2018 TECHNICAL BRIEF

A Data Quality Evaluation of Administrative Data

Using CEDARS Student Grade History Data as a case study

ERDC
Education Research & Data Center

AUTHORS

Vivien Chen

Education Research and Data Center

Karen Pyle

Education Research and Data Center

Andrew Weller

Education Research and Data Center

ABOUT THE ERDC

The research presented here uses data from the Education Research and Data Center, located in the Washington Office of Financial Management. ERDC works with partner agencies to conduct powerful analyses of learning that can help inform the decisionmaking of Washington legislators, parents, and education providers. ERDC's data system is a statewide longitudinal data system that includes de-identified data about people's preschool, educational and workforce experiences.

ADDRESS

Education Research and Data Center
106 11th Ave SW, Suite 2200
PO Box 43124
Olympia, WA 98504-3113

PHONE

360-902-0599

FAX

360-725-5174

EMAIL

erdc@ofm.wa.gov

Study Background

One purpose of administrative data is to support evidence-based policymaking. With good data, policymakers can make decisions that improve the viability and effectiveness of public programs and policies. However, administrative data is sometimes insufficient when addressing key policy questions, because of varying data quality. This is in part because most administrative data is collected mainly for recording and reporting the use of public services, or for providing basic information about service users, rather than research.

> "Data are the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible."[1]

Due to this difference in purpose, administrative data can present particular problems when used for evidence-based policy research. Previous research examining the quality of administrative data has identified several key issues[2]. In this report, as a state data and research agency, we address three commonly-found errors - reporting, nonresponse and measurement, using a research project that Washington's Education Research and Data Center (ERDC) is conducting.

Case Study

In 2015, ERDC received a statewide longitudinal data system (SLDS) grant from the U.S. Department of Education to fund research using preschool-to-workforce (P-20W) data, and to improve SLDS administrative data quality. One of the grant studies examines Washington students' high school course-taking trajectories and their postsecondary outcomes in Science, Technology, Engineering, and Mathematics (STEM). The purpose of this project is to understand how the Washington state public school system

1 United Nations, Secretary General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development. A World That Counts: Mobilizing the Data Revolution for Sustainable Development (New York, NY: United Nations, November 2014); http:// www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf (accessed August 10, 2017).

2 Groen, J. A. 2012. "Source of Error in Survey and Administrative Data: The Importance of Reporting Procedures." Journal of Official Statistics, 28(2): 173-198. Rothbard A. 2015. "Quality Issues in the Use of Administrative Data Records." In: Fantuzzo J., Culhane D.P. (eds) Actionable Intelligence. Palgrave Macmillan, New York.

has prepared students for a future STEM workforce. Ideally, this project will provide insight about whether students are better prepared for the workforce by taking more rigorous STEM courses in reaction to the implementation of the of Common Core State Standard (CCSS).[3] By design, the STEM project will first explore Washington public students' STEM course-taking in high school, and then track those same students' participation in college STEM majors and courses.

The high school course data used for this study is from the Comprehensive Education Data and Research System Student Grade History file (CEDARS-GH), provided by the Office of Superintendent of Public Instruction (OSPI). Student Grade History was originally developed as a transcript-like collection that would provide information on all courses taken for high school credit, including transfer credits obtained outside the reporting district. The collection allows for the submission of multiple years of course information for a student within a single school year submission for the district. With this framework, the data can be used to answer questions about credit accumulation, course-taking trajectories, or courses taken within a single school year.

Currently, OSPI uses Student Grade History data for both public reporting and internal analysis. OSPI's dual credit and Community and Technical Education (CTE) reporting both rely on this data. The Washington School Improvement Framework (WSIF) also includes a metric- 9th Grade on Track, which uses this data to identify the percent of first-time 9th graders who passed all credits attempted. Internally, OSPI uses this information to explore course-taking trajectories and examine course information focused on specific content areas such as math or art.

To collect this data, districts are required to report all public high school level courses where credit was attempted for every student served during each school year. It also records students' course credits earned and final grade, as shown in their transcripts. It is required for school districts to submit all students' course records, for courses of high school rigor. Based on the reporting requirements, we expect the CEDARS-GH file to include each public high school student's complete course-taking profile as long as they enrolled and took courses. However, the data diagnosis we conducted tells a different story.

In the following sections, we first describe the data reporting and processing procedures of CEDARS-GH data before and when it is loaded into P-20W DW, and where potential *reporting error* may occur. Then we introduce a series of data diagnosis approaches we used to identify *nonresponse bias* and validate data completeness, using the STEM study as an example. In the final section, we address potential *measurement error* as we construct valid measurements for math and science course-taking, and then conclude with recommendations to improve the data quality.

3 For more on the development of the Common Core State Standard in Washington state, see http://www.k12.wa.us/CurriculumInstruct/CCSS/default.aspx.

The reporting process for CEDARS Grade History data: a source of reporting error?

Before CEDARS-GH data is loaded into the P-20W DW, there are several data reporting and processing steps as shown in Figure 1. The initial reporting phase happens at each high school where a student took a course and has a course-taking result (e.g. credits earned). Then schools within the same district report their data to the district office, followed by the district's periodic (as often as monthly) submissions to OSPI CEDARS database system. These reporting processes involve several submissions of data to CEDARS for the same school year, and districts are allowed to update prior year data as well. The last step before the data is loaded to ERDC's P-20W DW involves a set of queries run by OSPI to extract data for ERDC.

Each step of data entry, coding, and loading presents risks for reporting errors which could produce inaccurate data records, including duplications, missing cases, and inconsistency over time. For instance, schools may not enter complete records to districts, and districts may not submit all records to the OSPI CEDARS data system due to some coding error, or fewer data entry personnel. In addition, the business rules and queries used by OSPI to extract CEDARS data for ERDC are subject to errors that might cause data issues around completeness and consistency.

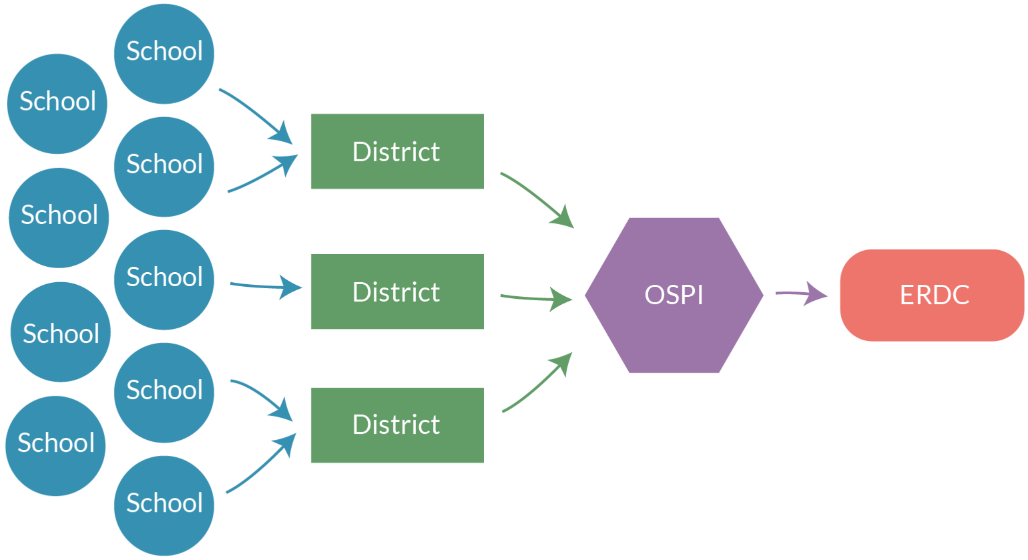

Figure 1. Data reporting process of CEDARS data

Mandatory reporting does not guarantee data completeness: non-response bias?

To look for possible data incompleteness that may have resulted from reporting errors and determine whether the course data in CEDARS-GH will work for the STEM research, we used several data diagnosis approaches. First, we analyzed year-to-year changes in course record counts at the district level. Two measures are used for this analysis: (1) course record count per district and year, and (2) percentage change in course record count per district between years. We started our analysis at the district level because districts collect data from schools and submit to the OSPI state data system (CEDARS). In an ideal world, a district collects all students' coursework records as long as a student takes a course within the district boundary. If a student was found missing a course record in some subject, it should simply mean that the student did not take the course in the school year. However, it is uncertain whether this assumption is true.

Our first analysis indicated that the data is incomplete, likely from reporting errors. The year-to-year record counts in Figure 2 demonstrates inconsistency in reporting all course records by districts. Some districts show a dramatic drop in course records in some years, while others do not. We then calculated the year-to-year percent changes in course record count for each district, which highlights the inconsistency for same-year cross-district comparisons. The results shown in Figure 2 (see also Table A1) help identify which districts we need to validate data quality with OSPI. For example, if the data quality has improved since 2013, why are there several districts with a large record count drop from 2012 to 2013?

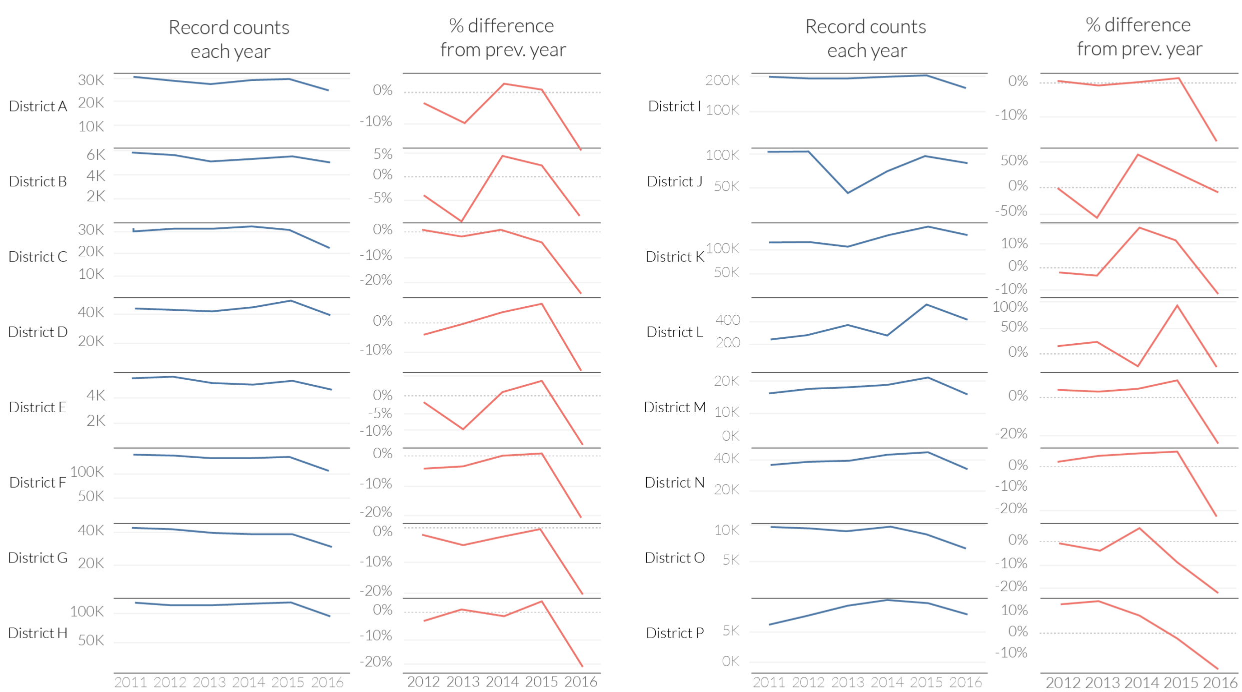

Figure 2. Course record count per district and year, and percent change across years. (See also Table A1)

Based on these analyses, we provided feedback to OSPI for further data quality validation, followed by several conversations in an attempt to obtain a more complete dataset. OSPI delivered a new CEDARS-GH file, in which we found some districts had more missing cases than in the prior data file,[4] so we decided to combine these two datasets for our subsequent analysis[5]. We appended two datasets (the first and updated version), and removed those that are complete duplicates by school year, grade level, student identifier, district identifier, school identifier, and state course codes.

Missing data analysis

Because schools and districts are required to report a complete educational history of each student, administrative education data should ideally include every course record. However, the CEDARS-GH final file we received was not complete nor of good quality. There were missing cases that needed to addressed, since they could potentially lead to selection bias. A missing record could mean one of two things: (1) students simply did not take the courses, or (2) students took courses but schools or districts did not report. We had to rule out reporting bias at the district or school level, so that we could be confident that missing cases were those who actually did not take the course.

To examine district and school reporting behavior, we analyzed course record completeness patterns by taking into account high school enrollments in each year. The following approaches were applied:

Data management

1. *Remove duplications*: Districts report each student's entire course taking history, each year. So, for a student who is in grade 12 during year X, courses they took in grades 9, 10, 11 and 12 are included in the file for year X and courses taken in grades 9, 10 and 11 are included in the file for year X-1 and so on. Courses are taught in terms and are reported by term. Since there is no unique key for each course, the annually-repeated course taking data loads result in many duplicate records. Our goal was to create a file with one record student per year per course per term. To de-duplicate the data, we first took the records where the reporting year is the same as the course taken year, for records where the term is 'ALLYR'. For all other term types, we used course title and course ID to take the last

4 In an attempt to fill in missing data, OSPI sent a more recent file that reflected corrections and updates made by districts, and corrections made to their process for extracting data for ERDC.

5 Although our goal is to get the consistent data OSPI received from districts and start our analysis from there, to meet the research and reporting timeline, we did not continue to examine missing data patterns to completely fill in the gap.

reported record for courses that appeared to be reported more than once.

2. *Create person-year-course summary file*: Aggregate records of the same course subject taken in the same grade and school year for each student. The course subjects extracted here are reading, math, and science. The CEDARS-GH variable used is State Course Code that OSPI suggested districts record, following the Secondary School Course Classification System[6].

3. *Calculate the percent of course completeness per district*: This is a rate to measure differential response rates and dispositions by district. Comparing this measurement by grade level and school year helps examine the consistency of district reporting behaviors.

$$\text{Percent course completness} = \frac{\text{Number of students with course record}}{\text{Number of students enrolled in the district}}$$

Visual analytics

We used Tableau as the visual analytic tool to demonstrate the distribution of course completeness rate from year to year. To determine whether or not course-taking patterns make sense, we used Washington high school graduation requirements and state assessment for 2013-2016 school years for reference. For example, to graduate from high school, students are required to take 4 credits of English and 3 credits of Math courses. Since state assessments are executed in 10th grade, it is very likely most students take a English course from 9th to 12th grade, and take Math courses in at least 9th and 10th grade. In Figures 2a – 2d, 25 random districts were chosen, to illustrate the incompleteness of the data.[7] Although there are few outlier districts that do not have close-to complete (90 percent to 100 percent) reading course records in 9th and 10th grade, most districts have 95 percent or more records and the trends are stable across years.

Math and science show different distributions. The majority of students took math and science courses in 9th and 10th grade. However, the 2013 data shows a wide range of record completeness (from 50 to 100 percent), and in 2014, most districts fall between 65 and 80 percent. In 2015 and 2016, record completeness is improved. There is no evidence that high school graduation requirements for math and science coursework was reduced during 2014 and 2015. Therefore, the inconsistency in earlier years raises concerns about data

6 Bradby, D., Pedroso, R., and Rogers, A. 2007. Secondary School Course Classification System: School Codes for the Exchange of Data (SCED) (NCES 2007-341). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

7 Although the STEM study does not examine reading course-taking patterns, including reading analysis here helps us to check whether the data quality varies by subject. In addition, to graduate from high school, for graduation class of 2018 and prior, students are required to take at least 4 credits of English. That said, each student is supposed to take English in every grade at high school. At minimum, we should observe that most high school students take English in 9th and 10th grade in each district. If the distribution chart shows large incompleteness in these two grade levels, it signals that data errors exist.

quality. One explanation is the use of state course codes, which we used to identify math and science courses in the data, in reporting by districts. OSPI started requiring districts to report state course codes for all "high school rigor" courses in 2012. However, there is no validation done by OSPI to ensure that the codes are applied completely, consistently or accurately by the districts. The finding here does suggest further investigation on data quality for these two years of data. It also provides a heads up for researchers who are going to use this data for analysis. (See Figures 3a - 3d, and Table A1.)

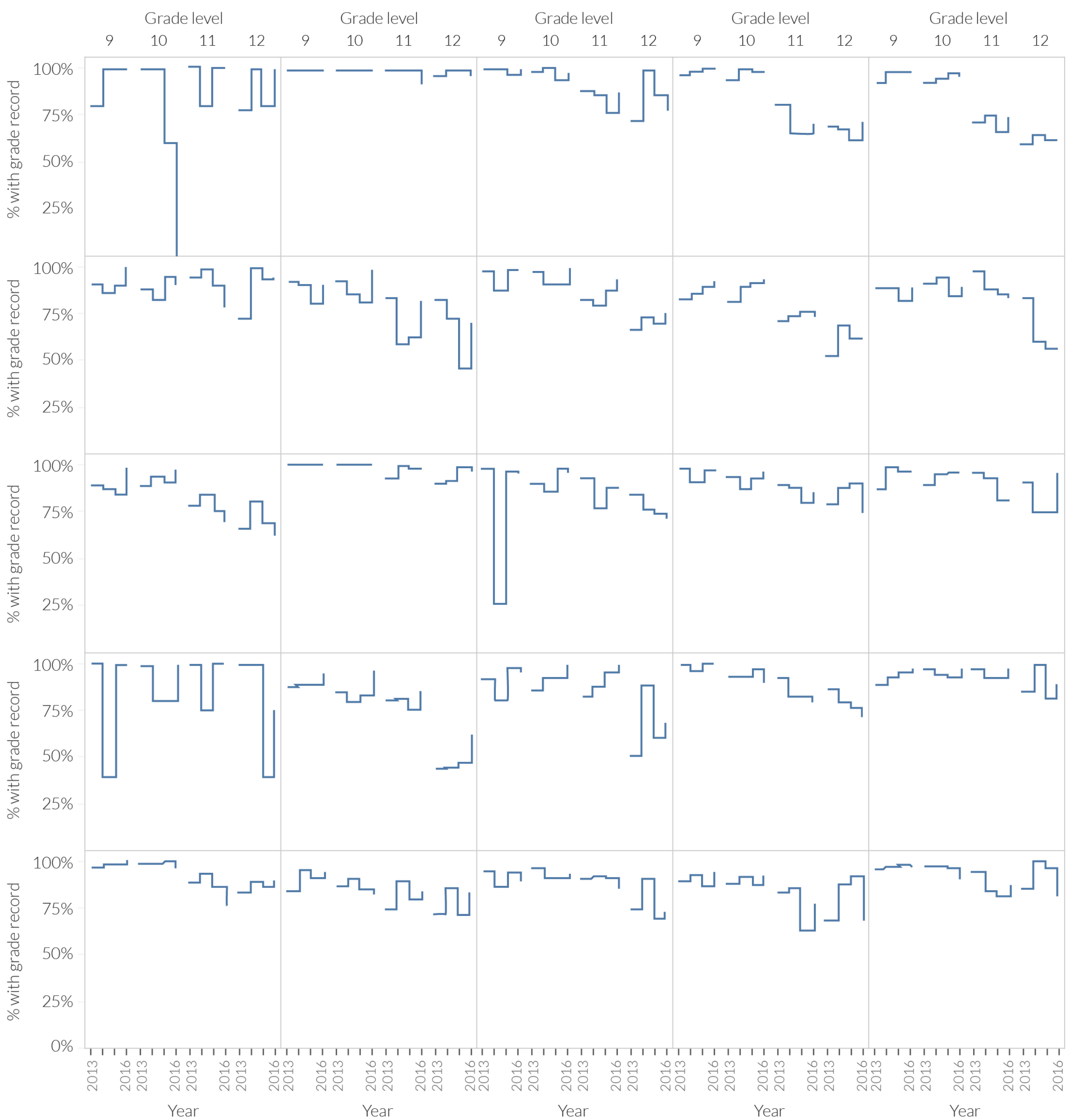

Figure 3a. All course record completeness.

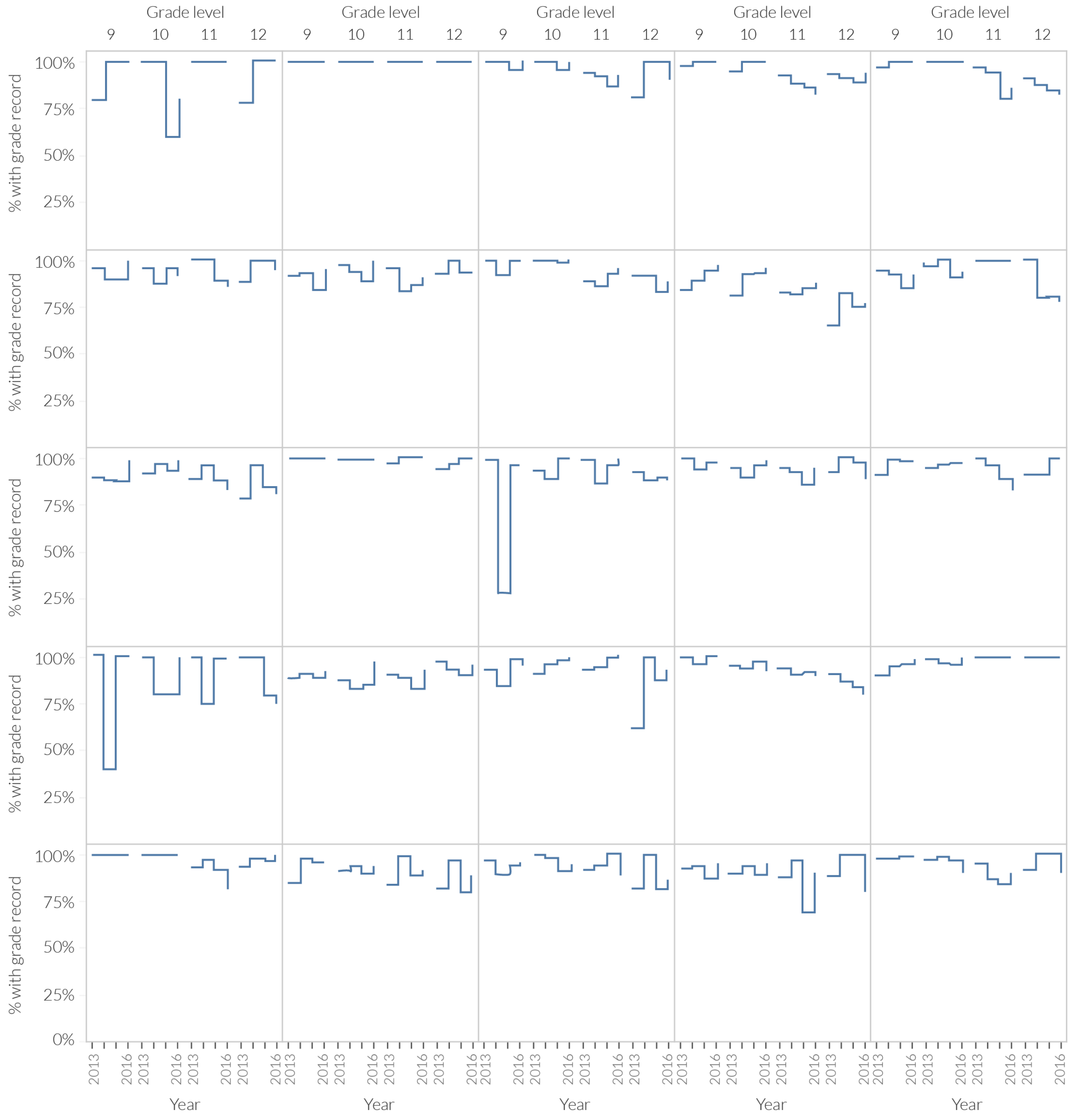

Figure 3b. Reading course record completeness

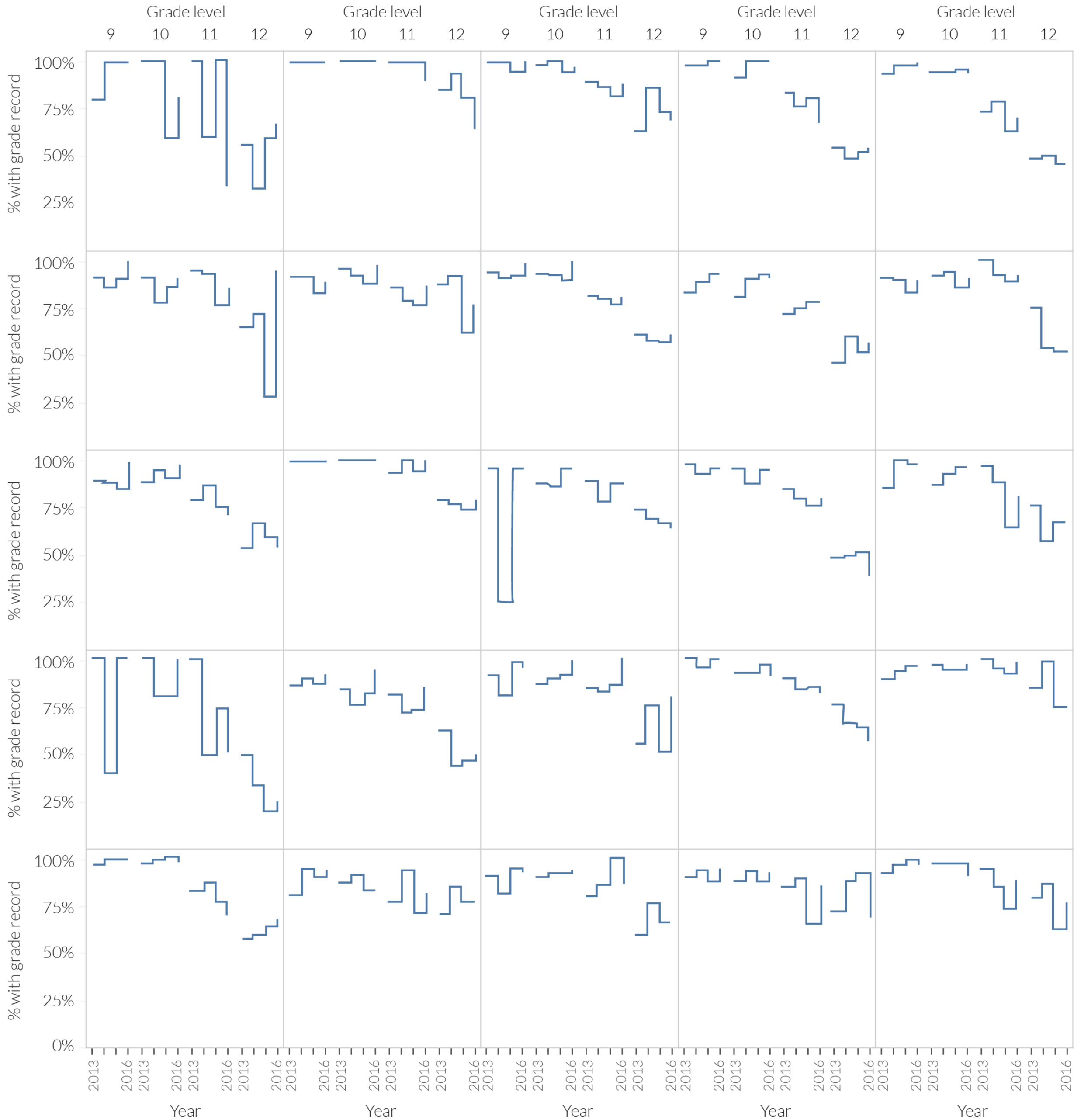

Figure 3c. Math course record completeness.

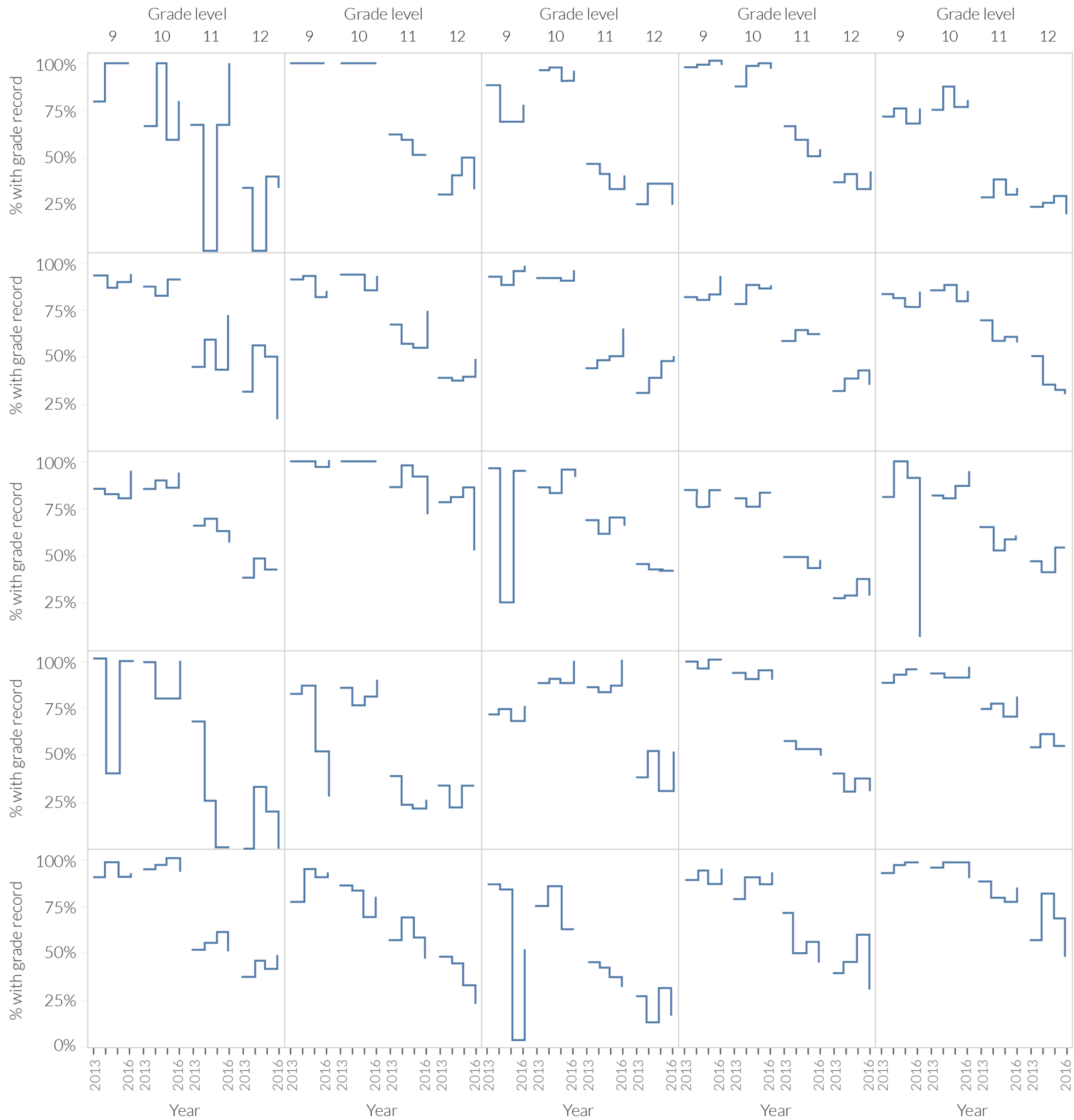

Figure 3d. Science course record completeness

Regression analysis

Finally, we used a regression analysis to explore the relationship between record completeness and reporting agencies' characteristics. We extracted district and school characteristics variables from OSPI Report Card data, and merged them with our CEDARS-GH research file. The simple regression equations are expressed for variation in district and school respectively:

$$Yd = \alpha + \beta\ Xd + \varepsilon d\ (1)$$

$$Ys = \gamma + \tau\ Xs + \varepsilon s\ (2)$$

Equation (1) is for district-level analysis, where each variable used was aggregated to the district level; whereas equation (2) is for school-level analysis. refers to percent of students with course record, d for district, and s for school. X refers to district or school characteristics controlled in the model, and in this analysis, we include student demographics composition, percent of students in various school programs, and school quality indicators. is error term, and represent the coefficient for each district or school covariate, and and are constant.

Table 3 shows the results from the district sample. Districts with a higher proportion of American Indian or Black/African American students are associated with fewer course records, whereas districts with more students of two or more races are slightly more likely to have complete course records, especially for math. Districts with a higher proportion of students from low-income families (measured by percent of students eligible for free- or reduced-price lunch) are less likely to have complete course records, and such association is especially strong for science course record completeness. On the other hand, districts with higher proportion of students with a Section 504 plan are better at reporting course records.

The analysis of the association between districts' school quality and course reporting completeness shows student-teacher ratio per classroom is the strongest predictor of course record completeness. The more students per teacher in a classroom, the fewer course records reported. Districts with a higher graduation rate are also better at reporting records. Among those district variables, student-teacher ratio per classroom is the strongest predictor. This might imply that human resources are positively associated with a district's capability for reporting administrative records.

At school level, the only demographic characteristic associated with course record completeness is race, in that the percentage of Black/African American students is negatively associated with record completeness (see Table 4). Schools with higher proportions of: 1) disadvantaged students from low-income families, 2) students in special education programs or 3) students on a 504 plan, were less likely to report complete course records. Compared to district results, the contradictory finding of Section 504 plan might indicate the fact that reporting Section 504 is more of district duty. It may be that in schools with a lot of students on 504 plans, staff time that might be used to ensure course data quality and completeness is instead used to work with students who need additional services.

Consistent with district findings, school quality indicators play an important role in record reporting. Among school-level characteristics, graduation rate is the strongest predictor of course record completeness, followed by student-teacher ratio.

Even within the same district, there was variation in course record reporting across schools. R-squares in Tables 3 and 4 show that school characteristics explained about 40 to 42 percent of variation in percent of course record completeness, while district characteristics explained about 20 to 28 percent. While schools and districts are supposed to report all courses of high school rigor, it is uncertain why there is such variation and why some schools did not report any course records. One explanation is that some schools included in the enrollment data are not of a regular school type (e.g. juvenile detention schools) and may not be required to report course records. This points to the need to resolve record completeness issues at individual school and district levels when a researcher considers using CEDARS-GH data.

Table 3. Regression analysis on the percentage of students with course records: district level

	% with English course	% with Math course	% with Science course
Demographic composition			
% American Indian	-0.175***	-0.171***	-0.107**
	(0.000)	(0.000)	(0.000)
% Asian	50.073	35.227	33.269
	(0.787)	(0.778)	(0.837)
% Pacific Islander	8.347	5.830	5.452
	(0.787)	(0.779)	(0.837)
% Asian and PI	-54.198	-38.077	-35.859
	(0.787)	(0.778)	(0.837)
% Black/African American	-0.160**	-0.142**	-0.083
	(0.002)	(0.002)	(0.002)
% Hispanic	0.146	0.075	0.291**
	(0.001)	(0.001)	(0.001)
% Two or more races	0.079*	0.132***	0.089*
	(0.001)	(0.001)	(0.001)
% Male student	0.010	0.010	-0.050
	(0.002)	(0.001)	(0.002)
In school program			
% Migrant Ed Program	-0.020	-0.006	-0.050
	(0.001)	(0.001)	(0.001)
% Bilingual Program	0.036	0.149	-0.105
	(0.001)	(0.001)	(0.001)
% Special Education	-0.032	-0.034	-0.048
	(0.001)	(0.001)	(0.001)
% Free/reduced-price lunch	-0.136**	-0.200***	-0.276***
	(0.000)	(0.000)	(0.000)
% Section 504 Plan	0.078*	0.099**	0.153***
	(0.002)	(0.002)	(0.002)
School quality indicator			
Student-teacher ratio	-0.339***	-0.302***	-0.361***
	(0.001)	(0.001)	(0.001)
% Teacher with MA degree	0.045	0.084*	0.058
	(0.000)	(0.000)	(0.000)
5yr graduation rate	0.167***	0.098**	0.089**
	(0.000)	(0.000)	(0.000)
R-Sqr	0.216	0.197	0.280
Observations	966	966	966

Source: 2013-2016 CEDARS grade history data and OSPI state report card data.

Note: Standardized beta coefficients; Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4. Regression analysis on the percentage of students with course records: school level

	% with English course	% with Math course	% with Science course
Demographic composition			
% American Indian	-0.029	-0.010	0.006
	(0.000)	(0.000)	(0.000)
%Asian	32.258	26.266	3.459
	(1.133)	(1.112)	(1.094)
%Pacific Islander	5.590	4.540	0.599
	(1.133)	(1.112)	(1.094)
%Asian and PI	-34.156	-27.734	-3.416
	(1.133)	(1.112)	(1.094)
%Black/African American	-0.097***	-0.073**	-0.084***
	(0.001)	(0.001)	(0.001)
%Hispanic	0.018	0.066*	0.037
	(0.000)	(0.000)	(0.000)
%Two or more races	-0.008	-0.011	0.011
	(0.001)	(0.001)	(0.001)
%Male student	0.008	0.017	0.023
	(0.001)	(0.001)	(0.001)
In school program			
%Migrant Ed Program	-0.002	-0.008	0.010
	(0.001)	(0.001)	(0.001)
%Bilingual Program	0.030	0.013	-0.016
	(0.001)	(0.001)	(0.001)
%Special Edu	-0.100***	-0.086***	-0.113***
	(0.000)	(0.000)	(0.000)
%Free/reduced-price lunch	-0.085***	-0.125***	-0.181***
	(0.000)	(0.000)	(0.000)
%Section 504 Plan	-0.060***	-0.068***	-0.063***
	(0.001)	(0.001)	(0.001)
School quality indicator			
Student-teacher ratio	-0.221***	-0.210***	-0.191***
	(0.000)	(0.000)	(0.000)
%Teacher with MA degree	-0.006	0.014	0.027
	(0.000)	(0.000)	(0.000)
5yr graduation rate	0.512***	0.512***	0.434***
	(0.000)	(0.000)	(0.000)
R-Sqr	0.415	0.421	0.400
Observations	2,131	2,131	2,131

Source: 2013-2016 CEDARS grade history data and OSPI state report card data.
Note: Standardized beta coefficients; Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Potential measurement error

Another caveat an administrative data researcher often runs into is measurement error. Administrative data is not designed to be collected for specific research purposes. Some particular concepts may not be accurately measured by available data, and that leads to potential measurement error.

Ideally, valid course information should measure course rigor and quantity. Based on the NCES SCED course-coding framework it is suggested that four elements be included in student course history data: course description, course level, Carnegie unit, and sequence (Bradby, Pedroso, and Rogers 2007). This framework provides a standard reporting guideline for local educational agencies for providing consistent course records. The goal is to conduct valid and reliable comparative analysis on student course-taking within and across districts. However, CEDARS-GH does not apply this framework completely.

The CEDARS-GH requires districts to submit a data field- State Course Code, based on the 5-digit NCES-SCED course codes. This coding scheme provides only information about subject area (i.e. English language and literature, mathematics, etc.) and course identifiers within each subject area. The other three crucial elements (course level, sequence and Carnegie unit) are missing. Without this information, it is challenging to analyze statewide course-taking patterns and outcomes across schools or districts, and across grades or years.

Using the STEM study as an example, the measurements we were most keen to construct were math and science course rigor, sequence, and the number of course credits. We tried to create standard and consistent measures to examine high school student course-taking progression across schools and districts, and through grade levels over years. Overcoming the limitations of insufficient course information has been a bumpy road. We could capture math course sequence from the 5-digit state course codes, but we could not identify the same for science. We used the variable "Course Designation Code[8]" to identify whether a course is more rigorous in terms of college preparation. But we could not distinguish the level of each course, per the district's definition. Given such limitations, there is a long way to go before we can precisely analyze the association between student course-taking and longitudinal outcomes.

8 For details about this variable, see CEDARS manual - http://www.k12.wa.us/CEDARS/Manuals.aspx.

Discussion

In this report, we conducted a data quality diagnosis using the CEDARS Student Grade History file, and provided recommendations for further data collection and research that are interested in using this data. Specifically, we addressed three common data issues found in administrative data: reporting error, nonresponse bias, and measurement error. The reporting error is associated with the nonresponse bias that normally comes from incomplete data collection. The measurement error is mainly from not collecting the crucial data elements necessary to measure the concepts of interest in the SLDS data collection system. We believe that feedback to data providers, managers, and policymakers can make high-quality data at scale possible. Several recommendations are addressed below:

Continual improvement of data quality. There is no perfect data, especially during the first few phases of data collection. Administrative data collection takes tremendous human resources and is time-consuming. To improve the capacity of informing policy with reliable evidence, data agencies are continuously dedicated to improving data quality.

Taking CEDARS data collection as an example, OSPI regularly interacts with districts around the data collection through a stakeholder group and monthly statewide webinars. Composed of district representatives and SIS vendors, the CEDARS Stakeholder group provides regular collaboration on pressing issues related to the collection, including clarifications on reporting guidance and input on proposed changes. Monthly statewide webinars offer trainings for all districts, with an emphasis on data review through various web applications that allow districts to view data submitted to CEDARS. OSPI also allows districts an opportunity to review their CEDARS data for quality and completeness prior to OSPI using it for public reporting. If they identify issues with their CEDARS data they can correct and resubmit it.

Any substantial change to CEDARS files that is not explicitly required by state law, needs to go through intensive discussions, a rigorous review process, and be approved by the K-12 Data Governance group. As part of that approval process, the CEDARS Stakeholder group provides input on the feasibility and effort required to implement the proposed change. The K-12 Data Governance evaluates that effort against the benefit of collecting additional information, and approves or denies the request.

Transparency.[9] This refers to whether metadata information is well documented and openly provided to data users. The information could include a data codebook, how and why the data were collected, changes in data elements from year to year, and limitations of the data. It is also essential for a data collection agency to describe how decisions on

9 Potok, N. 2018. "Standards and Guidelines for Combined Statistical Data." The ANNALS OF the American Academy of Political and Social Science, 675 January.

methods and procedures were made[10]. The more transparent information is available, the more data users understand how to use the data accurately.

Even though OSPI annually provides a detailed user manual for CEDARS data tables, targeted toward the districts submitting the data, we still rely heavily on back-and-forth conversations with OSPI to understand the key measures we attempt to construct, data collection and processing procedures, consistency of data elements across years, and so on. If technical notes could be well documented in a standard way, and not housed only in a data analyst's memory or work PC, it would substantially speed up data analysis and improve research accuracy.

Fitness for policy evaluation purpose. A vast collection of administrative data could be used for not only reporting purposes, but also for crafting policy. Often times, data agencies are required to report snapshot analyses on an annual or quarterly basis. However, these may not be sufficient to reliably evaluate and inform policy. We did not find strong evidence that the current CEDARS-GH data can be used to observe whether policy changes align with the implementation of more rigorous state assessments. This is probably due to the fact that the current CEDARS-GH file was not designed to observe whether policy changes align with the implementation of more rigorous state assessments, or other relevant policy evaluations.

Data collection with longitudinal scope. Although student administrative data is not collected for specific research purposes, it is expected the data could be used for evidence-based policymaking. The data should include: (1) each student's past and current schooling experience; and (2) policy changes during the same time period. A dataset designed and collected for only a short-term snapshots provides limited information to inform policy.

Useful longitudinal data requires not only collecting data over time, but also collecting meaningful, useful, consistent, and comparable data elements. The current CEDARS student grade history data includes inconsistent data elements and definitions. For example, state course codes are assigned to courses by districts or individual high schools, with no centralized system or guidance by OSPI, so there is no consistency across the state. Further, districts may have recorded same courses differently across time (i.e. math course was coded as art, or vice versa). It is important for OSPI to provide effective trainings for districts to report records accurately.

Quality control. High-quality longitudinal data is accurate, complete, coherent, comparable, reproducible, and sustainable. These features need to be evaluated and maintained over time. Sometimes data quality controls are not practiced for an administrative dataset because the data is not often used for research purposes. If a data

10 National Academies of Science, Engineering and Medicine. 2017. "Principles and Practices for A Federal Statistical Agency," 6th ed., Washington, DC: Committee on National Statistics. The National Academies Press.

collection agency does not produce routine reports using some data, it is less likely that data quality issues will be discovered and addressed in a timely manner. For example, the CEDARS-GH data was not thoroughly explored until external users, such as ERDC and other researchers, attempted to analyze students' course-taking. It is suggested that OSPI proactively analyze data completeness once districts submit records, provide timely feedback to districts, and request resubmission for complete records.

Content matters. Students' core schooling activity is course-taking. Detailed information about what type, level, and sequence of courses a student takes could provide significant insights into what best improves student outcomes. In the era that urges increased school accountability in Washington state, collecting detailed but relevant course information could help identify which courses meet standards, whether students are on track for educational outcomes, and how these courses prepare students for college and the workforce.

Appendix: Tabular data

Table A1. Course record count per district and year, and percent change across years.

Counts	2011	2012	2013	2014	2015	2016
District A	29945	29595	27243	28523	29414	24568
District B	5791	5598	5113	5367	5557	5151
District C	30764	31343	31108	31598	30665	23174
District D	44111	42875	42873	44837	47819	40558
District E	5559	5484	4993	5063	5297	4572
District F	139734	135179	131625	131969	133252	105911
District G	42658	41983	40190	39315	39380	31509
District H	116525	113102	114156	112308	116485	91809
District I	194663	195849	195493	195701	198129	163643
District J	102432	100726	43715	72600	95982	87222
District K	117723	114428	110335	130367	146619	129818
District L	244	285	356	290	565	416
District M	16975	17583	18212	19111	20932	15914
District N	37371	38342	40050	42621	45243	34420
District O	9720	9552	9163	9669	8901	6954
District P	6152	7010	8018	8701	8546	7122
Percent		2011-2012	2012-2013	2013-2014	2014-2015	2015-2016
District A		-1%	-8%	5%	3%	-16%
District B		-3%	-9%	5%	4%	-7%
District C		2%	-1%	2%	-3%	-24%
District D		-3%	0%	5%	7%	-15%
District E		-1%	-9%	1%	5%	-14%
District F		-3%	-3%	0%	1%	-21%
District G		-2%	-4%	-2%	0%	-20%
District H		-3%	1%	-2%	4%	-21%
District I		1%	0%	0%	1%	-17%
District J		-2%	-57%	66%	32%	-9%
District K		-3%	-4%	18%	12%	-11%
District L		17%	25%	-19%	95%	-26%
District M		4%	4%	5%	10%	-24%
District N		3%	4%	6%	6%	-24%
District O		-2%	-4%	6%	-8%	-22%
District P		14%	14%	9%	-2%	-17%

Table A2. Record count by subject, year, and grade level.

District	Grade	Year	Science	Math	English	Any
District 1	9	2013	80%	80%	80%	80%
District 1	9	2014	100%	100%	100%	100%
District 1	9	2015	100%	100%	100%	100%
District 1	9	2016	100%	100%	100%	100%
District 1	10	2013	67%	100%	100%	100%
District 1	10	2014	100%	100%	100%	100%
District 1	10	2015	60%	60%	60%	60%
District 1	10	2016	80%	80%	0%	80%
District 1	11	2013	67%	100%	100%	100%
District 1	11	2014	0%	60%	80%	100%
District 1	11	2015	67%	100%	100%	100%
District 1	11	2016	100%	33%	100%	100%
District 1	12	2013	33%	56%	78%	78%
District 1	12	2014	0%	33%	100%	100%
District 1	12	2015	40%	60%	80%	100%
District 1	12	2016	33%	67%	100%	100%
District 2	9	2013	100%	100%	100%	100%
District 2	9	2014	100%	100%	100%	100%
District 2	9	2015	100%	100%	100%	100%
District 2	9	2016	100%	100%	100%	100%
District 2	10	2013	100%	100%	100%	100%
District 2	10	2014	100%	100%	100%	100%
District 2	10	2015	100%	100%	100%	100%
District 2	10	2016	100%	100%	100%	100%
District 2	11	2013	61%	100%	100%	100%
District 2	11	2014	60%	100%	100%	100%
District 2	11	2015	52%	100%	100%	100%
District 2	11	2016	52%	88%	92%	100%
District 2	12	2013	31%	85%	97%	100%
District 2	12	2014	41%	94%	100%	100%
District 2	12	2015	50%	80%	100%	100%
District 2	12	2016	34%	64%	97%	100%
District 3	9	2013	88%	100%	100%	100%
District 3	9	2014	69%	100%	100%	100%
District 3	9	2015	69%	95%	96%	96%
District 3	9	2016	78%	100%	100%	100%
District 3	10	2013	96%	98%	99%	100%
District 3	10	2014	97%	100%	100%	100%
District 3	10	2015	91%	94%	94%	95%
District 3	10	2016	96%	96%	98%	99%
District 3	11	2013	46%	89%	89%	94%
District 3	11	2014	42%	87%	86%	93%

District	Grade	Year	Science	Math	English	Any
District 3	11	2015	34%	82%	77%	87%
District 3	11	2016	40%	88%	87%	93%
District 3	12	2013	24%	63%	73%	81%
District 3	12	2014	36%	86%	100%	100%
District 3	12	2015	36%	73%	86%	100%
District 3	12	2016	25%	68%	78%	90%
District 4	9	2013	98%	97%	98%	98%
District 4	9	2014	99%	97%	99%	100%
District 4	9	2015	100%	100%	100%	100%
District 4	9	2016	99%	99%	99%	99%
District 4	10	2013	88%	91%	94%	95%
District 4	10	2014	99%	100%	100%	100%
District 4	10	2015	100%	100%	99%	100%
District 4	10	2016	97%	99%	99%	100%
District 4	11	2013	66%	83%	81%	93%
District 4	11	2014	60%	76%	66%	89%
District 4	11	2015	51%	81%	66%	87%
District 4	11	2016	55%	66%	71%	82%
District 4	12	2013	37%	55%	69%	94%
District 4	12	2014	41%	49%	68%	92%
District 4	12	2015	34%	52%	61%	89%
District 4	12	2016	42%	53%	72%	94%
District 5	9	2013	71%	94%	92%	97%
District 5	9	2014	76%	97%	98%	100%
District 5	9	2015	68%	97%	98%	100%
District 5	9	2016	75%	98%	98%	100%
District 5	10	2013	75%	94%	93%	100%
District 5	10	2014	87%	94%	95%	100%
District 5	10	2015	77%	95%	98%	100%
District 5	10	2016	81%	94%	96%	98%
District 5	11	2013	29%	74%	72%	97%
District 5	11	2014	39%	78%	76%	95%
District 5	11	2015	30%	62%	66%	80%
District 5	11	2016	33%	69%	74%	86%
District 5	12	2013	24%	49%	60%	91%
District 5	12	2014	26%	50%	65%	88%
District 5	12	2015	29%	46%	63%	85%
District 5	12	2016	20%	45%	62%	83%
District 6	9	2013	92%	92%	92%	96%
District 6	9	2014	86%	86%	86%	90%
District 6	9	2015	90%	90%	90%	90%
District 6	9	2016	93%	100%	100%	100%

District	Grade	Year	Science	Math	English	Any
District 6	10	2013	88%	92%	88%	96%
District 6	10	2014	83%	78%	83%	87%
District 6	10	2015	91%	87%	96%	96%
District 6	10	2016	91%	91%	91%	91%
District 6	11	2013	45%	95%	95%	100%
District 6	11	2014	59%	94%	100%	100%
District 6	11	2015	43%	76%	90%	90%
District 6	11	2016	71%	86%	79%	86%
District 6	12	2013	31%	65%	73%	88%
District 6	12	2014	56%	72%	100%	100%
District 6	12	2015	50%	28%	94%	100%
District 6	12	2016	16%	95%	95%	95%
District 7	9	2013	91%	92%	92%	93%
District 7	9	2014	93%	91%	91%	94%
District 7	9	2015	82%	83%	81%	85%
District 7	9	2016	85%	89%	90%	95%
District 7	10	2013	93%	96%	93%	97%
District 7	10	2014	93%	93%	86%	94%
District 7	10	2015	86%	88%	81%	89%
District 7	10	2016	93%	98%	98%	100%
District 7	11	2013	67%	86%	84%	96%
District 7	11	2014	56%	79%	60%	83%
District 7	11	2015	55%	76%	63%	87%
District 7	11	2016	75%	87%	83%	90%
District 7	12	2013	39%	87%	83%	93%
District 7	12	2014	37%	92%	73%	100%
District 7	12	2015	38%	62%	47%	93%
District 7	12	2016	48%	77%	70%	92%
District 8	9	2013	92%	94%	99%	100%
District 8	9	2014	88%	91%	88%	93%
District 8	9	2015	96%	93%	99%	100%
District 8	9	2016	99%	99%	98%	100%
District 8	10	2013	92%	95%	98%	100%
District 8	10	2014	92%	92%	92%	100%
District 8	10	2015	91%	90%	92%	99%
District 8	10	2016	96%	100%	100%	100%
District 8	11	2013	44%	81%	83%	89%
District 8	11	2014	48%	80%	80%	87%
District 8	11	2015	49%	77%	88%	93%
District 8	11	2016	65%	81%	93%	96%
District 8	12	2013	31%	61%	67%	92%
District 8	12	2014	39%	58%	73%	92%
District 8	12	2015	47%	57%	71%	84%
District 8	12	2016	49%	61%	76%	88%
District 9	9	2013	82%	84%	84%	85%
District 9	9	2014	80%	89%	87%	90%
District 9	9	2015	83%	93%	90%	94%
District 9	9	2016	93%	93%	93%	97%
District 9	10	2013	78%	81%	82%	82%
District 9	10	2014	88%	90%	90%	92%
District 9	10	2015	87%	93%	92%	94%
District 9	10	2016	88%	91%	93%	95%
District 9	11	2013	58%	72%	71%	83%
District 9	11	2014	64%	75%	75%	82%
District 9	11	2015	62%	79%	76%	86%
District 9	11	2016	61%	79%	74%	88%
District 9	12	2013	31%	46%	53%	66%
District 9	12	2014	39%	60%	70%	82%
District 9	12	2015	42%	51%	63%	76%
District 9	12	2016	35%	56%	63%	77%
District 10	9	2013	84%	91%	90%	94%
District 10	9	2014	83%	90%	90%	93%
District 10	9	2015	77%	84%	84%	85%
District 10	9	2016	84%	90%	89%	92%
District 10	10	2013	86%	93%	91%	97%
District 10	10	2014	88%	95%	95%	100%
District 10	10	2015	80%	86%	85%	90%
District 10	10	2016	85%	91%	90%	94%
District 10	11	2013	70%	100%	100%	100%
District 10	11	2014	59%	92%	88%	100%
District 10	11	2015	60%	90%	86%	100%
District 10	11	2016	57%	92%	84%	100%
District 10	12	2013	50%	76%	84%	100%
District 10	12	2014	35%	54%	60%	80%
District 10	12	2015	33%	53%	57%	81%
District 10	12	2016	30%	52%	57%	78%
District 11	9	2013	85%	89%	89%	90%
District 11	9	2014	83%	88%	88%	89%
District 11	9	2015	81%	85%	85%	87%
District 11	9	2016	95%	99%	99%	99%
District 11	10	2013	86%	89%	90%	92%
District 11	10	2014	90%	94%	95%	97%
District 11	10	2015	86%	90%	91%	93%
District 11	10	2016	94%	97%	98%	99%
District 11	11	2013	66%	79%	78%	89%
District 11	11	2014	70%	86%	84%	96%
District 11	11	2015	63%	75%	76%	87%
District 11	11	2016	57%	71%	69%	82%

District	Grade	Year	Science	Math	English	Any
District 11	12	2013	38%	53%	66%	78%
District 11	12	2014	49%	66%	81%	97%
District 11	12	2015	42%	59%	70%	84%
District 11	12	2016	41%	54%	63%	81%
District 12	9	2013	100%	100%	100%	100%
District 12	9	2014	100%	100%	100%	100%
District 12	9	2015	97%	100%	100%	100%
District 12	9	2016	100%	100%	100%	100%
District 12	10	2013	100%	100%	100%	100%
District 12	10	2014	100%	100%	100%	100%
District 12	10	2015	100%	100%	100%	100%
District 12	10	2016	100%	100%	100%	100%
District 12	11	2013	87%	94%	94%	98%
District 12	11	2014	98%	100%	100%	100%
District 12	11	2015	91%	94%	99%	100%
District 12	11	2016	72%	100%	100%	100%
District 12	12	2013	79%	79%	91%	95%
District 12	12	2014	80%	77%	92%	97%
District 12	12	2015	86%	74%	100%	100%
District 12	12	2016	52%	79%	97%	100%
District 13	9	2013	96%	96%	98%	99%
District 13	9	2014	25%	26%	26%	29%
District 13	9	2015	95%	95%	97%	97%
District 13	9	2016	93%	95%	96%	97%
District 13	10	2013	86%	88%	90%	94%
District 13	10	2014	85%	86%	87%	89%
District 13	10	2015	95%	95%	99%	100%
District 13	10	2016	93%	96%	97%	100%
District 13	11	2013	69%	88%	94%	99%
District 13	11	2014	62%	77%	77%	86%
District 13	11	2015	70%	88%	89%	97%
District 13	11	2016	68%	87%	88%	99%
District 13	12	2013	45%	74%	84%	93%
District 13	12	2014	43%	69%	77%	88%
District 13	12	2015	42%	67%	75%	90%
District 13	12	2016	40%	64%	72%	89%
District 14	9	2013	85%	98%	99%	100%
District 14	9	2014	75%	93%	92%	94%
District 14	9	2015	84%	96%	97%	98%
District 14	9	2016	84%	95%	97%	98%
District 14	10	2013	80%	95%	94%	95%
District 14	10	2014	75%	87%	88%	90%
District 14	10	2015	84%	95%	94%	96%
District 14	10	2016	84%	95%	96%	99%
District 14	11	2013	50%	85%	90%	95%
District 14	11	2014	49%	79%	89%	93%
District 14	11	2015	43%	75%	80%	86%
District 14	11	2016	47%	80%	86%	95%
District 14	12	2013	27%	48%	79%	92%
District 14	12	2014	28%	49%	88%	100%
District 14	12	2015	38%	50%	90%	98%
District 14	12	2016	28%	39%	75%	89%
District 15	9	2013	81%	85%	87%	91%
District 15	9	2014	100%	100%	100%	100%
District 15	9	2015	91%	98%	99%	99%
District 15	9	2016	7%	97%	99%	99%
District 15	10	2013	82%	87%	90%	95%
District 15	10	2014	80%	93%	95%	96%
District 15	10	2015	87%	96%	96%	97%
District 15	10	2016	94%	95%	97%	97%
District 15	11	2013	65%	97%	97%	100%
District 15	11	2014	52%	88%	96%	96%
District 15	11	2015	58%	65%	81%	89%
District 15	11	2016	60%	80%	81%	83%
District 15	12	2013	47%	76%	90%	92%
District 15	12	2014	41%	58%	76%	92%
District 15	12	2015	55%	66%	76%	100%
District 15	12	2016	54%	64%	95%	100%
District 16	9	2013	100%	100%	100%	100%
District 16	9	2014	40%	40%	40%	40%
District 16	9	2015	100%	100%	100%	100%
District 16	9	2016	100%	100%	100%	100%
District 16	10	2013	100%	100%	100%	100%
District 16	10	2014	80%	80%	80%	80%
District 16	10	2015	80%	80%	80%	80%
District 16	10	2016	100%	100%	100%	100%
District 16	11	2013	67%	100%	100%	100%
District 16	11	2014	25%	50%	75%	75%
District 16	11	2015	0%	75%	100%	100%
District 16	11	2016	0%	50%	100%	100%
District 16	12	2013	0%	50%	100%	100%
District 16	12	2014	33%	33%	100%	100%
District 16	12	2015	20%	20%	40%	80%
District 16	12	2016	0%	25%	75%	75%
District 17	9	2013	83%	86%	87%	89%
District 17	9	2014	87%	89%	88%	91%
District 17	9	2015	52%	88%	88%	89%
District 17	9	2016	27%	92%	92%	92%

District	Grade	Year	Science	Math	English	Any
District 17	10	2013	85%	84%	86%	87%
District 17	10	2014	77%	77%	80%	83%
District 17	10	2015	81%	83%	83%	85%
District 17	10	2016	90%	95%	96%	97%
District 17	11	2013	38%	82%	81%	90%
District 17	11	2014	23%	72%	82%	89%
District 17	11	2015	22%	73%	76%	83%
District 17	11	2016	26%	85%	86%	93%
District 17	12	2013	34%	63%	44%	97%
District 17	12	2014	22%	44%	45%	94%
District 17	12	2015	33%	46%	47%	91%
District 17	12	2016	32%	50%	62%	96%
District 18	9	2013	72%	92%	93%	93%
District 18	9	2014	73%	82%	81%	84%
District 18	9	2015	68%	99%	99%	99%
District 18	9	2016	76%	96%	96%	96%
District 18	10	2013	88%	88%	87%	91%
District 18	10	2014	91%	91%	93%	96%
District 18	10	2015	90%	92%	93%	97%
District 18	10	2016	100%	100%	100%	100%
District 18	11	2013	86%	85%	83%	93%
District 18	11	2014	84%	84%	89%	94%
District 18	11	2015	87%	87%	96%	99%
District 18	11	2016	100%	100%	100%	100%
District 18	12	2013	38%	55%	52%	62%
District 18	12	2014	52%	76%	89%	100%
District 18	12	2015	31%	52%	61%	87%
District 18	12	2016	51%	80%	69%	93%
District 19	9	2013	99%	100%	100%	100%
District 19	9	2014	96%	96%	96%	97%
District 19	9	2015	100%	100%	100%	100%
District 19	9	2016	100%	100%	100%	100%
District 19	10	2013	93%	93%	94%	95%
District 19	10	2014	91%	93%	93%	94%
District 19	10	2015	95%	97%	97%	97%
District 19	10	2016	90%	91%	91%	92%
District 19	11	2013	58%	91%	93%	94%
District 19	11	2014	53%	84%	83%	90%
District 19	11	2015	53%	86%	83%	92%
District 19	11	2016	49%	83%	80%	90%
District 19	12	2013	40%	76%	87%	91%
District 19	12	2014	31%	67%	80%	87%
District 19	12	2015	38%	65%	77%	84%
District 19	12	2016	30%	57%	71%	80%

District	Grade	Year	Science	Math	English	Any
District 20	9	2013	89%	90%	90%	91%
District 20	9	2014	93%	95%	94%	95%
District 20	9	2015	95%	97%	96%	97%
District 20	9	2016	96%	98%	98%	99%
District 20	10	2013	93%	97%	97%	99%
District 20	10	2014	91%	95%	95%	97%
District 20	10	2015	91%	95%	94%	96%
District 20	10	2016	96%	98%	98%	99%
District 20	11	2013	75%	100%	97%	100%
District 20	11	2014	77%	95%	93%	100%
District 20	11	2015	71%	93%	93%	100%
District 20	11	2016	81%	98%	98%	100%
District 20	12	2013	54%	86%	85%	100%
District 20	12	2014	61%	100%	100%	100%
District 20	12	2015	54%	75%	82%	100%
District 20	12	2016	55%	75%	89%	100%
District 21	9	2013	91%	97%	98%	100%
District 21	9	2014	98%	99%	99%	100%
District 21	9	2015	90%	99%	99%	100%
District 21	9	2016	92%	100%	100%	100%
District 21	10	2013	94%	98%	99%	100%
District 21	10	2014	97%	99%	99%	100%
District 21	10	2015	100%	100%	100%	100%
District 21	10	2016	94%	97%	97%	100%
District 21	11	2013	52%	83%	90%	94%
District 21	11	2014	56%	88%	94%	98%
District 21	11	2015	61%	78%	87%	92%
District 21	11	2016	50%	70%	77%	81%
District 21	12	2013	38%	57%	84%	93%
District 21	12	2014	46%	59%	89%	98%
District 21	12	2015	42%	65%	87%	96%
District 21	12	2016	49%	68%	91%	100%
District 22	9	2013	77%	80%	84%	85%
District 22	9	2014	95%	95%	96%	98%
District 22	9	2015	91%	91%	91%	96%
District 22	9	2016	93%	94%	95%	95%
District 22	10	2013	86%	88%	88%	92%
District 22	10	2014	84%	91%	92%	94%
District 22	10	2015	68%	84%	85%	90%
District 22	10	2016	80%	84%	83%	93%
District 22	11	2013	57%	78%	75%	84%
District 22	11	2014	69%	94%	90%	100%
District 22	11	2015	58%	71%	81%	89%
District 22	11	2016	47%	82%	84%	91%

District	Grade	Year	Science	Math	English	Any
District 22	12	2013	48%	70%	73%	82%
District 22	12	2014	45%	86%	87%	97%
District 22	12	2015	33%	77%	72%	80%
District 22	12	2016	23%	78%	84%	89%
District 23	9	2013	86%	91%	96%	97%
District 23	9	2014	84%	82%	87%	89%
District 23	9	2015	3%	95%	94%	95%
District 23	9	2016	52%	93%	90%	96%
District 23	10	2013	75%	91%	97%	100%
District 23	10	2014	85%	92%	92%	99%
District 23	10	2015	63%	92%	91%	92%
District 23	10	2016	62%	94%	94%	95%
District 23	11	2013	45%	81%	91%	92%
District 23	11	2014	42%	87%	92%	95%
District 23	11	2015	37%	100%	91%	100%
District 23	11	2016	32%	86%	86%	89%
District 23	12	2013	27%	59%	75%	82%
District 23	12	2014	13%	77%	92%	100%
District 23	12	2015	32%	67%	70%	81%
District 23	12	2016	16%	67%	73%	86%
District 24	9	2013	89%	90%	90%	92%
District 24	9	2014	94%	94%	94%	94%
District 24	9	2015	87%	87%	88%	88%
District 24	9	2016	95%	95%	95%	95%
District 24	10	2013	79%	89%	88%	90%
District 24	10	2014	90%	93%	93%	94%
District 24	10	2015	86%	88%	88%	90%
District 24	10	2016	93%	93%	93%	95%
District 24	11	2013	70%	85%	84%	88%
District 24	11	2014	50%	90%	87%	97%
District 24	11	2015	56%	65%	63%	68%
District 24	11	2016	45%	86%	78%	91%
District 24	12	2013	39%	72%	69%	88%
District 24	12	2014	46%	89%	89%	100%
District 24	12	2015	60%	92%	92%	100%
District 24	12	2016	30%	69%	69%	80%
District 25	9	2013	93%	93%	97%	98%
District 25	9	2014	96%	96%	98%	98%
District 25	9	2015	97%	99%	99%	99%
District 25	9	2016	96%	97%	97%	99%
District 25	10	2013	96%	97%	98%	98%
District 25	10	2014	97%	97%	98%	99%
District 25	10	2015	97%	97%	97%	97%
District 25	10	2016	90%	90%	90%	90%

District	Grade	Year	Science	Math	English	Any
District 25	11	2013	88%	95%	95%	96%
District 25	11	2014	80%	85%	84%	87%
District 25	11	2015	78%	73%	82%	85%
District 25	11	2016	85%	88%	88%	89%
District 25	12	2013	57%	79%	86%	92%
District 25	12	2014	82%	87%	100%	100%
District 25	12	2015	68%	63%	97%	100%
District 25	12	2016	47%	77%	82%	90%

Education Research & Data Center

Trusted. Accurate. Objective.