K-12 Educational Data System Background

Prepared by Julie O'Brian, October 2007 for the Data and Accountability Subcommittee of the Governor's P-20 Council.

Context

History

Before 2000, the Colorado K-12 educational data "system" was a number of disconnected databases within the Colorado Department of Education (CDE), most created in response to state or federal data collection requirements. In 2000, SB00-186 created the School Accountability Reports (SARs) and appropriated significant funds to CDE to create a data-warehouse to allow CDE to publicly report all of the data elements include in the SARs. Then in 2001, when NCLB required states to develop data systems that would allow them to meet significantly more federal data reporting requirements, the CDE data-warehouse was adapted to meet this new use.

Neither the NCLB requirements nor the SAR requirements included all of the data that CDE had been collecting and storing in separate data files. Thus, not all the data currently collected by CDE from districts is stored in the data warehouse. Some data is still stored in separate files within the department.

In the last few years, CDE has added functionality to the data warehouse as funds have become available, including bringing more data into the warehouse and adding a reporting tool, CEDAR (Colorado Education Data Analysis Reporting). CEDAR, is a Cognos-based reporting tool that sits on top of the CDE data warehouse and allows users to access data from the warehouse in a variety of formats.

In the summer of 2007 CDE was awarded a $4.2 million grant from the US Department of Education to "support the development of a longitudinal state data system." With grant funding, CDE proposed to build on the current K-12 data warehouse and CEDAR (CDE, 2007).

How Colorado Compares to the Data Quality Campaign Criteria
(Data Quality Campaign, 2006)

Data Quality Component	Implications for data use	Progress to date
Unique statewide student identifier.	Necessary to link different types of student data over time, and to link student data to teacher/classroom, school and district data.	Colorado has a state student identifier. CDE implemented the SASID system in 2002.
Student-level enrollment, demographic and program participation information.	When used with growth data, allows questions like: Which groups of students perform well? Poorly? Are there systematic differences in performance based on individual student characteristics? Are there differences in student performance based on programs in which they participate?	Included in the CDE data warehouse. However, "program participation" is limited to federal program classifications.

Data Quality Component	Implications for data use	Progress to date
The ability to match individual students' test records from year to year to measure academic growth.	Supports answers to questions like: Which students are improving performance over time? How much? What is the "value added" for individual student participating in the learning experiences at a particular school? District?	The CDE warehouse has this feature. HB07-1048 required the development of a consistent longitudinal growth model (to use for accountability reporting, defining "safe harbor" in AYP, and for the Governor's Distinguished Improvement Awards program).
Information on untested students.	Makes it possible to answer questions like: Are there systematic patterns in the students that are not being tested?	Colorado collects this data.
A teacher identifier system with the ability to match teachers to students.	Because student's experiences in our educational systems are primarily at the classroom level, this is a critical element to determine what works for what students under what conditions. This is also a key element in determining the effectiveness of teacher preparation.	Colorado does not have a teacher identifier at the state level. Some districts have teacher identifiers. SB07-140 Created a teacher quality commission to explore the creation of a teacher identifier.
Student level transcript information, including information on courses completed and grades earned.	Questions that you can answer with this element include: Is there a relationship between grades and student performance on outcome assessments? How does student course completion relate to performance on outcome assessments? Is there a relationship between student completion of certain high school courses and their readiness for post-secondary coursework?	Not collected at the state level. Individual districts have this information. To collect this data at the state level would require either common definitions of courses or a schema for mapping courses to one another across districts.
Student-level college readiness test scores.	This may provide information that would help answer the question of which students are leaving K-12 "ready" for college?	All Colorado students take the ACT and their scores on this assessment are included in the CDE data warehouse.
Student-level graduation and dropout data	This helps us to answer which students are graduating or dropping out, a precursor to figuring out why.	CDE collects this data but definitions have been inconsistent across districts.
The ability to match student records	This helps with a number of questions including: Which students go on to start and then complete post-secondary degrees? Are	Not at this time. SB06-24 requires public post-secondary institutions to adopt

Data Quality Component	Implications for data use	Progress to date
between the P-12 and postsecondary system	there systematic patterns in which students graduating from Colorado high schools need remediation at the postsecondary level?	the student identifier assigned to students while they are in the K-12 system.
A state data audit system assessing data quality, validity and reliability	This is a critical component for the big question of whether or not the data clean enough to use for anything.	CDE has implemented a system that flags some errors during the automatic data transfer (from districts to state) process. CDE is reportedly implementing other validation rules and statistical checks. (CDE, 2007) HB07-1270 requires an external audit of K-12 state educational data systems. HB 07-1320 expanded the role of EDAC in reviewing CDE data collection.

Note: The above table reflects information from the Data Quality Campaign, Colorado Summary of the ten elements, (2006), the CDE Statewide Longitudinal Data Systems Grant Summary and the following bills from the Colorado General Assembly: HB07-1048, HB07-1270, HB07-1320, SB06-24, and SB07-140.

Issues

Below critical issues related to the K-12 data system are grouped into categories related to the different functions of the data system, including: Data Collection, Data Storage, and Data Access and Reporting. A final category relates to the capacity (in general) of the current system to meet these functions.

Data Collection (or the transfer of data from districts to the state)
1. The current system for district submission of data to the state is burdensome on districts. "Translating the data elements stored in district data systems into the format required by the state requires reprogramming every time the *definition of a data element* or the *file format* changes" (Mass, Felker & O'Brian, 2006). This is a significant expense for districts
   - Districts are required to submit data files with *different file formats* through the state "automatic data transfer system" more than 10 times each year.
   - Data submissions have inconsistent requirements for how the data must be formatted. Even though the same data elements are collected multiple times, data element definitions change across submissions. For example, "for the October Count submission, the *ethnicity* data field has 5 categories and corresponding codes, For the CELA submission, *ethnicity* has 8 categories. A *3* in October Count represents *Asian*, while a *3* in the CELA submission represents *Black*," (Mass, Felker & O'Brian, 2006).
   - Data element definitions are not consistent with external standards for data element definitions. "National resources are available to help with the standardization in how data

elements are defined including a data dictionary from the National Center for Educational Statistics (NCES) and element definitions from the Schools Interoperability Framework (SIF)," (Mass, Felker & O'Brian, 2006).

2. CDE is making progress towards consistent data and file definitions, but it is unclear whether what they are proposing to do will ultimately solve the problems.
   - CDE is creating a standard "Educational Data Dictionary" which will define elements commonly across all submissions. However, these definitions are not currently aligned with external standards for data element definitions. (CDE, 2007)
   - CDE has not indicated plans to consolidate the number of times districts must submit data, nor have they committed to standardized file formats. (CDE, 2007)

3. The current plan for how data is transferred from districts to the state doesn't reflect current technologies – it is not a 21st Century Data System.
   - Currently the transfer of educational data in Colorado is static in that data is transferred through single point in time submissions that occur repeatedly throughout the year. "All of the changes to district data that occur over a defined period of time are pooled and transferred to the state in one big [batch] submission. This means that there is a period of time when the data has been changed at the district level but has not yet been updated at the state level. Some of the state level data is out-of-date almost as soon as it is received," (Mass, Felker & O'Brian, 2067).
   - Moving to standard data element definitions and standardized file formats is the first step towards moving to a *transactional* data exchange system. In a transactional system, data is updated almost continuously. Data changes at the district level result in immediate, real time, updates to the data held at the state level. This is they type of data transfer that happens when you withdraw funds from an ATM. The data that the bank has about your account is updated immediately. The transfer of data from an ATM to the bank is transactional. (Mass, Felker & O'Brian, 2006) Colorado is not currently moving towards a transactional system. (CDE, 2007)

Data storage, organization and quality assurance
4. It is unclear whether the data currently in the CDE data warehouse is of sufficient quality to support the decisions that are made based on the data.

Data access and external data reporting
5. Districts don't use the data stored in the CDE data warehouse.
   - For the most part the CDE data infrastructure is devoted to external reporting for accountability purposes and some limited internal uses of the data among CDE staff to support program improvement efforts.
   - The data services currently available through CEDAR are limited to a series of pre-defined reports and some ability to customize data-representations within subsets of the data in the warehouse (cubes) that have been defined by CDE staff (Mass, Felker & O'Brian, 2006). Only four districts provided input into the pre-defined reports (CDE, 2007).
   - The state only has enough CEDAR software licenses to allow two login accounts per district. This means that most principals, for example, cannot use CEDAR.

- CDE has limited staff time available to create the "cubes" or to build links to other state-held data sources that might make the data available through CEDAR more useful (Mass, Felker & O'Brian, 2006).
- The only assessment data available in CEDAR is CSAP. It is not clear whether or not districts would be willing to provide additional district assessment data to the state which would make CEDAR more useful.
- "A much wider variety of data uses by many more educational stakeholders is possible." If data collection were streamlined and "transactional" then additional data services could be made available to districts, BOCES and other educational stakeholders. These data services could better support district use of data in decision-making and expand research on effective educational programs/practices within Colorado, (Mass, Felker & O'Brian, 2007).

6. There is a growing difference between the "Have" and "Have Not's" among districts with regard to data use.
- Many districts, especially small ones, have limited access to data services. Some districts don't even match student performance data to the teachers/classrooms that those students participated in during that year.
- Even in districts where there is access to data and data tools, use of data to drive decisions at all levels is limited by staff capacity.
- While many districts who participated in the C2D3 project have learned processes for collaboratively analyzing and interpreting data for decision-making at all levels (including the classroom). Participation has been limited by funding.

7. Transfer of data between districts when students move is inefficient and does not go through or make use of the state data system even though a state student identifier makes it technically possible to do so.
- When students move from one district to another within the state, districts are required to send the student's CSAP data to the new district. However, districts have difficulties sharing this data electronically because of different system requirements.
- Districts meet this legislative requirement by sending print copies of the data to be included in the transcript files at the records department in the new district.
- Although the state data warehouse includes this information, CDE has been unwilling to facilitate electronic transfer of this data (D. Maas, personal communication, October 9th, 2007).

8. Folks outside of districts don't have access to K-12 state educational data.
- Educational stakeholders such as higher education institutions, educational support organizations, community foundations and researchers have no way of accessing the state education data through CEDAR because licenses have only been made available to districts.
- This limits the research that could be done to improve the educational system.

9. There is an absence of policies/procedures related to how the state education data can be accessed externally (including by districts who want to use data from other districts).

Capacity
10. There are limited state resources available for the upfront investment necessary to develop a 21st Century Educational Data System for Colorado.
11. It is not clear whether or not the uses CDE has proposed for the new $4.2 million grant will solve the problems identified here. The current plan does not include newer technologies (CDE, 2007).
12. It is not clear what the "vision" is for K-12 educational data management in Colorado.
13. It is not clear whether or not CDE staffed appropriately to develop a 21st century data system.

Principles/values/beliefs

Short-term recommendations

Long-term recommendations

References

Colorado Department of Education (2007, September) *Statewide Longitudinal Data System (SLDS) Grant.* Denver, CO: Author.

Data Quality Campaign. (2006) *2006 NCEA State P-12 Data Collection Survey Results.* Retrieved August 7, 2007, from http://www.dataqualitycampaign.org/survey_results.

ESP Solutions Group (2005). *Optimal Reference Guide: Management of an Education Information System.* Retrieved August 7, 2002, from http://www.espsg.com/resources.php.

House Bill 07-1048, Colorado Legislature (2007).

House Bill 07-1270, Colorado Legislature (2007).

House Bill 07-1320, Colorado Legislature (2007).

Mass, D., Felker, J. & O'Brian, J. (2006). *Redesigning Educational Data Management in Colorado.* Paper presented at the Colorado Association of School Executives (CASE) data systems meeting, Denver, CO.

Reichardt, R., Paone, J. and Badolato, V. (2006) *Shining the Light: The State of Teaching in Colorado.* Denver, CO: Alliance for Quality Teaching.

Senate Bill 00-186, Colorado Legislature (2000)

Senate Bill 06-24, Colorado Legislature (2006)

Senate Bill 07-140, Colorado Legislature (2007).