



Researcher's guide to Massachusetts state education data

September 2018
Office of Planning and Research

Table of Contents

I.	Introduction	2
II.	DESE’s Research Agenda	2
III.	Available Data	3
IV.	Requesting Access to State Data	6
V.	Using SIMS Data.....	10
VI.	Using SCS Data.....	16
VII.	Using MCAS Data.....	19
VIII.	Using NSC Data	27
IX.	Using Other Student-Level Data	29
X.	Using Educator Data	30
XI.	Using Financial Data	32
XII.	Appendices	36
	1. MCAS test data availability by subject and grade	36
	2. Criteria for access to confidential data	37
	3. Sample SIMS file format.....	38
	4. Changes in SIMS data elements by year	40

I. Introduction

The Massachusetts Department of Elementary and Secondary Education (DESE) has a wealth of data on students, schools, districts and educators available for research purposes. The purpose of this guide for researchers and evaluators is to clarify what data is available, how to obtain and interpret it, and, ultimately, to generate better research projects and more accurate and useful results.

The Office of Planning and Research (OPR) is the first point of contact for any external researchers wishing to use DESE data, whether on a contract with the DESE or for their own research purposes. OPR can direct researchers to the non-confidential data that are available and shepherd researchers through the process of qualifying to receive confidential data, if appropriate. Carrie Conaway, Chief Strategy and Research Officer, is the best place to start. She can be reached at cconaway@doe.mass.edu.

II. DESE's Research Agenda

DESE's goal is to prepare all students for success in the world that awaits them after high school. We are implementing five strategies aimed at reaching that goal:

- Strengthen standards, curriculum, instruction, and assessment
- Promote educator effectiveness
- Support social-emotional learning, health, and safety
- Turn around the lowest performing districts and schools
- Enhance resource allocation and data use

Research focused on any one or more of these priorities, whether broad or narrow in scope, would be of potential value to DESE. The agency's research agenda is posted at <http://www.doe.mass.edu/research/agenda.html> and is updated annually.

III. Available data

We have four kinds of data for researchers:

- aggregate data at the school, district, and statewide levels,
- confidential student-level data,
- non-confidential student-level data, and
- data on educators.

Aggregate data

Most of the aggregate data that DESE collects on students, educators, schools, and districts can be found online at one of the following five links:

- The **Profiles** website: <http://profiles.doe.mass.edu/>. Profiles include school- and district-level data on enrollment, class size, graduation and college enrollment rates, dropouts, MCAS/PARCC and SAT results, school and district accountability determinations, plans of high school graduates, per-pupil expenditures, teacher data and salaries, technology, and more. It also provides directories of all the schools in the state by type (public, private, charter, collaborative, etc.). All Profiles tables are exportable in Excel format.
- The **Statistical Reports** website: <http://www.doe.mass.edu/infoservices/reports/>. This lists all the reports produced by the Information Services unit, many of which are in queryable and/or downloadable format. While much of this information is also posted on Profiles, some is available in more detail, and some also provide additional statewide summaries. These reports include the graduation rate, grade retention, mobility, and dropout reports.
- The **school and district finance** website: <http://www.doe.mass.edu/finance/statistics/>. Finance reports on the School Finance and District Support website includes per pupil expenditures, teacher average salaries, special education expenditures, and state totals. Other useful reports on charter schools, school choice, and state aid programs are available as well. Greater explanation about the financial data sets can be found in Section XII. Additional financial data are also available in the DART Detail: Staffing and Finance (see below).
- We have created **aggregate school- and district-level files** to make student assessment results, data used for accountability determinations, and higher education enrollment data from the National Student Clearinghouse available in a more accessible format for researchers. As these become available, we post them at <http://www.doe.mass.edu/infoservices/research/>.
- Researchers may also wish to review the Department's **District Analysis and Review Tools (DARTs)**: <http://www.mass.gov/ese/dart>. The DARTs offer snapshots of district and school performance, allowing users to easily track select data elements over time and make meaningful comparisons to the state or to comparable organizations. The DARTs for Districts and Schools cover a broad range of district and school interests including demographic, assessment, student support, educator, financial, and achievement gap data. The DART for districts and schools are available through the Profiles webpage. For instructions on how to access DART in School/District Profiles see:

<http://www.doe.mass.edu/dart/>. The Department also has a suite of DART tools that provide in-depth data on targeted topic areas including: DART Detail: Staffing and Finance, DART Detail: English Language Learners, Dart Detail: Success after High School. These are available at the same link. A reference listing data points contained across DARTs is posted at <http://www.doe.mass.edu/dart/dart-reference-guide.xlsx>.

Researchers may also wish to review OPR's Reports webpage:

<http://www.doe.mass.edu/research/reports/>. This page includes links to all of the Department's major report series as well as to recent program evaluations and legislative reports.

Confidential student-level data

Confidential student-level data are made available only to researchers who have submitted an approved research project proposal and have agreed to the terms of a data-sharing agreement with the agency. These data series are described in greater detail in later sections.

- **Student Information Management System (SIMS)**: Individual student-level records linked by a unique statewide identifier (SASID) and including student demographics, attendance, discipline, program participation, and plans after high school. 2002 to present. Currently collected in October, March, and end of year.
- **Student Course Schedules (SCS)**. Individual student-level records on which students are enrolled in which courses, along with course completion and performance. 2011 to present. Currently collected in October and end of year; the October file is generally not shared for research purposes.
- **Massachusetts Comprehensive Assessment System (MCAS)**. The Massachusetts Comprehensive Assessment System assesses student performance in mathematics, English language arts, and science and technology/engineering. Student-level records available back to 1998, linked to SASIDs from 2002 to present.
 - **MCAS student questionnaire**. As part of the MCAS administration, students in grades 8 and 10 (and from 2017 on, also grade 5) are asked to complete a questionnaire. The questions have changed considerably over the years. 2002 to present.
 - **English proficiency assessments**. Between 2005 and 2012, the Massachusetts English Proficiency Assessment (MEPA) was administered to all English learners (ELs). We switched to the ACCESS assessment in 2012.
- **School Safety Discipline Report (SSDR)**. Individual student-level records of each time an offense occurs on school property and/or a student is suspended or expelled. 2002 to present.
- **College admissions test data**. The Department receives SAT and Advanced Placement (AP) examination scores for students who take those assessments from College Board. This data is provided to the department without SASIDs, and the Department matches the student names to the corresponding SASID. Graduating classes of 2007 to present.

- **National Student Clearinghouse (NSC)**. The National Student Clearinghouse (NSC) is a non-profit institution that provides enrollment and degree verification from colleges and universities. NSC data is used by the Department to determine college-going, persistence and completion of its high school graduates.

Non-confidential student-level data

We make available certain non-identifiable student-level MCAS and SDDR data for research purposes. The files are available at <http://www.doe.mass.edu/infoservices/research/> and come in two forms: records with individual student demographic information but no school or district identifiers, and records with school and district identifiers but no student demographic information. Assessment files are typically posted in December each year; disciplinary data are typically posted in January or February.

Data on educators

Some educator-level data is public record, whereas others are made available only to researchers who have submitted an approved research project proposal and have agreed to the terms of a data-sharing agreement with the agency. These data series are described in greater detail in later sections.

- **Education Personnel Information Management System (EPIMS)**. Demographic and work assignment information about educators and administrative personnel in schools and districts. 2007 to present.
- **Educator Licensure and Renewal System (ELAR)**. Data on all current and formerly licensed educators, plus recent enrollees in and completers of Massachusetts educator preparation programs. Beginning in 2018 DESE will take a snapshot of the database on October 1 each year to facilitate research and data-sharing.

IV. Requesting Access to State Data

To receive data from the Department not already made available on our website—including but not limited to confidential student-level data—researchers must go through the following steps:

1) Contact the Office of Planning and Research (OPR)

The first step in any data request is to contact Carrie Conaway, Chief Strategy and Research Officer, at cconaway@doe.mass.edu. OPR will determine whether a project will meet our criteria for data-sharing. Appendix 2 of this document lists the criteria DESE uses to assess whether or not to provide data for a particular research project.

Current data handbooks for most data series can be found on DESE's [Data Collection](#) website. Older SIMS handbooks are posted on the DESE research page: <http://www.doe.mass.edu/infoservices/research/>.

2) Submit a project proposal

The next step is submitting a project proposal. The proposal must be on the letterhead of the sponsoring organization and must contain the following elements:

- 1) A brief description of the purpose of the study
- 2) A list of the research question(s) to be answered
- 3) The proposed methodology or analytical approach for answering those questions
- 4) The data needed from DESE to answer the research questions, specifying the particular data series, years, and schools/districts from which data are required
 - a. Please note whether you specifically need access to student or educator names, dates of birth, and/or local (not state) identification numbers, and if so, why. We only provide these data elements if specifically required for a study design, e.g., for data matching purposes.
 - b. Also please state if you specifically need educator's local evaluation ratings. We only give these data out if they are directly tied to a research question for the study, as they are subject to a higher confidentiality requirement than our other data.
 - c. Please note if you already have access to the data you are requesting from another approved project (or note which you do and don't have, if relevant).
- 5) Which [program office\(s\)](#) your work will support, and a description of any prior contact you may have had with them regarding their interest in the project. (Prior contact is not required; we just want to know if you've already had discussions or not.)
- 6) The anticipated timeframe and deliverable(s) for the project.
 - a. In the timeframe, please propose appropriate touch points for agency program staff to engage with you throughout your project, e.g., early on for framing research questions and providing background information, midway for reviewing

methodological decisions and/or draft findings, near the end for helping to interpret results, and so forth.

- b. In most cases the deliverable(s) should include a non-technical version oriented towards policymakers and/or practitioners along with whatever more technical work you may wish to produce for publication. Often this takes the form of a two-page summary in our [EdLines](#) series and/or a briefing for agency leadership and program staff.
 - c. Where appropriate, deliverables should also include individual reports and/or briefings for any individual districts or schools participating in your project.
- 7) The source of funding for the project, if applicable
- 8) Names, titles, email/phone, and organizational affiliations and addresses of the following:
- a. The principal investigator(s) for the project
 - b. The person who will serve as the liaison to DESE for the project
 - c. The person or people who will sign off on the MOU (usually the PI or another authorized signatory from the research organization; some universities also add on someone from their IRB/sponsored projects office)
 - d. The person or people who will require access to our secured drop box to obtain the data

3) Receive proposal approval or denial from DESE

Requests are reviewed by OPR to determine whether a proposed research project meets current DESE information needs and data confidentiality requirements. One of the primary criteria for granting outside researcher's access to confidential data is whether the research will address questions of interest to the Department in high priority areas for policy development and implementation.

The criteria the Department uses to evaluate data requests are provided in Appendix 2. In addition, all proposals are reviewed by the Department's Legal and Education Data Services offices. Given the rigor of the approval process and competing priorities of staff, it can take a substantial amount of time for a proposal to be approved or denied once it has been submitted. Please take this into account in your planning and scheduling.

4) Sign a memorandum of understanding

If the project is approved, OPR will draw up a memorandum of understanding (MOU) detailing the nature and terms of the data-sharing agreement. All parties must agree to the terms of the MOU before any data-sharing can occur. The language of the MOU is determined by federal guidelines for student and educator confidentiality. The only language in the MOU that is negotiable is the language describing the actual data request. The requesting researcher is responsible for managing any approval process at his or her home institution, such as an Office for Sponsored Research or legal department.

A key consideration in the Memorandum of Understanding is how the security of the state's confidential student data will be maintained. Recipients of state data are required to agree to many conditions, including:

- Not using the data for any purpose other than to conduct the specific approved research project;
- Not copying, reproducing, or transmitting the data to anyone else;
- Not reporting study results in a manner that could identify an individual student, including observing a minimum N size of 10 for any disaggregations;
- Storing data using industry-standard encryption and authentication;
- Implementing administrative, physical, and technical safeguards to protect data confidentiality, security, and integrity and to prevent redisclosure;
- Destroying all data once no longer needed for the purpose of the research project.

Researchers who cannot agree to these terms will not receive access to state data.

5) Receive access to data

Once the MOU is in place, OPR will request the confidential data from the Data Analysis and Reporting group. The primary focus of this group is completing the federal and state reporting required of the Department. Data requests for researchers are fulfilled as time is available after those reports are completed. Therefore, please anticipate that once we determine that once a proposal qualifies to receive our data, it could take several months for us to fulfill a data request, depending on the complexity of the request and the timing relative to our annual reporting requirements. We provide data in SPSS format only.

Projects done at the request of and on contract to the Department, rather than proposed by a third party, follow a similar process. These data requests are fulfilled faster than those for projects by third parties. However, these researchers must still submit a proposal as described above and must sign an MOU in order to have access to personally identifiable data.

Data that is covered in the MOU with the Department will be shared with researchers once it is available. Researchers do not need to contact the Department to request each file annually.

V. Using SIMS Data

About SIMS

The Student Information Management System, which launched statewide in 2002, contains individual student records, with a unique statewide identifier (SASID) for each student. SIMS collects data on student demographics, attendance, discipline, program participation, and plans after high school, among other information. Linking SIMS data files across years can allow for longitudinal calculations such as grade retention rates.

SIMS data are generally submitted three times per year school year: in October, March, and end of year.¹ These data usually become available for research purposes in January, June, and September, respectively. Most research projects use the October collection for the basic demographic information about each student, linking to the end-of-year collection for information on student dropout and graduation. The March series is collected primarily as preparation for administration of the spring MCAS and is not generally used for research purposes.

Which students are reported in SIMS?

Massachusetts public school districts must report all students in a district between the ages of 3 and 21 who are:

- Enrolled in the district, regardless of the reason;
- Enrolled in private schools or collaboratives (in state and out of state) and for whom the district is financially responsible;
- Not educated by the district but have active IEPs and receive their special education services from the district (may include home-schooled students or students educated in private schools); or
- Outplaced to the Department of Youth Services (DYS) or a correctional facility.

SIMS does *not* include:

- Home-schooled students (unless they receive special education services from the district);
- Students attending private schools paid by their parent or guardian (unless they receive special education services from the district);
- Students in alternative/adult programs working toward their GED; and
- Students traveling abroad for the school year.

What is a SASID?

A SASID, or State-Assigned Student Identifier, is required for all students reported in the SIMS data collection. It is a ten-digit number unique to each student, and it stays with the student throughout the student's public education in Massachusetts, including pre-K and higher education. The SASID is included in all SIMS submissions. SASIDs ensure that all the information associated with a student does not get associated with another student sharing similar identifying information.

¹ For several years in the mid 2000s, an additional administration of SIMS that only included special education students occurred on December 1, but this has been discontinued.

Common questions on SIMS business rules

Terminology for SIMS administrations

The Department has several ways of identifying a particular administration of SIMS. In addition to the date, an administration of SIMS is given a period number representing the consecutive administration of SIMS as well as a year and letter. For example, the October 1, 2004, administration of SIMS is also Period 13, meaning it is the 13th administration of SIMS since SIMS originated, and SIMS 05A, meaning it is the October administration in the 2004–05 school year. The letter B was used for December administrations, which no longer occur. Consequently, most years use only the letters A, C and D.

Variable names

The datasets for SIMS use variable headings that are shortened versions of full variable names. These full variables are linked to DOE data elements: the categories schools use when reporting their data. For example, the full variable name for the variable heading ENSTAT is “enrollment status at time of data collection” and is also DOE12. The handbook for a given period provides information, organized by DOE element, on the full definition of the variable as well as the possible responses.

Several variables in the SIMS datasets are not included in the DOE elements. The first is ORG_CODE, or organizational code. This is a four-digit code for the school district, and is the same as the first four digits of the school code, which is DOE015. The second variable not included in the DOE elements is PERIOD, which is explained in “Terminology for SIMS administrations” above.

Variable definitions

The current SIMS Data Handbook provides definitions, codes and usage information for each of the DOE data elements. Some of the variables have changed over time. **Before including a given variable, be sure to account for any changes in the definition or coding of that data element.** The total number of variables also changes from year to year, as some are discontinued and others added. Some of the important changes in variables are described below. Additional information on variable changes over time can be found in Appendix 4. Researchers should note that updated handbooks continue to be posted on an annual basis. Hence, all researchers should check the website for any updates that have been added after this version of the Researcher Guide was compiled and compare handbooks from the years relevant to their research.

Changes to students’ characteristics across SIMS administrations

Students sometimes have different codes for categories such as race, special education status, low income status, or limited English proficiency status in different SIMS administrations. Income and special education status can change over time; race represents whatever the parent or guardian indicates on school enrollment forms; and LEP reflects English proficiency status per school determination. Researchers using data from a single administration of SIMS should use the code that is current at the time of that administration. For researchers using data from more than one

year, determining which code to use will depend on the research question. The Department can help researchers decide how to address this issue for their particular research focus.

Multiple SASIDs

While the Department tries to avoid having multiple SASIDs for a single student, it does happen on occasion. In this case, use the SASID that is linked with an MCAS score when possible. Data made available to researchers will already have most multiple SASID issues resolved.

Multiple records for the same student

Once a student has been reported as enrolled in a district in a particular school year, he or she must be included in all district SIMS reports for the remainder of the school year even if he or she transfers out. As a result, students may have more than one record in any given SIMS administration. The enrollment status (ENSTAT) variable indicates whether a student is currently enrolled in a given school or is being reported as a transfer. The reason for reporting and reason for enrollment variables can also contribute to a student having multiple records. See below for the Department's process of ordering enrollment status to make a SIMS file unique.

Minimum N

DESE uses a minimum N of 6 students for reporting any student demographic information and a minimum N of 10 for reporting student test outcomes. Accountability determinations are made for schools and districts with a minimum N of 20 in the aggregate and for subgroups.

Graduation and dropout rates

Researchers conducting analyses that include graduation or dropout rates should read Annual Dropout Rate vs. Cohort Graduation Rate, available online at <http://www.doe.mass.edu/infoservices/reports/gradrates/dropoutvsgrad.html> and the Frequently Asked Questions on graduation rates, available online at <http://www.doe.mass.edu/infoservices/reports/gradrates/gradratesfaq.html>.

In 2006, DESE began to calculate both a four-year and a five-year graduation rate for each cohort. Prior to 2006, graduation rates were estimated from annual dropout data or from grade-level enrollment information. For each cohort since 2006, the standard graduation rate calculation counts students who move within or between districts in whichever school and district they are in when they graduate. However, in order to understand better how schools and districts are doing with students that started ninth grade with them and didn't transfer out, DESE also publishes an adjusted graduation rate. The adjusted graduation rate includes only those students who were in the original cohort and did not leave the cohort. It does not include the students that transferred into the school or district after October 1 of ninth grade.

Linking SIMS files across time allows for calculations like dropout and graduation rate. The process of attributing student outcomes to particular schools and districts can be complex in certain situations and may rely on the March SIMS collection which is not typically shared with researchers. If it is critical for a research project to be able to precisely identify all dropouts and/or members of a graduation cohort as DESE defines them, a graduate and/or dropout file can be requested.

Common questions on SIMS data elements

Race categories

Through the 2004–05 school year DESE collected a single data element covering both race and Hispanic ethnicity. Its categories were: American Indian or Alaskan Native; Asian or Pacific Islander; black; white; and Hispanic.

Beginning in 2005–06, DESE started collecting one data element for race and another for Hispanic ethnicity. It also separated the Asian or Pacific Islander category into two separate categories and allowed respondents to choose to identify with more than one race. The current data elements and possible categories are:

- *Ethnicity*: Not Hispanic or Latino; Hispanic or Latino
- *Race*: white; black or African American; Asian; American Indian or Alaska Native; Native Hawaiian or other Pacific Islander

Each student now has one designation for ethnicity (either Hispanic or not Hispanic) and at least one designation for race, resulting in 62 possible combinations.

We include any student reporting Hispanic ethnicity in the Hispanic category regardless of race. Students classified as not Hispanic are included in the race category they designated. Non-Hispanic students with more than one race are considered multi-race. For example, a student who is both black and Hispanic would count only as Hispanic when collapsing categories. A student who is black, white, and Hispanic would still be included in the Hispanic category. A non-Hispanic student who is both Asian and white would be in the multi-race category.

This results in seven racial/ethnic categories: Hispanic or Latino; White; Black or African American; Asian; American Indian or Alaska Native; Native Hawaiian or other Pacific Islander; and multi-race, non-Hispanic. Additional information about the race/ethnicity categories can be found on the Department’s website at:

http://www.doe.mass.edu/infoservices/data/guides/race_faq.html.

Bilingual education/English language learners

For the 2003–04 SIMS administrations, “Bilingual Education Program Status” changed to “English Language Learners Program Status,” and the response codes also changed. The new variable reflected the passage of legislation in Massachusetts changing the default model of second-language education from transitional bilingual education to sheltered English immersion. The variables for the scores of limited English proficient students on a range of standardized tests were also eliminated. Also, between the 2003–04 and 2004–05 administrations, the meaning of response code 04 changed from “receiving English as a second language services (for waived students only)” to “LEP student whose parent/guardian has consented to opt out of all ELL programs offered in the district.” In 2009, the LEP definition was modified and ELL Program status codes were clarified. Until 2018, a former EL student was defined as a student not currently an EL, but had been at some point in the two previous academic years. Beginning in the 2018–19 school year, former EL status was extended to the prior four academic years.

Low income/economically disadvantaged status

For the SIMS 2001–02 and 2002–03 administrations, the Department divided income status into those eligible and not eligible for free or reduced-price lunch. In the 2003–04 administration, the

codes were expanded to allow differentiation between eligibility for free lunch and eligibility for reduced-price lunch.

In 2014–15, DESE substantially changed its data collection on low income status in response to a policy change at the U.S. Department of Agriculture, which sponsors the free- and reduced-price lunch program. Under USDA’s new [Community Eligibility Program](#), many of the state’s largest and poorest districts may offer free lunch to all students rather than having to individually qualify them for the program. These districts could choose to report each student’s free and reduced-price lunch status in SIMS, or they could report the default 00 value for all students.

As a result, DESE no longer has systematic, statewide individual-level data on students’ free and reduced price lunch status, and DOE19 is no longer used by DESE for reporting and statistical purposes. Instead, beginning in the 2014–15 school year, DESE is using a new metric called “economically disadvantaged,” which is based on a student’s participation in one or more of the following state administered programs: the Supplemental Nutrition Assistance Program (SNAP), the Transitional Assistance for Families with Dependent Children (TAFDC), the Department of Children and Families’ foster care program, and MassHealth (Medicaid). This variable appears as ECODIS in the data. Due to the change in methodology, the number of economically disadvantaged students in most schools is lower than the number of low-income students reported in 2013-14 and prior years. Enrollment and achievement data for economically disadvantaged students cannot be directly compared to low income data in prior years. For more information, see <http://www.doe.mass.edu/infoservices/data/ed.html>.

Special education

In the 2003–04 SIMS administrations, the Department added six new special education categories. Some of these were discontinued in subsequent years. In 2007–08, the variable for special education status was divided by age, with one for children between ages three and five, and another for students between ages six and 21.

High needs

DESE calculates a “high needs” student variable based on the student’s characteristics across SIMS cycles. A student is high needs if he or she is designated as low income (prior to the 2014–15 school year) or economically disadvantaged (starting in the 2014–15 school year), an English learner or former English learner (note change in definition of former EL described above), or a student with disabilities. Because of the change from low income to economically disadvantaged in the 2014–15 school year, be cautious in interpreting or reporting changes in the high needs subgroup over time.

Days of membership/days of attendance: 555 code

Starting in the 2006–07 school year, if a student was reported as enrolled in the previous school year’s end of year report, the student must be submitted again by the district in October of the following school year. If a student’s status change occurred over the summer, districts use the 555 response code for the days in membership and days in attendance variables when reporting summer transfers, graduates, and dropouts. In prior years, the Department relied on a Missing Students Report for students who were enrolled in the end of year report but did not appear again in the following school year’s October submission.

Discontinued Variables

- **Suspension.** Beginning in 2012–13, the in- and out-of-school suspension variable in the SIMS data collection was discontinued, with the default code for those variables as 0.
- **AP courses.** Beginning in 2012–13, the AP course variables in the SIMS data collection was discontinued, with the default code for those variables as 500. Students taking AP courses can be identified through the Student Course Schedule data series.

Making a SIMS file unique

Students may be attributed to multiple schools within an SIMS data collection for a variety of reasons. For example, a student may transfer from public School A to public School B, where they are currently enrolled. In this case the student will appear in the sims file as an transfer in state (enstat=20) at School A and enrolled (enstat=01) at School B.

When a unique file is needed to merge data sets or attribute a student to one school for a given time period, the following enrollment hierarchy process is performed on the SIMS file.

Enstat	Enrollment Hierarchy	Explanation
01	1	Enrolled
06	2	Deceased
04	3	Graduated
10	4	Certificate of attainment
11	4	Completed grade 12 and district approved program
09	5	Reached maximum age, did not graduate or receive a CA
05	6	Permanent expulsion (expulsion)
31, 32...36	7	Dropout
20	8	Transferred in state public
21	9	Transferred - in state private
22	9	Transferred out of state (public or private)
23	9	Home school
24	9	Adult diploma program leading to MA diploma
40	10	Not enrolled but receiving Special Education services only
41	10	Transferred, no longer receiving special education services only

The file is then sorted by the sasid, enrollment hierarchy and days of membership (after removing the 555 days of membership coding used for summer transfers).

For example, SPSS coding would be:

RECODE ATTEND MEMBER (555=SYSMIS).

RECODE ENSTAT ('01'=1) ('06'=2) ('04'=3) ('10'=4) ('11'=4) (09'=5) ('05'=6) ('31'=7) ('31'=7) ('32'=7) ('33'=7) ('34'=7) ('35'=7) ('36'=7) ('20'=8) ('21'=9) ('22'=9) ('23'=9) ('24'=9) ('40'=10) ('41'=11) into ensort.

Sort cases by SASID (A) ensort (A) MEMBER (A).

Select if (sasid ~=lag(sasid)).

Note that SIMS DOE 12 expanded in 2007. Prior to 2007 enrollment codes were:

- 01 Enrolled
- 02 Transferred
- 03 Dropped Out
- 04 Graduate with Competency Determination
- 05 Permanent Exclusion
- 06 Deceased
- 09 Reached maximum age, did not graduate or receive Certificate of Attainment
- 10 Certificate of Attainment

What other resources are available to help researchers use SIMS data?

- SIMS website at <http://www.doe.mass.edu/infoservices/data/sims/>
- Current SIMS Data Handbook: available on the Department's website at <http://www.doe.mass.edu/infoservices/data/sims/SIMS-DataHandbook.pdf>
- Previous SIMS Data Handbooks (from 2002 to 2007): available at <http://www.doe.mass.edu/infoservices/research/>
- List of Massachusetts District and School Codes: available on the Department's website at <http://profiles.doe.mass.edu/search/search.aspx?leftNavId>. From the pull-down menu, select Public School or Public School District and click on the blue Get Results button. The list of either schools or districts will come up and to the right is an export button that will export the information displayed into an Excel file.
- SIMS Explanation of the SIMS Summary Reports, available on the Department's website at <http://www.doe.mass.edu/infoservices/data/sims/sumreports/FY2013Explanation.pdf> provides information on the reports provided to districts and clarifies how certain calculations are done with the SIMS data. For example, the school enrollment report notes what students are included in calculating enrollment numbers based on certain SIMS variables.

VI. Using SCS data

About SCS

The Student Course Schedule data series collects data on the courses students are enrolled in and their performance (marks) in those courses. By linking course assignments across SIMS, EPIMS, and SCS, we can identify which teachers serve which students. SCS is collected in October and at the end of the school year, though the October data is generally not shared for research purposes. The end-of-year data includes all the course data from October and has course completion and performance information (if applicable). End-of-year data is collected in mid-August and files are typically available to researchers around December.

Common questions about SCS data

Terminology of SCS administrations

Similar to SIMS, each administration of SCS is given a period number as well as year and letter. For example the collection for the end of the 2014 school year is titled SCS14D and is also SC_Period 13, meaning it is the 13th administration of the SCS.² SCS is collected twice each school year, once in October which is denoted by the letter A and once at the end of the school year which is denoted by the letter D. Typically only the end-of-year SCS data collection is shared with researchers.

Variable names

Similar to SIMS, the SCS dataset uses variable headings that are a shortened version of the full variable name.

Element	Name	File Heading
SCS01	LASID	LASID
SCS02	SASID	SASID
SCS03	School/Program ID Number	CS_LOC
SCS04	Local Course Code	LCC
SCS05	Subject Area-Course Code	COURSE
SCS06	Class Section	SECTION
SCS07	Course Term	TERM
SCS08	Course Enrollment Status	CS_STAT
SCS09	Course Level	CS_LEVEL
SCS10	Course Credit Available	CS_CAVAIL
SCS11	Course Credit Earned	CS_EARN
SCS12	Course Letter Mark	CS_LMARK
SCS13	Course Numeric Mark	CS_NMARK

² This includes two collections that were done during a pilot year. The first full year of SCS is for the 2010–11 school year.

The SCS data set also includes the ORG_CODE or organization code representing the school district, as well as the period which is explained in “Terminology of SCS administrations” above.

Minimum N

DESE uses a minimum N of 6 students for reporting any student demographic information and a minimum N of 10 for reporting student test outcomes. Accountability determinations are made for schools and districts with a minimum N of 20 in the aggregate and for subgroups.

Variable definitions

The current SCS handbook provides definitions, codes and usage information for each SCS data element. Additional information about some of the data elements are below.

Course types

Districts provide information on what courses the students are enrolled in through two data points, local course code (SCS04) and state-defined subject code (SCS05). For state analysis, Course Subject Area Course Code (SCS05) is used to determine the subject area.

Student performance and course marks

Information on student performance can be provided a number of ways. Students may receive a letter mark, a numeric mark, or both. As outlined in the SCS data handbook, letter marks include standard alphabetical marks (e.g. A-), Pass/Fail, or categorical assessment (e.g. Minimally Acceptable). Numeric marks are on 100-point scale and can go to two decimal places. Students may also be noted as “mark not required” or “ungraded course.” Mark not required and ungraded courses are frequent in the elementary grades, and to a lesser extent in middle school grades, particularly 6th graders in K–6 schools. For example, in the 2011 SCS data set nearly 20 percent of grade 6 students had all their core courses marked as ungraded or mark not required. In high school grades, mark not required and ungraded courses for core subject courses are most common in proficiency-based programs. When doing analyses that focuses on course performance, be aware of the prevalence of ungraded coursework in your sample and make sure you are appropriately accounting for them.

When students have both letter marks and numeric marks, the state uses the letter mark in determining passing and failing. Courses are deemed as passing for state calculations if the student earned a letter grade of D- or better, a rating of minimally acceptable or better, “Pass,” numerical mark of 59.5 or higher, or full credit awarded for a credit-bearing course that is ungraded/mark not required.

Students only receive grades for completed courses (based on SCS08) in SCS. Courses in progress are noted as 88 or 88888.

Course level (SCS09)

It is not clear that this data element is appropriately capturing full, comparable information on the level of difficulty of all courses at this time. This is not the variable the Department uses for calculating Advanced Placement or other advanced coursework. The state uses the NCES Course Subject Area Course Code (SCS05) to determine if it is an Advanced Placement course.

Matching SCS to other files

In many cases, the end-of-year SCS file is merged with the end-of-year SIMS file by SASID or by SASID, Org Code, and School Code, depending on the analysis. To merge SCS with EPIMS, match files on Org Code, School Code, Course, Section and Term.

Course term

The SCS data handbook outlines the array of options that may be used in referencing the length of a course. In aggregating the amount of course work taken in a given subject area, DESE assumes a semester course equals 0.5 of a full-year course, a trimester equals 0.34 of a full-year course, a quarter equals 0.25, a quinmester equals 0.2, etc. Multiple trimester and multiple non-consecutive trimesters are assumed equal to 0.667 of a full-year course, multiple quarters/multiple non-consecutive quarters are equal to 0.5 of a full-year course, and multiple quinmesters/multiple non-consecutive quinmesters are equal 0.4 of a full-year course.

Other data anomalies

- For a small number of high schools, all or most courses for graduating 12th graders were erroneously marked as withdrawn (and letter and numeric mark reflecting “Withdrawn”) in the SCS submission. DESE put validations in place beginning in the 2016 school year, but be aware of the anomaly if looking at 12th grade course completion or performance rates.
- A few districts have missing SCS data in one or more collections, as follows:

District	Data missing	Data reported but EPIMS cross-validation waived
Boston		SCS16A
MA Virtual Academy		SCS17D
Greenfield	SCS11D	
Lowell	SCS11D	
Medford	SCS16D	SCS16A, SCS17A, SCS17D
Sharon		SCS17D
South Hadley	SCS16A	
Watertown		SCS16A

What other resources are available to help researchers use SCS data?

- SCS website at <http://www.doe.mass.edu/infoservices/data/scs/>.
- Current SCS Data Handbook: available on the Department’s website at <http://www.doe.mass.edu/infoservices/data/scs/SCS-DataHandbook.pdf>

VII. Using MCAS Data

Who takes the MCAS?

The 1993 Education Reform Law mandates that all students in the tested grades who are educated with Massachusetts public funds participate in MCAS, including students:

- Enrolled in public schools
- Enrolled in charter schools
- Enrolled in educational collaboratives
- Enrolled at public expense in approved and unapproved private special education schools within and outside Massachusetts
- Receiving educational services in institutional settings
- In the custody of either the Department of Social Services or the Department of Youth Services
- With disabilities (see note below)
- With limited English proficiency (see note below)

Some students with disabilities may receive certain accommodations to facilitate their participation, such as changes in the test timing, setting, presentation, or how the student responds to questions. The small numbers of students with significant disabilities who cannot participate in MCAS even with accommodations are assessed using the MCAS Alternate Assessment (<http://www.doe.mass.edu/mcas/alt/>).

English learners (EL) students are exempted from the MCAS ELA test in their first year in the U.S., though they must take any mathematics and science assessments offered at their grade level. Their ELA results, however, are not included in performance aggregations for any subject area during their first year. Since 2006, EL students have been considered first-year students if they first enrolled in a U.S. school after March 1 of the prior school year.

See the [MCAS website](#) for more information on the testing program in general. The [Test Administration Resources](#), [Test Questions](#), [Results](#), [Technical Reports](#), and [Accessibility and Accommodations](#) links may be particularly helpful.

About MCAS data

State assessment data includes data from the English language arts and mathematics MCAS assessments in grades 3 to 8 and 10; the science assessments in grades 5, 8, and high school; the MCAS student questionnaire data in grades 8 and 10 (and grade 5, beginning in 2017); and the MEPA/ACCESS data on English learners' proficiency in reading, writing, listening, and speaking English.

- **MCAS.** Student-level MCAS data are available back to 1998 but can only be linked with SIMS data starting with the 2002 MCAS test administration. Prior to that time, some student demographic data were available from MCAS, but in the form of student self-reports rather than linked by unique student identifiers. The quality of the match between SIMS and MCAS improved substantially by the 2004 administration; for this reason, we recommend that research projects use student data from 2004 and beyond

where possible. MCAS files from the spring test administration typically become available for research purposes by December.

The two-year PARCC test drive in 2015 and 2016 and the transition to the Next Generation MCAS beginning in 2017 have made using MCAS data across these years challenging, but not impossible. Additional information on using the data from these assessments is below.

- **Student questionnaire.** As part of the MCAS administration, students in grades 8 and 10 (and from 2017 on, also grade 5) are asked to complete a questionnaire. The questions have varied over the years and have included topics such as post-high school plans, computer use, after-school activities, college preparation activities, hours spent on homework, self-efficacy on college and career ready skills, school culture and climate, and other items. Some questions have been asked consistently for years; others change each year. Use the search tool on the Department’s website to find the questionnaires for each year; suggested search term: “MCAS student questionnaire [year].”

Starting in 2015, many of the items on the questionnaire were developed to be reported with a Rasch index. The 2015 and 2016 questionnaires related to student perceptions of self-efficacy on the expectations of college- and career-ready standards in mathematics (2015) and ELA (2016). The 2017 and 2018 surveys focus on student perception of [school climate and culture](#), including sub-dimensions of engagement, safety, and environment. For technical reports and more information on how to use these data, please contact Shelagh Peoples in the Office of Planning and Research (speoples@doe.mass.edu).

- **MEPA/ACCESS for English learners.** Between 2005 and 2012, the Massachusetts English Proficiency Assessment (MEPA) was administered to all English learners (ELs): those whose native language is not English and who are not able to perform ordinary classroom work in English. MEPA scores are available from 2005 to 2012. More information on MEPA is available at <http://www.doe.mass.edu/mcas/mepa/>.

Starting with the 2012–13 school year, EL students have participated in the ACCESS test instead. The results for 2012–13 include a crosswalk that translates ACCESS and MEPA scores. More information about ACCESS for ELs is available at <http://www.doe.mass.edu/mcas/access/>.

Our policy is to release MEPA and/or ACCESS data only for projects specifically looking at the English learner population and requiring more detail than what is available through MCAS and SIMS.

The MCAS and graduation requirements

Students must pass the grade 10 English Language Arts (ELA) and mathematics and (for the class of 2010 and beyond) high school science/technology assessments in order to graduate from high school; this is called the “competency determination.” Students are permitted to retake each assessment as many times as they need to in order to pass; therefore, students may have multiple

assessment records for these tests. Data regarding the competency determination rate can be found at <http://www.doe.mass.edu/mcas/results.html>.

The next generation grade 10 assessments in ELA and mathematics will first be administered in spring 2019. For the graduating classes of 2021 and 2022, the state will determine the passing score in ELA and mathematics by equipercentile-equating the new test to the old test. The policies for future graduating cohorts and for the transition to next-generation science tests at the high school level will be set by the state Board of Elementary and Secondary Education in 2019.

Transition to the Next-Generation MCAS: 2014 through 2019

Massachusetts is now administering the Next-Generation MCAS: computer-based assessments with innovative item types aligned to the new curriculum frameworks. The transition to this test occurred over five years (2014 to 2019) and created some data availability and comparability issues that are important for researchers to understand.

In 2014, Massachusetts began a two-year “test-drive” of the PARCC test. In spring 2014, approximately 700 randomly selected schools participated in the Performance Based Assessment field test of math or ELA tests developed by PARCC. In most cases, PARCC field testing was done in one subject (math or ELA) in only some classes at some grade levels in the participating schools. No performance level information was provided for students who took the PARCC field tests.

The majority of these participating schools chose to have the students who took the PARCC field test also take the MCAS; these students have complete student assessment data for 2014. About 10 percent of schools who took part in the field test did not double-test, and have “PRC” noted as the performance information for field tested students in one subject and an MCAS score for the other subject area in the field tested grades and classes. Thirteen schools only took PARCC tests in both subjects and all grades. This results in just under 10,000 students in grades 3 through 8 who do not have MCAS performance information in one or both subjects for 2014, representing 2.3 percent of students tested in those grades.

In 2015 and 2016, Massachusetts school districts had the option of administering MCAS or PARCC to their students in grades 3 to 8 to fulfill their state testing requirement in English language arts and mathematics. Districts that selected PARCC also had the choice to administer that test on paper or on a computer and could make different test mode choices by school. Districts that had selected PARCC in 2015 could not switch back to MCAS in 2016, but those who selected MCAS in 2015 could switch to PARCC in 2016. See the section below, “Working with 2015 and 2016 Massachusetts assessment data,” for details.

In fall 2015, the Department and Board of Elementary and Secondary Education decided to pursue development of the Next-Generation MCAS assessments, building upon the best aspects of MCAS and PARCC, combining innovative items developed by PARCC with new items specifically designed to assess the Massachusetts learning standards.

In 2017, the state administered the first Next-Generation MCAS in ELA and math to students in grades 3 through 8. Although the new assessments are intended to be administered on computer, transition to primary online testing in grades 3 to 8 was introduced two grades at a time (grades 4 and 8 in 2017, grades 5 and 7 in 2018, and all 3 to 8 grades in 2019). The Next-Generation MCAS

ELA and math tests in grade 10 will begin in spring 2019, as will Next-Generation science for grades 5 and 8. The transition timeline to next-generation high school science tests will be determined by the state Board of Elementary and Secondary Education in 2019.

General MCAS data issues

Merging MCAS and SIMS databases

Both the MCAS and the SIMS databases use a student's SASID (unique student identifying code) to track student data. The SASID is, therefore, the best way to merge the two databases for researchers linking demographic data with assessment results. MCAS files from 2005 to the present have been checked for missing or incorrect SASIDs. MCAS files prior to 2005 may have a small number of duplicate or missing SASIDs. Duplicate SASIDs will not affect aggregation at the school or district levels. However, researchers tracking individual students should omit these cases from the data.

Additional variables in the MCAS data sets

Each year, the MCAS team merges in some variables from SIMS into the MCAS data sets to facilitate reporting. These often include calculated variables that can be useful to researchers, such as urban district designation; official school type (elementary, elementary/middle, middle, middle/high school, high school); number of years student has attended MA schools; number of years of continuous enrollment; whether the student was ever designated an English learner; and so forth. These variables vary from year to year depending on reporting needs. Check the MCAS file layouts posted on the Information Services [research files web page](#) for more information.

Minimum N

DESE uses a minimum N of 6 students for reporting student demographic information and a minimum N of 10 for reporting student test outcomes. Accountability determinations are made for schools and districts with a minimum N of 20 in the aggregate and for subgroups.

Attribution of scores

When a student attends a public school out of his or her district, that student's MCAS scores are included with the school and district the student attends, even if a different district pays for the student. However, if a student attends a private school and is paid for by a public school district, as is sometimes the case for special education students, that student's results are included in those of the sending district. The SPRP_DIS and SPRP_SCH variables represent the district and school to which a student's MCAS scores are attributed. The coding is the same as for the ORG_CODE and SCHOOL variables in SIMS, although the actual school and district may be different in SIMS than in the MCAS database for a given student.

Accounting for students with MCAS scores but no SASID

Despite the Department's best efforts, each year there are some students with MCAS scores but no SASIDs. These students' scores are counted in state, district, and school aggregate results. From 2004 on, students without SASIDs are *not* counted in any subgroup totals. Prior to 2004, these students were counted in subgroups based on information provided on MCAS test booklets.

Accounting for students with raw scores only

In some cases, students have raw scores but no scaled scores. This can occur for several reasons. English learners who have attended school in the United States for less than one year are required to take the math and science MCAS tests, but their scores do not count and they do not receive

scaled scores. Students who take only part of an MCAS exam—for example, if they are absent for one of the testing days—also receive only raw scores.

Accounting for student absences during testing

Before 2006, absent students were assigned scaled scores and performance levels based on the raw score earned (usually 200, or warning/failing). Since 2006, absent students receive a performance level of ABS, which counts in school and district participation rates, but not in aggregate performance levels.

Accounting for students who change schools after October 1

Beginning in 2006, the scores of students who enrolled in a school after October 1 of that year are counted in the district's scores only, not in individual school scores. The exception to this is single-school districts, such as charter or regional vocational-technical schools, for which all enrolled students' scores count. The OCTENR variable in the MCAS dataset identifies which students were and were not enrolled after October 1 each year.

Race codes

Before SIMS became available in 2004, information on race was reported directly on the MCAS booklets. Because of this, there may be some discrepancies on race prior to 2004. Additionally, the race/ethnicity variables for MCAS were not updated in the same year as for SIMS. In the 2006–07 school year, the previous year's SIMS race/ethnicity codes for the student were used.

Working with legacy MCAS data

Using scaled and raw scores from the legacy MCAS

Researchers comparing scores across years on the legacy MCAS (that is, data from prior to 2015) need to be aware of several important caveats. Raw scores represent the number of raw score points earned on the test. Scaled scores rescale the raw scores onto a common metric from 200 to 280 on two-point intervals, adjusting for test difficulty. Scaled scores from 200 to 218 represent a performance level of Warning/Failing; 220 through 238, Needs Improvement; 240 through 258, Proficient; and 260 to 280, Advanced. These scaled scores are common across all tests, but the raw scores associated with each scaled score vary across tests, years, grades, and subjects. To determine which raw scores translate to which scaled scores, simply crosstabulate the data for the particular test and year in question.

We strongly recommend that researchers *not* use scaled scores for descriptive or multivariate analysis on legacy MCAS data, in order to avoid estimating incorrect standard errors. Each performance level essentially has its own scale. The number of raw score points that translate into a single scaled score can vary considerably depending on the performance level, with error greater in the tails. As a result, the scale is not linear and the scaled scores are heteroskedastic (different variances at different parts of the scales) so are not appropriately used for calculating means or conducting significance testing.

Instead, we recommend that researchers either calculate statistics on raw scores and convert them to the associated scaled score when reporting or convert the raw scores to standardized units (mean 0, variance 1) and report results as effect sizes. More information on this issue is available in the MCAS technical reports, available online at

http://www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport_2007.htm.

Changes to statistical technique for equating

Beginning in 2007, the Department modified the statistical technique that it uses to ensure the stability of the measurement of MCAS performance standards from year to year. Details of the change are described in the [2007 MCAS Technical Report](#). Beginning in 2014, we modified the equating technique for the grade 10 ELA and mathematics tests. Details are available on page 6 of the [2014 MCAS technical report](#).

Working with 2015 and 2016 Massachusetts assessment data

In 2015 and 2016, Massachusetts school districts had the option of administering MCAS or PARCC to their students in grades 3 to 8 to fulfill their state testing requirement. This section provides details on the testing process for the 2015 and 2016 data and how researchers should handle the resulting complexities of data analysis to ensure comparability and interpretability of findings.

Test administration details

Among the state's school districts that serve grades 3 to 8, 46 percent administered MCAS and 54 percent PARCC in 2015; see [details](#). In 2016, 28 percent of districts administered MCAS and 72 percent administered PARCC; see [details](#). Almost all districts were required to choose a single test district-wide; only Boston, Worcester, and Springfield were permitted to choose school by school.

In addition, PARCC districts had the option to administer the test online or on paper and could choose test mode school by school. In 2015, 31 percent of districts administered the test entirely on paper, 50 percent entirely online, and 19 percent a mix of the two. In 2016, 39 percent administered entirely on paper, 44 percent administered entirely online, and 17 percent a mix of the two.

All grade 10 students continued to take the legacy MCAS, as it is currently the test used to meet the state's high school graduation requirement. In 2015, a small sample of districts chose to administer PARCC in grades 9 and 11, and some grade 8 PARCC-takers took an Algebra I test rather than a grade 8 mathematics test.

Representative samples

Districts had the option to choose which test to administer, so the districts that selected MCAS may be systematically different than those that selected PARCC. Differences in 2015 were small, but we took additional steps to ensure that our comparisons of data were fair and accurate.

Analysts selected representative samples of 2015 MCAS and PARCC test-takers that mirrored the test-taking population in 2014 and used those students' results to identify which MCAS score was equivalent to which PARCC score in terms of proficiency. We then applied this information to the entire sample to generate a statewide percent proficient. The Office of Student Assessment published a [white paper](#) detailing the methodology, along with [concordance tables](#) that equate PARCC and MCAS scaled scores by subject and grade level. Student-level data provided to researchers for 2015 includes a flag for whether the student was in the representative sample. Researchers who want to calculate statewide statistics should select on this flag and run analyses just on this group of students.

We reported MCAS equivalent scores for students taking PARCC in 2016 using refined [concordance tables](#). We did not report state results, however, because selection bias issues were

too large to be overcome through this method. As a result, we did not report 2016 state-level achievement and growth results in grades 3 to 8 ELA and Mathematics. School- and district-level results were reported specific to whichever assessment was administered locally.

Mode effects

In our online test administrations, we observed a small but meaningful difference in PARCC test results by test mode, with students who took the test on paper scoring somewhat higher than those who took it online. This is a common issue in the first year an assessment is administered online, particularly among students with limited access to computers at home and/or school. The Center for Analysis of Longitudinal Data in Education Research has published a [working paper](#) on this issue.

In 2017, the state used the representative sample and concordance work from 2015 to 2016 to render mode adjustments on PARCC thetas from those years. These mode-adjusted thetas are available for all grade 3 to 8 assessments in those two years and were used to calculate student growth percentiles from 2017 onward. (SGPs calculated in 2015 and 2016 did not adjust for mode.) Analyses conducted with growth percentiles indicated that the majority of mode effects were addressed through the method. Researchers who received 2015 or 2016 data files prior to 2017 should re-request those files to get access to the mode-adjusted thetas.

Working with Next-Generation MCAS data

In 2017, Massachusetts began testing students in grades 3 through 8 on the Next-Generation MCAS assessment in English language arts and mathematics. In 2017, students in grades 4 and 8 were required to take the test online, and 95% of students did so (a few got waivers). Districts had the choice of online or paper administration for the other grades. In 2018, students in grades 5 and 7 also tested online. In 2019, all students in grades 3 to 8 tested online unless they had an accommodation to take a paper form.

The next-generation ELA and math MCAS in grade 10 will begin in 2019, along with next-generation science tests for grades 5 and 8. The transition plan for the next-generation high school science tests will be determined by the state Board of Elementary and Secondary Education in 2019.

Unlike the PARCC scaled scores, the Next-Generation MCAS scaled scores already account for potential mode effects. Mode adjustment studies were conducted in each year in which schools were allowed to select the mode of administration. For grades in which online testing was the expectation, psychometric procedures were used to link the paper results to the online scale.

The scaled scores on the next-generation MCAS have also been designed to address the prior issues around non-comparability across grades and non-linearity of scales across performance levels. Studies using data just from the Next-Generation MCAS therefore should be able to work directly with the scaled scores, rather than using thetas or converting raw or scaled scores to z-scores. Nonetheless we recommend that researchers do the usual tests of model assumptions before reporting findings. For more details, see the [MCAS technical reports](#).

We plan to produce a crosswalk between the Next-Generation MCAS scaled scores in grades 3 to 8 and the 2014 legacy MCAS (the last year prior to the assessment transition) sometime during 2018–19.

Combining assessment data across systems

Researchers may wish to combine data from legacy MCAS, PARCC, and Next-Generation MCAS. This can be done but, particularly when significance testing will be used, should be done carefully to account for the psychometric properties of the tests and the sample selection issues during 2015 and 2016. See important notes above regarding using raw scores for analysis of legacy MCAS and adjusting for mode in PARCC data.

To combine data across assessments, we recommend the following:

- Up to 2014: Convert legacy MCAS raw scores to z-scores
- 2015 and 2016: Use mode-adjusted thetas
- 2017 forward: Use thetas

Researchers should consider the appropriate population from which to compute z-scores given their research question.

What other resources are available to help researchers use MCAS data?

- School and district-level MCAS data from 2003 onward are posted on the Department's website at http://profiles.doe.mass.edu/state_report/mcas.aspx.
- MCAS Technical documents, including technical reports, validity studies, and information about test item development are available at <http://www.doe.mass.edu/mcas/tech/>
- Background information on MCAS, including Frequently Asked Questions, is available online at <http://www.doe.mass.edu/mcas/overview.html>.
- The MCAS District File Layout provides variable names and descriptions for MCAS data for each year. To locate this, complete the Download Info Services Files Form located at: http://www.doe.mass.edu/InfoServices/research/download_form.aspx. After selecting either the MCAS Student-Level Files or the MCAS Aggregated Files, scroll down to the hyperlink for the Excel Workbook with an Excel icon placed just to the right of the hyperlink. This file provides a list of variables, their meanings, and the possible response codes in effect for that year.

VII. Using NSC data

What is the National Student Clearinghouse?

The National Student Clearinghouse (NSC) is a non-profit institution that provides enrollment and degree verification from colleges and universities.³ The Department has a contract with NSC to obtain these data and links them to other state data using the SASID.

How the Department requests data from NSC

Before 2012, the Department submitted batches of recent student graduates. For example, a file titled “2007 Graduates” contained all the students who were identified as graduates in 2007.⁴ These files provide a snapshot of student enrollment in postsecondary prior to the date they were created by NSC; they are not ideal for tracking students’ persistence for longitudinal studies. In spring 2012, the Department submitted a file to NSC containing all Massachusetts graduates from 2003 to 2010. We maintain these data for historical purposes.

Starting in 2012, the Department began submitting an updated file of graduates to NSC in late winter/early spring each year; these data are typically available to researchers in late spring/early summer. As new Massachusetts graduates are added to the file, earlier graduating classes are left off. The most recent submission of graduates from spring 2018 includes Massachusetts graduates from 2009 through 2017. The Department does not generally submit data on a graduating class until 16 months after graduation (i.e. information on the graduating class of 2015 will be available spring/summer 2017). This timeline aligns with our federal reporting requirements and provides the most complete picture of graduates’ enrollment in the year after high school graduation.

In 2013, 2015, and 2017, the Department submitted a file of students who were ever enrolled in the prior 10 school years in grades 8 through 12 or special education beyond 12th grade, but who were not currently enrolled or a graduate from a Massachusetts public school, to NSC for the purpose of determining the enrollment status and educational achievement at postsecondary institutions of these students who are not part of the graduate file. Because of NSC contract and budget constraints, we do not have a firm schedule for future submissions of the non-graduate file, though we do intend to continue to request these data occasionally.

SEARCH DATE is a variable in the NSC files which DESE includes in their submission to NSC. It is the date at which point NSC looks for enrollment in a postsecondary institution, and the format is YYYYMMDD. In 2016, DESE changed what SEARCH DATE was included to capture a student’s college enrollment during their high school years (e.g., dual enrollment). Prior to 2016, the SEARCHDATE was January 1 of the year the student graduated from high school. For example, if a student graduated from high school in 2012, the SEARCHDATE was 20120101. Beginning in 2016, the SEARCHDATE is three years prior to the year the student graduated from high school. For example, if a student graduated from high school in 2012, the SEARCHDATE is 20090101.

³ The 3,600 colleges and universities participating in NSC enroll over 98% of all students attending public and private postsecondary institutions nationally.

⁴ The 2007 graduate file would include all students identified as a graduate in 2007, not just those who were part of a 4- or 5-year graduation cohort. It does not contain information about graduates from previous years.

How NSC matches DESE to data

NSC uses a matching algorithm to determine if a student enrolled in a postsecondary institution based on the first name, last name, middle initial and student's date of birth submitted by DESE with the information submitted by postsecondary institutions. The Record Found indicator states if a match was made: Y means postsecondary enrollment information exists for the student, and enrollment begin and end dates indicates when the student was enrolled in postsecondary; N means the student is not found in NSC postsecondary records. This is not a perfect match; both false positives (students who did not continue on to postsecondary are mistakenly matched up with students who did go to college) and false negatives (students reported as N who did enroll at college) can occur.

What other resources are available to help researchers use NSC data?

- A guide for reading the NSC report can be found at http://www.studentclearinghouse.org/colleges/files/ST_DetailReportGuide.pdf.
- Additional information on the National Center for Education Statistics Classification of Instructional Program (NCES CIP) codes used to classify majors is available at <http://nces.ed.gov/ipeds/cipcode>.
- For additional information on the strengths and weaknesses of the NSC data set and lessons learned about using this data in research, please refer to [The Missing Manual: Using National Student Clearinghouse Data to Track Postsecondary Outcomes](#) and other resources on the NSC's [Notes from the Field](#) section on the Working with Our Data tab.

IX. Using other student-level data

School safety and discipline data (SSDR)

The SSDR report tracks each time an offense occurs on school property. Prior to 2012–13, this report only collected information on drug, violent, or criminal offenses for all students, along with all offenses by students with disabilities that resulted in suspensions or expulsions.

Beginning in 2012–13, the report expanded to also include any other suspensions or expulsions for non-drug, non-violent or non-criminal related offenses, irrespective of type of student, and the resulting disciplinary action. (Note: The in- and out-of-school suspension variable in the SIMS data collection was discontinued in 2012–13. SSDR is now the sole source of suspension information.) See <http://www.doe.mass.edu/infoservices/data/ssdr.html> for more information on SSDR.

The trend line in the state’s discipline data was affected by the implementation of Chapter 222 of the Acts of 2012, a policy change regarding suspensions which took place on July 1, 2014. See our [highlights of changes to the state law](#) and our [advisory to districts](#) for more details.

College admissions tests (AP, SAT, and ACT)

DESE receives student-level data on SAT and Advanced Placement (AP) examination scores for students who take those assessments from College Board. This data is provided to the department without SASIDs, and the Department matches the student names to the corresponding SASID. These are available for the graduating classes of 2007 to present.

DESE also receives student-level data from the ACT program. Very few students in Massachusetts take the ACT and the data are not linked to SASIDs, so they are not commonly used in research projects.

X. Using educator data

What data are available about educators?

DESE has two sources of information about educators: the Education Personnel Information Management System (EPIMS) and the Educator Licensure and Renewal data (ELAR). EPIMS collects demographic and work assignment information on individual public school educators twice per year (October 1 and end of year). ELAR collects educator preparation and licensure information on licensed educators and educators currently enrolled in preparation programs in the state, whether they are practicing in a public school or not. ELAR is a constantly updated transactional database; DESE takes a snapshot of it once a year on or about October 1 to generate data for analytical purposes.

Common questions about EPIMS data

EPIMS data structure

EPIMS collects two separate sets of records about each educator, linked by MEPID, the unique educator identifier. The staff roster collects one record per educator per district of employment, with demographic information such as age, gender, race/ethnicity, and employment status at the time of data collection. The work assignment data collects one record per educator per work assignment (i.e., a course assignment), with information about the specific course and section taught, grade level, course term, and full-time equivalency.

Linking to the Student Course Schedule data series

To merge SCS with EPIMS, match files on Org Code, School Code, Course, Section and Term. The student-teacher link is of higher quality in the 2011–12 years and forward, once school districts fully adopted the SCED codes at the elementary and middle school levels.

Data quality issues

At this time, the beginning teacher field in EPIMS does not appear to be consistently reported across districts. It is best to identify new teachers as those that have not previously appeared in EPIMS. This will correctly capture all teachers new to teaching in Massachusetts public schools since 2007 but will incorrectly classify experienced teachers whose prior experience was in private schools or out of state.

Also, hire date appears to be reported by most districts as the date that employment started, rather than the date an offer of employment was made.

A few districts have missing data for some EPIMS collections, as follows:

- Boston EPIMS16A (except the Dever and UP Holland schools)
- MA Virtual Academy EPIMS17D
- Medford EPIMS16A, EPIMS16D (data accepted but cross-validation waived), EPIMS17A
- Sharon EPIMS17D
- South Hadley EPIMS16A
- Watertown EPIMS16A (data accepted but cross-validation waived)

Common questions about ELAR data

Sharing ELAR data

DESE generally shares ELAR data only for projects specifically looking at issues related to educator preparation or licensure.

General data elements

ELAR collects first/middle/last name, self-reported race, and self-reported gender on all records. MEPIDs are attached to the records through EPIMS identification maintenance. Some older records, particularly for non-active educators, may not have MEPIDs attached.

Preparation program data elements

ELAR collects preparation program organization code and name, general program area (e.g., Academic: Administrator), program description (e.g., elementary), grade level the program covers (e.g., 5 to 8), the credential being earned (initial/professional, baccalaureate/post-baccalaureate, etc.), the dates of program enrollment and completion, and the program completion status for each candidate enrolled in a preparation program. We began collecting preparation program data in 2009–10, but initially on completers rather than all enrollees.

Educator license data

ELAR collects educator license number, license credential field (e.g., biology, elementary), license grade level (e.g., 1 to 6, 5 to 8), license stage level (e.g., initial, professional), license classification (e.g., Academic: Teacher), date the credential was issued, date the credential expires or expired (where available), and status of the credential (e.g., expired, licensed) for each license held by a Massachusetts educator.

MTEL data

ELAR collects MTEL test code, name, and date; test result (e.g, P=pass, F=fail), reading test result, writing test result, and numerical test scores for all tests taken by Massachusetts educators.

Self-reported degree data

Educators may choose to self-report their college major and degree level.

XII. Using Financial Data

What financial data are available?

Sources of expenditure and revenue finance data include the End of Year Financial Report (EOYR) for municipal and regional districts and the Charter School End of Year Report (CSEOYR). Additional information on state aid and reimbursements are reported by the School Finance office, including Chapter 70 aid calculations and allocations, calculations of tuitions to charter schools and school choice, state reimbursement for special education extraordinary expenses (“circuit breaker”), and state transportation reimbursement.

Expenditures are typically reported by expenditure function and funding source. Funding sources include general appropriations (district and municipal expenditures for education), grants (federal, state, and local), and local revolving funds (e.g. athletics fees, food service revenues). Expenditures are reported out as annual dollar and per pupil amounts from the EOYR used by municipal and regional districts and CSEOYR by charter schools. However, the two reports do not have the same chart of accounts and cannot be merged. The EOYR chart of accounts can be found at <http://www.doe.mass.edu/charter/finance/revexp/>; the CSEOYR chart of accounts is embedded in the [CSEOYR workbooks](#). Data from these different sources should only be compared at highly aggregated levels that include approximately the same expenditure categories.

End of Year Report (EOYR) for municipal and regional districts

The End of Year Financial Report (EOYR) is a comprehensive report of revenues and expenditures by fiscal year. Information on End of Year Financial Reports are available at <http://www.doe.mass.edu/finance/accounting/eoy/>. Data from this report allows for several spending comparison reports available as downloadable Excel files at <http://www.doe.mass.edu/finance/statistics/> and in the DARTs.

These reports include instructional and operational expenditures; capital, debt, community activities and regional assessments are not included. These reports typically are aggregated at the function level and by functional categories, and by funding sources (general appropriations and grants/revolving funds). The Chart of Accounts-Criteria for Financial Reporting used in EOYR is available at: <http://www.doe.mass.edu/finance/accounting/eoy/ChartOfAccounts.pdf>

These reports distinguish between expenditures for in- and out-of district students. In-district students either reside in the district or are tuitioned in from another district. Instructional and operational costs are divided by this number to get per-in-district pupil expenditures. Out-of-district expenditures include tuitions and transportation for students who attend charter schools, students who attend other districts through interdistrict school choice or other tuition arrangements, and students with disabilities served through out-of-district placements. An example of other tuition arrangements would be tuitions that a town might pay to send its students to high school in a nearby district because the town has no high school and doesn't choose to be a member of a regional high school district. The out-of-district total cannot be properly reported as a per-pupil expenditure because the cost of tuitions varies greatly depending on the reason for going out of district. School choice tuitions are approximately \$5,000 by state regulation, while charter school tuitions are approximately the same as a district's cost per pupil, and a special education placement for a student with complex needs could cost more than \$100,000. Two districts with one student

each in the same special education placement would have very different costs “per out-of-district pupil” if one of them also sent 100 students to other districts at \$5,000 each. School choice and charter opportunities are not equally available to every district’s students, and those tuitions vary a great deal by district.

Reports include:

- **[Per Pupil Expenditures, All Funds](#)** In addition to showing the overall cost per pupil, this report provides detail about how much districts spend in specific functional areas such as administration, teaching, and maintenance. These are available at
- **[Special Education Direct Expenditures Trends](#)** This report provides details on expenditures for students with disabilities, both in and out of district. Expenditures from general appropriations are reported by placement category, and expenditures from grants are reported as a lump sum.
- **[State Total Spending Trends](#)** This report provides statewide spending from all funds by functions and functional categories. It also offers calculations of year-over-year change and some charts of the data.
- **[Teacher Average Salaries](#)** This report includes total spending on teacher salaries divided by full-time-equivalent (FTE) teachers to calculate average teacher salary by district. The report explains which salary data from the EOYR, and which teachers, are included.
- **[District Analysis and Review Tools \(DARTs\)](#)** Other finance reports using EOYR data can be found in the DART Detail: Staffing and Finance. This tool provides a wide array of reports: district staffing, teachers, SPED staffing, finance summary, expenditure by source, per-pupil expenditure, out of district expenditures, district revenues, and revolving fund revenue and expenditure. See the [DART Finance User Guide](#) for descriptions of the data and guidance on potential uses, as well as staffing and finance definitions.

Charter School End of Year Reports (CSEOYR)

Charter schools are required to keep an accurate account of all of their financial activities and report their revenues and expenditures. This data is reported slightly differently than the EOYR for municipal and regional districts. Aggregated annual data of the Charter School End of Year Financial Report is available at <http://www.doe.mass.edu/charter/finance/revexp/>. Reports show dollar and per pupil expenditures, as well as revenue. Once the workbook is open, researchers can access worksheets with the raw data. Note: Charter schools do not enroll out-of-district pupils. Additional information about charter school finance available at <http://www.doe.mass.edu/charter/finance/>.

Additional revenue data, including state aid

Chapter 70

Chapter 70 state education aid of over \$4 billion is allocated to districts each year largely based on their foundation enrollment and their property and income wealth. The aid ensures that all districts can spend at or above their foundation budgets (the state’s estimate of the minimum needed in each district to provide an adequate educational program). The funds are deposited into local general funds, and are appropriated to and budgeted by school committees without any distinction

between local and state aid dollars. Information about how Chapter 70 is calculated, links to legislation, state aid and spending requirements and the following data resources are available at <http://www.doe.mass.edu/finance/chapter70/>:

- *Chapter 70 district profiles*: For each school district, yearly spending and state aid totals in comparison to the foundation budget, are available going back to FY93.
- *Chapter 70 Trends in Aid and Local Contribution*: The trend tool shows key factors influencing Chapter 70 required local contributions and state aid calculations going back to FY2007.

School choice and charter tuition

Some districts have tuition expenditures and/or revenues from interdistrict school choice. Details of these tuitions, by sending and receiving districts, are available at:

<http://www.doe.mass.edu/finance/schoolchoice/>. A significant part of charter school revenues are tuitions paid by municipal districts; details about tuitions by sending district and receiving charter district can be found at <http://www.doe.mass.edu/charter/finance/tuition/>.

Circuit Breaker (reimbursements for extraordinary special education expenditures)

The Special Education Circuit Breaker program reimburses districts for per student costs that exceed four times the state average foundation budget per pupil as calculated under the Chapter 70 program, with the goal of the state paying up to 75 percent of the costs above that threshold, subject to appropriation. It also includes a provision that allows districts to claim for extraordinary relief when claimable Special Education costs exceed 125% of the previous year's claimed costs. Worksheets available at <http://www.doe.mass.edu/finance/circuitbreaker/> report Circuit Breaker funding by year.

Transportation

Regional districts are reimbursed for a percentage of their transportation expenses set by the legislature in each year's budget appropriations. Transportation reimbursements data are available at <http://www.doe.mass.edu/finance/transportation/> by fiscal year.

Grant allocation and awards

Reports on state and federal entitlement and allocation grants, as well as competitive awards, are available at <http://www.doe.mass.edu/grants/awards.html> and posted in workbooks by fiscal year at <http://www.doe.mass.edu/finance/accounting/eoy/>. Expenditures from grants are reported in aggregated categories, rather than by individual grants, in the EOYR and CEOYR reports.

Common questions about finance data

Student enrollment calculation

Student enrollment has specific definitions in the finance area that differ from standard reporting. For end-of-year per pupil expenditure reports, FTE pupil numbers are calculated based on days in membership over the whole year. For the state's Chapter 70 aid program, foundation enrollment is all students enrolled in a district on October 1st of the prior year for which the district is financially responsible because they reside in the district. Student enrollment reported in Profiles is the number of students enrolled in a district (regardless of residence or tuition) as of October 1st each year.

Teacher FTEs

Teacher FTEs also have specific definitions in the finance area that differ from standard reporting. The FTE of teachers for the teacher average salary report is calculated from EPIMS (staffing data) but includes teachers who are on salary but not actually teaching for part or all of the year, and teachers that teach post-secondary, post-graduate vocational programs and are not collected through EPIMS. This differs from teacher FTEs reported in Profiles reports, which use data reported as of October 1 each year.

Fiscal year

Fiscal years refer to the same time period as school years, July 1 to the following June 30. FY16 or 2016 refers to fiscal year 2015–2016.

Caution in comparing across data sources

Finance data analysis requires clarity about what is and is not included across and within the various data reports. For example, municipal/regional districts and charter districts report in separate templates and charts of accounts. Charter data includes capital and debt, whereas municipal/regional expenditure per-pupil reports do not include debt, capital, community activities, and regional assessments. Please contact the Office of Planning and Research for questions about using these data for research purposes.

XIII. Appendices

Appendix 1: MCAS Test Data Availability by Subject and Grade

(For years in which state assessment data can be linked to SIMS)

Discipline	2002	2003	2004	2005	2006	2007 to present
<i>ELA</i>						
3	✓	✓	✓	✓	✓	✓
4	✓	✓	✓	✓	✓	✓
5					✓	✓
6					✓	✓
7	✓	✓	✓	✓	✓	✓
8					✓	✓
10	✓	✓	✓	✓	✓	✓
<i>Composition</i>						
4	✓	✓	✓	✓	✓	✓
7	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓
<i>Mathematics</i>						
3					✓	✓
4	✓	✓	✓	✓	✓	✓
5					✓	✓
6	✓	✓	✓	✓	✓	✓
7					✓	✓
8	✓	✓	✓	✓	✓	✓
10	✓	✓	✓	✓	✓	✓
<i>Science/technology</i>						
5		✓	✓	✓	✓	✓
8		✓	✓	✓	✓	✓
Biology						✓
Chemistry						✓
Physics						✓
Tech./Eng.						✓
<i>Retests</i>						
ELA grade 10	✓	✓	✓	✓	✓	✓
Math grade 10	✓	✓	✓	✓	✓	✓
Composition	✓	✓	✓	✓	✓	✓

Appendix 2: Criteria for Access to Confidential Data

Required criteria

- The researcher must be affiliated with an academic institution or independent, nonpartisan research organization.
- The researcher and any institutions with which s/he is affiliated, including third-party funders of the proposed research project, must not be in an advocacy role with respect to the topic of study or have a stake in the study's outcome.
- The study design and methodology must support an objective analysis of the proposed research question(s).
- The study must propose to examine a topic relevant for policymaking.
- The requester must be willing and able to comply with the requirements associated with being an authorized representative of the Department for the purpose of conducting the study, as defined in the MOU.
- If the study request is from a junior researcher such as a student or research assistant, the project proposal must be under the direction of a faculty advisor who already has an approved data-sharing agreement with the Department. The proposal must include a letter of support from the senior advisor.

Additional considerations: Study design and methodology

- The proposal includes specific, well thought out research questions that relate to a topic relevant for policy decision-making.
- The methodology aligns with the research questions (i.e., the research questions can be answered with the proposed methodology) and is consistent with current best practices in research design.
- The research questions can be answered effectively with data the Department has available.
- Previous work published by the researcher (and/or faculty advisor) suggests that he or she has expertise in the topic to be studied and the methodology to be used.
- Previous work published by the researcher (and/or faculty advisor) suggests that he or she has credibility with other researchers and experts.

Additional considerations: Benefit and relevance to DESE policy and programming

- DESE has the capacity to fulfill the request at this time.
- Gathering more information on the proposed topic is a high Department priority.
- The information gained will be relevant statewide or for high priority districts, schools, or students.
- The study is likely to provide information in time to be of use for anticipated decision-making.
- The study does not overlap undesirably with other work already in progress.

Meeting all listed criteria is not a guarantee that the Department will make individually identifiable data available to a researcher; these criteria are necessary but not sufficient. Final discretion to share identifiable data with a researcher rests with the Department.

Appendix 3: Sample SIMS File Format

Below are the variable names and corresponding data elements from SIMS17A. The exact order of the variables in the and variable names may slightly vary over time.

File Heading	Element	Name
PERIOD	-	Period
SASID	DOE002	State Assigned Student Identifier (SASID)
FNAME	DOE003	First Name
MNAME	DOE004	Middle Name
LNAME	DOE005	Last Name
CITY_OB	DOE008	City/Town of Birth
DOB	DOE006	Date of Birth
LASID	DOE001	Locally Assigned Student Identifier (LASID)
SCHOOL	DOE015	School Identification Number
ATTEND	DOE017	Days in Attendance
MEMBER	DOE018	Days in Membership
TOWN	DOE008	City/Town of Birth
RACE	DOE010	Race/Ethnicity
RFR	DOE011	Reason for Reporting
ENSTAT	DOE012	Enrollment Status at Time of Data Collection
RFE	DOE013	Reason for Enrollment
GRADE	DOE016	Grade Level
LOWINC	DOE019	Low-Income Status
LEP	DOE025	English Language Learner
PLAN	DOE033	High School Completer Plans
IMMIG	DOE022	Immigration Status
FLANG	DOE024	First (Native Language)
CTE	DOE035	Career/Vocational Technical Education (CVTE) - Type of Program
CTE_COMP	DOE031	CVTE- Competency Attainment
COUNTRY	DOE023	Country of Origin
ORG_CODE	-	District Code
SPED_DIS	DOE036	Special Education - Nature of Primary Disability
LVL_NEED	DOE038	Special Education - Level of Need
FPS_504_CODE	DOE039	504 Plan Status
C74_SPOP	DOE042	CVTE - Special Populations
C74_PROG	DOE043	CVTE - Chapter 74-Approved Vocational Technical Education Program Participation
N74_PRG	DOE044	CVTE - Non-Chapter 74 Career and Technical Education Program Participation
AP1	-	DISCONTINUED
AP2	-	DISCONTINUED

AP3	-	DISCONTINUED
AP4	-	DISCONTINUED
AP5	-	DISCONTINUED
TITLE1	DOE020	Title I Participation
ELP_PROG	DOE026	English Language Learners Program Status
IN_SUSP	-	DISCONTINUED
OUT_SUSP	-	DISCONTINUED
TRUANT	DOE052	Unexcused Student Absences
GENDER	DOE009	Gender
SPEDEVAL	DOE040	Special Education - Evaluation Results
GRAD_CORE	DOE037	Graduate, Completed Massachusetts Core Curr.
ALT_ED	DOE027	Program Code
SPED_3TO5	DOE032	Special Education Placement, ages 3-5
SPED_6TO21	DOE034	Special Education Placement, ages 6-21
T1_CHOICE	DOE028	Title I School Choice Participation
LEP_ARRIVE	DOE021	ELL Students in their First Year in U.S. Schools
PERIOD_NAME	-	Description of Period
ORG_ID	-	First four of School ID (DOE0015) - Numeric
MD_MILITARY_FAMILY	DOE029	Member of Military Family
TITLE_I_NON_INST	DOE030	Non-Instructional Title I Targeted Assistance Services
SLIFE_CODE	DOE041	English Learner Student with Limited or Interrupted Formal Education (SLIFE)
ECODIS	-	Economically Disadvantaged

Appendix 4: Changes in SIMS Data Elements By Year

Updated handbooks are posted to the [SIMS website](#) and the Information Services [research files web page](#) as they are available. Please check these sites for any updates that have been added after this version of the Researcher Guide was compiled.

2003

No changes from 2002. The same handbook covers both 2002 and 2003.

2004

DOE020

Change in title, definition, and codes

NEW

Title 1 Participation (28 possible codes – see handbook)

An indication of the type of Title 1 Services in which a student is enrolled at the specified time of reporting (e.g., October 1).

FORMER

Perkins Low Income Status (2 possible codes – see handbook)

An indication of whether the student meets ANY ONE of the following definitions of low income:

1. The student is eligible for free or reduced price lunch; or
2. The student receives Transitional Aid to Families benefits; or
3. The student is eligible for food stamps

DOE022

Change in title

NEW

Emergency Immigration Education Program Status

FORMER

Immigrant Status

DOE026

Change in title, definition, and codes

NEW

English Language Learners Program Status (5 possible codes – see codebook)

An indication of the type of English Language Learners Program a student is enrolled in at the time of reporting (e.g., October 1).

FORMER

Bilingual Education Program Status (6 possible codes—see handbook)

An Indication of the type of Bilingual Education Program a student is enrolled in at the time of reporting (e.g. October 1).

DOE027–031

Discontinued from the previous year.

DOE032

Change in title, definition, and codes.

NEW

Special Education – Private Placement (3 possible codes – see handbook)

The Enrolled in Public Schools Not Placed or Referred by Public Agency is defined as those students “who have been enrolled by their parents or guardians in regular parochial or other private schools and whose basic education is paid through private resources and who receive special education and related services at public expense from a local educational agency or intermediate educational unit under a service plan.” “A private institution or school is a school NOT under Federal or public supervision or control and may be non-profit or proprietary.”

FORMER

TBE – Exit Type (4 possible codes – see handbook)

An indication of why a student withdrew from a Transitional Bilingual Education program.

NOTE: Complete change between 02–03 handbook and 04. Refers to bilingual education in 02 and to special education in 04.

DOE036-052

New codes that were not in use during the previous year.

2005a (first data collection)

No code changes between 2004 and 2005a.

2005b (mid-year data collection)

DOE032

Title, definition and codes remain the same, but information is added to the “use” section. See handbook for full text.

DOE037

Change in title and to code 02 from 2005a.

NEW

Special Education – Eligibility Determination

FORMER

Special Education – Nature of Services

DOE039, 040, and 041
Discontinued in 2005b.

2006

DOE024
Changes to the “use” section only.

DOE031
Change in definition. See appendix for full text.

DOE044
Change in definition. See appendix for full text.

2007

DOE012
Expand codes to collect more detailed information.

DOE015
Expand codes to collect more detailed information.

DOE021, DOE032, DOE037
Discontinued.

DOE040
New code: Special Education Evaluation Results.

2008

DOE021
New field to capture LEP Student in 1st Year in U.S. Schools. (Unused field – was discontinued variable with 99 for all students)

DOE027
New field to capture Alternative Education (Unused field – was discontinued variable with 500 for all students)

DOE028
New field to capture Title I School Choice Participation (Unused field – was discontinued variable with 500 for all students)

DOE029
Discontinued.

DOE030

Discontinued.

DOE032

New field to capture Special Ed Placement, ages 3–5 (Unused field – was discontinued variable with 500 for all students)

DOE034

Modification — ages 6–21 only

DOE037

New field to capture Graduate, completed Massachusetts Core Curriculum (Unused field – was discontinued variable with 500 for all students)

2009**DOE021**

Clarifies definition and codes for LEP Students in their First Year in U. S. Schools for Pre-K through grade 1 students.

DOE022

Changes definition from identifying students eligible for the Emergency Immigration Education Program to identifying students meeting the legal definition of immigrant.

DOE023

Removes references to the Emergency Immigrant Education Program.

DOE025

Modifies Limited English Proficiency definition.

DOE026

Clarifies English Language Learner Program Status code descriptions.

DOE035

Discontinues several values and modifies explanatory text.

DOE052

Modifies Student Truancy definition.

Appendix C

Adds 16 First Language codes and modify titles for several languages listed.

2010

No changes to SIMS data elements.

2011

No changes to SIMS data elements. Student Course Schedule (SCS) collection added to provide the link between student and educator data.

2012**DOE031**

Four new values added to determine number of students completing the program in non-Chapter 74 programs.

DOE032

New codes added to meet new USED requirements.

DOE039

New field to capture 504 plans. (Unused field – previously discontinued).

2013**DOE045 and DOE046**

Discontinued. Suspension data will be collected in SDDR. Default value=0.

DOE047- DOE051

Discontinued. AP data will be collected solely in SCS. Default value=500.

2014**DOE012**

Changed description of 36 to also include status unknown

DOE027

Changed description to include Innovation and Charter Locations as well Alternative Education Programs

2015**DOE021- DOE025**

Updated language.

DOE029

New field to capture military family (Unused field – previously discontinued).

Appendix D, F, G

Updated.

2016**DOE 012**

Description of value 05 changed from “permanently excluded” to “expelled”

DOE019

Added notes on Low Income Status; Districts not required to enter this data.

DOE025

Added note about Foreign Exchange Students

DOE030

New data element to SIMS: Non-Instructional Title I Targeted Assistance Services

2017**DOE009**

Add non-binary gender option

DOE020

Allows for optional homeless flag in [SIF](#) (School Interoperability Framework, the way the state collects information from most local student information systems) for students receiving targeted assistance in a non-Title I School

DOE041

New field to capture English language learners with limited or interrupted formal education (SLIFE)

2018**DOE024**

Changed to NCES codes for language instead of MA legacy numeric codes.

DOE028

Title 1 School Choice Participation - discontinued

DOE029

Change to values: four values instead of just yes/no

DOE039

New value (02) to differentiate students who were on a 504 plan earlier in the year but not currently

DOE033

New value (10) to identify post-high school plans to start an apprenticeship

2019**DOE026**

New value (05) to identify participants in transitional bilingual education

DOE028

New data element to identify students who have earned the Seal of Biliteracy on their high school diploma

DOE031

Removed values and consolidated to three options: 01 – completed Chapter 74 program; 11 – completed non-Chapter 74 program; 500 – does not apply

DOE045

New data element to identify students enrolled in a High Quality Career Pathway program (either Innovation Pathway or Early College)

DOE046

New data element for industry concentration of students enrolled in a High Quality Career Pathway

DOE047, DOE048, DOE049

New data elements to identify students who have earned specific industry-recognized credentials