

Barriers to Effective Education Evaluation

An E³ Alliance White Paper Update

Prepared by E³ Alliance

SPRING 2008



For comments or questions, please contact:

Dr. Hannah Gourgey

Director of Analysis and Alignment

E³ Alliance

512/223-7243

hgourgey@e3alliance.org

About E³ Alliance

The E³ Alliance was formed in May of 2006 by founding partners Austin Area Research Organization, Austin Community College, District, and the University of Texas.

The mission of the E³ Alliance is to use an objective, data-based process to determine high-leverage alignment opportunities throughout the region's education system and to be the catalyst for systemic change. Much of the work of the E³ Alliance is based on compiling and analyzing data and information never before available for any region of the state (and perhaps the country) in order to ascertain key leverage points in driving systemic change in education. As a result, the organization relies heavily on the existing publicly available data provided by state agencies and on other key education researchers. The following whitepaper identifies and describes the data challenges that confound effective evaluation throughout this region and the state.

Introduction

Education pundit and activist Dr. Willard Daggett commented recently that the State of Texas “collects more data on its students than half the rest of the country combined.” In fact, Texas state agencies, education service centers, research organizations and local school districts collect more information about student demographic, performance, and outcomes than most states across the country. These data elements range from basic demographic information to highly-detailed performance and behavior records that include – for example -- refined definitions and criteria for different student “leaver codes” for students who do and do not graduate from Texas public secondary schools.

Given the breadth and – in many instances – depth of information concerning student characteristics, performance and behavior, Texas would seem to be the ideal environment in which to conduct comprehensive qualitative and quantitative longitudinal research. However, for the state agencies and entities responsible for data collection and warehousing, significant challenges to sustained data collection, comparable data sets, implementation of data standards combine to inhibit robust third-party research and evaluation. Furthermore, education-related policies at the state and local levels limit access to data and can also offer skewed outcomes without a qualitative overlay that drills down to nuanced variations in the interpretation of standards, trends and outcomes.

This paper addresses two kinds of critical barriers to effective evaluation:

1. Data-Centered Barriers that lead to inefficiency of effort, inconsistencies, and – in some cases – lack of availability of data, and
2. Policy-Centered Barriers that make effective data analysis difficult or impossible.

Good News about Data in Central Texas

Central Texas enjoys some strong advantages to data collection that have enabled the region to move forward with several ambitious research projects in recent years despite experiencing many of challenges that will be discussed in more detail in this paper.

Access to agency data is somewhat easier than in other regions in Texas because of proximity: researchers are able to work more directly with agency staff to address questions and provide strategies that allow for broader collection and extraction. For instance, in aggregated data requests to the Texas Higher Education Coordinating Board (THECB), the E³ Alliance has been able to work directly with THECB staff to come up with solutions to masking challenges that included combining cohorts, and creating grouped categories that allow for trend analysis with more detail than would be possible using standard data requests.

Also, Central Texas hosts The University of Texas at Austin and a number of affiliated education research institutions such as the Ray Marshal Center, the Charles A. Dana Center, then Annette Strauss Institute, and the National Center for Education Accountability. These groups are often able to leverage each others' work to deliver higher quality research. And as the state capital, the region has a much higher degree of scrutiny and research participation among its districts than many other areas of the state. The challenge is that these efforts are often uncoordinated and conducted without knowledge of existing or competing projects in the area.

Why What We Have Isn't Enough

Texas does not just have more data *collection* than most states, we also have more *dissemination*. The Texas Education Agency (TEA) collects PEIMS data from school districts, processes and audits it, then releases much of it in "raw" form. Data from PEIMS is combined with data on standardized testing (TAKS, ACT, SAT) and reported online through the AEIS system. The THECB collects and reports data on college matriculation, fields of study, graduation, and other statistics from public and (more recently) private colleges in online reports and queries. In addition, both agencies respond to specific data requests from the public, for free or relatively small fees. Why isn't this enough?

The bottom line rests with two endemic challenges that will be fleshed out in detail in the following pages:

1. *Data inconsistencies* that arise from challenges to the process of data collection, definitions and standards, incomplete data sets, and incomparable data.
2. *State and local policies* that limit access to data and prohibit developing effective analysis on the critical interplay between key factors and influences on student success. These policy challenges derive from FERPA regulations that will be

discussed in further detail – but can be summarized as preserving the privacy rights of students and their families.

This report will focus on data-centered and policy-centered challenges that fragment, disrupt, redefine and mask data sets and critical student records and impede the research and evaluation process. It is important to note that this whitepaper is *not* a lobbying effort or indeed an attempt to drive any specific outcome; rather it is an attempt to provide informed input from one key region of the state.

Data-Centered Barriers to Effective Evaluation

Central Texas public and private education institutions face similar challenges to the rest of the state with incomplete and inconsistent data because:

- 1. Data sets and data bases are disparate and disconnected.**
- 2. Many education institutions lack the capacity for data collection.**
- 3. Existing data are not necessarily available in a timely manner for public use or even designated for research.**
- 4. Data collection is not systemic.**
- 5. There are no permanent identifiers that enable longitudinal research and linkages across education agencies, employment, military or criminal justice records.**

1. Data sets and databases are disparate and disconnected.

In E³ Alliance’s Central Texas Education Snapshot, the scope of the research focused on 35 school districts that included the ACC service region and the Austin MSA, the region’s seven higher education institutions (both public and private), and 16 charter schools, totaling 58 education institutions. For some data such as the Texas Assessment of Knowledge and Skills (TAKS) required reporting to TEA’s Public Education Information Management System (PEIMS) database, the data were consistent because they derived from a single data source.

Other data, however, such as college matriculation rates by district, were far more difficult for E³ to analyze, because each district reported or captured these data differently. Furthermore, THECB tracks college matriculation but limits publicly available data to first-time college-going freshman matriculating the fall following high school graduation. Current research suggests that this type of “traditional” freshman is no longer the way the majority of high school graduates experience higher education. In addition, these rates only recently (in the last two years) have included private independent colleges in state and do not include any higher education institutions out of state.

Many school districts capable of tracking these data also include students who may have opted to begin college in the spring semester or who enrolled in an out of state university. Thus, the state matriculation rate may vary as much as 15% or more from the self-reporting of a given district. This discrepancy can be caught for those districts with the capacity for research and evaluation but many lack sufficient resources to conduct these efforts on their own.

In another example of data inconsistency, colleges may capture their enrolling freshmen's high school or district during the admission's process, but this "tracking" does not continue to graduation. Many of these institutions cannot track college graduation rates based on school districts. THECB has this capability, and can provide such data upon request, however, for many districts graduation rates dwindle and often cannot be subdivided across demographics or by higher education institutions without risking FERPA violations.

Inter-agency database and data collection inconsistencies pose another challenge to effective research and evaluation. Information at TEA, THECB, and the Texas Workforce Commission (TWC), is collected and warehoused differently making data linking difficult. Effective longitudinal research using these data sets can only occur by connecting student information to college and employment records. It is important to note that while some research institutions have authorization and access to a portion of these data (such as the Ray Marshall Center for the Study of Human Resources at the University of Texas at Austin, who has state authorization to access UI wage records and therefore can link to employment records) NO research entity in the state currently has authorization to access all these data sets.

In addition to inter-agency discrepancies in data, divisions in research, evaluation and accountability have also led to intra-agency discrepancies. In a given year, students are asked to provide demographic information both in the fall for PEIMS and then again in the spring for TAKS. These data are warehoused separately and depending on the student's responses may be completely different for the same individual. When this occurs PEIMS can make note of it. The general practice is to assume that the most recent data is the most accurate.

2. Many education institutions lack the capacity for data collection and evaluation.

Within the Central Texas region there is tremendous variation in district (and college) capacity to collect and analyze data. Austin ISD and the University of Texas at Austin by virtue of their size and resources are both standouts in this area and have conducted vital and transforming research to serve their students. Far more of our region's districts and colleges with the same aspirations simply cannot meet these standards due to insufficient financial and human resources.

The challenges arising from this disparity include both process and outcome variations. In Del Valle ISD, access to student records in compliance with FERPA's study exception may still be limited for any researcher because there is no automated process to retrieve those records. For Concordia University, some requests for data are limited by the restrictions of access and data retrieval required in the national database for all of the affiliated Concordia Colleges and Universities.

Other data may be considered important, but because it is not required, districts and colleges opt to direct resources elsewhere. One important higher education example is in defining "student success." Many smaller private, junior or community colleges consider a student successful if she or he graduates OR has transferred and graduated from another college or university. This second standard is not required by state or federal law to record and therefore those institutions with fewer resources do not track it. The result is that graduation rates as a measure for student success may be inaccurate and misleading.

In conference with many of Central Texas districts several factors are considered critical but not required to be tracked by any agency and therefore are not collected. Examples include:

- Parent's education level
- Parental involvement (admittedly very difficult to quantify and evaluate)
- Extra-curricular activities
- 1st Generation College Go-er
- Specific At-Risk Factors

We have found in our interviews that smaller districts are hungry for these data and analyses and simply have to make difficult resource decisions to allocate funds toward other programs and efforts. In some instances, even wealthier districts may opt to direct resources to other services in part because they feel student performance is either sufficient or more important than internal evaluation. The need for this research is compelling for both groups of districts, but these data cannot be examined consistently.

3. Existing data are not necessarily available in a timely manner for public use or even designated for research.

TEA's PEIMS database contains literally hundreds of data elements for individual students across the state of Texas. For several variables such as "at-risk" there are 14 criteria that a school district may use to identify a student as at risk. However, TEA requires only a "yes/no" response to this question, leaving the condition under which a student became at risk in question. Researchers are unable to obtain from the state agencies further details even though the defined criteria exist.

In addition to the kind of data available, the timing of availability also impinges on the research cycle. A district will not have to report any final numbers on the graduating class of May 2007 until December in order to capture any students who may need extra

courses in the summer to pass or who met all standards but failed TAKS and need to take it again. At that point, a number of processing and auditing steps are then undertaken by TEA. As a result – final numbers on graduation counts and rates are not computed until Spring of the year following graduation. As of April 2007, the last student retention rates available are for school year 2004-2005. Colleges and universities are not required to report fall matriculation rates until well into the year following matriculation, again slowing down the process of obtaining critical data on any given student’s progress and choices.

It is not realistic to expect all types of data to be available “near real time” after institutions have submitted them. However, for key decision-making drivers listed below, a timeline of no more than one-quarter post submission should be implemented to help school districts, college and other institutions better evaluate critical program interventions specific to sub-populations. These data include:

- Results from Standardized Tests
- College Entrance Exam Results
- Enrollment Rates and Counts
- Graduation Rates (secondary & post secondary)
- Transfer Rates (post secondary)

Furthermore, ad hoc requests should be given timelines for completion that consider agency capacity and priorities. Such information would assist third-party evaluators in determining what kind of support may help the respective agencies in fulfilling the request. If agency staff are unavailable to fulfill requests, a mechanism to “buy” staff time should be available to support better and in faster information usage.

4. Data collection is not systemic.

In general and by comparison, Texas state education agencies perform well in their ability to collect data from local school districts – especially given the wide range of district ability to capture these data. However, this range in ability also allows for data to be collected in a variety of ways. Districts are required to enter data on their students into the PEIMS database, but need not have electronic transcripts from which these data are downloaded or recorded. Human error in data entry becomes a factor that exacerbates these challenges.

In PEIMS, certain errors may result in a student being removed from computations of critical school performance measures such as TAKS passing rates or graduation rates – leading to higher margins of error in these rates. State agencies attempt to correct obvious errors, but it is very challenging to determine student ethnicity if he or she decides to report one heritage in the fall and another in the spring.

Students in Central Texas and around the state are also able to access services from community-based organizations that work directly with school districts and campuses on projects that include: tutoring, civic education, drop out prevention, and career-based experiences. Many of these organizations collect information about their students’ performance and behavior but often it is only to meet the requirements of funding

organizations or as staffing and time allow. Thus data coming from community-based organizations derive from different standards and levels of consistency and are therefore difficult or impossible to integrate into the formal student records for TEA.

Recently, funding from the 21st Century grant has required participating CBO's to provide data to the Texas Education Agency for evaluation and accountability, but access to these data is not currently given to outside groups. In any event, this requirement only applies to participants in 21st Century programs and many more students receive services through other funding streams. However, there are coalitions of community-based organizations in Central Texas such as Ready-by-21 that have constructed a data collection tool that can be used for internal use as well as by outside organizations to determine which interventions are in which schools. To date, its service area includes only Travis County.

5. There are no permanent identifiers that enable longitudinal research and linkages across education agencies, employment, military or criminal justice records.

In Colorado, the state education agency (Colorado Department of Education) assigns a ten-digit STATE student identification number (SASID) to every child upon entry into the public education system as a way to track from P-12 across any district, charter, or private school within the state. School districts may also assign their own individual ID numbers (LASID), but are required to use the state ID if a student receives any services through the district and for all state reporting.

This method allows a student to be tracked across a range of experiences and to ensure that information about that student can be transferred to new schools and districts as needed. It has the added advantage of offering researchers to study the effects of certain critical demographics or performance issues of a period of time. These ID's are used internally and in the process of research removed from any student identifying information thereby remaining in compliance with FERPA.

This year, the Colorado legislature passed a law that allows for the P-12 SASID to track beyond high school for a selected number of public higher education colleges and universities. As this pilot effort expands, public and private institutions will be included. For researchers, the Colorado Department of Education provides an "ESID" an artificial individual student ID that only the CDE can connect back to a specific student. The use of the ESID allows for the release of individual student records while protecting the identity of students.

Colorado is also launching a teacher identification number system in order to link student successes back to individual teachers and classrooms and to track teacher mobility. The effort is in part fueled by data that show concentrations of teachers teaching out of field in high-needs schools. Central Texas experiences this same problem.

Texas has student identification numbers that are used both at the district and state levels. However, these numbers are changed from district to district as a student moves making tracking of mobile students via ID impossible. These ID's also do not transfer from secondary to post secondary education – making the ability to track a student's progress from high school through higher education and the workforce far more difficult. Social security numbers are not an effective substitute: in some districts, social security numbers are either missing or incorrect for 33% or more of students.

In Summary

Overall, the data-centered challenges that Central Texas experiences extend to the rest of the state—systemic challenges to data collection and processing at the district-level, discrepancies and redundancies in student information across and within state education agencies, data dissemination timelines, variations in definitions and standards limit access and depth. These factors are compounded by state and local policies that further inhibit robust data evaluation.

Policy-Centered Barriers to Effective Evaluation

In recent years, the state of Texas has increased privacy protection surrounding individual student records based on state interpretation of restrictions due to the Family Education Rights Privacy Act (FERPA). In addition to these state restrictions to data access, local policies enacted by school districts vary and often make comparisons on certain performance indicators challenging or irrelevant.

This section will address state and local policies or policy guidelines that affect data collection and evaluation both in Central Texas and around the state. The specific topics we will cover include:

- 1. Individualized data are highly protected in compliance with FERPA.**
- 2. Several new policies on state standardized assessments implemented over the last four years have made data set comparisons difficult.**
- 3. Variations in district versus state policies can effect student behavior and therefore make district comparisons challenging.**
- 4. Ideological bias limits usefulness of some data sets.**

1. Individualized data are highly protected in compliance with FERPA.

The Family Education Rights Privacy Act or FERPA was established in 1974 for two purposes:

- A) to provide parents the right to and process for inspecting and reviewing their children's education records and
- B) to protect these sensitive individual records against public release for non-academic and non-research purposes.

Application, interpretation and enforcement of FERPA has varied widely over the last thirty years both at federal and state levels. Recent actions by the state of Texas have tightened restrictions surrounding the release of individual student data as a result of direct federal input in 2004.

The consequence of this high-level of protection is a limited ability to access individual student records with enough "unmasked" data to allow for more complex statistical analysis. "Masking" is the term used by state agencies to refer to deleting student records from open data requests that may have enough specificity as to render the student in question identifiable by the data requester. There are provisions in FERPA that allow for research to be performed by external groups in compliance with the privacy restrictions. Texas has chosen not to grant access using these provisions in most cases. (See Appendix A for more detail on FERPA) A study performed by the Ray Marshall Center in 2005 of 34 participating states determined that there is no single standard for interpreting FERPA, and that Texas remains one of the most restrictive in its interpretation of FERPA regulations across the country.

In addition, agency-to-agency data sharing is inconsistent and often restricted. Texas Education Agency will mask records sent to the Texas Higher Education Coordinating Board if TEA believes student privacy is in jeopardy. Similarly, the Texas Higher Education Coordinating Board has no authorization to release information to the Texas Workforce Commission or even back to school districts unless the students (18 or older) have authorized such release.

Some local higher education applications (such as ACC) in Central Texas include an option for applicants to authorize release of their data. Students are under no obligation to sign the release. Utilizing this method may offer college and universities greater latitude for research, however, this strategy can also further restrict access if interpretations of non-signature assume greater privacy protection.

2. Several new policies on state standardized assessments implemented over the last four years have made data set comparisons difficult.

Several new state assessments implemented to meet the standards required by the No Child Left Behind Act have made comparison of student performance over time difficult.

In 2002, the Texas Assessment of Knowledge and Skills (TAKS) replaced the Texas Assessment of Academic Skills (TAAS). TAKS was designed to be a far more rigorous assessment of progress in critical subject areas. However, because of the change in the

form as well as rigor of the exam, comparing student performance pre-2003 to post-2003 using these standardized assessments is difficult or ineffective.

Similarly, in 2003, the Texas Success Initiative (TSI) replaced the Texas Academic Skills Program (TASP) as a measure for student preparedness to enter and succeed in higher education. Under TSI there are several more options and tests by which a student may meet the standard. In addition, during the year in which TSI was first implemented, access to the new assessments was unavailable so the state opted to use a lower exit TAKS score to demonstrate suitable preparation for college. Any data for TSI's initial year therefore is already skewed. Finally, the law enacting TSI also allowed for a student to opt not to enroll in developmental courses, thereby deferring remediation. Whenever the standards have changed, interpretation of trend data becomes difficult or impossible.

This year the State Legislature is considering replacing high school TAKS tests with "end-of-course" exams. These tests will not allow for comparable measurements to current high school TAKS assessments, as well as across or within districts where the rigor of the courses may vary widely.

This last discrepancy leads to the next major policy-centered data challenge that pits local district policy against state standards.

3. Variations in district versus state policies can effect student behavior and therefore make district comparisons challenging.

In general, the policies set by the state are viewed as the minimum standard for all districts. However in some areas, the state allows latitude in interpretation in order to provide flexibility for the district to meet local population needs. For example, elementary school students who do not meet basic standards of English proficiency may be enrolled in bilingual programs and be tracked by the state throughout their primary school experience. However, enrollment eligibility versus requirements vary widely among school districts and often depend on availability, parental approval and teacher or administrator input.

A second example challenging district comparisons occurs at the high school level as students enroll in the Recommended High School Plan (RHP) or Distinguished Academic Plan (DAP). These two plans vary only by the addition of a foreign language credit in the DAP, however districts may add additional courses to their local definition of the DAP. This means that comparing the percent of students taking a DAP curriculum across districts may or may not be comparable.

4. Ideological bias limits usefulness of some data sets.

It is important to note at the beginning of this section that any process to establish data criteria and definitions will inherently carry with it some bias or presumptions. It is NOT the focus or interest of the E³ Alliance to align with any particular ideology but to provide a description of how these biases have affected data criteria and outcomes for Central Texas (and as applicable for the state).

For instance, the question concerning drop out rates may best exemplify such bias (at both ends of the spectrum). In the late 1990's the Governor's Business Council commissioned the Higher Education Coordinating Board to conduct a study of how many 1992 7th graders continued on in the Texas public education system to complete their high school diploma and obtain a higher education degree. The methodology used is an attrition model – those students who leave for any reason are taken out of the cohort, but new students coming in to the state are not added. This report indicated a high school “graduation rate” of 58%. This rate may be low in comparison to other methods because of inevitable attrition of students who moved out of state, or left the public education system but opted for private education or home schooling and did eventually complete a high school course of study.

Conversely, TEA's graduation rate was 82%. The TEA method uses a different starting point (9th grade cohort) and often removes students from the numerator or denominator who are considered “errors,” who left to get a “job,” who were listed in GED programs but not indicated as completing those programs, etc..

Other methods such as the Cumulative Promotion Index have been designed to account for students leaving or coming in, and will yield yet a different graduation rate (often in between the GBC and TEA reported outcomes). In addition, the discrepancy between graduation rates and drop out rates is tremendously high. For instance, Round Rock ISD claims a 2005 graduation rate of 90.4% and a drop out rate of .4%. Many students who leave the system and may be considered “drop outs” by the general public are assigned leaver codes that do not count in the drop out number.

While the drop out rate provides the best example of ideological bias in data definitions and criteria, there are other instances in which this occurs: disciplinary action, Bilingual and LEP status, student retention.

Opportunities Pending

This year represents a seminal moment in Central Texas data challenges AND opportunities.

Texas Education Research Centers

The state of Texas has recently authorized the creation of Education Research Centers designed to provide an appropriate amount of protection in compliance with FERPA but to allow for greater access for research and evaluation. Central Texas researchers and programs will likely have access to at least one of these ERC's as they come on line. E³ Alliance is incorporated in one of the proposals we anticipate will be an award recipient. The value of these ERC's will be their collaborative model involving multiple agencies. Access to individual student records within FERPA guidelines is likely to be greater, however the timeline for launch of these projects suggest that no real movement can occur before year's end. Ideally, the model will also provide for state identifiers (as used

in Colorado) that can be connected across P-16 and into the workforce. (*Please See Appendix C for updates to ERC Strategies and Recommendations*)

Non-Agency Student Performance Factors and Indicators

Throughout the spring of 2007, E³ Alliance has been completing an inventory of non-agency factors or indicators that districts believe matter greatly in student performance and success both in K-12 and in the higher education experience. The inventory is based on conversations with seven regional school districts: Austin, Bastrop, Del Valle, Eanes, Manor, Round Rock and San Marcos.

The plan for this inventory is to work with these districts to prioritize these factors and then begin collection by the summer of 2007.

Central Texas Models for Data Access, Sharing and Analysis

E³ Alliance has already developed models of data sharing that other regions can adapt rather than having to re-invent the wheel. The best example is the *Central Texas Regional Snapshot* that derives from replicable aggregate data requests to Texas Education Agency and the Texas Higher Education Coordinating Board. We are currently working on models of individual student record requests that we hope to serve as models for other regions as well.

Coordination, Collaboration and Convergence of Regional Data Evaluation Projects

Finally, and most importantly, the momentum that has been building for more detailed and rigorous data, evaluation and research has resulted in several regional projects that are coordinating their efforts:

1. ***E³ Alliance Regional Blueprint for Change***: a collaborative designed to be a catalyst for system education change through research, evaluation, and alignment. Currently in the process of developing a blueprint for change that includes:
 - An objective data map that details the portfolio of critical factors affecting Central Texas Student performance and achievement.
 - A series of performance rubrics that match to specific kinds of interventions and alignment opportunities for program evaluation to help channel private and public support into programs and systemic alignments with a record of success.
 - An alignment matrix that details opportunities for regional alignment within and across Pre-K-16, community based organizations, and industry
2. ***Student Futures Project (formerly the Central Texas High School Graduate Data Center)***: a research collaboration between Skillpoint Alliance and the Ray Marshall Center that surveys regional high school graduates and tracks their progress from high school through post secondary and the work place to determine what choices they are making and

why. This longitudinal study marks the first of its kind in Texas and serves at the vanguard of this type of research across the nation. The ultimate goal for the HSGDC is to provide regional school districts with a look at what their graduates are doing beyond high school and what interventions are most likely to lead to student success in college and careers.

- 3. *Success by 6 and Ready by 21 Coalitions*:** two consortia focused on early childhood education and preparation for life beyond high school respectively. Both coalitions are developing (or have developed) standards and process for more effective data collection.

Conclusion

Issues with inconsistent data sets, inability to access critical, decision-driving data, and the development of longitudinal data sets for robust research all severely limit the ability of Central Texas to conduct effective evaluations that can drive decisions in resource allocation, collaborative opportunities and institutional alignment. The result is that few effective interventions are given the opportunity to be taken to scale. Most solutions do not require policy changes but rather are focused on operational processes that may be streamlined, simplified or updated to ease the burden of data collection, exchange and access. E³ Alliance welcomes the opportunity to work with interested parties in helping to implement appropriate and effective change.

Appendix A

FERPA Research Provisions

“Masking” is the term used by state agencies to refer to deleting student records from open data requests that may have enough specificity as to render the student in question identifiable by data requester . Once a record is “masked” or deleted it will never return within the scope of that original request. As a result, information requests with multiple variables or data elements are returned to the requester with greatly reduces data sets and hinders analysis. (See Appendix B.) Without the ability to conduct studies that test the interdependency of key variables it is challenging for researchers to obtain a true picture of the key influencers on student performance.

The law provides two allowances for individual data release. The first allowance occurs if the data requestor serves as an “authorized agent” of the state or local education agency from which it is soliciting the request. In this instance, the requestor would be under the control of said agency and subject to its rules and guidelines regarding publishing and warehousing the data.

The second provision is called a “study exception” and allows third party research that serves the ultimate purpose of improving instruction. This allowance is most often implemented when the state or local education agency is unable to undertake the research or evaluation on its own due to a lack of human or financial resources. The latitude under this allowance is somewhat broader and does not hold the third party to the state or local education agencies’ rules pertaining to publishing and warehousing data. It does require that the third party demonstrate that it can properly protect individual records and that these records be destroyed upon completion of the specific study for which they were requested.

State agencies have made the decision not to grant “authorized agent” or “study exceptions” for data requests. As a result, access to individual student data is severely limited and subject to guidelines used to ensure that the identity of the student is protected while process data requests. Guidelines protecting student identity limit multivariate analysis by assigning “equal weight” to data elements that are easily discernable, school attending, race/ethnicity and not easily discernable such as performance on TAKS exams.

Appendix B
Texas Education Agency
FERPA Compliance Rules

PEIMS AD HOC REPORTING MASKING RULES

MASKING STUDENT AGGREGATE DATA (COUNTS)

- Any student count less than 5 but not 0 is masked at the district level and lower for any combination of variables (campus, grade). This applies to any program or demographic category that applies to a student. We do not mask percentages.
- We will not supply counts and percentages on the same report/file due to being able to apply the percentage of a total to deduce the masked value.
- We will not supply a file/report with masked counts by campus and the same counts by district due to being able to join the files to fill in blank cells by using district totals to deduce masked cells at the campus level.

MASKING STUDENT LEVEL DATA (REMOVING RECORDS)

- A student-level file will not contain any identifiable id's such as names, SSN's or PEIMS generated student id's. A student-level file will contain a scrambled unique student number and any other demographic or program category data that is requested after records are removed at the level of data that is requested. The scrambling method will be unique to a particular request and will not be used on any other request.
- There will be no way for the data to be aggregated at any level to deduce counts less than 5 but not 0. For example: if the universe for a request is 4.5 million records and 15 PEIMS variables (ethnicity, gender, campus, LEP at risk...) are requested, records will be removed at the level (15 variables) of data that is requested. This is done by creating a record count data set of the original student level data set by all variables keeping only counts greater than 4 and merging the count data set back into the student level data set if the counts are greater than 4. It has been our experience that requesting more than 7 or 8 fields removes a large amount of data.
- If a customer requests more data but different variables (10 variables) and wants to use the same scrambling method from a previous data request, the removal of records will be based on the previous request (15 variables) and the new request (10 variables) for a total of 25 variables. This is done so the requestor is not be able to fill in blank cells in the first request and the removal process is applied over all data that is requested using a specific scrambling method.
- When building a cohort that follows a group of students across multiple years, if a student record is removed from one year, it is removed from all years.

MASKING STAFF DATA

- There is no removal of records process with staff data. Names are available and SSN's will be scrambled using a scrambling method that is unique to a particular request. A customer may use the same scrambling method on another request as long as the method is owned by the customer. There is no sharing of scrambling methods from one request from an entity to another entity.

Appendix C Education Research Center Data Access and Effective Uses Strategies

Challenge

The State of Texas has authorized and awarded the establishment of three Education Research Centers (ERC) whose purposes are to serve as research coordination, collaboration and accessing nexus for studies on local and statewide levels in P-16 education. The basic design is that the TEA – and eventually other agencies (TWC, TDJS, etc) – would send multiple years of individual records to the THECB, where individual records from multiple agencies could be matched based on social security numbers and other data fields, then de-identified, encrypted and held in a common data warehouse for access by the ERCs. Conceptually, these ERCs will help to streamline research and data access processes to support improved education performance while diminishing complications due to FERPA restrictions and data accessibility issues.

Certain challenges to the successful implementation of these ERC's exist.

1. **Timeline:** Neither the RFP nor the Award process designated a timeline for when the ERCs would be operational. To date, essential features of the ERC process such as developing a state Commission to approve additional research projects, have not be developed or created. Much of the timing is dependent upon finalization of agency-to-agency MOUs that describe and delimit the data access, sharing and warehousing process and responsibilities, and such MOUs have been modified multiple times causing long delays in data sharing.
2. **MOU Criteria:** The structure of the MOU between TEA and THECB has been under negotiation for over a year with key areas of conflict around who has access the “identified” student level data and for what purposes. For example, there is concern over whether data sets will be retained for use by projects conducted beyond duration and scope of the original, designated research effort. As a result, a complete data exchange from TEA is still pending close to a year after ERC grants were awarded.
3. **Database Integration:** The MOUs require that all databases between agency and ERC be integrated to optimize research. Currently, no organization appears to have the capacity for this process without additional resources allocated for that purpose. In particular different protocols for identifying or tracking students without social security numbers often result a significant percentage (between 5-10%) of students in higher education not being tracked. Other resources such as the National Student Clearinghouse (NSC) have different protocols as well as out of state student that can help complete the picture. Until recently, plans were not in place to integrate any data from private institutions of higher education. In general, plans to provide integrated, wholistic data sets appear to be only now emerging.

Concerns

- Research already backlogged or underway requires urgent implementation of the ERC's
- Infrastructure for THECB to physically host the data warehouse with secure connections to all ERC's will take too long to implement
- Offsite researchers will have insufficient access to data warehouse (possible requirement that researchers are on site at ERC)
- Substitution strategies for de-identified data may not work to have data available from multiple agencies and other data sources in a usable, cross-linked data warehouse

Strategies

One possible strategy to comprehensively address these challenges is for the ERCs to request a defined set of needed outcomes (rather than means to achieve those). For example:

Requirement: *All data from TEA and THECB data sets aligned to the individual student level for XX years shall be in a data warehouse securely accessible in real time to all ERC-designated researchers within Y months. Data from primary state and federal agencies (TWC, Military, HHS, TCJS) is available in the data warehouse within Z months.*

If TEA and THECB are concerned about being able to meet the timeline, then the ERC's could collaborate to bring together other state education agencies from across the nation to discuss best practices and to address the process other states used to provide access to data within compliance of FERPA regulations.

Below is an inventory of essential features to any data system developed that offers effective use of statewide student information for the purposes of research, program evaluation, and performance management.

Essential Features for Effective Data Usage

1. ERC's have access to confidential student-level data as allowed by FERPA under the research exception.
 - a. (For examples of FERPA guidelines from Florida see: <http://www.fl DOE.org/ese/pdf/ferpa.pdf>) Note: This language is no different, the difference is in the **POLICY** that Florida's Department of Education has for granting the research exception more readily than Texas. In addition, Florida has incorporated K-20 under its perview eliminating the integration challenges of data from high school to college. For more information:

Jay Pfeiffer
Deputy Commissioner

Division of Accountability, Research and Measurement
325 West Gaines Street, Suite 814
Tallahassee, Florida 32399-0400
Phone: (850) 245-0437 Email: Jay.Pfeiffer@fldoe.org

2. Student-level data must have the ability to be integrated across several critical agencies and institutions including: TEA, THECB, NSC, ACT, SAT, , TWC, Armed Forces TDCJ, THHS, CPS. (Different states have employed different methods including permanent student identifiers, or social security numbers).
3. Individualized-de-identified data must be available longitudinally to track student performance over time.
4. The data system has the ability to integrate data and performance variables from community-based interventions and services.
5. Researchers need to have secured virtual access to student level data in order to conduct research and evaluation off-site.
6. Data entry and auditing must occur as close to “real time” as possible to expedite access to most recent records that help to determine new program effectiveness.
7. State exams results (by learning objective NOT simply by score) should be maintained in the SAME database as student enrollment records in order to make student tracking and progress monitoring easier.