

Consumer Guide to

MLDS Center

Data, Dashboards, and Research Reports

(April, 2015)



The purpose of this consumer guide is to provide more information about the Center’s research to help consumer’s better understand our data, dashboards, and research reports. This guide includes information about the structure of the Center’s data, the complexity of the data, and guidelines for interpreting our data. We outline the challenges associated with using the Center’s data and provide information about how we will handle those challenges. Throughout our research reporting, in dashboards and in research reports, we will inform the audience of how we are handling these challenges, including the statistical techniques used and the benefits and limitations of those techniques.

Table of Contents

[The Center Data](#)

[Where Does the Data Come From?](#)

[The Nature of Longitudinal Data](#)

[The Natural Multilevel Structure of Education and Workforce Data](#)

[Figure 1: Illustration of Clustered Data](#)

[Figure 2: Illustration of Multilevel Model Results](#)

[Limitations of MLDS Center Data](#)

[Confidentiality of MLDS Center Data](#)

[The Complexity of Examining Education and Workforce Outcomes](#)

[Figure 3: Bronfenbrenner’s Ecological Systems Theory](#)

[Contexts that Influence Student Outcomes](#)

[Contexts that Influence Employee Outcomes](#)

[Student Group Differences](#)

[Handling Complexity](#)

[Associations versus Causal Relationships](#)

[Research Quality Assurance Standards](#)

[Building High Quality Data Sets](#)

[Triangulating Data and Evidence](#)

[MLDS Center Research](#)

The Center Data

Where Does the Data Come From?

Although the MLDS Center is an independent unit of State government, the Center partners with other State agencies that provide the data used for the Center's website and studies. The Center receives data from the [Maryland State Department of Education](#), the [Maryland Higher Education Commission](#), and the [Maryland Department of Labor, Licensing, and Regulation](#). The partnerships of these agencies with the Center allow the Center to research issues that combine the interests of the individual agencies as well as all Marylanders, providing a more detailed picture of issues related to Maryland's education and workforce training policies and programs.

The Center is also authorized to import data from third party (non-agency) sources, such as the [National Student Clearinghouse](#) (NSC ;). The NSC provides information about some students that attend college in other states, but only when the out of state colleges that these students attend submit enrollment information on these students to the NSC. The NSC only provides data for the list of students that the Maryland State Department of Education (MSDE) submits, under its contract, to the NSC. For students not on that list, no new data will be discovered under this process. When the MSDE receives the data from the NSC, they share this information with the Center. This data is particularly helpful in determining which out of state colleges Maryland high school graduates attend and graduate from.

The Nature of Longitudinal Data

Longitudinal data means information on individuals collected and analyzed over multiple points in time. The longitudinal data currently used by the Center contains present data, as well as data collected since 2008. The Center will add data over time which will strengthen the overall research and reporting. It also allows the Center to answer questions about how groups of students change, and identify the influencing factors affecting that change. Directly examining change in the same groups of individuals over time is a major strength of longitudinal data, in comparison to data collected on groups of individuals at only one point in time (called cross-sectional data).

Longitudinal data are information about individuals as they progress through the education pipeline from Early Childhood into the Workforce. This journey is separated into the following four sectors:

1. Early Childhood, which includes children from the age of 3 up until kindergarten
2. K-12, which include kindergarten through the 12th grade of high school
3. Postsecondary, which includes 2 year colleges, 4 year institutions, and career and technical educations
4. Workforce, which includes worker employment, wages and periods of unemployment

Longitudinal data can be more difficult to collect and maintain because individuals may be missing from the data over time. For example, if a student attends middle school in Maryland and moves to Virginia for high school, the Center would have data for the student's middle school years, but not for the student's high school years. This means that the analysis of longitudinal data requires the use of complex statistical models to account for the collection of data on the same individuals over time and the issue of having missing data on some of those individuals over time.

The Natural Multilevel Structure of Education and Workforce Data

Multilevel data means data that exist at multiple levels within a hierarchy (see Figure 1). Individuals within the data set are grouped by some aspect of their experience, and there are typically multiple groups (called clusters in statistical analyses). For example, in the MLDS, students may be clustered by classrooms, schools, school districts, or universities; employees may be clustered by places of employment or job type; and individuals and organizations are clustered by county or zip code. When clustered data are used, the analysis must take the clustering into account. For example, students' experiences within the same school are likely to be more similar to one another than they are to students' experiences in a different school. This may be because of a shared curriculum, having teachers who share lesson plans, having school principals that operate under a single superintendent's vision, etc. Advanced statistical techniques are necessary to adjust for the similarities that result from data that are clustered

(see Figure 2). These advanced statistical techniques make our analyses more trustworthy than analyses that ignore the clustered nature of the data.

Figure 1: Illustration of Clustered Data

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

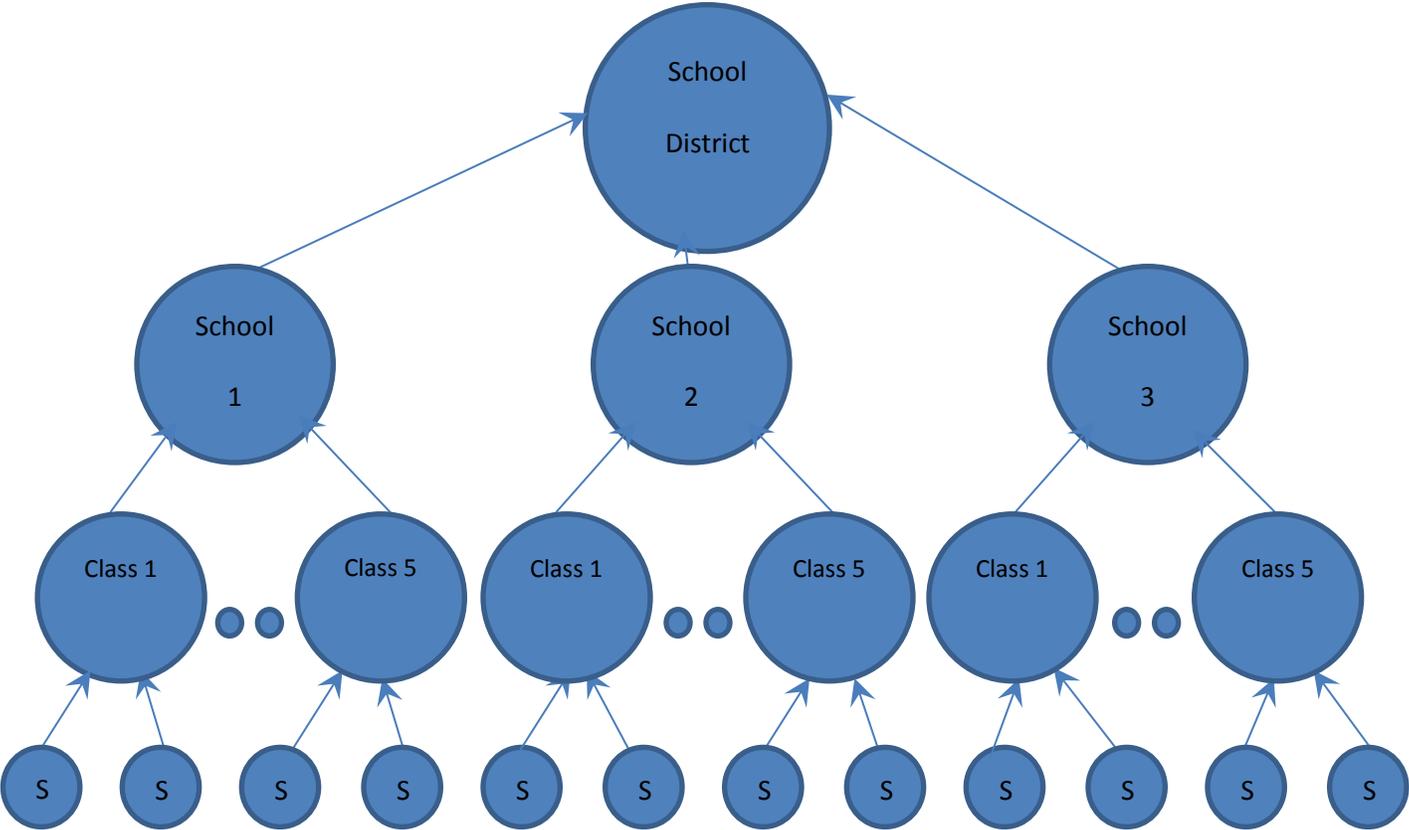
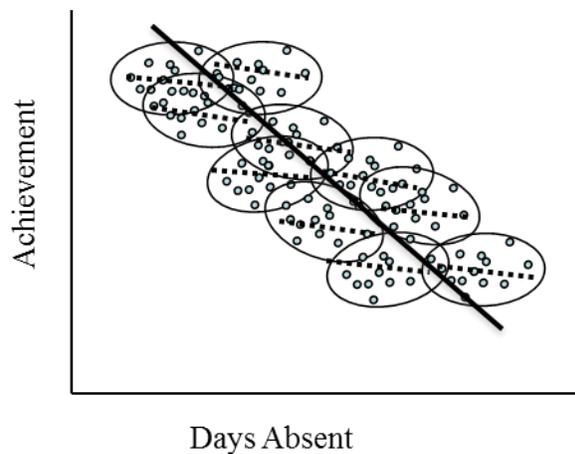


Figure 2: Illustration of Multilevel Model Results



In this illustration achievement is related to number of days absent. The overall trend, not accounting for the clustering (bold black line), indicates that achievement decreases as number of days absent increases. Each dot represents a student and each circled cluster represents a school. The dotted lines indicate the trend line for each school, and the trend line can differ for each school. Without accounting for the clustering, we may overestimate the trend for the relationship between achievement and number of days absent.

Limitations of MLDS Center Data

The data at the MLDS Center are unique because they come from multiple state agencies and because they are longitudinal. This structure creates several challenges and limitations. First, the individual-level data must be matched across agencies and over an extended period of time. The Center's research involves studying data from more than one of the partner agencies. If an individual-level data record from one agency does not match another from another agency, the researchers may exclude that data in a particular study. It is important to remember that the data the Center receives from the Maryland State Department of Education and from the Department of Labor, Licensing, and Regulation are specific to the state of Maryland. Information from other states is not included. For example, if a student attends middle school in Maryland and moves to Virginia for high school, the Center would have data for the student's middle school years, but not for the student's high school years. The

Center receives data for out-of-state postsecondary enrollments from the National Student Clearinghouse (NSC) through MSDE as detailed above, but certain institutions are not included (for example, military academies). According to the NSC website, more than 3,500 colleges and universities, enrolling over 98% of all students in public and private U.S. institutions, participate in the NSC. The PK-12 education data used by the Center only includes public school students within the State of Maryland, so K-12 students who attend private schools are not included in Center data. The Center's workforce data does not include certain groups of employees, such as self-employed individuals and employees of religious organizations. It also excludes data for employees of the federal government (for example Department of Defense employees), which is especially important to note given Maryland's close proximity to the District of Columbia and because Maryland has the second highest percentage of such employees among the states. To address these limitations, the Center has standards that it follows to ensure research quality (see section on [Research Quality Assurance Standards](#)).

Confidentiality of MLDS Center Data

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.*, as well as additional 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.*

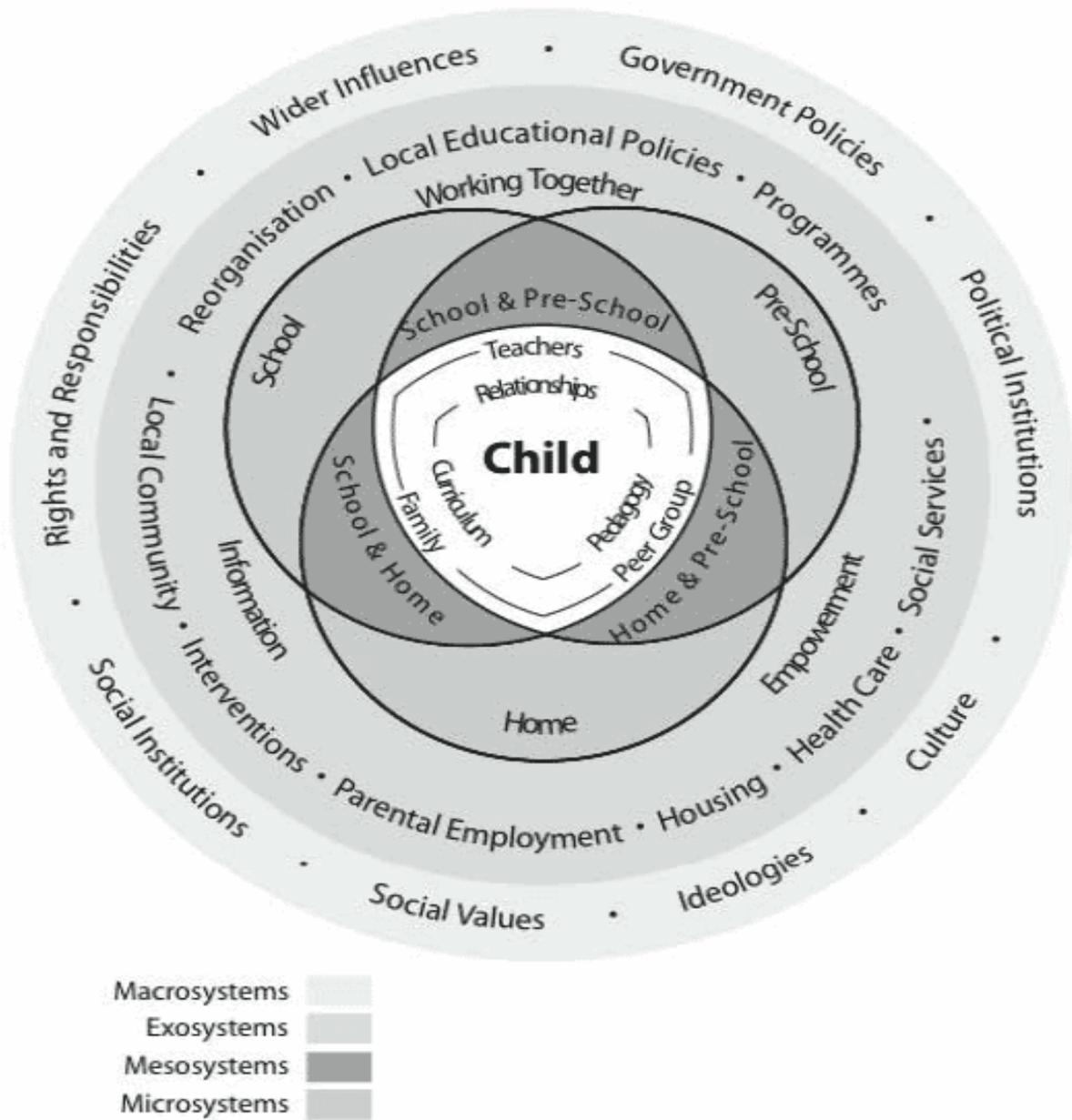
The records provided by the partner agencies to the MLDS Center do include personally identifiable information (PII) on either students or employees. Personally identifiable information includes, for example, name, address, date of birth, and other identification numbers or demographic data. The MLDS Center only uses the PII for the purpose of linking an individual's records across the different data sources. Once the records are matched, the PII data is removed so that the data is de-identified, and replaced by a randomly assigned token. The work performed by the research team at the MLDS Center will be done using the de-identified data. All reporting of the results of the research will be done using de-identified aggregate data. If any data standing alone or in combination with publicly available data could

lead to the identification of individuals, such data when reported will be suppressed in order to protect confidentiality.

The Complexity of Examining Education and Workforce Outcomes

Multiple factors influence education and workforce outcomes. Researchers in the Center will analyze the MLDS data to examine the influences of these factors on education and workforce outcomes. For example, a student's test scores may be influenced by the student's classmates, teachers, and home life. In this example, the classroom (classmates and teachers) and home are environmental contexts that influence the student. In order to fully understand these influences, Bronfenbrenner's (1979) ecological systems theory guides our way of thinking about how multiple environmental contexts influence outcomes. An individual is influenced by interactions within and between environmental contexts over time. Figure 1 depicts some of the environmental contexts that may influence a student's outcomes. In this example, the student (child) is located in the middle of the figure. The circles that are located closest to the student depict commonly experienced environmental contexts which may have the most influence on the student (for example, classmates, parents, and teachers). The circles that are located further away from the student depict more remote environmental contexts which may be less influential on outcomes (for example, culture and social values).

Figure 3: Bronfenbrenner's Ecological Systems Theory



Contexts that Influence *Student Outcomes*

A student's classmates, teachers, home life, classroom factors, and school factors can all influence student performance and achievement. Although this list is not exhaustive it illustrates the complex nature of how multiple environmental contexts might influence student outcomes. Below are some examples.

- A student's classmates can influence outcomes. For example, peer pressure may influence whether a student completes homework assignments, studies for a test, or regularly attends school. Having classmates that get good grades may influence a student to study harder in order to obtain better grades.
- A student's teachers can also influence outcomes. For example, a teacher with more experience may be better able to help a student who is performing poorly to raise test scores.
- A student's home life can also influence outcomes. For example, a student from a poorer family might need more help at school because the student's parents are working evenings and are unable to provide help with homework.
- A student's outcomes can also be influenced by classroom factors. For example, classroom size may influence performance. A student in a small classroom may receive a lot of individual attention from the teacher, enabling the student to learn better and get better grades. Alternatively, a student in a large classroom may receive less attention from the teacher and not perform as well.
- A student's outcomes can also be influenced by school factors. For example, the financial resources of the individual school or overall school district may influence outcomes. A school with more financial resources may be able to provide more extracurricular activities for students. A student attending a school with a lot of extracurricular activities may have access to help with homework after school, which may help the student to perform better than a student attending a school with few extracurricular activities.

Contexts that Influence *Employee Outcomes*

Employee outcomes are influenced by the employee's education and training. Adequate education and training are needed for the employee to be prepared for the job and succeed at the job. For a professional writer, this may mean having taken challenging writing courses. For the auto-mechanic, this may mean having an externship at an auto body shop. For a lawyer, this may mean having attended a highly ranked law school. Some employees who do not feel prepared for their jobs may return to school to get further education, which may make them more productive at work.

Student Group Differences

In education research, student group differences are often examined in order to see if outcomes are different between groups of students. Student groups are groups that differ based on a characteristic (for example, males and females). Student groups can be constructed by characteristics of the students themselves, their families, and their communities, schools and school districts. For example, for students, educational research often finds group differences by gender, race or ethnicity, special education status, and primary language spoken in the home. At the family level, socioeconomic status and neighborhood of residence are strong predictors of educational outcomes. At the school level, the economic and other resources available, characteristics of the teachers, and the community setting being urban, rural, or suburban are predictors of educational outcomes. Such differences are found in examining outcomes such as achievement, standardized test scores, grades, and attendance.

It is also important to note that such student groupings overlap. Many student group characteristics center around and overlap with the socioeconomic status of a student and his or her family. For example, a student growing up in a lower resourced family likely attends a lower resourced school, may have less experienced teachers and a less safe neighborhood to live in, and may speak a different language than English at home, all of which predict lower school achievement. Whereas, a student who lives in a better resourced school district may have access to more extracurricular activities at school, more academic supports when struggling, greater numbers of supportive adults, and a lower student-to-teacher ratio in the classroom, all of which predict better school outcomes. Historical factors are also important to consider when examining student group differences. For example, students who are members of demographic groups who have been historically discriminated against, particularly in the arena of education services or access, may show lasting effects of that discrimination.

It is important to interpret education research findings within the context of these important student group differences. For example, an analysis may indicate African American, Latino, and Native American boys are testing lower than other student groups. That finding should be interpreted in the context of the historical discrimination those groups of boys may have faced that impacted their educational outcomes. Therefore, it is hard to draw a direct conclusion about why the boys are testing lower ([see section on Association versus Causal Relationships](#)).

It is a core part of the MLDS Center’s mission to examine such group level differences in educational outcomes. When we find such student group differences our goal will be to the extent possible to examine those differences in light of this complex set of factors that impact those outcomes. Our findings will be interpreted in the context of such differences.

Handling Complexity

The complex nature of the Center’s data requires researchers to use advanced statistical techniques. These techniques will be explained when used in dashboards and research reports. We will also explain the benefits and limitations of the methods used. Also, see section on [Research Quality Assurance Standards](#).

Associations versus Causal Relationships

Most of the time, in policy research, we are interested in determining the “effect” of some policy on some outcome. For example, what is the effect of taking STEM-related courses in 9th grade on eventual earnings? The data that the Center has are observational data. With observational data, associations can be examined. An association is present when two variables are related in a way that is unlikely to have occurred just by chance. If there is an association between STEM-course taking and earnings, it does not mean that STEM-course taking causes greater earnings. It is possible that those students who have chosen to enroll in STEM-related courses in 9th grade are different from students who did not enroll in those courses. For example, students who enroll in STEM-related courses may have higher grades than students who do not enroll in STEM-related courses. It may be the characteristic on which they differ (grades) that is the reason for the different earnings after high school, rather than the enrollment in STEM-related courses.

There are statistical techniques that can be used to address the differences between students who receive a program, like STEM-related courses, and those who do not receive the

program. These techniques use data about the students and their schools in order to make the two groups approximately equivalent. When possible, the MLDS Center will use these sophisticated techniques when conducting analyses. The Center will include a detailed description of these statistical techniques in all Center dashboards and research reports.

Research Quality Assurance Standards

Given the complexity of the MLDS Center data and research, multiple quality assurance standards are in place to guide our research. We maintain standards that ensure that we are building high quality data sets and using valid and trustworthy data, methods, and reporting standards.

Building High Quality Data Sets

The MLDS Center works to gather, integrate, and analyze data from multiple state agencies. To do this accurately, it is important to verify that the individuals can be matched across agencies ([see section on The Nature of Longitudinal Data](#)). This verification is accomplished by the MLDS Center through a [standard process](#) for evaluating the accuracy of the information being reported. Accurate, valid data are needed in order to confidently inform the public and policymakers.

Triangulating Data and Evidence

Triangulation is a process that strengthens the validity of data by verifying information across multiple sources. In other words, one piece of information from one source is matched to the same piece of information from another source or multiple other sources. If the pieces match, then the information is considered to be valid. The MLDS Center will interpret findings in the context of other important information that has been provided by partner agencies, and other available sources, including but not limited to books, published journal articles, research projects, and presentations.

Given the complexity of influences on educational outcomes ([see section on The Complexity of Examining Education and Workforce Outcomes](#)), it is important to take care in drawing conclusions based on one set of data, one analysis, one policy or program.

Accumulating evidence is a key aspect of drawing valid and widely applicable conclusions in education research.

MLDS Center Research

The MLDS Center research will inform policy makers and practitioners on the topics outlined in the [Research Agenda](#) established by the MLDS Governing Board. Our longitudinal data analyses can be used to answer important questions about student and employee outcomes that have heretofore been unknown.