. Also, Stanford data managers need ongoing maintenance of their systems for organizing and exchanging SFUSD data.

Anticipate deviation from the standards for data exchange and use

Partnership leaders must anticipate that other personnel will deviate from the standards of the data exchange and use agreement because of pre-existing relationships and the desire to expedite the data exchange. When this deviation happens, the partnership director can help the supervisor of analytics and the data manager guide the personnel back on the right track, or in a politically correct way, explain to the research or practice personnel the policies guiding data exchange and the rationale for abiding by them.

References

Coburn, C. E. & Penuel, W. R. (2016). Research-Practice Partnerships in Education: Outcomes, Dynamics, and Open Questions. *Educational Researcher, 45*(1), 48-54.

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts. New York, NY: William T. Grant Foundation.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence use in districts: Mapping the terrain. *American Journal of Education*, *112*(4), 469-495.

Conaway, C., Kessler, V., & Schwartz, N. (2015). What research do state education agencies really need? The promise and limitation of state longitudinal data systems. *Educational Evaluation and Policy Analysis*, *37*(1S), 16S-28S.

Farley-Ripple, E. N. Research Use in School District Central Office DEcision-making: A Case Study. *Educational Management Administration and Leadership.* 40(6), 786-806.

Farrell, C. C. (2015). Designing school systems to encourage data use and instructional improvement: A comparison of school districts and charter management organizations. *Educational Administration Quarterly*, *51*(3), 438–471.

Honig, M.I., Venkateswaran, N., McNeil, P. (2017). Research Use as Learning: the Case of Fundamental Change in School District Central Offices. *American Educational Research Journal*, 54(5), 938-971.

Nelson, I. A., London, R. A., & Strobel, K. R. (2015). Reinventing the role of the university researcher. *Educational Researcher*, *44*(1), 17-26.

Turley, R. N., & Stevens, C. (2015). Lessons from a school district-university research partnership: The Houston Education Research Consortium. *Educational Evaluation and Policy Analysis,37*(1S),6S-15S.

Appendix A: Sample language in agreement to house SFUSD data at Stanford

AGREEMENT FOR CONFIDENTIAL DATA EXCHANGE BETWEEN SAN FRANCISCO UNIFIED SCHOOL DISTRICT AND STANFORD UNIVERSITY

This Data Exchange and Confidentiality Agreement ("Agreement") is entered into by and between San Francisco Unified School District (hereinafter referred to as "SFUSD" or "District") and the Board of Trustees of the Leland Stanford Junior University (collectively the "Parties"), and is dated for convenience [date]. This Agreement sets forth the means to be used by Stanford University to ensure the confidentiality and security of information and data to be exchanged between SFUSD and Stanford University.

A. THE FAMILY EDUCATIONAL RIGHTS AND PRIVACY ACT

The Family Education Rights Privacy Act ("FERPA") permits the release of personally identifiable student data without prior written parental consent if the release is to "organizations conducting studies for, or on behalf of, educational agencies or institutions for the purpose of developing, validating, or administering predictive tests, administering student aid programs, and improving instruction, if such studies are conducted in such a manner as will not permit the personal identification of students and their parents by persons other than representatives of such organizations and such information will be destroyed when no longer needed for the purpose for which it is conducted." 20 USC 1232g(b)(1)(F). Each party agrees to comply with FERPA, and other applicable laws, in its performance of this Agreement.

B. PURPOSE

SFUSD and Stanford University agree that SFUSD will provide Stanford University with confidential data pursuant to FERPA section 1232g(b)(1)(F), for the purpose of research as described in this Agreement. Stanford University will store the data on the District's behalf in a secure (that is, a system or computer using current industry standard technology to protect against unauthorized access including encryption software), password-protected computer, server or other electronic device at Stanford University's Center for Education Policy Analysis (CEPA). CEPA will share data with other researchers at Stanford University only with the prior approval by SFUSD as to specific data. Individual researchers must establish approved data use agreements with SFUSD and approval from Stanford University's Institutional Review Board (IRB) for their specific research projects to request data for specific research projects. CEPA will provide the approved data to a Stanford University researcher only once the data use agreement has been approved by SFUSD officials, and CEPA has received written permission (via email or other means) from SFUSD to give the data to Stanford University researcher or researchers based on the signed data use agreement.

Data provided by SFUSD for the purposes of this Agreement will include only the data legally permissible to provide to outside agencies for the purposes of research.

Background:

Many Stanford Research Organizations/Stanford Researchers at Stanford University make requests for quantitative data from SFUSD based on data use agreements established between the Stanford Research Organizations/Stanford Researchers and the District. In the past, SFUSD had to fulfill multiple requests for data from Stanford University, and often these requests were for similar data within very similar parameters. To avoid duplicating efforts on the District's part, Stanford University and SFUSD entered into this Agreement for Stanford University to receive and store SFUSD data. This also facilitates prompt transfer of data from SFUSD to Stanford University Stanford Research Organizations/Stanford Researchers who have data use agreements with SFUSD.

C. SCOPE

Part 1: Cooperation in Providing the Data

1.01 SFUSD agrees to provide CEPA with downloads of data that is legally permissible to share with a research institution. These downloads happen within approximately the first two months of each semester of each school year of this Agreement.

1.02 CEPA agrees to provide back to SFUSD extracts of District data when requested by District administrators in the Research, Planning, and Assessment office. CEPA agrees to provide this data according to the same protocols outlined for data exchange in this Agreement.

1.03 SFUSD and Stanford University will make sure that all appropriate data is transferred and stored pursuant to the terms of this Agreement. CEPA data manager/s will maintain a spreadsheet tracking the exchange of SFUSD data between SFUSD and CEPA, and will make such spreadsheet available to District RPA administrators upon request.

1.04 Prior to receiving any District data, a requesting Stanford Research Organization/Stanford Researcher must have established a data use agreement with SFUSD that outlines the data requested with a high degree of specificity.The data use agreement will be prepared by SFUSD and appropriately signed by an authorized SFUSD representative and the Stanford Principal Investigator, and such agreement shall include an appropriate list of requested data elements.

Part 2: Agreement to Provide Data to Other Organizations within Stanford University

2.01 The Faculty Director and the Executive Director of CEPA agree to devote personnel time to managing the SFUSD data and fulfilling data requests from other Stanford Research Organizations/Stanford Researchers with funding support from the Stanford University Graduate School of Education. Data requests from other Stanford Research Organizations/Stanford Researchers will only be approved by CEPA data managers once the Stanford Research Organization/Stanford Researcher has an approved data use agreement with SFUSD and CEPA receives a written approval/request to fulfill the data request by SFUSD.

Part 3: Protection and Security of Personal Information

3.01 CEPA will use appropriate and reasonable efforts to store all SFUSD data on a password-protected, secure server with encryption connected to a desktop computer in a locked office. Back-up copies of the data (on CD) will be stored in a locked file cabinet in a locked office. Only CEPA's Faculty Director, the Executive Director and data manager will have access to the SFUSD data set stored at CEPA.

3.02 CEPA will transfer data to Stanford Research Organizations/Stanford Researchers through a Secure-FTP (file transfer protocol) site upon receipt of a written approval/request from SFUSD pursuant to the terms of this Agreement. Identification numbers included in the data from SFUSD will be de-identified by CEPA before being shared with any Stanford Research Organization/Stanford Researcher, unless approval has been granted by SFUSD to share unscrambled identification numbers and such approval is reflected in the data use agreement with the Stanford Research Organization/Stanford Researcher. Stanford will limit access to the algorithm to convert the de-identified numbers to those with a legitimate educational interest in accessing the data in order to carry out this Agreement.

3.03 CEPA will maintain a spreadsheet tracking the exchange of data between CEPA and other Stanford Research Organizations/Stanford Researchers, and make such spreadsheet available to District RPA administrators upon request.

Part 4: Benefits to School and/or District

4.01 SFUSD has a limited amount of resources and wants to make sure it has a productive partnership with Stanford University. Therefore, SFUSD wants to expand its capacity to fulfill data requests from Stanford Research Organizations/Stanford Researchers in a timely manner and set up and maintain a secure and legally compliant data center on the Stanford University campus. This will diminish the time and resources SFUSD currently devotes to these data requests. This will also allow SFUSD to fulfill Stanford Research Organizations/Stanford Researchers' data requests in a timely manner.

Part 5: Timeline

5.01 CEPA agrees to provide Stanford Research Organizations/Stanford Researchers with data as approved in writing by SFUSD within 30 days of receipt of a written approval/ request by SFUSD to transfer data. CEPA will not transfer any data that has not been specifically requested in writing by SFUSD to be transferred. This timeline includes the initial submission of a proposed data use agreement by a Stanford Research Organization/Stanford Researcher to SFUSD as well as any back and forth discussion needed between the Stanford Research Organization/Stanford Researcher and CEPA to make sure the Stanford Research Organization/Stanford Researcher has all of the approved data elements he or she needs. This timeline does not include additional requests for data made by the Stanford Research

Organization/Stanford Researcher after the initial request. SFUSD will make every effort to meet this timeline, however, during periods of peak workload, some delays may occur.

D. NATURE OF DATA

To address the purpose and scope of work detailed in this Agreement, Stanford University requests the specific SFUSD records or data listed in Attachment A ("Specific Records on Data Elements") which shall be incorporated into this Agreement by reference as though fully set forth herein.

E. TRANSFER OF DATA

SFUSD and Stanford University shall use a Secure, mutually agreed upon means and schedule for transferring confidential information. At no time will SFUSD or Stanford University send data electronically in an unsecured format to or from the parties. The parties agree that the data may be electronically transferred via secure file transfer protocol (SFTP) accessible to authorized users secure login only.

First, the data will be transferred from SFUSD to Stanford University. Stanford University's Graduate School of Education has agreed to Securely store the data at CEPA, under the leadership of the CEPA Faculty Director and Executive Director.

Second, CEPA will devote personnel to manage the incoming data, make sure the data are in a readable format with guidance from SFUSD, organize the data, and work with SFUSD to understand the data elements. This includes maintaining a codebook (i.e., detailed documentation – source, date, and definition) of all the data from SFUSD.

Third, any Stanford Research Organization/Stanford Researcher that would like to conduct research with SFUSD data must complete and submit a proposed data use agreement to the District's Office of Research, Planning and Assessment. SFUSD will assess the proposed data use agreement, and if such agreement is approved, will assess what data elements are appropriate to be shared with the Stanford Research Organization/Stanford Researcher for the purpose of the project outlined in the data use agreement.

Fourth, for approved data use agreements, Stanford University personnel will contact the CEPA data manager to request a secure transfer of data elements outlined in the approved and fully executed data use agreement to the Stanford Research Organization/Stanford Researcher, and provide the CEPA data manager with a written request (via email or other means) to supply the approved data to the Stanford Research Organization/Stanford Researcher.

Fifth, the CEPA data manager will securely transfer these data to the Stanford Research Organization/Stanford Researcher within 30 days of receipt of the written request from SFUSD.

Sixth, once the data is transferred via SFTP to Stanford Research Organization/Stanford Researcher, the CEPA data manager will delete SFUSD data from the SFTP after receipt by the project and/or by the end of 14 days.

F. PERIOD OF AGREEMENT; EFFECTIVE DATE

The period of this Agreement shall be from [date] through [date]. This Agreement shall be effective when fully signed by the parties, and shall continue in full force and effect through the period of the Agreement unless terminated earlier by either party pursuant to Section I ("Termination") herein.

G. STANFORD UNIVERSITY RESPONSIBILITIES

Stanford University agrees to the following confidentiality statements:

1. Stanford University acknowledges that these data are confidential data and proprietary to SFUSD, and agree to use reasonable and appropriate measures to protect such information from unauthorized disclosures and to comply with all applicable District policy and Local, State and Federal confidentiality laws and regulations including but not limited to the California Education Code and the Family Education Rights and Privacy Act (FERPA).

2. Stanford University shall designate the CEPA Faculty Director as the person responsible for the security and confidentiality of the data and will notify SFUSD immediately in writing of any change in designee.

3. Stanford University will use appropriate safeguards to prevent the use or disclosure of the information other than as provided by this data use agreement.

4. Stanford University shall instruct all staff with access to data provided by District to Stanford University pursuant to this Agreement in the requirements for handling confidential student information, and require each person who will have access to such information to sign an agreement to comply with the confidentiality provisions of this Agreement.

5. Stanford University shall not assign this Agreement or any portion thereof to a subcontractor or other third party without the prior written consent of SFUSD, and any attempted assignment without such prior written consent in violation of this Section shall be without effect and shall automatically terminate this Agreement.

6. Stanford University agrees that the research shall be conducted in a manner that does not permit personal identification of students, parents or staff by individuals other than representatives of Stanford University who have legitimate educational interests in the information.

7. Stanford University will report only aggregate data and will not report any individual data, nor will data be reported in a manner that permits indirect identification of any individual.

8. Stanford University will not contact the individuals included in the data sets without obtaining advance written authorization from SFUSD.

9. If a research proposal from a Stanford Research Organization/Stanford Researcher includes a survey of students, or a survey of any other subject, then the Stanford Research Organization/Stanford Researcher shall obtain advance written approval from SFUSD's Office of Research, Planning, and Assessment ("RPA) and Stanford University's Institutional Review Board (IRB), and if such approval is granted by each of those, then Stanford Research Organization/Stanford Researcher shall also obtain prior written informed consent from the parent or legal guardian of each student participating in such survey where students are the survey subjects. If the use of surveys will be proposed, the surveys must be attached to the research application and subsequently approved by SFUSD's RPA in writing prior to use.

10. Stanford University shall not re-disclose any individual-level data with or without identifying information to any other requesting individuals, agencies, or organizations without prior written authorization by SFUSD, and Stanford University shall not disclose to any Stanford Research Organization/Stanford Researcher any personally identifiable student information (e.g. name, date of birth, address…) unless express prior written permission to do so is granted by SFUSD for a specific project only, as provided for in this Agreement.

11. Stanford University shall use the data only for the purpose described in Section A above and according to the terms and conditions of this Agreement. These data shall not be used for personal gain or profit.

12. Stanford University shall use reasonable and appropriate efforts to keep all information furnished by SFUSD on encrypted machines and in a space otherwise physically and electronically secure from unauthorized access, including using such efforts to prevent unauthorized persons from retrieving, or altering the information by means of a computer, remote terminal, or other means. No data will be stored on unencrypted laptop or desktop computers or servers or other unencrypted portable computing devices or media, e.g., flash drives, etc.

13. Stanford University shall permit examination and on-site inspections by SFUSD upon reasonable advance notice for the purpose of ascertaining whether the terms of this Agreement are being met.

14. Stanford University agrees that the confidential data shall be securely destroyed when no longer needed for the purposes of this Agreement.

15. Stanford University agrees that SFUSD shall not be named or otherwise identified in the study, unless written permission to do so is granted by SFUSD for a specific project only.

H. LIABILITY

1. Stanford University agrees to be responsible for, and assumes all liability for, any claims, costs, damages or expenses (including reasonable attorneys' fees) that may arise from or relate to Stanford University's sole intentional or negligent release of personally identifiable student, parent or staff data (collectively "Claims"). Stanford University agrees to hold harmless SFUSD and pay any costs incurred by SFUSD in connection with any such Claim.

2. SFUSD agrees to be responsible for, and assumes all liability for, any claims, costs, damages or expenses (including reasonable attorneys' fees) that may arise from or relate to SFUSD's sole intentional or negligent release of personally identifiable student, parent or staff data (collectively "Claims"). SFUSD agrees to hold harmless Stanford University and pay any costs incurred by Stanford University in connection with any such Claim.

3. In the event of concurrent negligence of Stanford University, its Board, officers, employees and agents, and SFUSD, its Board, officers, employees and agents, the liability for any and all claims for injuries or damages to persons and/or property shall be apportioned under the California theory of comparative negligence as presently established or as may hereafter be modified. Nothing in this Agreement shall constitute a waiver or limitation of any rights that Stanford University may have under applicable law in the event of concurrent negligence of persons or entities other than District.

4. Stanford University and SFUSD agree to cooperate with each other in the investigation and disposition of third-party liability claims arising out of any services provided under this Agreement. It is the intention of Stanford University and District to fully cooperate in the disposition of all such claims. Such cooperation may include joint investigation, defense and disposition of claims of third parties arising from services performed under this Agreement. Stanford University and District agree to promptly inform one another whenever an incident report, claim or complaint is filed or when an investigation is initiated concerning any service performed under this Agreement.

5. The provisions of this Section shall survive the termination or expiration of this Agreement.

I. TERMINATION

1. This Agreement may be terminated as follows, after notification via the United States Postal Service (certified mail or registered mail) or recognized overnight delivery service (e.g., UPS, DHL or FedEx):
   a. By either party immediately in the event of a material breach of this Agreement by another party.
   b. By either party after 30 days advance written notice to the other party, for any reason or no reason.

2. The confidentiality provisions of this Agreement shall survive the expiration or termination of the Agreement. If this Agreement is terminated by either party for material breach or for any other reason with 30 days written notice, the confidential information shall be returned or securely destroyed within 7 days of the termination. If the Agreement terminates at the end of

the term described in Section D, Stanford University shall return or securely destroy all confidential information when it is no longer needed for the purposes of this Agreement. Such return or secure destruction shall occur within 7 days after the data are no longer needed for the purposes of this Agreement.

3. Secure destruction of the confidential information shall be accomplished by utilizing an approved method of confidential destruction, including shredding, burning or certified/witnessed destruction for physical materials and verified erasure of magnetic media using approved methods of secure electronic file destruction.

J. GENERAL UNDERSTANDING

1. This Agreement contains the entire understanding of the parties and may only be amended in writing signed by the parties.

2. This Agreement shall be governed by and construed under the laws of the State of California, without regard to its conflict of law rules.

3. Either party's failure at any time to enforce any default or right reserved to it, or to require performance of any of the Agreement's terms, covenants, or provisions by the other party at the time designated, shall not be a waiver of any such default or right to which the party is entitled, nor shall it in any way affect the right of the party to enforce such provisions thereafter.

4. If any term or provision of this Agreement shall be found illegal or unenforceable, this Agreement shall remain in full force and effect, and such term or provision shall be deemed stricken.

5. Original copies of this Agreement shall be executed by the respective party's authorized signatory(ies). This Agreement may be executed in one or more counterparts, each of which shall be deemed an original agreement, but all of which shall be considered one instrument and shall become a binding agreement when one or more counterparts have been signed by each of the parties and delivered to the other.

[Field for signatures]
[List of variables covered by the agreement are listed on a separate page called "Attachment A"]

Appendix B: Sample language in amendments to data agreements

SAN FRANCISCO UNIFIED SCHOOL DISTRICT

FIRST AMENDMENT TO AGREEMENT

THIS FIRST AMENDMENT TO AGREEMENT ("Amendment"), dated for convenience [date], is made in San Francisco, California, by and between the **San Francisco Unified School District ("District")** and **Stanford University Graduate School of Education ("Stanford Research Organization").**

WHEREAS, the District and Stanford Research Organization have entered into the Agreement (as defined below), whereby Stanford Research Organization would perform the agreed-upon data exchange between the District and Stanford Research Organization; and

WHEREAS, the District and Stanford Research Organization desire to modify the Agreement on the terms and conditions set forth herein to add additional data elements and additional years of the data to the data that Stanford Research Organization will receive from SFUSD pursuant to the terms of the Agreement,

NOW, THEREFORE, be it agreed between the parties as follows:

1. **Definitions**. The following definitions shall apply to this Amendment:
   a. **Agreement**. The term "Agreement" shall mean the "Agreement for Confidential Data Exchange Between San Francisco Unified School District and Stanford Research Organization" dated **[date]**.
   b. **Other Terms**. Terms used and not defined in this Amendment shall have the same meanings assigned to such terms as in the Agreement.
2. **Modifications to the Agreement.** The Agreement is hereby modified as follows:
   a. Attachment A of the Agreement is hereby deleted and replaced in its entirety to read as follows.

[List of variables included in the amendment]

Signatures of the Parties.

IN WITNESS WHEREOF the District and Stanford Research Organization have executed this Amendment to the Agreement.

[Field for signatures]