



Consumer Guide to

MLDS Center

Data and Reporting

(Revised June 2017)



Statement of Purpose

The purpose of this consumer guide is to provide information about the Center’s data and reporting. The first section describes the data, where they come from, and considerations for using and interpreting the data. The second section discusses the complexities of longitudinal data and the complexities of examining education and workforce outcomes. Finally, the third section details how the Center uses the data for reporting.

Contents

Statement of Purpose	2
The Maryland Longitudinal Data System (MLDS) Data.....	3
What Data are in the MLDS and Where are the Data From?	3
What are the Limitations of MLDS Data?	4
How is Confidentiality of the MLDS Data Maintained?	5
Complexities of the Data.....	5
What is Longitudinal Data?.....	5
What if Data are Missing Over Time?	6
How are the Data Structured?	6
What are the Complexities of Examining Education and Workforce Outcomes?	8
How is Student Group Membership Related to Outcomes?	9
Use of the Data for Reporting.....	9
What is the Goal of MLDS Center Reporting?.....	9
What are the Data Reporting Standards?.....	9
How does the MLDS Center Build High Quality Data Sets?	10
How are the Results of the MLDS Center Verified?.....	10
What Types of Relationships are Examined?.....	11
How do I get More Information?	12

The Maryland Longitudinal Data System (MLDS) Data

What Data are in the MLDS and Where are the Data From?

The MLDS Center is an independent unit of the Maryland state government. The Center collects and organizes statewide student and workforce data. The Center partners with other state agencies to receive data on Maryland students and employees. Specifically, the Center receives data from the Maryland State Department of Education (MSDE), the Maryland Higher Education Commission (MHEC), and the Maryland Department of Labor, Licensing, and Regulation (DLLR). By receiving data from these agencies, the MLDS Center is able to study topics that combine the interests of the individual agencies and provide a more detailed picture of issues related to Maryland's educational systems and workforce training policies and programs over time.

The Center is governed by a twelve-member [Governing Board](#), which includes seven members that are designated by statute, such as the State Superintendent of Schools and the State Secretary of Education, as well as five members appointed by the Governor. The MLDS Governing Board guides the reporting of the Center and has established a [Research Agenda](#) for the Center. There are three branches of the Center: (1) the Reporting and Information Services Branch; (2) the System Management Branch; and (3) the Research Branch, which includes faculty and graduate students from the [University of Maryland School of Social Work](#) and [College of Education](#). Each of the partner state agencies has a liaison to the Center, whose role is to help guide the data and reporting to be most relevant to our partner agencies.

The MLDS contains data on Maryland public high school students (academic characteristics, school characteristics), Maryland college students (academic characteristics, degree information), and Maryland workers (wages, industry worked) who attended a Maryland high school or college. Additionally, the MLDS contains data on out-of-state college enrollments for students who attended a Maryland public high school in 12th grade (data are from the National Student Clearinghouse; NSC).

What are the Limitations of MLDS Data?

The MLDS Center receives and links data from multiple state agencies over time, which creates a number of limitations. These limitations include:

- There is a potential for data mismatch between agencies. If records do not match, the individual may need to be excluded from the particular study.
- Data from state agencies only include individuals in Maryland. If individuals move in or out of Maryland, their long-term records may be incomplete. In particular, any data on employment while an individual is out of state is missing, so those individuals may need to be excluded from certain studies.
- Data for out-of-state students are not as complete as in-state records. The Center receives some information on Maryland students who attend out-of-state postsecondary institutions, but those records are not as thorough as those for in-state students. Individuals with missing data may need to be excluded from a particular study.
- K-12 data from MSDE only include public school students within Maryland, so students who attend private schools are not accounted for. Studies conducted by the Center will only be able to include public school students.
- Data from DLLR exclude certain groups, such as self-employed, religious, or federal employees. These exclusions may limit the scope or generalizability for a study involving employment. The exclusion of federal employees in particular is important because Maryland has one of the highest percentages of federal workers in the United States (according to the US Office of Personnel Management; see <https://www.opm.gov/>)

Such limitations mean that in certain cases, individuals or groups may need to be excluded from a particular study due to a lack of complete information. Alternatively, other methods may be applied to account for missing data (see section on [What if Data are Missing Over Time?](#)). The Center has set standards for research quality (see section on [What are the Data Reporting Standards?](#)) to address these limitations and challenges.

How is Confidentiality of the MLDS Data Maintained?

Maintaining confidentiality of the data is an important part of the Center's data management responsibilities. When data are received from the partner agencies, the data include personally identifiable information (PII) on students and employees. PII can include name, address, date of birth, identification numbers, and demographic data. The Center only uses PII to link individuals across different data sources. Once the data are linked, the data are de-identified by removing the PII and replacing it with a randomly assigned token ID number. All Center reporting uses de-identified aggregate data, and if there is any possibility of the reported data being used to identify individuals, either on its own or in combination with publicly available data, the data will be suppressed to protect confidentiality.

These policies and procedures are in response to state and federal data privacy laws, including the Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. §1232g, and its corresponding regulations, 34 C.F.R. §99.1 et seq., the Federal State Unemployment Compensation (UC) Program, 20 C.F.R. Part 603, Md. Code, Labor & Employment Article §8-625, and the Maryland Public Information Act, Md. Code, General Provisions Article, §4-101, et seq. Additionally, the Center's use of the data provided by the Center's agency partners (Maryland State Department of Education, Maryland Higher Education Commission, and Department of Labor, Licensing and Regulation) is governed by the MLDS Center's authorizing statute, Md. Code, Education Article §24-701, et seq.

Complexities of the Data

What is Longitudinal Data?

Longitudinal data include information collected on the same individuals over multiple points in time. For the longitudinal data in the MLDS, information is collected on individuals in the Maryland educational system or workforce on a quarterly or yearly basis, depending on the reporting institution and the data point. The longitudinal data currently used by the Center ranges from the 2007-2008 academic year to present. The Center continually adds data as they become available.

The longitudinal research at the MLDS Center is guided by the [Research Agenda](#), which is divided into four sections aimed at examining key transitions in education and the workforce.

The four sections of the [Research Agenda](#) are:

1. K-12 Readiness
2. Postsecondary Readiness and Access
3. Postsecondary Completion
4. Workforce Outcomes

Over time, individuals can progress through transitions as they attend school, graduate, then move to postsecondary and/or the workforce. By using longitudinal data, we can examine patterns that emerge in groups of people over time and identify factors that may influence those groups, a strength of longitudinal data.

What if Data are Missing Over Time?

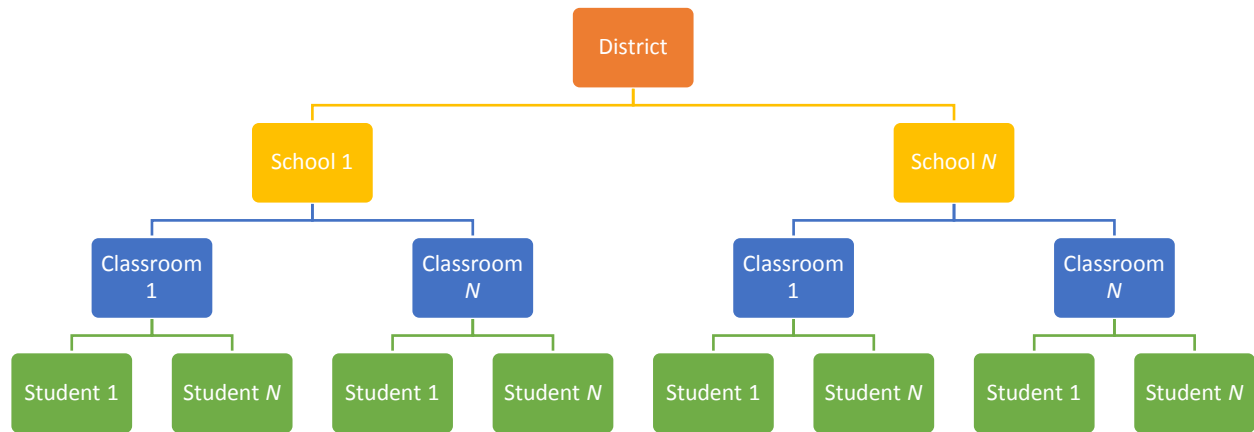
One issue that can occur in longitudinal data sets is missing data. Missing data may occur when individuals move into or out of Maryland, or the data are simply not reported to the collecting partner agency. It is important to account for missing data over time to maintain a data set that is as complete as possible for accurate reporting. This requires the use of complex statistical models.

How are the Data Structured?

The MLDS data are multilevel (clustered) data. Students are clustered within classrooms, classrooms are clustered within schools, and schools are clustered within districts (see Figure 1). Because of the nature of clustering and shared experiences within a cluster, individuals within one cluster are likely to be more similar to one another than they are to individuals within another cluster. For example, students within one school will likely have more in common with each other than with students within another school because of shared experiences like curriculum, teaching, and administration in the school.

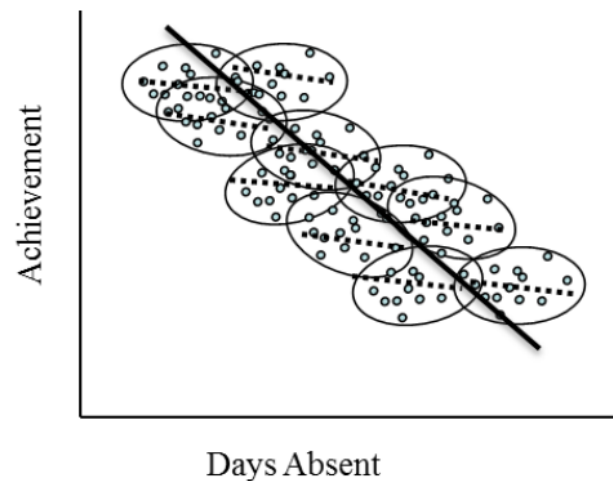
Figure 1: Illustration of Clustered Data

In this illustration, students are clustered within classrooms, classrooms are clustered within schools, and schools are clustered within school districts.



When using clustered data, it is important to use statistical techniques that account for the nature of the clustering. These techniques adjust for the similarities that result from data that are clustered and make the reporting more trustworthy than reporting that ignores the clustered nature of the data. In the example in Figure 2, individual student achievement is related to the number of days absent. Each dot represents a student and each circle, or cluster of students, represents a school. The dotted lines are the trend lines for each school while the solid line is the overall trend line without taking clustering into account. If clusters are not considered in this example, the relationship between achievement and days absent would likely be overestimated. Accounting for the clustering will make the findings about the relationship more accurate and trustworthy.

Figure 2: Illustration of Results from Multilevel Data with Students Clustered in Classrooms
In this illustration, each cluster of students (dots) has a trend line which differs from the overall trend. If only the overall trend is considered, the relationship between absences and achievement may be inaccurately estimated.



What are the Complexities of Examining Education and Workforce Outcomes?

Many of the studies conducted by the Center examine the multiple factors that influence outcomes for individuals, either in education or the workforce. The specific definition of “outcome” varies according to the research question guiding the study. For example, an outcome might be college enrollment for one study, whereas another study might examine whether college enrollment is associated with the outcome of full-time employment.

A seemingly simple study of outcomes becomes more complex when the variety of potential influences on any given outcome is considered. For example, student college enrollment and employment may be influenced by school-related factors like test scores and teacher characteristics; home-related factors like parent employment and education; and personal factors such as special education status and motivation.

How is Student Group Membership Related to Outcomes?

The MLDS Center [Research Agenda](#) specifies that it is a core part of the Center’s mission to examine how education and workforce findings vary by different critical subgroups and backgrounds, including race or ethnicity, gender, socioeconomic status, language, ability, and setting. At times these groupings can overlap or be strong predictors of membership in another group. A student in a lower income home, for example, likely attends a lower resourced school, may have less experienced teachers, and so on. Historical factors, such as prior discrimination against certain demographic groups in education services or access, may also have lasting effects on education outcomes. Additionally, workforce and employee outcomes are influenced by employee education and training. Employees can be grouped in any number of ways, such as by demographic characteristics, training, income, or industry. It is important to use statistical methods that can account for group differences and interpret the findings of any analysis within the context of those groupings.

Use of the Data for Reporting

What is the Goal of MLDS Center Reporting?

The MLDS Center reporting will inform policy makers and practitioners on the topics outlined in the [Research Agenda](#) established by the MLDS Governing Board. These topics require longitudinal data analyses to answer questions about students and employees in Maryland over time. The reporting is complex given the nature of workforce and education data (see section on [What are the Complexities of Examining Education and Workforce Outcomes?](#)) and vary according to the specific needs of the research question and the data available to answer the question.

What are the Data Reporting Standards?

In response to the complexity of the MLDS data and reporting, the Center has developed a set of data reporting standards in order to ensure that data sets, methods, and reporting are of high quality and use valid and trustworthy data. These standards are a step-by-

step process, where each step builds on the previous and the steps can be individually evaluated and refined. The seven steps are:

1. Identify the question
2. Determine available data
3. Refine the question and determine the population
4. Determine value-added benefit
5. Sufficiency of identity matching
6. Assess robustness of data elements
7. Verify results

In-depth descriptions of each step can be found in the [MLDS Data Reporting Standards](#).

How does the MLDS Center Build High Quality Data Sets?

The MLDS Center works to gather and integrate data from multiple state agencies to build high quality data sets. To do this accurately, it is important to verify that the individuals can be matched across agencies (see section on [What is the Nature of Longitudinal Data?](#)). The Center evaluates and reports the percentage of submitted records that can be successfully matched (contain valid data elements in the necessary fields) when data are received from a partner agency. The MLDS Center also examines the data it receives from partner agencies to assist them in identifying and improving potential data quality issues, including data reporting inconsistencies and missing/invalid data that occurs at a high rate or with a consistent pattern. The MLDS Center continuously works to improve its matching and data review procedures.

How are the Results of the MLDS Center Verified?

The MLDS Center reporting is conducted within the context of past and current research on education and workforce outcomes. All findings are evaluated to determine whether they are consistent with other reporting from our partner agencies and from national sources, such as books and journal articles. The MLDS Center aims to verify its findings and interpret them in the context of current research.

What Types of Relationships are Examined?

The MLDS data are observational data, which is the type of data typically used to examine associations rather than causal relationships. An association is present when two variables are related in a way that is unlikely to occur by chance. In an association, two variables may change together, but it cannot be concluded that a change in one caused the change in the other. This is different than determining causation, when a change in one variable can be shown to cause a change in another.

In policy research, the research question often asks “What is the effect?” For example, “what is the effect of taking STEM-related courses in 9th grade on eventual earnings?” Because the MLDS data are observational, this question could be answered in terms of whether there is an association between STEM-related course taking and eventual earnings, rather than whether STEM-related course taking causes greater earnings. If an association is found, it is possible that there are additional variables influencing both course taking and earnings, such as grades, parent education, school resources, and so on.

Some research questions can be more thoroughly answered if causal inferences can be made. The MLDS Center will use advanced statistical methods to strengthen the ability to draw causal inferences about education and workforce outcomes. Typically, causation can be determined through a randomized experiment, where individuals are randomly assigned to treatment and control groups, the treatment is applied to the treatment group, and the effect is measured and compared to the control group. If a difference is found between the two groups, it can be attributed to the treatment because the two groups were generated randomly and are considered to be equivalent. The MLDS data are observational, so the researcher cannot randomly assign treatment and control groups. However, to determine causation, the researcher can approximate random assignment using a quasi-experimental design. These methods mimic randomization well enough to allow causal inferences to be made. These methods will be explored by the MLDS Center to more fully address research questions about education and workforce outcomes.

How do I get More Information?

For more information, to view research reports and dashboards, request data, or contact the MLDS Center, please visit <https://mldscenter.maryland.gov/welcome-index.html>.