



## **Researcher's guide to Massachusetts state education data**

---

May 2016  
Office of Planning and Research

# Table of Contents

I. Introduction .....	2
II. ESE’s Research Agenda .....	3
III. Data Availability .....	4
IV. Requesting State Data .....	9
V. Using SIMS Data .....	12
VI. Using SCS data .....	17
VII. Using MCAS Data.....	20
VIII. Using NSC data.....	26
IX. Appendices .....	28

## **I. Introduction**

The Massachusetts Department of Elementary and Secondary Education (the Department) 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 ESE data, whether on a contract with the ESE 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, Associate Commissioner for Planning, Research and Delivery Systems, is the best place to start. She can be reached at [cconaway@doe.mass.edu](mailto:cconaway@doe.mass.edu).

## **II. ESE's Research Agenda**

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

- Strengthening standards, curriculum, instruction, and assessment
- Developing all educators
- Turning around the lowest performing districts and schools
- Using data and technology to support teaching and learning
- Supporting students' social and emotional health

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

### III. Data Availability

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 ESE 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 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 additional information is available at the statewide level. These reports include the graduation rate, grade retention, and dropout reports.
- The school and district finance website: <http://www.doe.mass.edu/finance/statistics/>. This includes data on per-pupil expenditures for districts across major categories of expenditures, such as administration, classroom teachers, professional development, instructional materials, guidance, etc. Additional financial data are also available in the DART Detail: Staffing and Finance (see below).
- Aggregate school- and district-level files have been created to make student assessment results 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/ease/dart>. These tools turn the Department's data into valuable, easily consumable information. 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 contains data elements that 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.mass.gov/edu/docs/ease/accountability/dart/dart-access-instruction.pdf>. 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 through <http://www.mass.gov/ese/dart>. A reference listing data points contained across DARTs is found at <http://www.mass.gov/edu/docs/ese/accountability/dart/dart-reference-guide.pdf>.

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**

Information on the data series collected by the Department can be found on the Data Collection website: <http://www.doe.mass.edu/infoservices/data/>. Please note: All dates listed below regarding when data are available to researchers are estimates. Circumstances may delay data being shared with researchers.

#### ***Student Information Management System (SIMS)***

This data 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, plans to complete high school, and program participation status, among other information. Linking SIMS data files across years can allow for calculations like grade retention rates. SIMS data are generally submitted three times per year school year: in October, March, and end-of-year.<sup>1</sup> These data usually become available for research purposes in December/January, May/June, and August/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. See Section V or <http://www.doe.mass.edu/infoservices/data/sims/> for more on the SIMS data set.

#### ***Student Course Schedules (SCS)***

In 2010–11, as part of the SIMS expansion, ESE began statewide collection of data on which students are enrolled in which courses and information on their course completion and performance. This allows us to link our student and educator databases through course assignments. SCS is collected in October and at the end of the school year. The October data is generally not shared for research purposes. The end of the 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 the following winter. See Section VI or <http://www.doe.mass.edu/infoservices/data/scs/> for more information on the SCS dataset.

#### ***School Safety Discipline Report (SSDR)***

This report tracks each time an offense occurs on school property. Prior to 2012–13, this report only collected information on drug-, violence- or criminal-related offenses for all students and any 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

---

<sup>1</sup> 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.

was discontinued in 2012–13. SSSR is now the sole source of suspension information.) See <http://www.doe.mass.edu/infoservices/data/ssdr.html> for more information on the SSSR.

### ***Massachusetts Comprehensive Assessment System (MCAS)***

The Massachusetts Comprehensive Assessment System assesses student performance in mathematics, English language arts, and science and technology/engineering. The table in Appendix 1 shows when each subject/grade tests were initially offered, and the table in Appendix 2 shows which tests are available in which years for each graduating class.

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 if possible.

MCAS files from the spring test administration typically become available for research purposes by December 1 of the test administration year. See the MCAS website at <http://www.doe.mass.edu/mcas/> for more information on the testing program in general. The [Test Administration Resources](#), [Test Questions](#), [Results](#), and [Technical Reports](#) links may be particularly helpful.

In 2014, some schools took part in the Partnership for Assessment of Readiness for College and Careers (PARCC) field testing. In 2015, 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 for English language arts and mathematics. (All grade 10 students continued to take the MCAS.) About half of the districts serving grades 3 to 8 administered PARCC for ELA and mathematics. Districts also had a choice for the 2015–16 school year.

ESE is in the process of upgrading MCAS to better measure critical thinking skills needed for success in the 21<sup>st</sup> century and will include innovative items developed by PARCC along with new items specifically created to assess the Massachusetts learning standards. Information on development of the new assessment can be found at <http://www.doe.mass.edu/mcas/nextgen/>.

See Section VII for more information on using MCAS data, including guidance on working with the 2015 MCAS and PARCC data and more information on the 2014 field testing.

### ***MCAS student questionnaire***

As part of the MCAS administration, students in grades 8 and 10 are asked to complete a questionnaire that includes questions about subjects such as post-high school plans, computer use, after-school activities, college preparation activities, hours spent on homework, and other items.

Some questions have been asked consistently for years; others vary from year to year. The questionnaire was revised substantially in 2011 to ask more questions about the school environment (e.g., “Students stay busy, and time is not wasted in class” and “In my classes, teachers encourage me to do my best”), and revised again in 2013 to focus on college and career readiness in various content areas.

Since 2004, the questionnaire has been administered in the eighth and tenth grades; earlier years also

included a questionnaire in the fourth grade. Data from the questionnaires can be linked to SIMS beginning in 2002. Use the search tool on the Department’s website to find the questionnaires for each year; suggested search term: “MCAS student questionnaire [year].”

### ***Massachusetts English Proficiency Assessment (MEPA) / ACCESS for ELLs***

Between 2005 and 2012, the Massachusetts English Proficiency Assessment (MEPA) was administered to all limited English proficiency (LEP) students: 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/>.

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

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

### ***National Student Clearinghouse (NSC)***

The National Student Clearinghouse (NSC) is a non-profit institution that provides enrollment and degree verification from colleges and universities.<sup>2</sup> The Department has a contract with NSC and is limited in the number of student records it can submit every year to be matched to their postsecondary institution information. NSC data is used by the Department to determine college going, persistence and completion of its high school graduates.

See Section VIII for more information on using NSC data and how the Department submits and receives these data.

### ***Other Data Sources***

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. We wait to share the AP and SAT files with researchers until that SASID crosswalk is completed so AP and SAT files are provided with SASID.

## **Non-confidential student-level data**

We make available certain non-identifiable student-level MCAS and student disciplinary 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. Research files are typically posted in December each year; disciplinary data are typically posted in January or February. Please note the change in SDDR data collection practice beginning with the 2012–13 school year noted above.

---

<sup>2</sup> The 3,300 colleges and universities participating in NSC enroll over 96% of all students attending public and private postsecondary institutions nationally.

## **Data on educators**

In the 2007–08 school year, the state launched the Education Personnel Information Management System, or EPIMS, to collect demographic and work assignment information about educators and administrative personnel in schools and districts. More details about this data collection are available at <http://www.doe.mass.edu/infoservices/data/epims/>. Through this data source, we know which teachers are teaching which courses, which can be linked to student data through the Student Course Schedules (SCS) collection. Several statewide data summary reports from EPIMS are available on the teacher tab of the state profile at <http://profiles.doe.mass.edu/>. Most of the data fields in this collection contain public information, but several are confidential and are only available to researchers working on projects where those details are specifically needed to answer the research question.

## IV. Requesting 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, Associate Commissioner for Planning, Research and Delivery Systems, at [cconaway@doe.mass.edu](mailto:cconaway@doe.mass.edu). OPR will determine whether a project will meet our criteria for data-sharing. Appendix 3 of this document lists the criteria ESE uses to assess whether or not to provide data for a particular research project.

Data handbooks for each data series can be found on ESE's [Data Collection](#) website. Older SIMS handbooks are posted on the ESE 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 ESE 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
  - b. Also please state if you specifically need educator's local evaluation ratings. We only give these data out if they are directly and explicitly tied to a research question for the study, as they are subject to a higher confidentiality requirement than our other data.
- 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. The deliverable(s) should include at least one 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
  - 8) Names, titles, email/phone, and organizational affiliations of the following:
    - a. The researchers who will be working on the project
    - b. The person who will serve as the liaison to ESE 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)

### **3) Receive proposal approval or denial from ESE**

Requests are reviewed by OPR to determine whether a proposed research project meets a current ESE information need 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.

The criteria the Department uses to evaluate data requests are provided in Appendix 3. In addition, all proposals are reviewed by the Department's Legal and Education Data Services offices. Given the rigor of the approval process, it can take a significant 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 student-level 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

### 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.

### Clarification of SIMS data issues for researchers

#### *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 13<sup>th</sup> 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.

Two 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. It is important to be aware, however, that 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 Appendix 5 for the Departments process of ordering enrollment status to make a SIMS file unique.

### ***Minimum N***

ESE 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, ESE 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, ESE 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 ESE defines them, a graduate and/or dropout file can be requested.

## **Clarification of SIMS data elements**

### ***Race categories***

Through the 2004–05 school year the Department 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, the Department 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](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.

### ***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, ESE 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 could 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, ESE no longer has systematic, statewide individual-level data on students’ free and reduced price lunch status, and DOE19 is no longer used by ESE for reporting and statistical purposes. Instead, beginning in the 2014–15 school year, ESE 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’ (DCF) 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, however. 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***

ESE calculates a High Needs student variable based on the students 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 Language Learner or former English language learner, or a student with disabilities. A former ELL student is a student not currently an ELL, but had been at some point in the two previous academic years. 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.

### **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

### Clarification of SCS data issues for researchers

#### *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 13<sup>th</sup> administration of the SCS.<sup>3</sup> 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 use 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

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*

ESE uses a minimum N of 6 students for reporting any aggregate student demographic information and minimum of 10 for reporting aggregate student performance information.

#### *Variable Definitions*

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

---

<sup>3</sup> This includes two collections that were done during a pilot year. The first full year of SCS is for the 2010-11 School year.

### ***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/Grades***

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 course are frequent in the elementary grades, and to a lesser extent in middle school grades, particularly 6<sup>th</sup> graders in K–6 schools. For example, in the 2011 SCS data set nearly 20 percent of the 6<sup>th</sup> 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 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. When analyzing Advanced Placement courses, 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 the year SCS file is merged with end of the year SIMS file by SASID or by SASID, Org Code and School Code, depending on the analysis. To merge SCS with EPIMS, the file is matched 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, ESE 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 quinesters/multiple non-consecutive quinesters 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 12<sup>th</sup> graders were erroneously marked as withdrawn (and letter and numeric mark reflecting “Withdrawn”) in the SCS submission. ESE is working to avoid this in the future but be aware of the anomaly if looking at 12<sup>th</sup> grade course completion or performance rates.
- The end-of-year SCS data collection for the 2010–11 school year (SCS11D file) does not have data for Lowell and Greenfield. Those two districts did not submit data.

**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 for additional details)
  - With limited English proficiency (see note below for additional details)

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/>).

Limited English proficiency (LEP) 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, LEP students have been considered first-year students if they first enrolled in a U.S. school later than March 1 of the prior school year.

Additional information about the participation in MCAS of students with disabilities and LEP students can be found online at <http://www.doe.mass.edu/mcas/participation/?section=gen>.

### **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>.

### **Clarification of MCAS data issues for researchers**

#### ***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.

### ***Minimum N***

ESE 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, 20 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* *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 or 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. There are several reasons why this might occur. Limited English proficient students 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 sick 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 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.

### ***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*, available online at [http://www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport\\_2007.htm](http://www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport_2007.htm).

### ***Using scaled and raw scores***

Researchers comparing MCAS scores across years need to be aware of several important caveats. Raw scores represent the number of raw score points earned on the test; the difficulty of items varies from test to test. Scaled scores rescale the raw scores onto a common metric from 200 to 280 on two-point intervals. 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. However, the raw scores associated with each scaled score vary across tests and years. 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, 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 estimates of standard errors calculated from scaled scores may be incorrect. 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](http://www.mcasservicecenter.com/documents/MA/Technical%20Report/TechReport_2007.htm).

### **Transition to next generation assessment**

In 2014, Massachusetts began a two-year “test-drive” of the PARCC test. In spring 2014, approximately 700 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, 2.3 percent of students tested in those grades.

For the second year of the test drive, in 2015, 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. The following section provides in-depth information on that testing administration and how to use work with that data.

In Fall 2015, the Department and Board of Elementary and Secondary Education decided to upgrade MCAS, building upon the best aspects of MCAS and including innovative items developed by PARCC along with new items specifically designed to assess the Massachusetts learning standards.

Because of the time required to contract with a vendor for the new assessment, 2016 will be a transitional year. Districts that administered PARCC tests in ELA and math in spring 2015 will administer PARCC again and will continue to have the option to select the computer-based or paper-based version. Districts that administered MCAS tests in ELA and math in spring 2015 will administer MCAS again unless the district elected to switch to PARCC.

In 2017, the state will administer the next generation MCAS to all districts, with some students taking the test on a computer and others using pencil and paper. Our goal is for all students to take the assessments on a computer by 2019.

## **Working with 2015 Massachusetts Assessment Data**

In 2015, 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 that year 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 408 school districts, 46 percent administered MCAS and 54 percent PARCC in 2015; see details posted at <http://www.doe.mass.edu/news/news.aspx?id=13541>. Almost all districts had 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. 31 percent of districts administered the test entirely on paper, 50 percent entirely online, and 19 percent a mix of the two.

All grade 10 students continued to take the MCAS, as it is currently the test used to meet the state's high school graduation requirement. 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 it is possible that the districts that selected MCAS are systematically different than those that selected PARCC. Indeed, the MCAS and PARCC takers in 2015 were similar but not equivalent in terms of prior performance and student demographics. Therefore, we took 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. Additional information on the methodology is published in a [white paper](#) issued by the Office of Student Assessment Services.

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.

### **Combining assessment data**

Researchers may also wish to combine data from MCAS and PARCC districts for the purpose of analyzing the impact of programs across districts or schools participating in both assessments. 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.

For MCAS, as noted above, we always recommend running analyses either on the raw scores or a z-score transformation of the raw scores. The scaled scores constitute ordinal data and are heteroskedastic (different variances at different parts of the scales) so are not appropriately used for calculating means or conducting significance testing. The PARCC scaled score data do not suffer from this problem so can be used in significance testing. However, combining results across the two assessments requires some transformation.

We recommend that researchers work with the student-level thetas for each assessment, provided in the assessment data files, and transform them to z-scores. The z-scores can then appropriately be combined across assessments and used to measure student achievement results in standard deviation units. Researchers should consider the appropriate population against which to compute z-scores. For those interested in the statewide population, we have provided lookup tables of thetas to z-scores for the statewide representative sample data at <http://www.doe.mass.edu/infoservices/research/><sup>4</sup>. We will add these transformations to the student-level data in our data files shared with researchers (pending internal staff availability). However, if a research design pertains to a particular group of districts, schools, or subgroup of students, researchers may wish to calculate their own z-scores just for that subgroup. Statewide de-identified student-level data that allows calculations for districts and schools or major subgroups is available at <http://www.doe.mass.edu/infoservices/research/>.

### **PARCC mode effects**

In our first year of online test administration, 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, and we will continue to study it carefully. We recommend that researchers using PARCC 2015 data consider

---

<sup>4</sup> Standardized scores were adjusted for grade 8 students taking the PARCC Algebra I exam through the use of a matching analysis (coarsened exact matching comparing students taking Algebra I in grade 8 to those taking MCAS in grade 8 in 2015) and a multiple regression analysis. This adjustment provided a better comparison to the matched students taking the grade 8 math test in MCAS in 2015.

test mode in their analyses and conduct robustness checks to ensure that their findings do not differ based on test mode.

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

- School and district-level MCAS scores from 2003 onward are posted on the Department's website at [http://profiles.doe.mass.edu/state\\_report/mcas.aspx](http://profiles.doe.mass.edu/state_report/mcas.aspx).
- MCAS Technical documents, including technical reports, validity studies guidance on MCAS 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](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 hyper link for the Excel Workbook with an Excel icon placed just to the right of the hyper link. This file provides a list of variables, their meanings, and the possible response codes in effect for that year.

## **VIII. 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.<sup>5</sup> The Department has a contract with NSC and is limited in the number of student records it can submit every year. All records that the Department submits to NSC are tied to the unique 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.<sup>6</sup> 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 mega-file containing all Massachusetts graduates from 2003 to 2010 to NSC. Currently, the Department plans to annually submit an updated mega-file of graduates in late winter/early spring which will be available to researchers in late spring/early summer. As Massachusetts graduates are added to the mega-file, earlier graduating classes are left off. The most recent submission of graduates from spring 2016 includes Massachusetts graduates from 2008 through 2015. The Department does not generally share 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 spring 2013 and spring 2015, the Department submitted a file of students who were ever enrolled from school year 2003 to date in grades 8 through 12 or special education beyond 12<sup>th</sup> 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 restraints, there is no current plan to resubmit a non graduate file in the future.

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. In previous years, 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. The reason for changing the SEARCHDATE was to capture a student’s college enrollment during their high school years (e.g., dual enrollment).

---

<sup>5</sup> The 3,300 colleges and universities participating in NSC enroll over 96% of all students attending public and private postsecondary institutions nationally.

<sup>6</sup> 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 ESE 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 ESE 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](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/Default.aspx?y=55>.
- 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](#) other resources on the NSC's [Notes from the Field](#) section on the Working with Our Data tab.

## IX. Appendices

### Appendix 1: MCAS Test Data Availability by Subject and Grade

(For years in which MCAS can be linked to SIMS)

Discipline	2002	2003	2004	2005	2006	2007 to present <sup>7</sup>
<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	✓	✓	✓	✓	✓	✓

<sup>7</sup> In the 2013-14 school year, some grades/schools took part in the PARCC field test. For the 2014-15 school year, about half of the districts serving grades 3-8 administered PARCC for ELA and Math. See section VII for guidance on using 2015 assessment data.

## Appendix 2: MCAS Test Data Availability by Class Year

(For years in which MCAS can be linked to SIMS)

		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015 <sup>8</sup>
<i>Class of 2004</i>	<i>Grade</i>	10	11	12											
	English Lang. Arts	✓													
	Composition	✓													
	Mathematics	✓													
	Science														
<i>Class of 2005</i>	<i>Grade</i>	9	10	11	12										
	English Lang. Arts		✓												
	Composition		✓												
	Mathematics		✓												
	Science														
<i>Class of 2006</i>	<i>Grade</i>	8	9	10	11	12									
	English Lang. Arts			✓											
	Composition			✓											
	Mathematics	✓		✓											
	Science	✓													
<i>Class of 2007</i>	<i>Grade</i>	7	8	9	10	11	12								
	English Lang. Arts	✓			✓										
	Composition	✓			✓										
	Mathematics		✓		✓										
	Science		✓												
<i>Class of 2008</i>	<i>Grade</i>	6	7	8	9	10	11	12							
	English Lang. Arts		✓			✓									
	Composition		✓			✓									
	Mathematics	✓		✓		✓									
	Science			✓											

<sup>8</sup> For the 2014-15 school year, about half of the districts serving grades 3-8 administered PARCC for ELA and Math. See section VII for guidance on using 2015 assessment data

		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015 <sup>8</sup>
<i>Class of 2009</i>	<i>Grade</i>	5	6	7	8	9	10	11	12						
	English Lang. Arts			✓			✓								
	Composition			✓			✓								
	Mathematics		✓		✓		✓								
	Science	✓			✓		✓								
<i>Class of 2010</i>	<i>Grade</i>	4	5	6	7	8	9	10	11	12					
	English Lang. Arts	✓			✓	✓		✓							
	Composition	✓			✓			✓							
	Mathematics	✓		✓		✓		✓							
	Science		✓			✓		✓							
<i>Class of 2011</i>	<i>Grade</i>	3	4	5	6	7	8	9	10	11	12				
	English Lang. Arts	✓	✓			✓	✓		✓						
	Composition		✓			✓			✓						
	Mathematics		✓		✓	✓	✓		✓						
	Science			✓			✓		✓						
<i>Class of 2012</i>	<i>Grade</i>	2	3	4	5	6	7	8	9	10	11	12			
	English Lang. Arts		✓	✓		✓	✓	✓		✓					
	Composition			✓			✓			✓					
	Mathematics			✓		✓	✓	✓		✓					
	Science				✓			✓		✓					
<i>Class of 2013</i>	<i>Grade</i>	1	2	3	4	5	6	7	8	9	10	11	12		
	English Lang. Arts			✓	✓	✓	✓	✓	✓		✓				
	Composition				✓			✓			✓				
	Mathematics				✓	✓	✓	✓	✓		✓				
	Science					✓			✓		✓				
<i>Class of 2014</i>	<i>Grade</i>	<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	<i>12</i>	
	English Lang. Arts				✓	✓	✓	✓	✓	✓		✓			
	Composition					✓			✓			✓			
	Mathematics					✓	✓	✓	✓	✓		✓			
	Science						✓			✓		✓			

		2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015 <sup>8</sup>
<i>Class of 2015</i>	<i>Grade</i>		<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>11</i>	
	English Lang. Arts					✓	✓	✓	✓	✓	✓		✓		
	Composition						✓			✓			✓		
	Mathematics					✓	✓	✓	✓	✓	✓		✓		
	Science							✓			✓		✓		
<i>Class of 2016</i>	<i>Grade</i>		<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>		
	English Lang. Arts					✓	✓	✓	✓	✓	✓			✓	
	Composition						✓			✓				✓	
	Mathematics					✓	✓	✓	✓	✓	✓			✓	
	Science								✓			✓			✓
<i>Class of 2017</i>	<i>Grade</i>				<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>		
	English Lang. Arts						✓	✓	✓	✓	✓	✓			
	Composition								✓			✓			
	Mathematics							✓	✓	✓	✓	✓	✓		
	Science									✓			✓		
<i>Class of 2018</i>	<i>Grade</i>					<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	
	English Lang. Arts								✓	✓	✓	✓	✓	✓	
	Composition									✓			✓		
	Mathematics								✓	✓	✓	✓	✓	✓	
	Science										✓				✓
<i>Class of 2019</i>	<i>Grade</i>						<i>K</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>
	English Lang. Arts									✓	✓	✓	✓	✓	✓
	Composition										✓			✓	
	Mathematics									✓	✓	✓	✓	✓	✓
	Science											✓			✓

### Appendix 3: Confidential Student Data Request Review Criteria

<b>Required criteria</b>					
The researcher must be affiliated with an academic institution or independent, nonpartisan research organization.	No	Yes			
The researcher (and any institutions with which he or she is affiliated, including third-party funders of the proposed research project) must not be in an advocacy role with respect to the topic of the study or have a stake in the study's outcome.	No	Yes			
The study design and methodology must support an objective analysis of the proposed research question(s).	No	Yes			
The study must propose to study a topic relevant for policy decision-making.	No	Yes			
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.	No	Yes			
If a request 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.	n/a	No	Yes		
If the request includes data from the MEPA or ACCESS assessment, the research question relates specifically to English language learners and requires more detail on their performance than what is available in other state data.	n/a	No	Yes		
<i>Comments:</i>					
<i>Evidence that the proposal can fulfill this criterion is:</i>	<i>Very weak or absent</i>	<i>Weak</i>	<i>Moderate</i>	<i>Strong</i>	<i>Very strong</i>
<b>Study design and methodology</b>					
The proposal includes specific, well thought out research questions that relate to a topic relevant for policy decision-making.	0	1	2	3	4
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.	0	1	2	3	4
The research questions can be answered effectively with data the Department has available.	0	1	2	3	4
<i>Comments:</i>					

<i>Evidence that the proposal can fulfill this criterion is:</i>	<i>Very weak or absent</i>	<i>Weak</i>	<i>Moderate</i>	<i>Strong</i>	<i>Very strong</i>
<b>Objectivity and credibility</b>					
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.	0	1	2	3	4
Previous work published by the researcher (and/or faculty advisor) suggests that he or she has credibility with other researchers and experts.	0	1	2	3	4
<i>Comments:</i>					
<b>Benefit and relevance to ESE policy and programming</b>					
The Department has the capacity to fulfill the request at this time.	0	1	2	3	4
Gathering more information on the proposed topic is a high Department priority.	0	1	2	3	4
The information gathered will be relevant either statewide or for high-priority districts or schools.	0	1	2	3	4
The study is likely to provide information in time to be of use for anticipated decision-making.	0	1	2	3	4
The study does not overlap undesirably with other work already in progress.	0	1	2	3	4
The authors agree to produce both a technical and non-technical version of the report for ESE staff.	0	1	2	3	4
<i>Comments:</i>					

Proposals must meet all required criteria and score *strong* or *very strong* on most criteria in order to be approved. Note that meeting all criteria listed 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 4: Changes in SIMS Data Elements, By Year

Updated handbooks are posted to the ESE web site as they are available. Please check the web site for any updates that have been added after this version of the Researcher Guide was compiled.

### **2003**

---

*No changes from 2002. There is one handbook for 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**  
*Discontinued.*

**DOE032**  
*Discontinued.*

**DOE037**  
*Discontinued.*

**DOE040**  
*New code: Special Education Evaluation Results.*

---

**2008**

**DOE021**  
*New field to capture LEP Student in 1<sup>st</sup> 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 SDR. 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 12**

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

**DOE19**

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

**DOE25**

*Added note about Foreign Exchange Students*

**DOE30**

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

## Appendix 5: 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)).
```