

# Washington State ERDC P20W Data Handbook

## Author

Education Research & Data Center (ERDC)

## About the Author

The Education Research and Data Center, located in the Washington Office of Financial Management, works with partner agencies to conduct impactful analyses of education and workforce data that can help inform the decision-making of Washington legislators, parents, and education providers. ERDC’s Preschool-to-Grade 20-to-Workforce (P20W) data system is a statewide longitudinal data system that includes de-identified data about people’s preschool, educational and workforce experiences.

### Address

Education Research and Data Center  
 500 Jefferson St SE  
 Olympia, WA 98501  
 Mailstop: 43124 Phone  
  
 360-902-0599

### Email

[erdc@ofm.wa.gov](mailto:erdc@ofm.wa.gov)

## Revision History

Date	Version #	Author	Overview of Changes
October 2020	1.0	Megan Schoor	Original Version
October 2024	2.0	Liz Dehlbom, Megan Schoor	Major textual and formatting revisions throughout Moved detailed data element lists and availability to separate Data Dictionary document Removed publication lists throughout Added several data sources to Part I: Core P20W Source Data Removed Part II: Research Methods section Added Part II: Community and Organizational Characteristics section Added Part III: Restricted Data section Added Part IV: Data in Development section Added Other Resources appendix

## Acknowledgements

ERDC would like to recognize the contributions of the ERDC and Forecasting and Research Systems (FRS) staff who wrote sections of the Handbook and consulted on its structure. The information about available data contained in this Handbook is immense and this resource could not have been created without the efforts of many individuals.

*Annie Pennucci, Forecasting and Research Division Director*

*Rachel Hughes and Jordan Elwanger, Forecasting and Research Administrative Team*

*Katie Weaver-Randall, ERDC Director*

*Tim Norris, FRS Manager*

*Dani Fumia, ERDC Senior Research Scientist*

*Megan Schoor, FRS Senior Data Operations Specialist*

*Tess Grayson, ERDC Senior Data Scientist*

*John Sabel, FRS Senior Data Scientist*

*George Hough, ERDC Senior Data Scientist*

*Andrew Weller, FRS Senior Data Visualization Specialist*

*Tom Aldrich, ERDC Senior Data Scientist*

*Heather McCabe, ERDC Research Scientist*

*Melinda Wengrin, ERDC Senior Data Scientist*

*Liz Dehlbom, ERDC Senior Data Scientist*

*Matt Borda, ERDC Research Scientist*

ERDC is also grateful for the various staff at our partner agencies who reviewed the Handbook and provided valuable feedback. Their input and support of this project helped to improve the comprehensiveness of this resource.

Council of Presidents (COP)

State Board for Community and Technical Colleges (SBCTC)

Department of Children, Youth, and Families (DCYF)

Washington Student Achievement Council (WSAC)

Employment Security Department (ESD)

Workforce Training and Education Coordinating Board (WTECB)

Labor and Industries (LNI)

Office of Superintendent of Public Instruction (OSPI)

## About this Data Handbook

ERDC has created this Data Handbook as a resource for audiences to understand the data available through ERDC and the appropriate use of that data for different purposes. This Handbook provides a high-level description of each data source including details on the state agency that manages the data collection, substantive changes in data collections over time, an overview of the type of data available, any relevant policy or program changes, a description of how the data could be used, and any impacts to the data from COVID-19. This is done so that researchers, analysts, and data requesters can understand the data lineage and the breadth and depth of the information available through the Preschool-to-Grade 20-to-Workforce (P20W) data system. This Handbook categorizes the data in three ways:

- **P20W Core Data** – data available for redisclosure under ERDC’s data request process; stored and integrated into the P20W data system.
- **Restricted Data** – data shared with ERDC for specific purposes and typically not available to be redisclosed for other uses. This data is stored outside of the P20W data system to ensure it is not co-mingled with the P20W Core Data and it is typically not included in the ERDC Master Data Management Hub used in identity matching.
- **Data in Development** – data new to ERDC, or data that has been used minimally. Data in this category often needs to go through quality assurance processes and exploratory analysis before it can be fully utilized. Often data in this category is not yet ready for redisclosure through ERDC’s request process.

This Handbook can be used in conjunction with other data resources to gain a more comprehensive and detailed understanding of the data available through ERDC. The complementary resources are:

- The ERDC Data Dictionary which provides information about every data element contained in the files documented in the Handbook. The Data Dictionary is a resource for people who are requesting data from ERDC and who need to know which specific data elements are available from each file.
- The [ERDC P20W Data Availability webpage](#) which provides an overview of the availability of specific data files that are available for analysis or request.

This Handbook is intended to be used with the Data Dictionary and Data Status Overview to provide a comprehensive understanding of the data available through ERDC and the necessary details about the data to facilitate requests for data through ERDC’s request process.

While not a comprehensive look at any individual source data set, this Handbook is a body of knowledge that directs researchers to helpful resources to support an informed understanding of the data that ERDC has available and how it can be used to study the Washington education system.

Please contact the [ERDC](#) directly for detailed information about any of the subjects covered in the handbook. Research reports and data dashboards are available on the [ERDC website](#).

## Contents

About this Data Handbook.....	4
About the Education Research & Data Center (ERDC) .....	7
P20W Data Contributors .....	8
P20W Administrative Data Limitations .....	9
Flow of Contributor Data.....	9
Identity Resolution Process .....	10
Part I: Core P20W Source Data .....	13
Early Learning and Childcare.....	13
Childcare Subsidy Programs (CCSP).....	13
Early Support for Infants and Toddlers (ESIT) Program.....	15
Early Childhood Educational Assistance Program (ECEAP) .....	17
Early Learning Assessment .....	20
Early Learning Workforce .....	20
K-12 Public School.....	21
Kindergarten Assessment.....	21
Annual Statewide Assessment.....	23
School and District Characteristics .....	27
Reengagement Program.....	28
Student, Teacher, and Course Data.....	29
Comprehensive Education Data and Research System (CEDARS) .....	29
Core Student Record System (CSRS) .....	39
Staff and Educators.....	39
Educator Certification .....	39
K-12 Personnel Reporting.....	42
Retention and Mobility .....	44
Postsecondary Data .....	44
Educator Preparation Program (EPP) .....	44
Public Four-Year Baccalaureate Institutions.....	48
Public Community and Technical Colleges .....	51
General Equivalency Development (GED) .....	54
Private and Out-of-State Institutions .....	54
College Bound Scholarship .....	55
Apprenticeships .....	56
Registered Apprenticeships.....	56

Workforce .....	57
Unemployment Insurance (UI) .....	57
Part II: Community and Organizational Characteristics.....	61
American Community Survey / Census .....	61
Part III: Restricted Use Data .....	62
Criminal Justice .....	62
Early Childhood Educational Assistance Program (ECEAP) Boolean questions responses .....	63
Financial Aid data .....	63
Foster Care.....	65
Homelessness .....	65
Institutional Education .....	66
Part IV: Data in Development .....	67
Dual Credit Exam Scores.....	67
English Learner Annual Assessments .....	67
K-12 Graduation Pathways.....	67
Washington Private Vocational Schools.....	68
Workforce Innovation and Opportunity Act (WIOA).....	68
Appendix: Other Resources .....	69
Guide to Acronyms .....	69
Technical Reports and Documentation .....	69

## About the Education Research & Data Center (ERDC)

ERDC was established by legislation ([RCW 43.41.400](#)) in 2007 and works with partner agencies to conduct analyses to inform the decision-making of Washington legislators, educational institutions, researchers, families, and students. ERDC's mission is to develop longitudinal information spanning the preschool to workforce system to facilitate analyses, provide meaningful reports, collaborate on education research, and share data in ways that protect the privacy of students. ERDC researchers analyze data and answer critical questions involving the Washington state education system that are raised by partners and interested parties. Data is available to external researchers through [ERDC's data request process](#).

Originally, ERDC's major priorities were to:

- Coordinate with other state education agencies to compile and analyze education data
- Collaborate with the [LEAP Committee](#) and education and fiscal committees of the Legislature to compile and analyze data, to ensure that legislative interests are served
- Track enrollment and outcomes through the Public Centralized Higher Education Enrollment System (PCHEES), within the Office of Financial Management
- Develop a long-term higher education enrollment plan with other state educational agencies
- Conduct research that focuses on student transitions within and among the early learning, K-12, and higher education sectors of the Preschool-to-Grade 20 (P-20) system
- Share data from collaborative analyses with education agencies and institutions that contribute data to ERDC

In 2009, the Washington legislature expanded the mission of ERDC to include:

- Identify key education research and policy questions to address and the data needed to do so
- Lead the governance of expanded Preschool-to-Grade 20-to-Workforce (P20W) data
- Serve on the K-12 Data Governance Group and provide them a list of data elements and improvements necessary to ERDC's work
- Monitor and evaluate the versatility and quality of education data collection systems of the organizations and agencies that contribute to ERDC and ensure that the data they provide is relevant
- Provide recommendations to the legislature that meet the goals and objectives of the comprehensive K-12 data improvement system and K-12 Data Governance Group

The P20W data system is a statewide administrative record database that is populated by data collected by ERDC and other state agencies and which connects education data across early learning, K-12, postsecondary, and workforce sectors. The P20W data system is housed within the Washington Office of Financial Management (OFM) and managed by the ERDC in the Forecasting and Research Division. Longitudinal and cross-sector data make it possible to measure student and system progress across education sectors and time, which informs education policy and practice.

## P20W Data Contributors

ERDC receives a variety of administrative datasets from agency partners<sup>1</sup> that are incorporated into the P20W data system. These administrative datasets are outlined in Table 1, based on the category of data and data source. These datasets vary in subject matter, from early learning programs to K-20 public school data to workforce data. As such, ERDC’s P20W data system is the most comprehensive longitudinal education data system in the state.

**Table 1: P20W Core Data Sources**

Sector	Data Contributor (Source) and Agency Description
Early learning and childcare	<a href="#">Department of Children, Youth, and Families (DCYF)</a> DCYF is the lead agency for state-funded services that support children and families, which include childcare, foster care, and juvenile justice.
Public K-12 education	<a href="#">Office of Superintendent of Public Instruction (OSPI)</a> OSPI is the primary agency that oversees public K–12 education in Washington. OSPI coordinates the collection of data from the state’s public school districts and state-tribal compact schools.
Public community colleges and technical colleges	<a href="#">State Board for Community and Technical Colleges (SBCTC)</a> SBCTC advocates, coordinates and directs Washington state’s system of public community and technical colleges. The Board collects student data from each member school.
Public 4-year universities	<a href="#">Office of Financial Management (OFM)</a> OFM provides information, fiscal services, and policy support to the Governor, Legislature, and state agencies. OFM works with the state public four-year institutions to track public higher education enrollment and graduation outcomes.
Financial aid	<a href="#">Washington Student Achievement Council (WSAC)</a> WSAC provides strategic planning, oversight, advocacy, and program administration to support increased levels of educational attainment in Washington. WSAC collects financial aid award data from Washington public colleges and universities.
Apprenticeships	<a href="#">Labor and Industries (LNI)</a> LNI is the state agency that administers the state’s workers’ compensation system and collects applicant information from US Department of Labor apprenticeship programs.
Workforce	<a href="#">Employment Security Department (ESD)</a> ESD provides services related to unemployment benefits and insurance for employees in Washington.

<sup>1</sup> For details about partner agencies, please review the [ERDC website](#).



## P20W Administrative Data Limitations

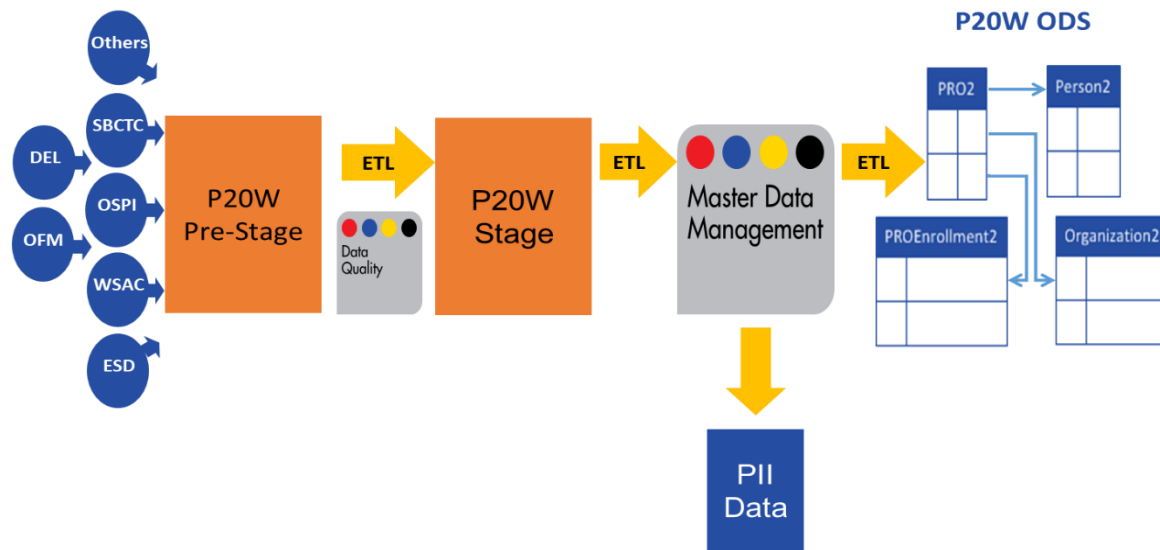
While all the datasets above are processed to the highest quality standards by the source agencies, it is important to recognize that inaccuracies may exist within administrative data. Unlike other data, where both cross- and within-subject controls are possible, such measures are often unfeasible and impossible to incorporate in administrative data. Administrative data is also not typically collected for research or evaluation purposes but to meet the administrative needs of specific programs and specific state or federal reporting or monitoring requirements. Administrative data is collected as both transactional and summative datasets by local administrators and submitted to an agency authority, making variance among data collectors a potential source of bias in each dataset. Quality control processes may be imposed after data is submitted to agency authorities, which could impact data quality in ways that are difficult to detect within the final dataset. For a more detailed discussion of the data quality of administrative data, see [ERDC's report](#).

The limitations described in this Handbook are not meant to suggest that the administrative data loaded into the P20W data system is unreliable. ERDC advises researchers to keep these potential concerns in mind as they request data and conduct research. Administrative data must always be thought of as the combination of both the collected data and the process used to collect the data. The data summaries in this Handbook delve into these processes, but they are not the only resource. Researchers who use ERDC data for analysis purposes should review all the available data documentation and adjust their models according to the research question and the administrative data collection procedures.

## Flow of Contributor Data

Figure 1 illustrates how ERDC loads source data from contributing agencies. Once data is received through a secure file transfer process, the data is loaded to a pre-stage database, then it undergoes a series of quality checks before it is transferred to a stage database. Personally identifiable information (PII) is separated at that point from the rest of the data and used for identity resolution. Once the identity resolution process is complete, the revised crosswalk of P20IDs and token IDs is incorporated into P20W Education Data System Operation Data Store (ODS). Once in the ODS, data are de-identified and available for analysis. Identifiers used in the identity resolution process do not advance beyond the Master Data Management (MDM) hub.

Figure 1: Flowchart of data through stages of the ERDC loading process



## Identity Resolution Process

The core of the P20W data system is the linking of cross-sector data. Through an identity resolution process, ERDC links individuals and entities across data files from contributing agencies and institutions to facilitate longitudinal and cross-sector analysis. Identity resolution is the process of identifying records that belong to the same entity (e.g. person or household). The purpose of identity resolution is to create linkages across multiple data sources so that students' records are linked across early learning, K-12, postsecondary, and workforce data. For ERDC's P20W System data system, this involves linking individual-level data, such as names and birth dates, across multiple sources to create unique person identifiers. These identifiers are referred to as the "P20ID." P20IDs are assigned to all individual-level data received by ERDC from our data contributors. As additional linking activities occur, P20IDs will be updated to reflect the most recent data available.

It is important to understand the identity resolution process so the researcher can evaluate whether it may impact the analysis, especially if the research includes matching the P20W data to additional data.

### Creation of TokenIDs and Assignment of P20IDs

Before a person-level P20ID can be created, an identity resolution "token" is created for each record in a dataset. Identity resolution tokens, known as "pkeys" on the P20W data system, are concatenated sets of identifiers that are guaranteed to be unique to an individual. For example, in workforce data that comes from the Employment Security Department (ESD), more than one individual can have the same Social Security Number (SSN). As a result, ERDC cannot rely on SSNs to uniquely identify individuals in this source of data. Consequently, each ESD identity resolution token is composed of the SSN, the employer account number and the employee name. ERDC has found that this set of identifiers is guaranteed to be unique to an individual in ESD data. The data is then loaded into the MDM system, the system generates integer based surrogate key, called a TokenID for each new pkey. TokenIDs are used

to merge and unmerge P20IDs when over- or under-matching occurs. Over-matching occurs when P20IDs contain TokenIDs for two or more people. Under-matching occurs when two or more different P20IDs actually refer to the same person.

Since every source of data has its own set of identifiers and its own set of data challenges, each source’s identity resolution token definition is different.

**Table 2: Identity token components in data sources**

Sector	Source System	College ID	Person ID	Birthdate	First Name	Middle Name	Last Name	SSN
Early Learning	DCYF-ELMS		✓					
	DCYF-ESIT		✓					
K-12	OSPI-CEDARS		✓	✓	✓	✓	✓	
	OSPI-WaKIDS		✓	✓	✓		✓	
	WSAC-CB		✓	✓				
	WSAC-Fin Aid			✓	✓		✓	✓
	WSAC-Unit			✓			1 <sup>st</sup> Initial	✓
Postsecondary	SBCTC	✓	✓	✓	✓	✓	✓	
	PCHEES	✓						
Educators and Staff	OSPI-ECert		✓	✓	✓	✓	✓	
	OFM-EPP	✓	✓					
	OSPI-S275		✓	✓	✓	✓	✓	

No single set of identifiers is common to all data sources, so the identity resolution process and match rules are tailored to the source of data being matched. For example, OSPI K-12 data has names and birth date, whereas ESD wage data has names of mediocre quality and SSNs. As a result, there is no way to directly match these together, but they can be indirectly matched by involving other sources of individual data. In this example, OSPI K-12 data and ESD wage data can be linked by using SBCTC as a matching bridge between the two data sources because SBCTC contains both high quality name and SSN fields.

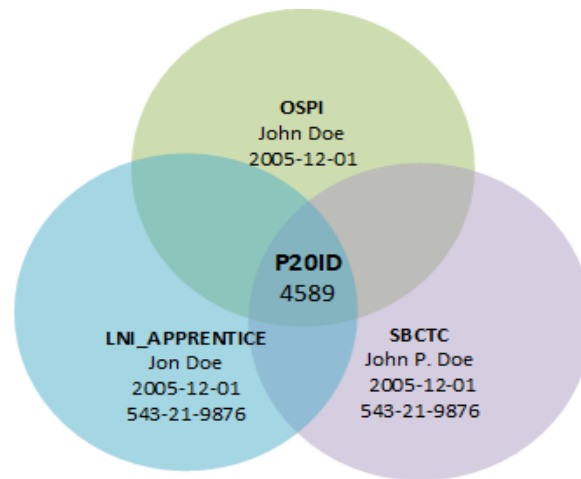
**Phases of Identity Resolution**

ERDC’s identity resolution process has two phases.

1. High probability match phase - P20IDs are provisionally matched using “automerger” match rules. These are conservative rules for matching P20IDs. In this phase inaccurate matches are not much of a concern, as the MDM auto-merge match rules and SQL based rules are designed to ensure extremely low false positive rates. These match pairs are not manually reviewed, though they are subject to cardinality analysis, after which the match pairs are then merged.

2. Mid-probability match phase – After the high probability matches have been merged, looser match rules are used to create additional match pairs. The advantage of looser match rules is that they can potentially create additional true-positive match pairs. But this comes at the cost of also creating additional false-positive match pairs. Given that there is a mixture of true-positive and false-positive match pairs, this set of match pairs is manually reviewed. The results undergo cardinality analysis, and then the true-positive match pairs are merged.

**Figure 2: Example of a common P20ID for same person across multiple source data systems**



### High Probability Match Phase: Match Rules and the Creation of Match Pairs

ERDC’s identity resolution process involves matching new P20IDs against the pre-existing P20IDs in MDM using a set of match rules. An example of a match rule is “Same names, same birth dates, same SSNs.” When new P20IDs are matched against the pre-existing P20IDs, the result is a set of prospective P20ID match pairs.

The quality of the match pairs depends on the match rule. For match rules involving the exact match between many ID fields, such as the “same names, same birthdates, and same SSNs”, the false positive rate is low and the matches can be provisionally accepted, subject to cardinality analysis described below. The quality of the match is lower, for example, when the match rules involving a small number of fields, fuzzy logic is used (such as SOUNDEX), or matching the first three characters of the last names. These lower quality matches pairs are manually reviewed.

### Mid-Probability Matched Phase - Manual Review

In manual review, each match pair is assigned a probabilistic score of the likelihood of a positive match. The probabilistic score, match pairs, and associated identifiers (names, dates of birth, etc.) are brought into the same dataset. The match pairs are then reviewed, and the match pairs that are deemed to be actual positive match pairs are flagged, and the results are integrated into the prospective match pairs

within MDM. At this point, MDM contains the set of provisional match pairs. These prospective match pairs are then subject to cardinality analysis.

### Cardinality Analysis Phase

Cardinality analysis is a key component of the identity resolution. It allows for more aggressive matching, while at the same time, it improves the quality of the existing P20IDs in MDM. In cardinality analysis, the provisional match pairs are merged on a trial basis. Then the cardinal relationships are determined between the P20IDs. These relationships can be 1:1, 1:Many, Many:1, or Many:Many. The 1:1 relationships, where P20ID matches exactly, are accepted. The relationships involving non 1:1 relationships need to be manually reviewed to resolve them as accurately as possible.

Once the cardinal relationships are verified, the results are fed back into MDM. The result is that some P20IDs might be unmerged, some could be merged, and some provisional match pairs potentially are shown not be actual match pairs. At the conclusion of this iterative process, the P20IDs are integrated into the data contained in the ODS for use in analysis.

## Part I: Core P20W Source Data

This section provides a set of descriptions or quick references to the core data files that feed into the P20W data system. This information is not an exhaustive list of data in the system, nor does it provide the detail needed for a researcher to sufficiently complete an ERDC Data Request Form. Rather, these descriptions are designed to:

- guide researchers toward data that are relevant to their research questions
- provide metadata that will inform research design
- provide examples of how the data is used in research

This Data Handbook is designed to be used in conjunction with the ERDC P20W Data Dictionary and the P20W Data Status tool to understand the specific data elements and years of data available. To utilize this additional information, see ERDC's [Data Resources](#) website.

## Early Learning and Childcare

### Childcare Subsidy Programs (CCSP)

Working Connections Child Care (WCCC) and Seasonal Child Care are collectively referred to as Child Care Subsidy Programs (CCSP). CCSP assists eligible low-income working families by promoting access to childcare and after-school programs that help prepare their children to succeed in school. DCYF is the designated lead for setting CCSP eligibility and payment thresholds, while the Department of Social and Health Services (DSHS) is responsible for delivering CCSP services, determining children's CCSP eligibility,

and authorizing payments for CCSP services.<sup>2</sup> Children served by this program range from birth to 13 years old, or up to 19 years old for those with a verified special need or under court supervision.

CCSP has been loaded to the P20W data system but is not yet available in the Early Learning/Kindergarten data mart. DCYF provides ERDC with an annual feed of participant data that contains CCSP eligibility and service data. The CCSP data extract, often referred to as Subsidy data, includes monthly records from January 2012 through December of the most recent available year for each child covered by the program. It is structured so that information is included for each child and each month of eligibility or service. Children served by CCSP can see multiple providers over the course of one month. Providers may be billed for multiple service codes. All providers who receive a payment are required to submit a claim for each invoice. As a result, there is a record for every distinct occurrence of child identifier, month of service, service provider, and service code. Key variables in this dataset include case number, authorization number, authorization line, service provider number, the month of service, service name, service unit type, service code, and provider type.

One major limitation of the CCSP data extract is the lack of information on sociodemographic characteristics of the participating children and their families. Researchers can overcome this issue by merging CCSP data with other available data from the Washington Kindergarten Inventory of Developing Skills (WaKIDS) and/or the Comprehensive Education Data and Research System (CEDARS) databases.

There are many opportunities to explore how CCSP may help low-income families improve their financial stability and expand access to childcare resources for their children. In particular, the CCSP data extract may be useful for exploring the following research topics:

- Explore school readiness among children who received childcare subsidy, compared to low-income children who did not receive the subsidy
- Examine the effect of childcare instability on the development outcomes and school readiness of childcare recipients
- Investigate the continuity of childcare and child's development outcomes

### **COVID-19 Impacts**

The COVID-19 pandemic and response in 2020 impacted recent years of subsidy data, magnifying a trend of decreasing caseloads. Caseloads for WCCC had fallen from 30,672 in state fiscal year 2017 to 25,203 in 2020. Caseloads fell further to 19,711 in 2021 and remained low at 20,648 in 2022. The Caseload Forecast Council (CFC) noted that the pandemic affected work patterns, childcare utilization, and childcare service delivery, leading in turn to a substantial decline in the WCCC caseload.<sup>3</sup>

While overall caseloads had declined, there were other programmatic impacts during this time. COVID-19 also impacted utilization of the Quality Rating and Improvement System (QRIS). Subsidy providers must achieve a minimum score on the QRIS to participate in the Early Achievers program, Washington's

---

<sup>2</sup> See [WAC Chapter 110-15](#) for more information, including eligibility criteria.

<sup>3</sup> CFC Handout for Working Connections Child Care (WCCC), November 8, 2023.

quality recognition and improvement system for child care programs. While the QRIS ratings are not part of the Subsidy data, ERDC has access to other data provided by DCYF that allows us to determine the QRIS ratings for providers. During the pandemic, the ratings were suspended and in the summer of 2021, they were reinstated with revisions.<sup>4</sup>

Policy changes during this period had a direct impact on both eligibility and payments. Permanent post-COVID policy changes include extending the Homeless Grace Period, excusing Vocational Single Parents from work activities, waiving copayments for Teen Parent/High School Students, and on-going changes related to the [Fair Start for Kids Act](#).

## Early Support for Infants and Toddlers (ESIT) Program

The Washington State Early Support for Infants and Toddlers (ESIT) program provides individualized and quality early intervention services to young children (birth to three years old) who have disabilities or developmental delays. The Department of Children, Youth, and Families (DCYF) oversees the ESIT program, in accordance with the federal Individuals with Disabilities Education Act (IDEA), Part C. The ESIT program has two main objectives: 1) find eligible children through screening, tracking, monitoring, and referral services for at-risk children, and 2) provide early intervention services, including developmental and therapeutic services for children identified as developmentally delayed or who have an established condition for delay. At the beginning of each calendar year, DCYF provides ERDC with an annual snapshot of the ESIT database (excluding some data tables) for all prior years.

To be eligible for ESIT, a child must have a 25 percent delay or perform at 1.5 standard deviations below their age group in one or more of the five developmental areas.<sup>5</sup> Early intervention services may include but are not limited to specialized instruction, speech therapy, occupational therapy, or physical therapy, which can be provided in a variety of settings including home, childcare, preschool or school programs, and communities. Early intervention services end on the child's third birthday, or upon achieving satisfactory results before the age cutoff.<sup>6</sup> Each ESIT program participant has an Individualized Family Service Plan (IFSP) developed for them to access the intervention services and resources specified in the plan. ESIT also uses a Child Outcomes Summary Form (COSF) to record specific outcomes at program entry and exit.<sup>7</sup>

The ESIT database includes all children who were ever referred to the ESIT program from calendar year 2009 to the most recent year with complete data. The ESIT database contains over one hundred tables and over one thousand unique columns. ERDC staff summarized key information into six data views: Child/family information, Referral, Eligibility, COSF, Services, and Transition.

---

<sup>4</sup> [The fair start for kids act 2023 evaluation report \(wa.gov\), page 30.](#)

<sup>5</sup> Eligibility details at [Early Intervention Eligibility | Washington State Department of Children, Youth, and Families](#)

<sup>6</sup> ESIT services can also end voluntarily based on the family's decision, if the family moves out of Washington, or if the family is out of reach from the current service providers.

<sup>7</sup> COSF outcomes are measured in the areas of Positive Social/Emotional Skills, Acquiring and Using Knowledge and Skills, and Use of Appropriate Behaviors. Results indicate if a child has Age-Expected Skills, Decreasing Degree of Age-Expected Skills, or No Age-Expected Skills, as well as Decreasing Degree of Immediate Foundational Skills.

Key information provided in the data views are:

- Child: Child gender, ethnicity, language, birth date, ESIT organization location
- Referral: Referral sources, reason, and received date
- Eligibility: Eligibility determination and date, evaluation domains and results, eligibility basis and provider
- COSF (Child outcomes summary form): outcome names, descriptions, and scores
- Services: service setting, start and end dates, service type, funding source
- Transition: transition type/description

Since it excludes children who do not receive ESIT services, the ESIT database represents a limited subgroup of early learners in Washington. The quality of ESIT data is generally consistent across the available years. While some data have a considerable number of missing values, most ESIT data is well-suited for examining participants' characteristics, eligibility, and services received. ESIT data on child characteristics are fairly complete, while a substantial amount of data is missing on family characteristics and children's diagnosed medical conditions. ERDC staff found discrepancies within ESIT records indicating transition to ECEAP, when compared to actual ECEAP participation records extracted from ERDC linked data. Missing values do not appear to correlate with specific factors like time or report source.

How race is defined within the ESIT dataset is also unique, compared to definitions in other educational databases like CEDARS for K-12 students. ERDC recommends using the CEDARS demographic fields when combining ESIT with CEDARS tables, since CEDARS data on race and ethnicity is generally more detailed. Furthermore, ESIT column names are not always explicit or do not consistently align with other educational databases. Data columns like Name, Description, and "Dosage" of services may also be ambiguous; while these columns have generic names, they refer to more specific information within the data.

Some children did not participate in the program after eligibility determination, or out of parents' choice. Children who were determined eligible but do not participate only appear in the referral data. Some participants may exit the Washington State public education system entirely over time. In some instances, a participant's reason for leaving the program is missing or unknown, which could suggest the presence of uncorrected bias within the data. Reporting an unknown reason for leaving may stem from both administrative or participant sources, with no clear way to determine if the missing information is systematic (i.e., reason is not recorded, reason for exit is not represented in the database) or random (i.e., participants move without notifying ESIT, parents' decision to withdraw was not communicated to ESIT, changes to eligibility status, etc.).

ERDC researchers have analyzed the connection of ESIT data and medical data and how health factors are associated with children's early learning outcomes. Exploring the potential connections between childhood health and early learning can help researchers and policymakers establish a solid foundation for children's school-readiness, including at-risk groups. ESIT data on program participants and services is also useful for addressing research questions like those below:



- Explore service recipients by demographic characteristics, reason for eligibility, services received, and outcomes.
- Identify program trends over time based on the critical steps that participants complete, including referral, evaluation, service plan development and review, and program exit.
- Investigate children’s pathways post-program participation, when combined with future educational attainment levels.

### COVID-19 Impacts

The COVID-19 pandemic impacted recent years of ESIT data, including decreased caseload, referrals and service delivery. Impacts may have extended to related phases in the ESIT journey such as evaluation and diagnosis, and the transition out of ESIT.

One of the notable impacts of COVID-19 was a significant drop in caseload. Caseload fell from 9,991 in the 2019-2020 school year to 8,808 in the 2020-21 school year. It climbed back to 9,849 in 2021-22 and then increased further to 11,116 in 2022-23.<sup>8</sup> Referrals also decreased during the pandemic.

COVID-19 impacted service delivery as education and other services moved from in-person to a virtual environment. Data entry was also impacted. In some cases, providers were not aware of the DCYF guidance that telehealth and video conferencing counts towards a service delivered in the natural environment.

Access to medical and other evaluation and diagnostic services may have been impacted by COVID-19, which would in turn have affected referrals to ESIT.

Finally, as with other programs such as ECEAP, we do not yet know the extent of COVID-19’s full impact on children starting from the moment of conception as the mother and child navigated a medical landscape severely impacted by COVID-19 during pregnancy and after birth.

### Early Childhood Educational Assistance Program (ECEAP)

The State of Washington provides preschool to three- and four-year old, low-income children through the [Early Childhood Educational Assistance Program \(ECEAP\)](#). DCYF oversees the program, which serves over 10,000 children a year at hundreds of locations. DCYF provides ERDC with annual data that includes all children who participate in the state funded ECEAP, which ERDC loads into the P20W system. Additionally, the data contains some information about the providers, site curricula, and teachers. The data does not currently include information on Head Start or privately funded early childhood programs.

There are two sources for ECEAP data. ECEAP data between the 1999-00 and 2011-12 school year is referred to as “Historical ECEAP” data. DCYF provided Historical ECEAP data to ERDC as a collection of files, which ERDC loaded into the P20W data system. Historical ECEAP data is contained in one table, which is structured with one record per child per enrollment segment. Historical ECEAP data fields

---

<sup>8</sup> Caseload Forecast Council Handout for Early Learning Support for Infants and Toddlers (ESIT), November 8, 2023

contain demographic information (gender, age, race, ethnicity, family income, disability indicator, and primary language) and enrollment information (site, contractor overseeing the site, start date and end date).

ECEAP data after the 2011-12 school year comes to ERDC from DCYF's Early Learning Management System (ELMS) database. While the ELMS database contains over one hundred data tables, ERDC only receives a subset of the data. DCYF's data team prepares an annual extract of the ELMS database and sends it to ERDC at the beginning of each calendar year. ERDC staff create a series of four "data views" that pull key data from the ELMS extract that feed into the P20W data system. These four data views are ELMS Eligibility, ELMS Enrollment, ELMS Organization, and ELMS Site. Each data view has its own unique level of analysis.

The ELMS Eligibility view includes data collected during determination of the child's ECEAP eligibility, so this information is only collected once (though some exceptions apply). The ELMS Enrollment view contains a record for every time a child enrolls in ECEAP. The ELMS Organization view outlines information about ECEAP providers, including contractors, subcontractors, and program sites for each year. The ELMS Site view includes data on program sites for each year, including location details, facility licensing and operations information, and characteristics about the curriculum used.

- The **Eligibility** view contains demographic information of potential value to the researcher: gender, age, race, ethnicity, family income, disability indicator, and language spoken at home.
- The **Enrollment** view contains a field for site, class within the site, start and end dates of enrollment, language of instruction, and whether the class is part day, full day, or extended day.
- The **Organizational** view contains information about sites and the subcontractor-contractor associations for each year of ELMS data.
- The **Site** view is not loaded to the P20W data system at this time but contains key characteristics of sites such as location (including latitude and longitude), whether the site is a licensed childcare facility or is operated by a tribal organization, and the curricula used at the site.

Depending on the nature of the research study, the Historical ECEAP data of 1999-00 to 2011-12 is potentially problematic, while the ELMS data that covers the 2012-13 to current years is considered reliable. Missing end dates during the 2010-11 and 2011-12 school years, for example, were identified during the transition from the Historical ECEAP system to the ELMS data management system. This adjustment complicates the process for computing ECEAP "dosage" when measured by the number of days enrolled in ECEAP. ELMS data comes from a relational database, and the quality has improved as a result of this structure and DCYF's continued commitment to data accuracy. ERDC's data views capture all values of key variables. For example, if a student has multiple race codes, each of the codes is captured and entered into the P20W system. It is up to the researcher to determine how to use all the codes, which are captured as multiple records.

This data is valuable to researchers who want to examine various aspects of state-funded ECEAP, including program-level and individual-level characteristics. Most of the existing research on early childhood programs explore the outcomes of ECEAP graduates as they transition into kindergarten and

future grade levels. To address questions like these, researchers must also work with K-12 data. For example, researchers at the Washington Institute of Public Policy (WSIPP) have conducted research using ECEAP data to explore the progress of ECEAP students within Washington's K-12 system.

Because ECEAP data includes information about participants and programs, it may be especially suited to address the following types of research questions:

- How does the ECEAP's effect on K-12 outcomes such as WaKIDS scores vary among different racial or ethnic groups?
- Do K-12 outcomes for ECEAP participants vary by gender?
- Does an increased dosage of ECEAP help children do better on WaKIDS?
- How do WaKIDS scores vary across contractors and/or sites?

ECEAP data does not account for participation in other early childhood education programs, so researchers must interpret their comparison results with caution. If a researcher uses a comparable low-income group as a comparison group, then this comparison group could include children that participated in other early childhood education programs, like Head Start. This distinction could ultimately dilute the measured effect of the ECEAP program on participant outcomes. If the comparison group is drawn from K-12 students (other kindergartners, for example), then demographics from the K-12 data system must be used to account for that group. However, it is important to note that K-12 demographics may not be at the level of granularity needed for all potential methodologies. Given that ECEAP data reflect annual snapshots in time, researchers may also find the data insufficient for more complex, longitudinal trend analyses. The ECEAP program has steadily expanded over time, so trends in the data could be due to the program effect changing over time and/or potential changes in the population of participating children and families.

### COVID-19 Impacts

COVID-19 impacted recent years of ECEAP data, including caseload drops, missing values and the benefits conferred by ECEAP. COVID-19 impacts may continue in more subtle ways moving forward.

One of the notable impacts of COVID-19 was a significant drop in caseload. Caseload fell from 13,401 in the 2019-20 school year to 10,619 in the 2020-21 school year. It climbed back to 13,103 in 2022-23, which was still short of pre-pandemic levels.<sup>9</sup> For attending children, serving them outside of a classroom setting could have impacted the quantity and quality of data collection. In particular, helping children interact positively with other children in a face-to-face situation was impossible in a virtual environment and virtual social interaction may have been challenging for both students and staff. Other data could have been impacted in less obvious ways. While data collection itself did not stop, data quality and completeness were impacted.

---

<sup>9</sup> CFC Handout for Early Childhood Education and Assistance Program (ECEAP), November 8, 2023.

Analysts and researchers can anticipate COVID-19 impacts in their analysis on the presence and benefits of ECEAP. For example, ELMS data during COVID-19 is not as representative of the entitlement population as it is at other times.

Children born during the spring or summer of 2020 could have entered ECEAP as 3-year olds in September of 2023. How COVID-19 impacted their development prior to entering ECEAP may have an effect on how they benefit from ECEAP, and the COVID-19 impact on ECEAP staffing may impact ECEAP site data and site impacts. The change in the nature of the interaction of ECEAP staff and children could affect how well children in ECEAP during COVID-19 years perform as they move into the K-12 system. ERDC will be looking at COVID-19 impacts into the future.

## Early Learning Assessment

Teaching Strategies GOLD (TSGold) is a formative observational assessment tool used in the Early Childhood Education and Assistance Program (ECEAP). ECEAP providers assess children on a quarterly basis. This age-appropriate instrument involves observations by the teacher over a period of time. These observations are captured in a number of numerical objectives. The objectives are then combined to form scores for the child over a range of six domains of child development: Social-Emotional, Physical, Cognitive, Language, Literacy, and Mathematics. These are the same domains that researchers will see in the WaKIDS data because the WaKIDS assessment is based on a subset of the TSGold assessment. However, comparisons of TSGold and WaKIDS assessment scores should be made with caution due to differences in the assessments and scoring.

TSGold data is supplied annually to ERDC. Each record corresponds to the scores for a given child and contains assessment data for both spring and fall. Additional fields contain information about the child and the site, but ERDC uses ELMS/ECEAP as the system of record for that information. Only the assessment information and the child's PII necessary for identity matching are moved into the ERDC P20W data system. The impact of COVID-19 on TSGold was that data received by ERDC was spotty or non-existent between 2018-19 and 2021-22 school years.

## Early Learning Workforce

Managed Education and Registry Information Tool (MERIT) is a system managed by DCYF for early learning professionals to log their employment, training, degrees and certifications in the field of early learning. It also allows them to find training and get recognized for their training achievements. While it was designed for professionals' use, the data in the MERIT system also provides information about the early learning workforce, including the providers. In particular, it includes the Quality Rating and Improvement System (QRIS) levels for providers.

As with other data sources for early learning, ERDC receives a subset of the MERIT data. This allows DCYF to ensure privacy of data where appropriate, but it may mean that ERDC does not have tables or fields that some researchers may need. Also, the data itself may not be complete and may need to be supplemented with workforce data from other sources, such as another early learning system like ELMS.

Finally, we should note that usage of this data to-date has been limited and it is not loaded into the P20W data system. ERDC has not done in-depth profiling and data quality analysis of this data source and COVID-19 impacts are not known.

## K-12 Public School

### Kindergarten Assessment

Washington Kindergarten Inventory of Developing Skills (WaKIDS), administered by the Office of the Superintendent of Public Instruction (OSPI), is a legislatively mandated assessment of students in state-funded, full-day kindergarten. It is based on a subset of the TSGold assessment. It began as a pilot program in the 2010-11 school year and was implemented statewide in the 2011-12 school year. The WaKIDS evaluation is an observational assessment. When children enter kindergarten, their teachers conduct the assessment to determine if each child is kindergarten-ready (i.e. whether they exhibit the common characteristics of children entering kindergarten). WaKIDS captures a range of observational evaluations that rate students in six developmental areas: Social-Emotional, Physical, Cognitive, Language, Literacy, and Mathematics. Kindergarten teachers are expected to complete these observations by October 31<sup>st</sup> of each school year.

At the beginning of each calendar year, ERDC receives a data file from OSPI with kindergarteners' WaKIDS scores, collected by teachers during the fall of the prior school year. ERDC then loads the data into the P20W system and matches students in that data file to K-12 CEDARS data. The WaKIDS data contains records for all kindergarteners who have completed the assessment since the 2011-12 school year. Initially, assessments were limited to students who attended high-poverty schools that were required to administer WaKIDS as part of their funding for full-day kindergarten. Starting with a small pilot in 2010-11, the program has expanded to full state-wide implementation by 2017-18 so that all incoming kindergartners are now assessed with WaKIDS, unless parents choose to opt their child out. WaKIDS data includes one record per kindergartner per year.

The six domains of the WaKIDS data assessment align with the tool that kindergarten teachers use to evaluate students in the six developmental areas. Within each domain, the assessment data captures the developmental level for each child and if they were considered "kindergarten-ready." Once all observations are made in a domain, the domain score can be determined. The development level of each child is the most fundamental measure in the WaKIDS data. Kindergarten readiness is determined based on whether the child's level of development reaches that of a four-year-old or older. Each level of development is classified by color, as outlined in Table 3. Children who reach the blue or purple development level are typically considered kindergarten-ready, though not all students at the blue development level are kindergarten-ready. Developmental levels for performance above kindergarten level were instituted in the 2018-19 school year. Performance levels for first, second, and third grades were assigned to the color bands Pink, Silver, and Brown respectively, adding to the existing color bands.

**Table 3: WaKIDS Developmental Level Color Bands**

Color	Level
Brown	3 <sup>rd</sup> grade
Silver	2 <sup>nd</sup> grade
Pink	1 <sup>st</sup> grade
Above Purple	Not Applicable
Purple	Age 5
Blue	Age 4
Green	Age 3
Yellow	Age 2
Orange	Below Age 2
Red	Not Applicable

Readiness flags are populated for all available years. The date that each domain was finalized is available from 2017-18 onward. Completion flags were populated for 2011-12 and 2012-13 only and are not loaded to the P20W data system. Child birthdates, which are important for identity matching, are fully populated from 2012-13 onward.

Given ERDC's review and processing of the WaKIDS data, the fields that are used in reporting and in loading to the P20W data system can be fully utilized. The readiness flags and developmental levels are considered to have the most policy-relevant value. This data is valuable to researchers who want to examine how pre-kindergarten programs influence children's kindergarten readiness and how kindergarten readiness influences children's experience in their future education. Existing research has primarily focused on the ECEAP program's effects on children's educational outcomes. However, the WaKIDS data has also been used to inform other early childhood education programs such as KinderCare or Head Start.

By merging WaKIDS data with OSPI enrollment records, researchers can analyze annual data at the teacher, school, or district level. WaKIDS source data includes information about the child, the school, and the teacher, however, ERDC considers CEDARS the source of record for K-12 information. In conjunction with other data sources like CEDARS, WaKIDS data can address research questions about kindergarten readiness and other educational outcomes, like those below:

- How might income level, special education status, or type of pre-kindergarten program impact kindergarten readiness?
- How does kindergarten readiness in turn affect [performance in later grades](#)?

One challenge to using WaKIDS data is its continued evolution since the program began. WaKIDS started as an assessment tied to full-day kindergarten. Full-day kindergarten began as a program within low-income schools. As a result, the initial years of WaKIDS reflected kindergarten readiness in the poorest school districts, based on the percentage of students who received free or reduced-price lunch. Full-day kindergarten was required of districts starting in the 2017-18 school year at which time WaKIDS was also required statewide. Researchers should account for this distinction when looking at the data over time.

Cut-point scores for determining kindergarten readiness also changed prior to the 2016-17 school year, which may explain some of the visible changes in the trends at that time. Another issue to note is that observational assessments may be impacted by bias, along with the extent of training and re-training of educators who conduct the assessments.

There are no known patterns or reasons for missing values, other than those resulting from structural changes in the program. While it is rare, duplicate student records may exist in the data files, where it appears that the same student took the assessment twice. ERDC considers the most recent assessment to be valid, so duplicate entries are likely an issue with early files when dates of assessments were not available. Other known quality issues are corrected during the data loading process. Students may have more than one record if they changed schools between the start of the school year and October 31<sup>st</sup>. However, if children transferred schools, the data collected from the first enrollment should follow the child to their new enrollment.

### **COVID-19 Impacts**

COVID-19 impacts on WaKIDS were three-fold. First, because COVID-19 impacted in-person school attendance in K-12, some data for WaKIDS that required in-person observation was difficult to collect. Second, COVID-19 impacts on a child's PreK experience could in turn impact WaKIDS scores. Third, kindergarten enrollment was reduced in the school years following COVID-19. Further study will be required to determine the extent of this impact.

Due to WaKIDS typically being administered early in the school year, WaKIDS data was collected in full during the 2019-20 school year. However, due to remote learning in the 2020-21 school year, the WaKIDS data for that year was unusually and systematically incomplete. For objectives that were more difficult to observe/assess remotely, data was around 85% complete versus the usual 99%. Nearly a quarter of students had at least one out of six areas incomplete where typically this was 3%

While the data for the 2020-21 school year was collected, OSPI did not provide this data to ERDC as it would not have been an adequate representation of the knowledge of entering kindergartners in the fall of the 2020-21 school year.

In the 2021-22 school year, WaKIDS data collection was improved and ERDC has WaKIDS data from 2021-22 onward. However, the impact of COVID-19 on WaKIDS scores from this point onward is a current topic of study in ERDC as we explore COVID-19 impacts on early learning.

### **Annual Statewide Assessment**

The Office of the Superintendent of Public Instruction (OSPI) provides ERDC with an annual file of results of standardized state testing in the K-12 public schools. It covers the 2005-06 school year forward and is loaded into the P20W data system. Statewide assessments measure the progress of students in third grade through 11<sup>th</sup> grade and within the educational system as a whole. Results are one measure of accountability in the Washington School Improvement Framework—each school is measured on the

framework and OSPI uses results to identify schools for additional support. Assessments are also required for federal accountability purposes.

Included in the assessment data are annual student level assessment results for all standardized, statewide tests in Reading, Writing, English Language Arts (ELA), Math and Science. Assessments are typically administered in the Spring, but can occur at other times of the year. In addition to public school students, the file also includes any home school or private school students who took assessments. This file does not include the Washington Kindergarten Inventory of Developing Skills (WaKIDS) or English Language Proficiency Assessments.

The types of assessments and grade levels in which they were administered changed multiple times based on state academic achievement standards (see Table 4). Various alternative tests, for students with significant cognitive challenges, were also administered during this timeframe (see Table 5). OSPI maintains a [timeline of the changes](#) to the statewide assessments for each subject. Care should be taken when comparing scores and passing rates between tests. For more information, see [ERDC documentation](#) on assessment comparisons.

**Table 4: Statewide Assessments by Grade Level and School Year**

Grade Level	2004-05 to 2007-08	2008-09	2009-10	2010-11 to 2012-13	2013-14 to 2015-16	2016-17 to present
3	WASL	MSP	MSP	MSP	SBA	SBA
4	WASL	MSP	MSP	MSP	SBA	SBA
5	WASL	MSP	MSP	MSP	SBA	SBA
6	WASL	MSP	MSP	MSP	SBA	SBA
7	WASL	MSP	MSP	MSP	SBA	SBA
8	WASL	MSP	MSP	MSP	SBA	SBA
10	WASL	HSPE	HSPE; EOC-Math	HSPE; EOC-Math/Sci	EOC-Sci	SBA
11					SBA	WCAS

WASL = Washington Assessment of Student Learning; MSP = Measurements of Student Progress; HSPE = High School Proficiency Exam; EOC = End of Course (Math: Year 1-Algebra 1 or Integrated 1, Year 2-Geometry or Integrated 2; Science: Biology); SBA = Smarter Balanced Assessment; WCAS = Washington Comprehensive Assessment of Science

**Table 5: Statewide Alternative Assessments School Year**

2004-05	2005-06	2006-07 to 2007-08	2008-09 to 2012-13	2013-14	2014-15 to present
DAWL	DAWL	DAWL	DAPE	DAPE	AIM
PORT	PORT	PORT	PORT	PORT	SBA-Basic
	WAMO	WABA	HSPB	AIM	
				SBA-Basic	

DAWL = Developmentally Appropriate WASL; PORT = Washington Alternate Assessment System Portfolio; WAMO = WASL-Modified; WABA = WASL-Basic; DAPE = Developmentally Appropriate Proficiency Exam; HSPB = High School Proficiency Exam – Basic; AIM = Access to Instruction and Measurement; SBA-Basic = Smarter Balanced Assessment - Basic



High school ELA and Math assessments are tied to graduation requirements, which have changed over time. A notable change relating to assessments was the shift from requiring End of Course (EOC) assessments to requiring comprehensive assessments beginning in 2015-16. The phasing out of EOC assessments lasted a few years. See the [OSPI Graduation Pathways webpage](#) for more information on legislative changes to graduation requirements and how these impact the assessment requirements. Additional high school graduation requirement details can be found on the [Washington State Board of Education \(SBE\) website](#).

Data elements include:

- Student data – school, district and grade level when tested as well as flags indicating if the student attends private school or is home-based
- Assessment information – test type, test grade level, test administration time period (e.g. Spring Test Administration), date the student took the assessment, and subject name
- Assessment results – scale score, performance level, standard met indicator for whether minimum standards were met, and attempt code (whether the student did or did not take the test and the reason why)

The granularity of this data is each assessment taken by each student. Assessments are uniquely identified within a year by test administration time period, test type, subject name, test grade, reporting grade, attempt code and score. Students typically have multiple assessment records in a year, usually one for each subject. In a small number of cases, students take the same subject tests multiple times in a year. In such instances, OSPI will typically report on the results from the first date the student tested.

ERDC receives one data file from OSPI annually, covering the previous school year. After a quality review, all records are loaded into the P20W data system, regardless of whether ERDC has an enrollment record for the student.

Data quality is generally good, since assessment results are required for federal reporting and are used, along with other metrics, by the state to determine school eligibility for additional support. District and school coverage is comprehensive each year, although not every enrolled student will have an assessment record as not all grades are tested each year.

Assessment results, such as scores, performance level and standard met indicator, are typically missing for 10-20% of records. This level of missing data is expected, since there are cases where students did not complete the assessment during that school year due to refusal, being exempt or absent, being no longer enrolled, or they passed the assessment previously. Some years have higher rates of missing results data. However, this is not a concern as most of these students have assessment results reported in another school year.

For high school assessments, missing results are caused primarily by students taking and passing the assessment (i.e., meeting standards) when they were enrolled in a grade level lower than the grade level of the test. In such cases the assessment record with results is reported for the school year when the

student took the test. An additional record should be included for the student in the school year when the assessment would be required, with no results data and an indicator in the Attempt field indicating that the student 'previously passed'. For example, if a student completed a grade 10 Algebra assessment when they were in 8<sup>th</sup> grade in 2010-11 and passed, then their score and other results data would be in a 2010-11 record with a Test Grade value of 10 and Reporting Grade value of 8. The student would also have a record in the 2012-13 school year, for the 10<sup>th</sup> grade math assessment, but with no results data and an attempt code indicating that they had previously passed. Researchers should therefore look across all high school years and middle school years to find assessment score and performance level results for the 10<sup>th</sup>-11<sup>th</sup> grade tests for each student, rather than focus on one reporting or test grade level. To determine if students met standards on an assessment, it is important to include cases where the Met Standard Indicator is 'yes,' or the Attempt code indicates that they previously passed.

When the state transitioned to using the Smarter Balanced Assessment (SBA) test in the 2013-14 school year, schools were allowed to choose whether to participate in a pilot of the new Smarter Balanced Assessment (SBA) or take the legacy test. The SBA pilot was only offered in third grade through eighth grade in that year and 91% of those assessments have missing results data. Researchers are advised to not use data for SBA assessments taken in the 2013-14 pilot year.

The entire data set is fully usable for research, with some caveats. Researchers typically want to know if students met standards in a subject area for a particular grade level of the test. Data for the 2005-06 and 2006-07 school years should be used with caution due to entirely missing Test Grade data. Reporting Grade (i.e., student's grade level when they took the test) is well-populated for those years, and OSPI suggests using Reporting Grade as a substitute when Test Grade is missing. However, students may test when they are in either an earlier or later grade than the grade level of the test. This scenario is most common in high school, especially for alternative test types (i.e., students take a test that is at a test grade level lower than their enrollment grade) and EOC Math and Science tests (i.e., students take a high school-level test when they are enrolled in middle school).

Another limitation is that indicators for home-school and private school students are missing for years 2005-06 through 2007-08. This information is only necessary if researchers are using the data as a standalone dataset and not merging it with other K-12 data.

Raw and scale scores are not useful for comparisons between students who take different test types within a subject or in different years. These data elements are not standardized across test types, subject or years. Therefore, scale scores should be used only for comparisons made within a single school year, test type (e.g. WASL or SBA), and subject area. Scale scores are not recommended to directly measure student growth over time. Performance levels, however, can be used to compare growth over time.

Assessment data is typically joined to an enrollment cohort and not used as a standalone file. Assessment results measure student learning and academic performance and, as such, are considered outcomes of K-12 or early childhood education. OSPI uses assessment data as a key indicator of school and district success, publishing results data in their [Report Card](#).

## COVID-19 Impacts

Due to the COVID-19 pandemic, statewide assessments were not administered during the Spring of the 2019-20 school year. During the 2020-21 school year, many schools were still [operating remotely or with hybrid schedules](#), and OSPI delayed the Spring assessments to the Fall of the 2021-22 school year. No assessment data from the 2019-20 and 2020-21 school years are available from ERDC. The normal schedule of Spring administration resumed during the 2021-22 school year. Therefore, the data for the 2021-22 school year contains the typical Spring assessments as well as the delayed assessments that were taken in the fall. Usually there are only a small percentage of assessments taken in the fall. Given the unique situation surrounding the administration of these assessments shortly after students returned to in-person school, it is not recommended for the Fall 2021-22 data to be used in analysis.

## School and District Characteristics

ERDC compiles school characteristic data from the Office of Superintendent of Public Instruction (OSPI) school-level student demographic and teacher data files as well as urban-rural locale data from the National Center for Education Statistics (NCES) Education Demographic and Geographic Estimates (EDGE) program.

For the 2004-05 through 2016-17 school years, all data for school-level student demographics, teacher characteristics and classroom size were contained in the OSPI 1\_2\_Demographic Information by School file. For the school years 2017-18 and following data files are extracted from the [OSPI Data Portal](#) website:

- Report Card Enrollment
- Report Card Teacher Demographics
- Report Card Teacher Ratio Program Certificate

NCES locale data were extracted from the [Common Core of Data](#) for the 2004-05 through 2014-15 school years and [EDGE](#) for the 2015-16 through current school years.

Data elements in the compiled school characteristic file contain records at the school and school year level. Variables include:

- School identifiers – ESD name, District and School Codes and names (OSPI and NCES)
- School descriptors – grade span serviced, school level, school type
- Enrollment characteristics – total enrollment count, enrollment and percent by student demographics (gender, race/ethnicity) and program participation (migrant, multilingual, special education, free/reduced price lunch, foster care, homeless, military parent, mobile, disability, and highly capable)
- Teacher and classroom characteristics – number of classroom teachers, ratio of students to classroom teachers, average years of teacher experience, and number and percentage of teachers with at least a master's degree

- Geographic indicators from NCES – metropolitan locale, urban/rural locale indicator, geographic setting, latitude and longitude, and Spatially Interpolated Demographic Estimates (SIDE) program income-to-poverty ratio estimate

As a caveat, some of these data elements are not complete for all years and are calculated for some years. Interested researchers are advised to refer to the “Variables Data” tab of the compiled file.

## Reengagement Program

As the Office of the Superintendent of Public Instruction (OSPI) describes on their [Open Doors Youth Reengagement](#) website, “Open Doors Youth Reengagement ([WAC 392-700](#)) is a reengagement system that provides education and services to older youth, ages 16-21, who have dropped out of school or are not expected to graduate from high school by the age of 21. Open Doors reengages disconnected youth through programs that encourage community partnerships, create multiple pathways for students to realize success, and provide an on-ramp to post-secondary achievement through a performance-based, individualized support model.”

Open Doors youth reengagement programs and schools were developed to provide multiple pathways to demonstrate career and college readiness. Students can earn their GED and participate in postsecondary or work readiness education (GED plus), earn a high school diploma or earn college credits/certificates or a two-year degree. It was the Legislature’s intent to encourage partnerships among school districts, community and technical colleges and community-based organizations to provide appropriate instruction and services that enable students to become productive members of their community.

OSPI annually provides ERDC with an end-of-year data file encompassing the experience of the reengagement students over the past school year as reported by the Youth Reengagement service providers. Files include data since the 2015-16 school year. For information on the user guide application for the end-of-year file, see [Open Doors Youth Reengagement End-of-Year Reporting](#).

The data files contain records for each student’s Open Doors enrollment/program participation segment during the school year. Variables include:

- Serving and Resident school and district
- Student demographics – age, race/ethnicity, gender
- Student academic characteristics – grad requirement year (GRY), Number of Days Enrolled, High School Equivalency Certificate, High School Diploma, Certificates Earned, College Degree, Industry Recognized Certificate, etc.
- Student program participation – Program Exit Date, Exit Description, Migrant, Special Education, FRL, Homeless, 504, LAP, Limited English Proficiency

During the COVID-19 pandemic, enrollment in Open Doors declined. At the end of the 2019-20 school year and through the 2020-21 school year, most schools were [operating fully remotely or with hybrid schedules](#). The impacts have not yet been analyzed by ERDC.

ERDC has published reports utilizing the yearly cohorts for following the annual cohorts from three to five years using outcome information from the P20W data system. These reports can be found on ERDC's website under the Publications section.

## Student, Teacher, and Course Data

### *Comprehensive Education Data and Research System (CEDARS)*

The [Comprehensive Education Data and Research System](#) (CEDARS) contains data reported to the Office of the Superintendent of Public Instruction (OSPI) by each public school district in Washington since the 2009-10 school year. This information covers the majority of administrative data intake for all public school districts, covering topics ranging from enrollment to discipline. Every student who enrolls in any Washington public K-12 school is represented across the files contained within CEDARS and can be tracked over time through their Washington public education career. Students who participate solely in home and/or private school will not be represented in this data. See the [OSPI CEDARS Manuals](#) for a full list of data elements collected. Note that not all elements are provided to ERDC.

#### **Data Loading**

School districts have their own student information systems or vendor-operated systems that are managed at the district level. Districts transfer their local data into the CEDARS system periodically throughout the year, in addition to submitting prior school year data updates. OSPI extracts data from CEDARS and provides it to ERDC for loading into the data system twice per year – once for the preliminary data files for the current school year and a second time for the final data files for the prior school year. Final data is typically available for research by December.

#### *CEDARS Absence Data*

The CEDARS Absence file provided to ERDC is an event history file representing the occurrences of student absences in Washington public K-12 schools. The information, collected from districts in this data, describes the type of absence (full-day or partial-day, excused or unexcused) and date of absence for the school years since 2012-13. According to WAC 392-401-015, a student is absent when they are: (a) not physically present on school grounds; and (b) not participating in the following activities at an approved location - (i) instruction; (ii) instruction-related activity; or (iii) any other district or school approved activity that is regulated by an instructional academic accountability system, such as participation in district-sponsored sports.

The Absence data includes one record for each student served in the district during the current school year, for each absence type associated with the student for each school the student is enrolled. A school will report the student's absences while they were enrolled, even if the student withdraws prior to the

end of the school year. Absences must be reported for students in grades K–12. If attendance is tracked for preschool students, then those absences may also be reported in this file. A student can have multiple absence records per day at multiple schools, including combinations of excused and unexcused full- or partial-day absences, so care should be taken when aggregating. For example, a student with an unexcused full-day absence record and an excused full-day absence record for the same date should not have the records sum to two days absent.

The base analytic unit in the data is at the student level, but additional aggregations at the date, school, and district level are possible. Aggregating data by absence type is discouraged, but can be used to flag if a date is associated with a full-day absence. Types of absence and absence date are the primary analysis components of the dataset.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. This heavily impacted the absence data, with [definitions and reporting guidance changing](#) during the 2020-21 school year to address remote instruction.

#### *CEDARS Discipline and Exclusionary Discipline Data*

The discipline data provided to ERDC and collected in the CEDARS Student Discipline File (2012-13 through 2017-18) and the Student Exclusionary Discipline File (2018-19 forward), captures information regarding behavior and discipline actions for students involved in incidents during school or school-related activities that resulted in removal from their regular education setting. The data collection standards were developed by a special [student discipline taskforce](#), including “elements of education services, petitions for readmission, credit retrieval, and school dropout as a result of disciplinary action”, and were incorporated into the CEDARS collection.

These data are a collection of administrative records reported to OSPI based upon the discretion of each school district to determine when a behavior is an incident for which disciplinary action was taken. For the 2012-13 through 2017-18 school years, only the final or most serious disciplinary action (intervention) is included in a single record. For all years, the most serious behavior is entered into the Behavior Code field, and any other behaviors are listed separately. For 2018-19 forward, the Exclusionary Discipline data include a separate record for each exclusionary action for a single incident. If multiple students are associated with the same incident, then one record will be submitted for each action for each student with an exclusionary action.

In loading the Discipline data to the P20W data system, ERDC combines the data from both the old and the new file formats into one file structure. Incident IDs connect all students associated with a particular incident and could serve as a unit of analysis. Data could also be aggregated by behavior or type of intervention.

- Each record includes: District Code, School Code, Incident ID, Incident Date, and School Year.
- Key fields for research: Behavior Type, Type of exclusion/intervention, Date of exclusion/intervention, and number of days or total time of exclusion/intervention.

Researchers should be aware of the state-level policy and rule changes that impact the discipline data collection as well as limitations with this data relating to the file change in 2018-19. Analyses using all available years will require special handling. Between 2012-13 and 2014-15, several changes were also made to improve the data collection process and better address reporting requirements. Fields impacted by these changes include intervention, emergency expulsion, behavior fields, and staff involvement.

Most of the research centered on discipline data aims to address equity issues in education, such as the discipline rate and discipline days rate reported for the [OSPI Report Card Data for Schools and Districts](#). Additional data must be joined with these data to examine inequities in student discipline, e.g., race/ethnicity, program participation. Due to small numbers of exclusions for many schools and districts, only non-redacted data or top/bottom coded data can be reported by OSPI (see [Suppression Rules for Public Reporting](#)). Most of the analyses and group comparisons may only be performed at the state level, or for the largest and most diverse districts.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. This heavily impacted the discipline data.

#### *CEDARS English Learners Data*

The CEDARS English Learners (EL) data provided to ERDC includes administrative information on students receiving PreK-12 services since 2009-10 from Washington K-12 public school districts. The primary program is the State Transitional Bilingual Instructional Program (STBIP) and, starting in 2016-17, Native American Title III English Language Development Services data were also provided to ERDC. STBIP is also a federal Title III program.

As OSPI describes on their [Bilingual Education Program webpage](#), “both programs share the same goal: develop language proficiency that enables meaningful access to grade level curricula and instruction.” To determine program eligibility, students’ English language proficiency is tested. Students who were screened and did not qualify for services are also included in the data. Information about students’ primary language is not covered by this table, but those details can be found in the CEDARS enrollment table.

There is one record for each enrollment of a student into an English language program at a school for each school year.

- Key variables: program designation, program exit reason, start and exit dates, school year, instructional model, placement test, placement status, and grade level at placement.
- Each record contains: start date, district code, school code, school year, and program code.

Not all columns of this table are included during all years. For example, Placement Test fields and Grade Level at Placement are available from the 2012-13 school year forward. Some students have multiple records in a school year due to changes in Instructional Model.

Overall, the EL table contains high quality data. At the district level, year-to-year program enrollments are generally stable. Most columns are highly complete, but some fields should be used with caution due to high missing rates or non-conforming logic. Initial US Placement Date, Number of Months of US Attendance, and Number of Months of Non-US Education have high missing rates for the two years they were collected. All other columns are considered relatively complete. A small percentage of records violate CEDARS Manual business rules, such as exit dates occurring before start dates, an exit code with no exit date, or an exit date with no exit code. This issue occurs most frequently in the early years of CEDARS and improves after the 2011-12 school year.

Researchers should be aware that the meaning of the value for test level status varies based on the placement test taken by a student. For details, refer to appendix N in the [CEDARS manual](#). *The test data contained is used for screening/placement purposes only and should not be considered as a proxy for proficiency assessment results.*

This dataset is typically used to explore the characteristics and participation of students who receive English Language Learner services. It could also be used to study the effectiveness of particular instructional models, examine postsecondary outcomes for Native American students who receive Title III English Language Development Services or study the relationship between postsecondary completion and the timing of bilingual program exit.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. As a result of reducing and delaying in-person assessments, districts reported a provisional placement status with the expected state adopted placement assessment value until the actual placement assessment could be administered. If found to be eligible based on formal screening, a new record would be entered for the student. This guidance was continued in the 2021-22 school year. The impact of this change and the COVID-19 pandemic impacts on program participation, more generally, have not yet been analyzed by ERDC.

### [CEDARS Enrollment Data](#)

The K-12 enrollment table provided to ERDC is extracted by OSPI from a number of CEDARS Files: Location, School Student, District Student, Student Attributes and Programs, Race and Ethnicity. The core of the file includes student enrollment information by school and student characteristics. It covers the PreK-12 grades from the 2009-10 school year forward. Information on whether a student is documented as having a disability can be found here, as well as the type of disability.<sup>10</sup> This data also contains an indicator of student homelessness as defined by the McKinney-Vento Act, Section 725(2).

This is a student-level data set that includes every student enrollment segment in Washington K-12 public schools for each school year, including non-primary enrollments. The enrollments are the basis

---

<sup>10</sup> See [WAC 392-172A-01035](#) for a description of disability codes.



for loading other CEDARS-sourced data into the P20W system, in that only the records with a corresponding school enrollment are loaded from these tables (e.g., absence, program, discipline).

Data are generally complete in all records for gender, race/ethnicity, date enrolled, primary school, grade level, and days present. Other columns have missing data as expected, based on CEDARS Manual business rules or data collection changes. For example, not all students have a disability, withdrawal code only applies if a student leaves a school, and military parent/guardian status was not collected until the 2016-17 school year. When new data elements are collected, initial years may have higher rates of missing data.

During the COVID-19 pandemic, enrollment in public K-12 schools declined by 3% in the 2020-21 school year, with some families opting for home school, private school, delaying entry into kindergarten, or moving out of state. Enrollment declines were focused in younger grades, Running Start, and Open Doors. At the end of the 2019-20 school year and through the 2020-21 school year, most schools were [operating fully remotely or with hybrid schedules](#), which heavily impacted the data surrounding attendance. There may have been additional impacts on student mobility and student characteristics, such as homelessness, but they have not yet been analyzed by ERDC.

### *CEDARS Grade History*

The CEDARS Grade History File (GH), as provided to ERDC, includes student course-level records from Washington K-12 public schools for all courses taken for high school rigor credit. Transfer courses obtained outside the reporting district were included as of the 2016-17 school year. High school rigor course data has been collected annually since 2009-10. However, inconsistent data completeness across districts and school years has been identified in the early years of data collection, so researchers should exercise additional caution if using data prior to 2012-13. Generally, these data can be used to answer questions about credit accumulation, course-taking trajectories, or courses taken in a single school year. Note that this data is not equivalent to a student's transcript.

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

High school-level students are the major cohort for this dataset. However, there are course records indicating students of earlier grade level (i.e. 8<sup>th</sup> grade) who completed high school rigor courses prior to progressing to a high school grade level. This dataset also provides information about coursework for students in dual credit programs (i.e. Career and Technical Education-Dual Credit, Advanced Placement, College in the High School). Running Start is considered a transfer course and is therefore available since 2016-17.

The data collection process requires districts to report each student's entire high school rigor course-taking history each year (current year and all previous years). Since all course records are reported each year, duplicates are common. Researchers are strongly encouraged to examine missing patterns and conduct record deduplication before beginning any data analysis to address these inconsistencies. There are several other potential issues relating to analyzing course records.

- **Credits Attempted/Earned:** Outlier records show an abnormally high number of credits earned in a single school year. It is unclear whether some students actually earned a large amount of credits, or if the elevated amounts are the result of data entry or loading errors.
- **Subject Area:** The variable "state course code" is used to categorize complex coursework structures, based on the five-digit NCES-SCED course codes. This coding scheme provides only information about subject area (i.e. English language and literature, mathematics, etc.) and course identifiers within each subject area. However, these are assigned to courses by the districts, with no centralized system, so course code assignments are not consistent across the state. Course codes and titles recorded by each school or district may provide proximate information about course content. Additionally, districts may have recorded the same courses differently across time (i.e. a math course was coded as science). Comparisons across cohorts over time should be taken cautiously. There is currently no crosswalk between state course code and the course information provided by each school and district.
- **Specialized Courses:** The variable "course designation code" is used to identify type of courses (e.g. dual credit, Honors, etc.). Those codes can be automatically set in the district database and may not get updated systematically when changes occur. The lack of systematic updates may result in inaccurate counts of specialized class types and should be compared to official counts.
- **Course Sequence:** GH also has no direct measures of course level, sequence, or Carnegie unit.
- **Prior Coursework:** GH only collects high school rigor courses. It does not provide information for studies attempting to track course-taking from earlier grade levels, nor does it include summer school courses.
- **Alternative Learning Experience (ALE):** ALE relates to courses that serve students outside a traditional classroom setting, such as an alternative site or remotely. This data has been collected since 2013-14 with values and reporting guidance changing over time. ERDC has not thoroughly analyzed this data.

The current GH file was not designed to observe whether policy changes align with the implementation of more rigorous state assessments. Researchers looking to examine associations between high school coursework and student outcomes should use this data with caution. Proxy measures for course level or sequence may be constructed with careful articulation of the data collection process and the definitions of each course coding logic. It is important to test the validity and reliability of each construct by considering the consistency and completeness across districts and years.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. New Letter Grade values were implemented to indicate Incomplete or Credit Waived due to COVID-19 school facility closures which

were only to be used for the graduating class of 2019-20. Given the broad differences in the learning environment during the 2019-20 and 2020-21 school years, grading may exhibit differences compared to other years. There was also an influx of new online program codes during these years. The impact of these changes and the COVID-19 pandemic impacts on course-taking, more generally, have not yet been analyzed by ERDC.

### *CEDARS Program Data*

The CEDARS program data provided to ERDC contains administrative records from school districts on students who participated in or received services from specific PreK-12 programs, eligibility for Free and Reduced-Price Lunch (FRPL) and selected student attributes from the 2009-10 school year forward. There have been several changes over time. Some programs have begun and/or ended, while some programs have been moved to or from other CEDARS collections. For example, Title III language instruction for English learners and immigrant students was moved to the English Learners collection in 2011-12. Therefore, not all OSPI programs are included in this table in all years. Further, Special Education and Career and Technical Education programs are not collected in this table. Commonly requested programs that can be found in this table are FRPL, gifted/highly capable, Section 504, Career Launch, and reengagement.

All variables in this table are reported in all school years (see program-specific notes below), with the Program ID variable being the key component. Since funding streams are often tied to program participation, this data is regularly used for state and federal reporting, which makes it subject to audit and review. While most schools are represented in the Program data, not every school will have applicable data to report. Data contain information at the student level on:

- Name of the program in which the student participated
- School and district in which they participated in the program
- Date that the student began receiving services in the program
- Reason why the student qualified for the program
- Date that the student exited the program and the reason for exit

A program record contains an exit date if the student exited the program. However, districts have different practices around entering these data, so researchers should be aware of possible inconsistencies. A small number of records that have exit dates in the future are not actual exits. Future school year exit dates are also used differently among districts, due to differences in district practices for rolling over to a new CEDARS year. Of the most frequently researched school programs, the FRPL program and Section 504 program have high percentages of exit dates occurring in future school years. Note that records will also commonly not have an exit date, but most records with exit dates will have an exit reason code. Records that have an exit reason code but no exit date are rare.

The Qualification code is only required for certain programs (see [CEDARS Manual](#)). While the qualification codes are collected for FRPL, they are not provided to ERDC. The utility of the qualification code differs by program. For example, the Seal of Biliteracy program uses the code to designate the

language(s) in which the seal is associated while the Reengagement program uses the code to describe where the student is receiving services.

**Program-Specific Notes:**

- FRPL eligibility extends from the previous school year to the first 30 serving days of the new school year, or until eligibility is determined. Not all schools participate in the program, including virtual academies and skill centers. As part of the Community Eligibility Program, entire schools can be deemed eligible if enough of the enrolled students are individually eligible. Additionally, the income eligibility thresholds for this program have changed over time, so use caution when comparing populations over time.
- When utilizing disability status and Section 504 information, it is important to recognize that very few students with a 504 plan for accommodations have a disability, and few students with disabilities have a 504 plan. Eligibility for protection under Section 504 is a physical or mental impairment that limits at least one major life activity. Impairments can be permanent or temporary, so it is possible for a student to have a 504 plan for only a few months, only a few school years, or for their entire K-12 career.
- The Migrant Education Program data was collected in CEDARS through the 2011-12 school year. Since the 2012-13 school year, this data was collected in the Migrant Student Data Records System (MSDR). Low record counts were seen in the initial years of this change and are not an accurate reflection of actual program enrollment during those years. Additionally, school code is not available for the Migrant Education Program in all years, but can be identified through cross referencing enrollment records.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. Temporary policy changes increased access to FRPL broadly during the 2020-21 and 2021-22 school years. There are likely additional impacts to participation for other programs, but they have not yet been analyzed by ERDC.

***CEDARS Race and Ethnicity Data***

The CEDARS Race and Ethnicity data include parent/guardian-reported (self-report) or observation-reported data on the race and ethnicity of each student in each district, using a two-part question, starting with the 2009-10 school year.

OSPI implemented this two-part race and ethnicity question as part of the CEDARS data collection process. Schools were first required to submit detailed race and ethnicity information for each student during the 2010-11 school year, using 49 race codes and ten ethnicity codes, though schools could submit it in the new format during the 2009-10 school year.

For students that did not self-report, federal guidelines specify that for Washington state, a student (or the parent/guardian on behalf of the student) is not required to identify their race and/or ethnicity on school forms. However, if a student (or parent/guardian on behalf of the student) does not complete

the two-part question on race and ethnicity, by law, school personnel must use ‘observer identification’ to select the race and ethnicity of the student.

In 2018-19, expanded ethnic and racial categories were added to CEDARS, with full implementation required by the 2021-22 school year. See the OSPI publication, [“Race and Ethnicity Student Data Task Force Guidance for the Washington State Public Education System”](#) for further details. The gradual implementation of the expanded categories is evident in the data. For the 2018-19 school year, about 10% of students had reported racial and ethnic information based on the expanded categories. The remaining 90% had not been resurveyed, and their data was carried forward from earlier years.

Each student in a school district is represented by an annual record for each racial and ethnicity category they identified with in their responses. Every record includes District Code, school year, and Ethnicity or Race Descriptions as well as Collection Method (beginning in 2018-19). The data is the source for the Federal Rollup Race/Ethnicity field included in the CEDARS Enrollment Data file that ERDC receives. For this field, student race and ethnicity responses are aggregated into a single field. If the student identified with any type of Hispanic code, then they are assigned to the ‘Hispanic’ category. If the student identified with two or more major non-Hispanic race groups, then the student is assigned to the ‘multiracial’ category. If the student identified with a single non-Hispanic race category, they are assigned to that category (American Indian/Alaska Native, Asian, Native Hawaiian/Pacific Islander, Black/African American, or White). For a deeper analysis of these data tables, see the ERDC publication [“Reporting on the racial and ethnic identities of Washington public K-12 students”](#).

This data is considered complete and sufficient for analysis at the state level, as race and ethnicity data are required to be submitted in conjunction with student enrollment data. In any given year, a very small number of students lack data in the Federal Rollup Race/Ethnicity categories in the Enrollment file. Each of the ethnic and racial groups are large enough to be analyzed at the state level. At the school district level, numerous heterogeneous communities will have data available for many of the racial/ethnic subgroups.

Numerous reports from ERDC and research partners have used earlier versions of this data to investigate subgroup differences in Racial/Ethnic categories. By creating a dataset with binary flags for each subgroup category ERDC and external researchers can construct various meaningful aggregations relevant to their research goals. For example, a multi-racial student can be used for calculations for each of the subgroup categories selected. Researchers should take care to construct numerators and denominators to include the proper subset of students as well as exercise caution with small subsets of ethnic/racial groups.

Despite the gradual implementation of new race and ethnicity questions starting in 2018-19 and some bias in the data due to over-responses (i.e., selection of all possible options for both the ethnicity and racial categories), these enhanced data elements will enrich the research conducted by ERDC and affiliated researchers. This data enables more in-depth, detailed analysis of student outcomes for racial and ethnic subgroups than the federal multiracial category. Response bias may be enough to limit

analysis of small subgroups at the school district level, but the minimal noise introduced at the state level should not significantly alter any results for groups or subgroups.

### *CEDARS Special Education Data*

The CEDARS Special Education data provided to ERDC has information on students who receive special education services from Washington K-12 public school districts from the 2009-10 school year forward. Special Education is a district-level program. Data cover information on the level of support and services students receive, how long students receive services during their K-12 career, and why they exit the special education program. Note that not all students with a disability will have records in the Special Education table. For data on student disability and disability code, see the CEDARS enrollment section. For information on Section 504, see the CEDARS program section.

There is one record for each enrollment of a student into the special education program at a district for each school year. Typically, there are not multiple entries for a student within both a district and a school year. The key components of this dataset relate to the Least Restrictive Environment (LRE) where a student receives their education, program exit reason, start date, and exit date. LRE offers information on the level of support and services being provided to students.

Every record has a school year, an LRE code and description, a start date, and a district code. Since records only contain an exit date when a student leaves the special education program, it is common for this field to be missing. Consistent with the business rules in the [CEDARS manual](#), all exit dates occur after start dates. A small number of records have an exit code with no exit date, or an exit date with no exit code. Due to collection requirements being optional or situational, missing values are common for some variables, and not all variables in this table are collected during all years.

There have been several changes to LRE codes and exit reason codes in CEDARS over time, with some codes discontinued and new codes introduced. Generally, LRE codes can be grouped by age ranges 0-2, 3-5, and 6-21 years old. The definition of the start and exit dates have also changed over time, notably in the 2013-14 school year when they shifted from statewide service dates to school year specific dates. A separate Initial WA Service Date field was collected for the statewide information. There are typically not multiple entries per student per year at a location. In each of the date variables, there is a small percentage of unrealistic dates that are likely to be errors.

To measure the length of time that students receive special education services (i.e. dosage), the definition change for start dates may present a challenge. For analysis purposes, researchers should expect to treat pre-2013-14 data differently than data from 2013-14 onward.

Since funding is tied to special education program participation, this data is regularly used for state and federal reporting, which makes it subject to audit and review. For this reason, the overall quality of the CEDARS special education data is considered reliable. Data are consistent with CEDARS Manual business rules and the data elements of interest for use in research are highly complete and allow for longitudinal analysis.

Most districts have consistent enrollment trends from year to year. However, enrollment changes can occur if the state statute governing the level of special education funding changes or services are moved to a different agency. For example, state funding for children under the age of three shifted from OSPI to DCYF in 2020-21 and the special education records in CEDARS for this age group dropped significantly. The majority of the remaining records pertain to school districts that are also early intervention service providers, separate from education services.

During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. This may have had an impact on program participation, but it has not yet been analyzed by ERDC.

### *Core Student Record System (CSRS)*

The Core Student Record System (CSRS) data contains records reported to the Office of the Superintendent of Public Instruction (OSPI) by each public school district in Washington between the 2004-05 and 2008-09 school years. Broadly, this data source collected much of the same data as CEDARS, such as enrollments, absences, and program participation, but with some important differences.

In contrast to CEDARS, CSRS did not have a robust process for data validation, so data elements were more likely to be misaligned in addition to difficulties resolving student identification across districts. This data can still be used with the understanding that many data values are different than what was collected after the 2008-09 school year in CEDARS and not all data elements will have a comparable field between the two sources.

## Staff and Educators

### *Educator Certification*

The Office of Superintendent of Public Instruction (OSPI) is the authority that issues educator certificates and maintains a database containing records of all teachers receiving teaching certificates per [RCW 28A.300.040](#). OSPI's E-Certification (E-Cert) data system allows a Washington state educator to apply for, renew, or have reissued their educator certificates. Links on the [OSPI E-Certification website](#) to user guides provide helpful information in understanding the relationships between educators, administrators, educator preparation programs, the Professional Educator Standards Board (PESB), and OSPI.

There are many different types of certifications for K-12 educators in the state. The types of certificates and the conditions that must be met for the issuance of each are set by PESB per [RCW 28A.410](#). PESB is an independent Board, housed under the umbrella of OSPI.

Educator certification data includes information about OSPI-issued educator certificates, including endorsements in various areas of concentrated instruction, e.g., science, math, history, music, etc. Data for all individuals with certificates issued in the calendar year 1970 or after are included in these files,

regardless of whether the individual has worked in public K-12 education. Each spring an update file is sent to ERDC with the past year's certifications, to be appended to the historical file.

To use certification data effectively, it is helpful to understand the laws and administrative code behind the credentialing of educators.

### Certificate Roles

Educator certificates are classified by role and are defined in [WAC 181-79A-140](#). Roles include various types of teachers, positions within the arena of Career and Technical Education (CTE), educational staff associate positions (ESA) that provide support services to students, and positions authorized to work in tribal schools.

### Certificate Types

Just as educator roles are defined in law, so are the accompanying certificate types that authorize the people in those roles for service in the schools.

The OSPI [Teacher Certificate](#) website provides a full list and description of the certificate types specified in [WAC 181-79A-142](#). Certificate types include those for full teaching duties, substitute teachers, limited or conditional certificates, tribal school teachers, and a variety of CTE certificate types.

Emergency teacher certification outlined in [WAC 181-79A-228](#) allows educator preparation programs to recommend a candidate for this emergency certificate if the candidate has completed all program completion requirements except for one or more of the assessments. As a result of closed testing centers during the COVID-19 pandemic, PESB issued guidance to let programs waive the assessments and issue an emergency certificate. A high number of emergency certifications were issued in 2020. This option was in effect until July 2023.

### Endorsement Data

Endorsements for teachers are related to the subject matter they are qualified to teach and are specific to certain grade levels. [WAC 181-82A-202](#) lists available certificate endorsements. [WAC 181-82-110](#) describes situations where teachers employed by a school district may be assigned to classes other than in their endorsement areas. There are no such provisions for CTE teachers. See [Endorsements](#) on the PESB website for additional information.

Currently, at least one endorsement is required for educator certification per [WAC 181-82A-215](#). A second endorsement is required for holders of bilingual education and English language learner endorsements (see [WAC 181-79A-132](#)). Some endorsements may only be added to an existing endorsed teacher certificate (see [WAC 181-82A-208](#)). Some older-issued certificates are not associated with an endorsement.



## CTE Endorsements: Endorsement codes and V-Codes

CTE endorsements are associated with V-Codes, 7-character codes with a 'V' prefix. Endorsements for CTE teachers qualify them for teaching specific CTE courses. There are two routes to CTE endorsements described on the OSPI [Initial Career and Technical Educator \(CTE\) Certificate](#) website:

- The College & University route ([WAC 181-77-031](#)) leads to endorsement in one of four areas:
  - CTE Agriculture Education (V010000),
  - CTE Business and Marketing Education (V078000),
  - CTE Family & Consumer Science Education (V200002), and
  - CTE Technology Education (V210100).
- The Business & Industry route ([WAC 181-77-041](#)) leads to endorsements in specialty areas within Agriculture, Business & Marketing, Health & Human Services, and Skills & Technology Sciences. Applicant must complete a PESB-approved preparation program.

College & University route endorsements are not available to those following the Business & Industry route. Applicants must apply for the Residency Teaching Certificate and use it to add the CTE endorsement.

Each recent CTE endorsement record is associated with a 4-character endorsement code, which is, in turn, associated with a specific V-code. Before 2008, the V-code was entered as the certification code; in almost all cases, no 4-digit endorsement code was entered. Starting around 2008, the V-code was made part of the description associated with the endorsement code. All CTE endorsement codes are in the 8XXX series, where the XXX is part of the associated V-code.

### Certification Data Files

All records contain fields for certification code and the corresponding issue date, effective date, and expiration date. If an endorsement code is part of the certificate record, fields are expected to be filled for endorsement issue date, effective date, and expiration date. A candidate will have a record for each endorsement they have earned, with the teaching certificate information attached that was current at the time the endorsement was awarded.

There are a few details to keep in mind when working with the certification data.

- Certificates are valid for periods ranging from a partial school year to seven or more years. All time periods of activity for a certificate will be reflected in the data file with appropriate dates, therefore a single person can have several records.
- It is common for an educator to have several certifications/endorsements in effect in a single school year. There is a field for the status of a certificate or endorsement that indicates whether it is active, expired, suspended, or any number of other statuses.
- There are three different identifiers included in the dataset to choose from when the file needs to be joined to another data source. These identifiers are:

- **Certification Number.** This is the official number assigned by OSPI to identify a person as an educator in Washington, regardless of what kind of certificates or endorsements they are qualified for. It is a 7 byte field, with the first six bytes as numeric and the last byte as an alpha character (letter). This is a permanent, lifetime identifier, much like a social security number.
- **Personnel ID.** This is a number assigned in the OSPI dataset that tracks an educator's work history in the state K-12 system. This ID follows the person from job to job, district to district, role to role.
- **Educator ID.** This is another identifier used in other OSPI datasets that can be used as a linking variable if the need arises.
- Demographic data can be obtained by linking to OSPI's employment history dataset, known as the S-275, which ERDC receives along with the E-cert file every year. It contains the above three identifiers also to facilitate linking.

### COVID-19 Impacts

In 2020, in an effort to mitigate the progression of the COVID-19 pandemic, testing centers where educator candidates could go to take their required assessments were closed, offering no alternative method for taking the assessments. PESB issued guidance allowing educator preparation programs to waive assessment requirements for their candidates and provide them with recommendations for emergency certificates, good for one year, renewable until they could take and pass the required assessments. In 2023 the waiver option was rescinded and anyone still on an emergency certificate was required to complete the assessment requirements and obtain their program's recommendation for full certification. In the data these actions will be noted as an increase in emergency certifications from 2020 through mid-2023. As testing centers opened gradually, the increase in issuance of full teacher certifications would be more gradual throughout the years of 2021-2023.

### K-12 Personnel Reporting

The S-275 Personnel File, collected by the Office of Superintendent of Public Instruction (OSPI), is an annual snapshot of public K-12 staff, both certificated and classified, who are employed or under contract for employment any time during the school year (September 1 – August 31) as of October 1 of that year. ERDC receives annual S-275 file updates each spring covering years 2002-03 through the most recent school year.

Employees of school districts, educational service districts, charter schools, and tribal compact schools are included. Employees leaving the district after October 1 are retained in the file. Employees hired after October 1 may be included in the final version of the file, as district reporting does continue throughout the year. However, in practice it is generally accepted that the file is a snapshot in time with an as-of date of October 1 of the reported school year.

Districts submit the S-275 October 1 snapshot data to OSPI during November. Four primary categories of information about an educator are submitted to the S-275 system: demographics, degree(s) held and

experience of the individual, contract specifics, and assignment information. Every district employing an individual submits this information to the S-275 system. Final corrections and updates to total salary and contract items are completed by the following September 30 after the end of the school year being reported.

Data collected in this system feed a variety of reporting processes that are either mandated by state law, necessary for calculating state funding, or are needed for responding to requests by the federal government, the Legislature, or other organizations. While teaching certificates relevant to the position held by an individual are recorded in this system, it is not the system of record for all certifications earned by Washington teachers. The E-cert system is a separate system maintained by OSPI that houses detailed information about the certifications held, regardless of employment status.

For researchers who will not be linking S-275 data with other data sources, files are publicly available through OSPI's [SAFS Data Files](#) website. For more information and guidance in requesting or using the data, contact the Educator Growth and Development team at OSPI or [visit their website](#).

Instructions for submitting data to S-275 system for the current and previous four school years are available on the OSPI [Personnel Reporting](#) website. Given the complexity and comprehensive scope of these data, it is useful for researchers to familiarize themselves with the information they contain. These documents provide a detailed explanation of all the data elements ERDC receives.

Things to keep in mind when working with this data source:

- Each individual employed by a district can have multiple records within a school year. Scenarios resulting in this include, but are not limited to, multiple part-time assignments, working for more than one district, or changing assignment mid-year.
- How categories of educators are defined can vary with reporting needs. Most categorization schema are based on combinations of duty codes, activity codes, and program codes. The Educator Growth and Development group (EGAD) at OSPI have documentation for how they use these variables to fulfill their specific reporting needs. Contact their office at OSPI for more information if needed.
- The same three identifiers for linking as are in the E-cert data set are used in the S-275 so the datasets can be joined. Those identifiers are:
  - **Certification Number.** This is the official number assigned by OSPI to identify a person as an educator in Washington, regardless of what kind of certificates or endorsements they are qualified for. It is a 7 byte field, with the first six bytes as numeric and the last byte as an alpha character (letter). This is a permanent, lifetime identifier, much like a social security number.
  - **Personnel ID.** This is a number assigned in the OSPI dataset that tracks an educator's work history in the state K-12 system. This ID follows the person from job to job, district to district, role to role.
  - **Educator ID.** This is another identifier used in other OSPI datasets that can be used as a linking variable if the need arises.

- Employees who are assigned within a district but are on some sort of approved leave will still have a record for each of their active assignments, but the associated FTE will be 0.

## Retention and Mobility

The Retention-Mobility files are constructed using the S-275 as the base and are enhanced with the incorporation of teaching staff information from OSPI's Comprehensive Education Data and Research System (CEDARS) student information system. The CEDARS system includes information about teacher assignments.

Information from the S-275 data that are used to populate the Retention-Mobility file records include each employee's assignment(s) for a school year, including the school(s), specific role (i.e., teacher, principal, educational staff associate), and full-time equivalent (FTE) of each assignment, as well as demographic, education and experience. One person may function in a variety of roles, and the case-data structure of the S-275 captures these situations and is replicated in the Retention-Mobility files.

The S-275 is first reported as a snapshot of the workforce as of October 1 of each school year. However, educators do get hired throughout the year, and these later hires are captured for the Retention-Mobility files from the CEDARS system. It is unknown if there were COVID-19 impacts to this data source.

## Postsecondary Data

### Educator Preparation Program (EPP)

The Educator Preparation Program Data Collection is a collaborative effort between ERDC and the Public Educator Standards Board (PESB) to collect and report on student level data from any organization or entity in Washington approved by PESB to offer training to candidates in preparation for certification to work as a public K-12 educator, administrator, counselor, or school psychologist in the state of Washington. [RCW 28A.410.210](#) and ERDC's governing statute ([RCW 43.41.400](#)) outline the regulation of educator preparation programs by PESB, and ERDC's duties relating to processing and reporting on this data.

The expressed intent of the legislature was to facilitate a better understanding of the training by, and best practices of, educator preparation programs, and the career paths that educators who complete those programs follow to ensure those who seek to become educators in the state of Washington are adequately prepared, and in sufficient supply. In particular, article (9) of this law charges PESB with *"maintain(ing) data concerning educator preparation programs and their quality, educator certification, educator employment trends and needs, and other data deemed relevant by the board."* ERDC is to hold and analyze the data in service of PESBs business needs, and eventually make the data available for other research studies conducted by qualified third party researchers.

Data collection began in 2018 with the collection of data for reporting years 2016-17 through 2018-19. Because this was the first collection attempt and programs were asked for historical data, the resulting data quality was not sufficient for research purposes. The following year, 2019-20, the collection elements were expanded and ERDC again received three years of data. Data quality had improved, but 2019-20 was deemed to be the first year of high-quality data this is suitable for general research purposes.

### Collection Structure

The EPP data collection comprises seven different data files that are submitted by preparation programs annually. Four of these files contain candidate level data: Demographics, Admissions, Assessments, and Clinical Experience. The remaining three files cover program meta-data, institution term information, and advisory board meeting information.

The elements contained within the candidate level files are reflective of the set of standards used by PESB to evaluate programs for initial and ongoing approval to offer training to potential public educators. The data are primarily quantitative and can be characterized as administrative. Some qualitative fields were built into the collection during its development, but are not widely utilized.

The candidate level files provide a comprehensive look at the activities and progression through a preparation program by a candidate. Each file can be linked via a file key of seven fields that identify the candidate, their credential role and type, and the specific program(s) they are enrolled in. The information in these files is described as follows:

- **Demographic file:** This file contains basic demographic information, as well as enrollment information including start dates, withdrawal and exit information, program completion milestones, and endorsement recommendation information. Race and ethnicity are recorded using the detailed list of federal census codes, which are rolled up for reporting purposes to the federal categories in use at the time of this publication. Sex identification encompasses a choice for non-binary identification in addition to male or female.
- **Admissions file:** This file collects program and demographic information on all viable applicants to the preparation program, regardless of the admission decision. While every program has its own process for the evaluation and selection of applicants, the collection provides a basic framework for what information PESB expects programs to consider in their individual processes. Characteristics of the applicants that are considered assets or deficiencies are asked about in this file, in addition to the decision outcome for each applicant. Status details regarding basic skills testing, which is required of all applicants by state mandate, are recorded in this file as well.
- **Assessments file:** Teacher candidates are required to demonstrate content knowledge proficiency through successfully passing certain tests. The test information and outcomes are recorded in the Assessments file. If candidates are not able to demonstrate a passing score on their test, they may be eligible to demonstrate proficiency by other means. This program is referred to as the case-by-case exception option. Fields are included in the assessment file to

record information for candidates who take advantage of this opportunity for proficiency demonstration.

- **Clinical Practice file:** Most candidates for any certification are required to do some sort of practical experience spanning a certain number of hours. Information about the location, duration, and outcome for each practical experience a candidate participates in is included in this file.

The remaining files do not contain information about candidates. They detail other aspects of the program or institution offering the preparation program. They are described as follows:

- **Institution file:** This file provides information on the term structure of the institution offering the preparation program, the start and stop dates of its instructional periods, and the date census is taken during each instructional period.
- **Program file:** This file contains meta-data about each program offered by a program provider. A “program” for the purpose of this collection is any course of instruction that leads to a certification in a clear path. If within a larger program, options to branch off into one area versus another are offered, the providers identify each option or choice as a separate “program.” Identifying program paths to certification in this way sometimes results in hundreds of programs if the provider is a large institution with many institution-defined programs. Information about award level, credit requirements and programs length are included in this file, as well as index information to categorize the area of expertise addressed by the program through use of CIP codes.
- **PEAB file:** PESB expects each program area offered by a provider to be monitored by a professional education advisory board (PEAB). Requirements for advisory board membership and composition are set in administrative rule by PESB and are incorporated into PESB’s regulatory standards. PEABs are expected to meet at least once a year to review, discuss, and make recommendations for the program they are convened to advise. Responses by the program to the recommendations they make are captured each year, as well as the meeting times, discussion content, satisfaction with data products, and further recommendations of the advisory board. This file is not typically used for analysis by ERDC. Each program’s PEAB file is shared directly with PESB, who does their own evaluation / analysis of the information.

Because of the regulatory nature of the purpose of this collection, there is very little missing data in required fields. Data requests are subject to the terms and allowable uses outlined in the current data sharing agreements between ERDC and the providers, and proposed projects using these data may require approval from the data providers themselves. It is possible that some but not all providers may grant approval, meaning a dataset delivered under these conditions may not contain data from all providers.

### Data Collections and Processing

The reporting year is defined by PESB as beginning September 1<sup>st</sup> of each year, running through August 31<sup>st</sup> of the next year. This does not necessarily match each institution’s academic year. If an institution

runs an academic year with summer as the lagging term, the timeframes are aligned. For those with a summer leading term, their academic year crosses two reporting years. For this reason, there is both a “year” field and an “academic\_year” field in the collection to identify those activities that occur in a summer leading term.

The data files are submitted annually, and the collection cycle begins September 1<sup>st</sup>. The collection window is open through November, during which providers work with ERDC to submit their files and correct any errors in their data. Once all the data is cleaned and validated, the aggregate reports are calculated, verified by the programs, and sent to PESB. This typically happens by March, while at the same time work proceeds to revise the collection documentation in preparation for the next cycle. Data from a collection year would typically be available for use by third party researchers in the spring following the close of the collection in November. There are a variety of other reporting needs that are fulfilled for PESB during the spring and summer.

### Documentation

There are four primary pieces of documentation available for this collection. First is the Data Submissions Manual, available on [ERDC's website](#). This is a technical manual that is set up in question and answer format that tells the user what is new or what has changed in the collection from the previous year, provides instructions on how to construct the submission files, enter data, use the secure file transfer system, and a host of other topics related to the submission of accurate data. The file formats for each file are included, with definitions for each field and detailed instructions on how to enter data correctly for that field. An appendix with links to other supporting documents is included.

The Appendices Workbook is a companion document to the data manual that contains all the lookup tables and valid values needed for the collection.

Templates for each file that are preformatted for each field's data type are available for download so it is easy for the data administrators to enter their data and submit the files. These templates are received by ERDC, checked for quality, and eventually stacked in a master database for the year, which is then added to the P20W data system.

The last piece of documentation that is very helpful for the programs is the Data Cleaning Checklists workbook. In this workbook is a tab for each of the files that details the edit checks done by ERDC to quality check the data. Data administrators for a program are able to do the same edit checks ERDC does so they can fix as many errors as possible before submitting their data.

### COVID-19 Impacts

In 2020, in an effort to mitigate the progression of the COVID pandemic, testing centers where educator candidates could go to take their required assessments were closed, offering no alternative method for taking the assessments. Lacking the ability to complete that part of the requirements for certification, candidates' progress through their education program was stalled. In response, PESB issued guidance for educator preparation programs to waive assessment requirements and recommend their candidates for

one-year emergency certificates. The impact of this action can be seen in the assessment data for reporting years 2020-21 through 2022-23, as an entry of DEFER for many assessments. The impact is most heavily noted in 2020-21, and tapers off each year until the waiver was rescinded in 2023.

## Public Four-Year Baccalaureate Institutions

The [Public Centralized Higher Education Enrollment System](#) (PCHEES) collects data for enrollment and awards (degrees and certificates) from Washington public baccalaureate institutions. Data are available beginning with academic year 2007-08 forward. PCHEES was originally developed for state budget and accountability purposes and has evolved into a comprehensive database supporting a variety of research and reporting needs. PCHEES data is a critical resource in the production of tools and reports widely used by stakeholders, legislative staffers, and the general public. PCHEES was created under [RCW 43.62.050](#) and [RCW 28B.10.784](#), which require the Office of Financial Management (OFM) to collect and report higher education enrollment data. PCHEES data is extracted after each term and loaded into ERDC's P20W data system.

PCHEES enrollment information is collected at the student record level in three major files:

- The **Admissions** file includes information about the student at the time of initial enrollment, including type of student, term of admission, county of origin, high school attended/graduated, high school GPA, dual credit affiliation, standardized test scores, previous institution attended, previous degrees, and previous credits earned.
- The **Registration** file contains a summary of each student's enrollment in a single term, relating each student to their specific courses taken. This file includes institution, campus, specific course and section information, funding source, credit hours attempted and earned, and letter grade received.
- The **Student** file contains identifying information that may change from term to term including name, birthdate, gender, race/ethnicity, class standing, field of study, term GPA, cumulative GPA, fee-paying status, and financial aid indicators.

Four additional files are submitted by the institutions: Program, Term Dates, Course, and Completions. The first three of these files are submitted on a term basis and provide supporting details related to the data elements in the Admissions, Student, and Registration files. The Completions data is integrated into the P20W data system once annually and contains elements at the student record level about the awards conferred by the institution to the student during the academic year. While PCHEES data is collected at the student level, analysis at either the school or student level is appropriate for these data.

A variety of detailed documentation is available on the [ERDC website](#) for researchers to better understand the PCHEES fields, what data they represent, and how the fields can be used. A detailed submissions guide is maintained by the data stewards; updates are reflected in the release notes document provided in the guide. Additional valid values tables are also available by request for some fields that are not included in the submissions manual.



## Data Collections

Data are collected on all students enrolled at any of the six public baccalaureate institutions at two different points in time each term: Enrollment Day 10 and the Final Enrollment Day. Data representing a snapshot of enrollment around the 10<sup>th</sup> day of each term are submitted by the institutions, depending on their term structure and whether they enroll students during that term (e.g., semester institutions typically do not enroll students for a winter term).

Day 10 enrollment, sometimes referred to as “census day,” is also required for [Integrated Postsecondary Education Data System \(IPEDS\)](#) reporting. IPEDS is the mandatory federal data collection system for all schools approved for Title IV student financial aid funding.

## Data Submission Schedule & Availability

Data file submission for Day 10 and Final enrollment data and completions data follow a preset schedule that is published annually by ERDC. The schedules for enrollment and completions differ slightly.

Day 10 enrollment data is submitted for each term individually, by the end of the first month of each term, with the exception of summer data, which is due by mid-October following the end of summer term. Day 10 data are available shortly after the files are submitted and a data quality review is completed.

Final enrollment data is submitted for each term individually. Fall and winter term data are submitted by the end of the first month of the next subsequent term (e.g., fall term final is reported by end of month of the following January). Spring term final data is delayed until July 31<sup>st</sup> to accommodate graduation activities. Final data files may be available a month after submission, depending on data quality assessments.

Completions data for summer term is reported by October 15 of the same academic year. Completions for the subsequent fall, winter, and spring terms are reported by July 31<sup>st</sup> following the close of spring term. Completions data are available with the fall publication of the ERDC [Statewide Public Four-Year Dashboard](#).

## Data Processing

After data files are submitted, there is a process for institutions and OFM to review and resolve potential errors. If needed, the institution adjusts and reloads their data. After OFM confirms the data quality to be sufficient for research purposes, users with authorized access to the PCHEES application can download reports and predefined datasets. ERDC staff are then also able then to access PCHEES data, integrate it into the P20W data system, and perform analyses to facilitate research purposes or fulfill approved data requests from external requesters.

An institution can reload their data at any time, even after a submitted file has been released for all users. While that submission is in the review process, the previously approved submission is considered

the active submission and is the only version of the data that is released to all users and available for research. There is no simple way for a researcher to know if a more recent version of the data may become available for use in the near future. If having the very latest data is critical to a project, then prospective users should contact a PCHEES data steward to determine if the data will be updated during the preferred timeframe.

### Data Considerations

The Day 10 data for 2007-08 through 2009-10 was migrated from an earlier version of the PCHEES system and does not contain the complete set of data elements and enrollment types as seen in the current system. The Final collection data for the same academic years was submitted through the current system, so it does contain all the additional data fields and completion fields. The fully featured set of data approved for research is from 2010-11 forward. The following data elements for prior years are only available in the Final collection data: credit hours earned, PELL grant and Washington College Grant, term GPA, cumulative institutional GPA, cumulative overall GPA, and the degrees awarded fields in the completions file.

Overall, the PCHEES data is substantially complete for most fields. Each institution partially establishes its own policies about what to report and what not to report. As a result, some fields are mostly complete for some institutions, and mostly incomplete for others. Across all six institutions, fields that reflect student K-12 information like K-12 ID, last high school attended, and high school GPA, are mostly incomplete. SAT and ACT score records are also mostly incomplete across all institutions. GPA information is not available in the Day 10 files. Within the Final files, overall cumulative GPA (which includes transfer credits) is largely incomplete, while institutional cumulative GPA is generally complete. The veteran's military affiliation and benefit type fields are generally missing, along with the field for teacher certification ID. The Evergreen State College is an in-state institution that uses a Pass/Fail system instead of awarding grades, so GPA is not computed for students enrolled at that institution.

Some characteristics of PCHEES reflect its original objective: to support the state budget. The PCHEES academic year begins with the summer term and runs through spring, and this is how institutions organize their data for submission, regardless of their specific institutional academic year. Information about what constitutes the academic year for each institution for their own purposes can be accessed in the Term Dates file. Within PCHEES, undergraduate state-funded full-time-equivalent enrollment is 15 credit hours, which typically leads to degree completion in four years (excluding summer enrollment). The federal financial aid definition is 12 hours for undergraduate students. For graduate state-funded enrollment, full-time-equivalent enrollment is 10 credit hours.

In the spring of 2018, a gender category option for those who identify their sex as non-binary was loaded to the system for the 2017-18 academic year and beyond. Values are present for all institutions for all terms beginning with that academic year. Note that in practical analytic use, depending on the question asked or the desired configuration of the data, this field may be subject to heavy redaction requirements due to small counts.

For race/ethnicity reporting, institutions use a subset of the census race codes in addition to some specific codes designed especially for the PCHEES system. These codes map to the federal race category classification system currently used across the country and which is typically seen in IPEDS. A data requester may receive these census codes in their dataset, or they may ask to have the roll-up to the federal categories done for them. It is also not uncommon for a data requester to ask for a “Persons of Color” variable to be calculated that typical sums all categories other than White, Multi-Racial, and Not Reported.

### COVID-19 Impacts

At the start of the COVID-19 pandemic response in the spring of 2020, institutions transitioned to primarily remote operations. Courses were taught online and data reporting requirements were adjusted. Enrollment trends show limited impact on enrollment for the spring 2020 quarter. However, the numbers over the following year suggest many students did not come back to school right away. It should be noted that even though institutions transitioned to virtual classrooms, and later, to largely hybrid models, data reporting did not reflect this change in the fields that capture delivery mode. If a class was intended to be a fully in-person class, it was reported as such, even if it was all or in part delivered virtually. This is the main COVID-19 impact on the data. Data continued to be reported, although late due to technology challenges, and there were no data quality concerns of note during this period.

### Public Community and Technical Colleges

The [State Board for Community and Technical Colleges \(SBCTC\) Data Warehouse](#) collects enrollment and awards data (degrees and certificates) from Washington public community and technical colleges. The Data Warehouse supports research and reporting conducted by the SBCTC and by colleges in the state system. It is the primary source of information supporting policymaking and allocation of funds for the community and technical colleges (CTCs). The warehouse is built to enable each of the 34 colleges in the system to submit data for each quarter.

The SBCTC academic year begins with summer quarter and ends with spring quarter, in line with the Department of Education’s definition. Full-time equivalent enrollment is 15 credit hours. The federal financial aid definition is 12 hours for undergraduate students. All colleges operate on a quarterly system.

SBCTC enrollment and course completion information is reported in several files each term. SBCTC extracts data from these files and provides them to ERDC on a quarterly basis for loading in the P20W data system. These files capture information about students, students’ coursework, credits, and degree completion, as outlined below:

- The ***Student*** file includes one record for each student for each quarter they are enrolled at a college. It contains student demographic information and some summary items for the reporting term, including credits and full-time equivalent (FTE) enrollment counts by funding

source, total FTE and credits earned to date, GPA, residency information, dual credit standing, veteran's benefits status, and whether the student took any eLearning course in the quarter. A variety of descriptive flags are also in this file to indicate aspects of student intention and planning or to describe their purpose for enrollment.

- The **StuClass** file includes one record for each student for each course they are enrolled in at a college for each quarter of the academic year. Designations that pertain to the type of course the student is enrolled in are included here as well. Information on converting from clock hours to credit hours is included in this file and is important for research using data prior to 2008. Prior to summer 2008, many programs were reported in terms of clock hours rather than credits. The transition to credit hours across the system began in summer of 2008, with programs continuing to report clock hours until those enrolled students matriculated from the system. The conversion process took over two years to fully accomplish. When using data from this timeframe, programs with courses that are recorded as clock hours must be converted to credit hour courses. This approach is essential for appropriately calculating GPA and FTE. Data elements from the StuClass file are often used to examine enrollment by funding source (i.e., state-, contract-, and/or student-funded), since funding is based on the student's enrollment in a course instead of the individual student or course on its own. CTC students have multiple records in the StuClass file based on their SBCTC-assigned data warehouse key (DW\_KEY), which tracks students longitudinally across all Washington CTCs.
- The **Class** file provides detailed characteristics about each CTC course such as Classification of Instructional Programs (CIP) designations, clock hour / credit weights, funding source, day, time, and location information, instructor identification and Full-time Equivalent (FTE), student enrollment FTE, 10-day enrollment counts, clinical activity information and workforce indicators.
- The **Transcripts** file includes one record for each student for each course for which the student has received a grade at a college. This file does not include records for non-graded courses or courses that are not yet graded. It provides information on credits earned and reported letter grades for students enrolled in the current quarter or the eight previous quarters. Enrolled students are uniquely identified based on a combination of their DW\_KEY, college of enrollment, year of enrollment, and quarter of enrollment. Course data elements like Item number, course number, and department/division identify each course for which the student has received a grade.
- The **Completions** file contains one record for each student for each degree or certificate they earned in a term. Students will appear in the file more than once if they complete more than one degree or certificate at a college during a specific academic year or academic year-quarter.

SBCTC data is available from academic year 2004-05 forward. The data extracts that ERDC receives from the SBCTC Data Warehouse follow a standard quality review process and are generally available for research 3-4 months after the end of the academic quarter. ERDC does not load any fields from SBCTC

Data Warehouse files that can be calculated using other fields or fields that were determined by SBCTC to be unreliable. SBCTC's Data Services Team produces a variety of data dictionaries and other resources that clarify important details about the data stored in the SBCTC Data Warehouse and any changes made to the data. For more information on each SBCTC data file, review SBCTC's [Coding and Reporting Guidelines webpage](#) and materials referenced on their [Data Warehouse Documentation webpage](#). Please note that ERDC receives many but not all fields from each SBCTC data file.

## Data Considerations

During the 2015-16 academic year, SBCTC began transitioning their data system and the data collection application used by the colleges from their HP Legacy system to a [new centralized ctcLink PeopleSoft system](#). Colleges were gradually transitioned to the new PeopleSoft system, and as of March 2023 all SBCTC colleges were successfully migrated to PeopleSoft. The 2022-23 academic year is the first academic year that all CTCs are on ctcLink.

While much of the data remained stable during the transition, there were some changes in definitions and availability of some SBCTC data fields over time. Examples of notable changes include:

- Legacy PROGRAM\_CODE is equivalent to ctcLink's ACAD\_PLAN and its supporting Plan Code Description (DESCR) field.
- [Guided Pathways](#), Meta-majors, and Program/Plan Codes were piloted at five colleges in 2016 and expanded in 2020 to all SBCTC member colleges.
- Race/Ethnicity data was collected in three Legacy fields (CENSUS\_HISP\_CD, CENSUS\_RACE\_CD, and CENSUS\_RACE\_CD2). In ctcLink, each of the standard U.S. Census race/ethnic categories are reported as a seven-digit indicator in RACE\_ETHNIC\_IND2.
- Legacy ECON\_DISAD\_IND was manually determined by each college as to whether a student is economically disadvantaged. In ctcLink, the definition is standardized to reflect a student's receipt of need-based financial aid or their participation in a need-based program.
- Legacy DISABILITY and HEALTH\_LIM were self-reported at the time of admission or registration. In ctcLink in 2023-24, information about a student's health limitations is collected by Disability Services staff if a student requests accommodation in REGISTERED\_DISABILITY.
- Family status was collected in Legacy FAM\_STAT at the time of admission or registration. During the 2015-16 academic year, SBCTC began collecting this information through financial aid applications or via the student's registration data in ctcLink.
- The PROGRAM\_CIP field is a six-digit code that indicates a student's major or course subject area according to the US Department of Education's CIP code taxonomy. In December 2022, SBCTC CIP codes were updated from the 2010 CIP code series and descriptions to the newest federal 2020 CIP code series and descriptions, starting with the 2020-21 academic year.
- The DIST\_ED field has captured whether a course is taught via distance education and has expanded several times since the last 1990s, which SBCTC tracks in a Change Matrix available upon request. The shift to remote learning during the COVID-19 pandemic was not consistently

reported in this field. Researchers should not assume that the DIST\_ED values are reliable during the 2020-21 and 2021-22 academic years.

## General Equivalency Development (GED)

The General Equivalency Development (GED®) Snapshot file provides information on students who successfully pass the GED exam and are awarded a General Educational Diploma. The State Board for Community and Technical Colleges (SBCTC) manages the collection of Washington State GED data, which is saved in a separate data warehouse than the SBCTC enrollment and course completion data. SBCTC manages Washington State GED data on behalf of the [American Council on Education](#) and Pearson, who partner to offer GED testing services nationwide. ERDC receives the GED Snapshot file annually each fall from SBCTC. This file includes new records for students who pass the GED exam between July 1<sup>st</sup> of the prior year and June 30<sup>th</sup> of the current year. ERDC has GED records for students who passed the exam since July, 1 2011.

## Private and Out-of-State Institutions

National Student Clearinghouse (NSC) collects data from nearly 3,600 colleges and universities across the county. This accounts for over 97% of all students in public and private institutions. These institutions regularly provide enrollment and graduation data to NSC. Subject to FERPA privacy regulations, NSC makes this data available through their StudentTracker service. ERDC uses StudentTracker to obtain enrollment and graduation data for selected students at private institutions and well as at out of state institutions. These data contain information relating to dates of enrollment, institution identification, program of study, and degree awarded, if applicable. ERDC incurs a fee per student submitted to StudentTracker. As a result, ERDC can obtain NSC data for only a limited number of students each year.

Since 2005, each summer ERDC obtains NSC data for the previous eight years of high school completers and students who exited high school in 11<sup>th</sup> or 12<sup>th</sup> grade. Completers are defined as having one of the following enrollment withdrawal values: Graduated with a regular High School Diploma, Graduated with an IEP, Graduated with Associates Degree, Graduated with International Baccalaureate High School Diploma. Consequently, the number of years of follow-up of completion and enrollment data varies. For the most recent high school completers or exiters ERDC has a one-year follow-up of NSC data. For high school completers of two years ago, ERDC has two years of follow-up data, and so on.

ERDC also has additional student data resulting from ad-hoc requests to StudentTracker.

### Data Processing

ERDC stacks all the data resulting from every request made to NSC. Since the same person can be submitted to NSC more than once in different years, ERDC deduplicates based on the recency of the NSC submission and the recency of the P20ID. In the end, each P20ID should have one set of NSC data for each student's most recent NSC submission. The date of the most recent data is maintained in the

DetailReportDate field. This reflects the date that NSC processed and produced the data for each student in the NSC stacked dataset.

Another important field is the RecordFound\_Y\_N flag. If NSC is able to match a person that was submitted, this flag is set to “Y”, and there will be additional enrollment and completion data. If not, then this flag will be set to “N.” ERDC has previously found that the data from NSC is generally complete, but limitations relating to identity resolution and reporting bias remain.<sup>11</sup>

Additional information about the NSC data can be found at the [StudentTracker website](#). The FAQ website includes information on submitted files as well as a data dictionary for the returned Student-level Detailed Report.

### Data Considerations

While the data received from NSC is of high quality and can be confidently used in research, there are some limitations to the data<sup>12</sup>:

- Students can choose to keep their student level data private (FERPA block), so not all data provided to NSC is available for research purposes. For example, ERDC’s August 2023 submission StudentTracker resulted in 393,940 matches, but there was a FERPA block on 5,129 of these.
- Some institutions block student level reporting. In ERDC’s August 2023 submission, student level data for 78 students was blocked due to this reason.
- Though NSC collects over 97% of all students in private and public institutions, not all institutions provide data to NSC. These include:
  - Within Washington State, of colleges with a thousand or more students (e.g. Digipen)
  - The military and Coast Guard academies
  - Most tribal colleges
  - Many very small institutions
- International students and undocumented (non-U.S. citizens) are often not reported to NSC.
- The NSC match rate is not 100%. Because of FERPA blocks, typographical errors in data ERDC submits to StudentTracker or within the NSC data itself, and ambiguousness of some of the matches, NSC estimates that the match rate for a typical high school cohort is 88% to 90%.

## College Bound Scholarship

Established by the Legislature in 2007, the College Bound Scholarship program provides financial assistance to low-income students who pledge by the start of high school to graduate and meet program requirements. This early promise of financial aid is intended to alleviate the financial barriers

---

<sup>11</sup> ERDC [NSC-Paper.pdf \(wa.gov\)](#)

<sup>12</sup> For more information about the data NSC collects, see [Working With Our Data | National Student Clearinghouse Research Center](#)

that prevent low-income students from considering higher education as a possibility. The scholarship covers the average tuition at comparable colleges, some fees, and a small book allowance.

The Washington Student Achievement Council (WSAC) College Bound Scholarship Data contains student demographic and K-12 enrollment information for all students who have completed a College Bound Scholarship application prior to high school graduation or were automatically enrolled.

Students in public school who meet one of the following requirements are automatically enrolled in College Bound:

- Students eligible for free-and-reduced price lunch in 7<sup>th</sup>, 8<sup>th</sup>, or newly eligible in 9<sup>th</sup> grade.
- Students who are in state foster care, or a dependent of the state between 7<sup>th</sup> grade and high school graduation.

Alternatively, students may apply directly to the program. WSAC submits completed enrollment and application data to ERDC annually. Data<sup>13</sup> is available since the 2015-16 academic year.

Use and disclosure of these data are done in collaboration with WSAC and can be used in the planning, policy development, and performance monitoring of these programs or in other research. Records pertaining to students who do not complete the application is not available to request.

## Apprenticeships

### Registered Apprenticeships

Overseen by the Washington State Department of Labor and Industry (LNI), the Registered Apprenticeship program provides on-the-job training with related classroom instruction under the supervision of journey-level professionals. Apprenticeships are paid positions and, after completion, apprentices receive a nationally recognized professional credential in their career field. Prospective participants must first contact the apprenticeship program of interest to determine if applications are being accepted. In general, apprentices must meet the four criteria below, but each program has its own set of admission requirements:

- Be at least 16 years or older.
- Be able to perform the work, with or without reasonable accommodation.
- Have the knowledge, skills, and abilities needed to learn the job.
- Provide proof of age, high school diploma or equivalency (GED), and/or honorable military discharge.

Employers seeking to offer apprenticeship positions can either create and sponsor apprenticeship programs or partner an existing apprenticeship program as a training agent. Each apprenticeship

---

<sup>13</sup> Data relating to foster youth that are auto enrolled into College Bound are not available to request.



program requires written program standards reviewed by LNI and approved by the Washington State Apprenticeship & Training Council (WSATC).

Apprenticeship programs contain educational requirements and approximately 75% of programs partner with a community or technical college (CTC) to offer these credits. However, the length and requirements of the program vary by program, the structure of the apprenticeship data more closely resemble typical postsecondary internships structures than other structured postsecondary degree trajectories.

Employers submit apprenticeship data to LNI which is maintained with the Apprenticeship Registration & Tracking System (ARTS). LNI provides a full extract of the Registered Apprenticeship dataset annually to ERDC starting in 2000. LNI does not evaluate changes across data extracts.

The apprenticeship participation data is provided at the individual person and apprenticeship program level of detail. Key apprenticeship program data elements include employer information, the occupation apprentices are training for, the Structural Occupation Code (SOC) and employer name as well as the apprenticeship start date, apprenticeship status, and apprenticeship status date. The data also includes information on the apprentice that includes gender, birthdate, race, ethnicity, county of residency and prior educational level.

These data can be effectively utilized to track and evaluate apprenticeship program activity, including participants' initial program enrollment through completion or non-completion, including any potential breaks, transfers, or suspensions among participants. It can answer questions regarding apprentice demographic characteristics, historical participation trends by program or occupation, rates of completion, and/or breaks in program continuity. Cross sectional linkages can identify if apprentices pursue postsecondary education or workforce opportunities during or after the apprenticeship period.

ERDC advises caution on the use of the apprentice demographic information due to the number of missing values. Researchers should account for this systemic data limitation in their analysis.

## Workforce

### Unemployment Insurance (UI)

The most widely used data about employment comes from a state's unemployment insurance program. Unemployment insurance program data includes information about workers and their employers, and about the previous occupations and industries of employment for those who file for unemployment payments.

The Unemployment Insurance (UI) Program is a federal-state program financed by payroll taxes paid by employers, and in a few states paid by the employee. The U.S. Department of Labor sets broad criteria for the eligibility and coverage, but states determine the specifics of the implementation. In Washington

State, the Employment Security Department (ESD) is responsible for the administration of the UI Program.

Nearly all employers with employees are required to participate in the UI Program if they pay wages to employees, regardless of the dollar amount. Participation includes registering with the state, submitting quarterly reports, and paying unemployment taxes or reimbursing the department for benefits paid for all part-time or full-time employees. There are participation exceptions. In Washington, those exceptions include the following:

- Spouse, children under 18, and student workers of small farm operators – those with payroll less than \$20,000 and fewer than 10 employees.
- Employees performing domestic services in a private home, college clubs, fraternities, or sororities, if the total wages paid are less than \$1,000 per quarter. If payroll exceeds \$1,000 in any quarter, wages must be reported for the entire year and the following year.
- Non-profit preschool staff, if fewer than four staff.
- Business owners are not reported. Sole proprietors do not report themselves, their spouses, or unmarried children under 18.
- Corporate officers are required to cover themselves for UI unless they opt out by January 15<sup>th</sup> each year.

There are additional types of employees that an employer may not be required to report, depending upon the circumstances. They include the following:

- Self-employed workers
- Church employees
- Work-study students, as long as the employer is a non-profit 501(c)(3), state government or local government
- Licensed insurance agents, real estate agents, brokers, and investment company agents
- Federal employment, such as U.S. Postal Service (USPS), federal civilian employees, and active duty and retired military are covered under a federal unemployment insurance system
- Railroad employees are covered under a separate unemployment insurance system, established under the Railroad Unemployment Insurance Act

Additional information regarding the Unemployment Insurance Program in Washington is available from the Employment Security Department.<sup>14</sup>

Employers are required to submit two files quarterly to ESD; a wage file at the employee level and a summary file at the employer level. The summary file is used widely for monitoring and forecasting employment and economic trends.

---

<sup>14</sup> [ESDWAGOV - Unemployment taxes](#)

### Employee Wages

The wage file is provided by the employer and contains information at the employee level. In Washington state, the quarterly wage detail report filed by employers includes the following information:

- Employer identifiers
- Employee identifiers
- Wages paid during quarter
- Hours worked during quarter
- Occupation of employee

In most states, employers do not report the hours worked for employees. However, in Washington employers are required to report hours because it is used in determining UI eligibility. A UI claimant must have 680 hours of covered employment in their base year to meet the initial requirements to qualify for a claim. The occupation of the employee became a reporting requirement beginning in the first quarter of 2024 in Washington state.

Wages paid and hours worked are reported for the entire quarter. It is not possible to break them out by month or determine when an employee worked during the quarter. Hours and wages can be expressed as averages per week within a quarter, but there is no way to determine how the hours and wages are distributed throughout the quarter. For example, if an employee earns \$5,000, the data does not indicate if those wages were paid by the employer in the first month, in the last month, or spread amongst all three months of the quarter. Hourly wage can also be calculated by the simple division of total wages for the quarter divided by the hours reported by the employer.

### Employer Characteristics

The employer summary file contains characteristics of the employer that can be associated with each of the employees provided in the wage file. Characteristics available for employers include:

- Industry – North American Industry Classification System (NAICS) code
- Ownership – Private or public (federal, state, local governments)
- Size of firm – number of employees (available for each month within the quarter)
- Location of the firm

It should be noted that multi-industry and multi-site employers are not required to break out their employment data by site and industry. Even when employers do provide these breakouts, it is not clear which employees are working in which locations/industries. Though the percentage of firms this applies to is relatively small, they account for approximately 30 percent of the UI wage records.

The industry is based on the North American Industrial Classification System (NAICS) and is kept up to date through industry verification surveys to capture changes in business activity.<sup>15</sup>

NAICS is a hierarchical classification system expressed with 2- through 6-digit codes. The 2-digit level, shown in Table 6, is the most general level. In many instances, the 2-digit categories are grouped into "supersectors," patterned after those used by the U.S. Bureau of Labor Statistics (BLS) in statistical summaries.<sup>16</sup>

**Table 6: North American Industrial Classification System**

Supersector	NAICS 2-digit code	NAICS description
Natural resources and mining	11	Agriculture, Forestry, Fishing and Hunting
	21	Mining, Quarrying, and Oil and Gas Extraction
Construction	23	Construction
Manufacturing	31-33	Manufacturing
Trade, transportation, and utilities	42	Wholesale Trade
	44-45	Retail Trade
	48-49	Transportation and Warehousing
	22	Utilities
Information	51	Information
Financial activities	52	Finance and Insurance
	53	Real Estate and Rental and Leasing
Professional and business services	54	Professional, Scientific, and Enterprises
	55	Management of companies and Enterprises
	56	Administrative and Support and Waste Management and Remediation Services
Education and health services	61	Educational Services
	62	Health Care and Social Assistance
Leisure and hospitality	71	Arts, Entertainment, and Recreation
	72	Accommodation and Food Services
Other services	81	Other Services (except Public Administration)
	92	Public Administration

While the industry sector provides useful information about the nature of work being done by the company. The typical occupational staffing pattern for an industry is available through [Bureau of Labor Statistics](#). These help to identify the major occupations in an industry and the percent of employment for that industry for those occupations. State and local area estimates of staffing patterns are also [available](#).

There is a lag between the time the employer files the report and the time the associated administrative data become available for research use. Both UI tax payments and wage reports are due by the last day

<sup>15</sup> NAICS was adopted by the U.S. Office of Management and Budget to replace the Standard Industrial Classification (SIC) system in 1997. The U.S. Census Bureau website contains detailed information about the [NAICS industry classification system](#).

<sup>16</sup> [NAICS Supersectors : U.S. Bureau of Labor Statistics \(bls.gov\)](#)

of the month following the last day of each quarter. Incorporating the wage data into administrative databases takes the remaining two months of the quarter. Data are ready for use for research purposes early in the subsequent quarter.

### Data Considerations

As with many administrative data sources, there are known missing values and unknown missing values in the UI Wage dataset. Known missing values are the result of specific groups of employees that are excluded from submitting data, as described above. ESD estimates the rate of coverage at 92%, but the rate may vary by industry. Researchers must account for the exclusion of some paid employees, such as work-study students, especially when analyzing education and wage data concurrently. Similarly, individuals who receive non-wage income (i.e., an ownership stake in a corporation or partnership, receipt of royalty income) will not have their total income reported because certain types of income are excluded from this data.

Beyond the known missing, there is also an issue of unknown missing. While not specifically excluded, gig workers have operated as both self-employed and wage-paid workers, depending on the employer. As rules change, gig workers may show artificially deflated wages and, in some cases, appear in the data intermittently. The structure of some organizations can limit the ability to determine an employee's industry or location. Comparing and analyzing wages by region can also be difficult because employers may only report by their corporate office address and not the location(s) where employees physically earn wages. This is particularly important in areas that border other states, because where an employee works determines where the UI tax is paid. For example, if a company is based in Washington, and their employee works in California, then UI taxes are paid (and data is reported) only in California instead of Washington. See [ERDC's Technical Report 2012-01: Employment Data Handbook, July 2012 for additional information](#).

At the onset of the COVID-19 pandemic, Washington state enacted several emergency rules in an effort to mitigate its spread. An effect of this was a decrease in the level of economic activity in the state, which can be seen in the UI data as a sharp decrease in the employment rate in mid-2020. Employment gradually increased over the next two years as emergency rules were loosened. For more information, see the [ESD COVID-19 economic data website](#).

## Part II: Community and Organizational Characteristics

### American Community Survey / Census

As described on the U.S. Census Bureau's [American Community Survey \(ACS\) website](#), "The ACS is a nationwide survey that collects and produces information on social, economic, housing, and demographic characteristics about our nation's population every year. This information provides an important tool for communities to use to see how they are changing. When people fill out the ACS form, they are helping to ensure that decisions about the future of their community can be made using the

best data available. Decision-makers require a clear picture of their population so that scarce resources can be allocated efficiently and effectively.”

Weighted ACS survey data are produced for 5-year intervals at the school district level. ERDC has created files for the intervals 2005-2009, 2006-2010, ... 2015-2019, etc. ERDC has constructed school district files from the four basic ACS Demographic Profiles produced by the US Census Bureau:

- **DP-02** covers social data (household type, marital status, grandparent(s) in household, school enrollment, educational attainment, disability, migration, nativity, language, ancestry, and computer and internet use).
- **DP-03** contains economic data (employment status, commute to work, occupation, industry, income and benefits, health insurance, and poverty).
- **DP-04** files contain housing data (housing occupancy, units in structure, tenure, plumbing, value, mortgage status, rent, and owner and rental costs).
- **DP-05** files store the demographic data (age and sex, race, ethnicity, and voting age population).

In addition to the data, a data dictionary/variable list is available for all the DP files.

ACS data are sample data and therefore subject to sampling error. As such, the Census Bureau produces 90% confidence intervals for each of the estimates, designated as Margins of Error (MOE). The MOE are not included in the ERDC DP data files but are included in the original data tables. For further information on ACS estimates and measures of sampling error consult the [2020 ACS General Handbook, Chapter 7](#).

## Part III: Restricted Use Data

Data listed in this section can only be used for specific purposes and may have restrictions on redisclosure. Restricted data is not stored in the P20W data system or integrated into the MDM identity matching processes.

### Criminal Justice

#### *Administrative Office of the Courts (AOC)*

Administrative Office of the Courts (AOC) maintains a statewide electronic court records database for all cases filed in courts in Washington State since 2010. This includes criminal, juvenile offender, and infraction cases filed within Superior, District, and Municipal courts. The case-level data contain information on the court, charges, case outcomes, and sentences. Person-level data contain information on demographics and case information.

#### *Department of Corrections (DOC)*

Department of Corrections (DOC) maintains information for people incarcerated in Washington State facilities and for people under community supervision in Washington State since 1953. Data relates to

person demographics, admission/release spans, program participation, offense/sentence obligations, custody levels, and risk-need assessment information. Individuals under DOC custody have been convicted of adult criminal offenses and sentenced to state confinement with a duration of a year and a day. This data does not contain individuals sentenced to Juvenile Rehabilitation or local detention.

## Early Childhood Educational Assistance Program (ECEAP) Boolean questions responses

This data set features a long list of yes-or-no questions that are asked/answered during the ECEAP application process. Some of the questions are highly sensitive in nature, addressing incarceration of parent(s), homelessness, and domestic violence, while other questions are not sensitive and collect demographic information. ERDC is restricted in how this data can be used and shared. Not all of the questions are asked in all years (some are retired and some are added), so completeness for specific questions vary.

## Financial Aid data

### *College Bound Scholarship Data*

The Washington Student Achievement Council (WSAC) provides ERDC with College Bound application and enrollment data annually that include both completed and incomplete application records as well as any student automatically enrolled in the program. The data is provided to ERDC in order to assist WSAC in determining eligibility in and monitoring and evaluation of the College Bound program. Data are linked to other P20W data through ERDC's identity resolution process and then ERDC provides linked data to WSAC. The data relating to students who do not complete the application is restricted to WSAC use only.

### *Financial Aid Family Income File*

The Washington Student Achievement Council (WSAC) provides ERDC student financial aid family income information annually. Data includes family income, household size, dependency status, and number of family members in colleges for all students that have applied for financial aid in Washington State from the 2012-13 academic year to present. ERDC matches data with the P20W data system and provides linked data to WSAC. This data is restricted to WSAC use only.

### *Unit Record Data*

Washington Student Achievement Council (WSAC) Unit Record Financial aid data is provided to ERDC by WSAC on behalf of Washington's public colleges and universities. This information reflects the administrative records of students who receive need-based aid at one or more of Washington's forty public postsecondary institutions (34 community and technical colleges and six four-year institutions) for each term during each academic year.

Data elements include demographic information including race, ethnicity, and gender. Data relating to students' finances are supplied by the institutions and include students calculated need, family

characteristics, expected financial contribution as well as the cost of attendance. Aid totals include loans, work study, scholarships, and waivers from all funding sources. Funding sources include federal, state, institutional and any other private funding sources. In addition to financial totals in these categories, specific awards of interest (see Table 7) are also collected.

**Table 7: List of federal and state financial aid awards**

Federal Awards	State Awards
Federal Pell Grant	State Washington College Grant
Federal Supplemental Educational Opportunity Grant	State Need Grant
Federal Academic Competitiveness Grant	State College Bound Scholarship
Federal Science and Mathematics Access to Retain Talent (SMART) Grant	State Workforce Training
Federal Perkins Loan	State Board for Community and Technical Colleges Opportunity Grant
Federal Direct Subsidized (Stafford) Loan	State Passport to College Scholarship
Federal Direct Unsubsidized (Stafford) Loan	State Opportunity Scholarship
Federal Unsubsidized PLUS or Stafford Loans	State Education Opportunity Grant
Federal Parent PLUS Loan	State GET (Guaranteed Education Tuition) Ready for Math and Science Conditional Scholarship
Federal Graduate PLUS Loan	Washington State College Grant Apprenticeship
Federal Grants or Scholarships for Nursing or Other Health Professional Training	State Passport to College Scholarship Incentive Grant
Federal Nursing Student Loan or Health Professions Student Loan	Washington State Opportunity BA Scholarship
Federal Teacher Education Assistance for College and Higher Education (TEACH) Grant	Washington State Opportunity Career and Technical Scholarship
	State Teacher Grant Programs
	State Teacher Conditional Scholarship Programs

WSAC provides data annually for all students receiving financial aid during the academic year. Data is available for use approximately eight to nine months after the end of the academic year, after ERDC quality assurance checks and identity resolution process is complete.

Use and disclosure of these data are done in collaboration with WSAC and can be used in the planning, policy development, and performance monitoring of these programs or in other research.

These data are considered generally reliable and accurate and can be used to understand the impact of financial aid awards and awarding practices on student outcomes.

**Data Considerations**

This data represents only students who received need-based aid at institutions participating in state aid programs. Students receiving only merit-based aid, students not eligible for federal aid (such as international students), or students who pay for school fully out-of-pocket, are not captured in these



data. Family finance data obtained from students' FAFSA is not required by some financial aid programs, which establish financial need by other means. If students only receive financial aid from such programs, family finance data may not be available. Furthermore, award rates differ by institution and depend on participation or eligibility in each program. This data should not be considered a reliable proxy for ability to pay or student income research. ERDC advises researchers to consider institutional influences when comparing across institutions. ERDC also advises researchers to use total aid award columns wherever possible when making cross-school comparisons.

## Foster Care

The Department of Children, Youth, and Family (DCYF) extracts from their Famlink system a dataset that contains foster care records for all children and youth who were in Division of Child and Family Services (DCFS) placement care and authority between ages 3-21, in an open removal episode anytime since 2012. Youth involved in extended foster care are also included in this data. Per DCYF's website, to be eligible for extended foster care, youth must be involved with education or employment on their 18<sup>th</sup> birthday.

This dataset covers removal episode beginning and end dates, as well as information on placements including start and end dates, the type of placement, and placement location. There is one record per child, per placement. There are often multiple placements per removal episode and some children have multiple removal episodes. Child level summary measures are also included: how many total placements and removal episodes per child, a child's first removal date, and the total number of days removed that includes all removal episodes per child.

The dataset is highly complete. No missing values for removal date, placement begin date, or first removal date. Missing values occur for discharge date and placement date, indicating that a child has not yet been discharged or is still in placement at the time that the data was extracted. Some removal and placement dates occur before 2012, but the records are tied to a child or youth who was in an open removal episode from 2012 and on. Some youth have high numbers of placement events.

The number of Washington students in foster care is low, which can limit data disaggregation. For instance, cell sizes become too small to present outcomes for some racial groups. A temporary policy change occurred during the Covid-19 pandemic allowing those participating in extended foster care to remain in the program after their 21<sup>st</sup> birthday during January 27, 2020 to September 30, 2021. Removal and placement of children into foster care may have been impacted during the COVID-19 pandemic.

## Homelessness

The Washington State Department of Commerce's data system called the Homelessness Management Information System (HMIS) covers individuals who were identified as receiving housing services. This data can be used to flag students and youth who experience homelessness. Information is collected via survey. An individual may decline to consent to having their information shared, in which case they would be excluded.

This dataset provides information about the length of time youth experience homelessness, including how long someone stays in temporary housing, when they exit to permanent housing, and if they return to homelessness. Information on key components, record level, and data quality will be available in the future. See section on OSPI Student Attributes and Programs for details on data collected from schools on students who are identified to receive services under the McKinney-Vento Act.

## Institutional Education

The Office of Superintendent of Public Instruction (OSPI) and the Department of Children, Youth, and Family (DCYF) provides institutional education data to the ERDC for legislatively mandated research and reporting including the work that is overseen by the Project Impact Education group.

This data does not include the entire story of a youth's criminal justice involvement. Other data sources would need to be used to provide a more complete picture such as Washington State Department of Corrections (DOC), Washington State Administrative Office of the Courts (AOC), Washington Association of Sheriffs & Police Chiefs (WASPC), and Washington State Patrol (WSP).

### *Juvenile Rehabilitation Roster*

The Department of Children, Youth, and Family (DCYF) Juvenile Rehabilitation data covers administrative records DCYF collects from facilities under their governance in a system called Automated Client Tracking (ACT). Youth up to age 25 are eligible to be sentenced to these facilities, though this eligibility has changed over time. Youth residing in these facilities can have adult or juvenile sentences. Information on key components, record level, and data quality will be available in the future.

This data can be used to identify youth who were involved with state-run juvenile justice facilities. Juvenile rehabilitation facilities operate year-round, while schools often operate on a school year. DCYF reports their data by state fiscal year. This should be kept in mind when making comparisons with the school system. Youth may have multiple stays in a year, which can include returns to facilities for parole violations. There are different timeframes that youth stay at different types of facilities. Residential stays at juvenile rehabilitation facilities may have been impacted during the COVID-19 pandemic due to health and safety practices.

### *K-12 Institutional Educational Services*

This dataset is a subset of enrollment records extracted from CEDARS that covers students who received K-12 educational services for at least one day at an institutional education facility as defined in [RCW 28A.190.005](#). OSPI is obligated to provide educational services to youth housed in any type of institutional education facility, which includes juvenile rehabilitation facilities under the governance of DCYF, county jail and detention facilities, and rehabilitation centers under the governance of DSHS. Young adults residing at Department of Corrections facilities have basic education services provided through SBCTC. Data describes when and where the educational services occurred for students in grades 6 through 12, starting in the 2013-14 school year. Information on key components, record level, and data quality will be available in the future.

This data can be used to identify students who are current or former residents of institutional education facilities. Youth who are eligible to receive K-12 education services, but are not enrolled are not represented in this dataset. During the COVID-19 pandemic, most schools were operating fully remotely or with hybrid schedules at the end of the 2019-20 school year and through the 2020-21 school year. This may have had an impact on educational programming in these settings, but it has not yet been analyzed by ERDC.

## Part IV: Data in Development

Data listed in this section are still in development and are not currently available to request. In some cases, the ERDC is working with the data contributing agency to establish an initial delivery. In other cases, ERDC has received the initial data and is in the process of developing understanding of the data and its appropriate uses. ERDC will continue to work with the contributing agencies to close these gaps and incorporate these data into the P20W data system, as appropriate.

### Dual Credit Exam Scores

Office of Superintendent of Public Instruction (OSPI) receives data relating to the registration and resulting scores of Advanced Placement, International Baccalaureate, and Cambridge dual credit exams. Data provided to ERDC is governed by the data share agreements between OSPI and the three entities offering those exams. This data will be under restricted use.

### English Learner Annual Assessments

Office of Superintendent of Public Instruction (OSPI) collects data relating to the annual proficiency assessments given to all students in kindergarten through 12<sup>th</sup> grades who qualify for the English language development services to determine if the student will remain eligible for these services. Alternative tests are available for students with significant cognitive disabilities. These annual assessments are different from the screener tools used to determine initial eligibility for English language services and do not relate to the annual assessments in English Language Arts, Math, or Science.

### K-12 Graduation Pathways

Under [HB 1599](#) (2019), [graduation pathways](#) were established as a requirement of high school graduation, starting with the class of 2020. There are multiple ways for students to meet the requirements of graduation pathways, including test-based, course-based, and performance-based options. The pathway must be aligned with a student's postsecondary goals as described in their [High School and Beyond Plan](#). For reporting on pathways, Office of Superintendent of Public Instruction (OSPI) compiles data across multiple sources, including:

- Course-based pathways (including college transition courses, dual credit, and CTE course sequences), identified using OSPI's student-level data collection.
- SAT exam and AP test score pathways, identified using files obtained from the College Board, the proprietary exam company offering those assessments.
- ACT, IB, and Cambridge International exam-based pathways identified using files from ACT, IB, and Cambridge respectively.
- ASVAB Armed Forces Qualification Test (AFQT) score provided by school districts/STEC and submitted to OSPI.

From these sources, it can be determined which graduation pathways a student completed but not which one was identified in the student's High School and Beyond Plan and used for graduation. Because of the flexible nature of these pathway options, many students complete more than one pathway option. This data will be under restricted use.

## Washington Private Vocational Schools

The Workforce Training Board (WTB) sets the minimum standards for more than 300 private vocational schools to ensure that the career-focused trainings provided by each school comply with the relevant licensing requirements in the state. Private vocational schools (PVS) provide student record-level data to WTB for research purposes, as outlined in [RCW 28C.18.060](#). WTB maintains the collection of this data, which includes information on student demographics, PVS student enrollments, and PVS student completions. Beginning in late 2023, WTB has shared PVS data with ERDC on an annual basis. ERDC matches the PVS data with other education records and provides the linked data back to WTB and OSPI to meet federal and state reporting requirements.

## Workforce Innovation and Opportunity Act (WIOA)

The Workforce Innovation and Opportunity Act (WIOA) is the federal law that passed in 2014 which funds programs and services that help job seekers gain access to employment services, basic education for adults, vocational rehabilitation support, and other workforce development resources<sup>17</sup>. Services provided under WIOA began in July 2015, with additional eligibility and program requirements established one year later. Washington's Employment Security Department (ESD) collects state-level WIOA program eligibility and participant information, including participant names, service types, and dates associated with participants' start and end of service receipt. As of late 2022, ESD provides ERDC with this data twice per year so that it can be linked with other data to study educational outcomes and workforce transitions of WIOA participants.

---

<sup>17</sup> [ESDWAGOV - Workforce Innovation and Opportunity Act \(WIOA\)](#)

## Appendix: Other Resources

### Guide to Acronyms

ERDC's [Guide to Acronyms](#) includes various acronyms used in the Handbook as well as other ERDC documentations and reports.

### Technical Reports and Documentation

[A Data Quality Evaluation of Administrative Data Using CEDARS Student Grade History Data as a case study](#) (2018)

[Q&A About the Use of OSPI Assessment Data](#) (2013)

[Using National Student Clearinghouse Data for Measuring Public Postsecondary Outcomes: Washington Case Study](#) (2018)