Feasibility Study: Multilingual Learners

At the October 1, 2024 meeting, the Data and Tools Advisory Board advanced a proposal from Jennifer Orlick and Laura Owen to include information on multilingual learners in the Cradle to Career Data System (C2C) analytical data set. Specifically, the California Department of Education (CDE) and the Department of Social Services (CDSS) were requested to provide information on:

- Students with Long-term English Learner status (LTEL)
- Students with At-Risk English Learner status (ARLTEL)
- Students with newcomer status
- Students who were in an immersion K-12 classroom
- Students with Dually Identified status (in both preschool and K-12)
- Students with Dual Language Learner (DLLs) status for children in Title V subsidized early learning programs prior to 2025
- Variables from the California Preschool Data Collection (CAPSDAC) system, including:
    - Language the Lead Teacher Uses
    - Languages the Lead Teacher is Proficient in
    - Languages Other Staff Use
    - Languages Other Staff are Proficient In
    - Language Program Type

The full text of the proposal can be found at this link.

When conducting feasibility studies, the Office of Cradle-to-Career Data (Office) considers four factors: data availability, data reliability and data validity, cost, and compliance. Information on each aspect is included below.

Note on Terms

The data points being considered are defined as follows:

- Students with Long-term English Learner status (LTEL)
  - Students in grades 6-12 who have been enrolled for 7 or more years and are still designated as English language learners.
- Students with At-Risk English Learner status (ARLTEL)
  - Students in grades 3-12 who have been enrolled in a U.S. school and are still designated as English language learners for 4 to 6 years.
- Students with newcomer status
  - Students in their first three years of enrollment at a K-12 school in the U.S. who are designated as English language learners.
- Students who were in an immersion K-12 classroom
  - 50/50 program or Other Language Allotment majority non-English program
- Students with Dually Identified status (in both preschool and K-12)
  - Students who are English language learners who also have an Individualized Education Program to address a disability
- Students with Dual Language Learner (DLLs) status for children in Title V subsidized early learning programs prior to 2025
  - Children ages birth to 5 who are learning two or more languages at the same time, where this data is reported.

While the proposal used the term "dual language learner" for children in early learning programs, children are usually described as being multi-language learners before starting kindergarten. This designation acknowledges that in their first five years, people are still gaining fluency in all languages to which they are exposed.

Data Availability

Data Sources and Historical Range: *Where the data originates from and the span of years for which data are available.*

Long-Term English Learners and At-Risk English Learners
CDE does not have information on students' long-term English language status from birth to age 5, given that children of all backgrounds are still gaining English proficiency during their early years.

Starting in kindergarten, the agency tracks Students with Long-term English Learner (LTEL) and At-Risk English Learner (ARTEL) status, but the definitions do not align with the ones proposed for the C2C analytical data set. For example, when CDE calculates LTEL and ARTEL status, they include variables such as prior test scores, in addition to the number of years the student has been classified as an English learner.

However, CDE provides information for the C2C analytical data set that indicates the year in which a student was identified as an English language learner. The Office staff could calculate LTEL and ARTEL status according to the proposed definitions and clarify that this varies from ways these variables are calculated by CDE.

Alternatively, CDE could provide flags on students who meet its definition of LTEL[1] students as part of its submission to the C2C analytical data set.

Newcomers
The concept of newcomer status is defined by California Assembly Bill 714 in the same way as the U.S. code for immigrant children and youth:
- aged three through twenty-one;
- not born in any U.S. state (each of the 50 states, the District of Columbia, and the Commonwealth of Puerto Rico); and
- not attending one or more schools in the United States for more than three full academic years.[2]

Subcategories of newcomers include English learners, immigrant children and youth, migratory children, refugees, students who may have experienced

---

[1] See https://www.cde.ca.gov/ds/ad/lteldef.asp
[2] Section 7011(5) of Title 20

limited or interrupted formal education, and unaccompanied children who are refugees or undocumented.[3]

Currently, CDE only collects information on English language learners and children who have participated in migrant education programs. It already provides information on English language learner status and children in migrant education programs for the C2C analytical data set. The data CDE provides to the P20W cannot be used to identify Newcomer students as defined in Section 7011(5) of Title 20 of the United States Code. CDE is currently reviewing existing data elements related to identification of these students. Currently,CDE is unable to provide further information on the feasibility of adding new data to the P20W for this purpose.

Given the proposed definition, the Office staff could calculate students designated as English language learners who have been enrolled in California public schools for three years or less.

Immersion K-12 Programs

This information is not available. CDE provides information on children who are receiving English language services and language of instruction, but it does not track information for immersion classrooms that contain both native-English language and English-language learner students. CDE does not collect data from LEAs that lend themselves to meeting the definition as included in the proposal.

Dually Identified

CDSS does not track information that could be used to calculate this metric.

Currently, CDE does not identify preschoolers who are flagged both as dual language learners and having an Individualized Education Plan (IEP) to address a disability at the student level. While work is underway to collect this information, it will not be available for several years.

Starting with children of kindergarten age, the Office staff could identify dually enrolled students using the data points that CDE provides on English language learners and children with IEPs, but these figures may not align with data posted

---

[3] See the CDE website at https://www.cde.ca.gov/sp/ml/newcomerstudnts.asp

by CDE. CDE currently makes their reporting on this information available via [DataQuest](#).

Language Learners in Title V Subsidized Programs

CDSS has comprehensive information on the language spoken at home by people who receive public benefits for childcare, foster care, Medicaid, CalWORKS, CalFresh, and In-Home Supportive Services, which can be used to identify English language learners. This information is already included in the C2C analytical data set. The data points that CDSS [already provides](#) include:

- Early learning and care primary language
- Race/ethnicity
- Sex/gender
- Foster youth
- Child with a disability

CDE's California Preschool Data Collection System

CDE is currently building the second phase of the preschool data collection system and will not have responsive data by the end of 2025. Until the system is fully operational, LEAs have adapted to reporting in the new system, and the data have been deemed valid and reliable, CDE will not be able to provide preschool data.

Future Relevance and Data Availability: *Evaluate whether the new data will remain useful as the analytical dataset evolves and if the information is likely to be collected in the future.*

The issue of language skills is likely to continue to be important, but with a growing emphasis on how language skills relate to immigration status. The definitions proposed do not expressly identify factors like documentation or refugee status, but language skills may become conflated with immigrant status.

Many of the requested data points on English language learner status and disability status are already available in the C2C analytical data set or could be calculated by the Office. This information is likely to remain available.

CDE is unlikely to collect information on dual-immersion classrooms, as this is not an area where they have authority to require reporting from LEAs.

Data Reliability and Data Validity

Institutional Variability: *Whether there are variations in administrative practices and data recording across institutions at the local level.*

Many of the requested data points could be calculated using data that are already collected at the state level. Therefore information is likely to be reasonably consistent across local institutions. This includes data points from CDE on Long-Term English Learners, At-Risk English Learners, Newcomers, Dually Enrolled Children (starting in kindergarten), and from CDSS on English Language Learners.

However, information on newcomers may be incorrect at the local level due to lack of data sharing agreements across state lines. For example, a child who enrolls in a California school for the first time in grade 3 and is not fluent in English but who was born in and previously attended school in Arizona might be flagged as a newcomer.

Similarly, if the Office calculates LTEL and ARTEL status, the measures would only be based on enrollment in public California K-12 institutions. If, for example, a student in the 11th grade is an English language learner, but had moved back and forth between California and Arizona multiple times, they would not be included in the LTEL or ARTEL metrics.

Agency Variability: *Data and metric definitions across various agencies follow a uniform format and standard.*

Data definitions for CDE and CDSS may not always align. For example, CDE uses a combination of surveys and standardized assessments to determine English language proficiency. CDSS evaluates language skills by household declaration[4][5].

---

[4]
https://www.cdss.ca.gov/Portals/9/Additional-Resources/Forms-and-Brochures/2020/A-D/CCD26.pdf?ver=2023-10-06-141046-190

[5]
https://cdss.ca.gov/inforesources/child-care-and-development/contractor-resources/cdss-cdmis-support/cdmis-manual/appendix-a-data-definitions/child-is-english-learner

For CDSS, the term "newcomer" is not based solely on language skills because it is used in the context of immigration and social services and may take into account a child's refugee status or the date of entry into the country.

Data Integration Across Agencies: *The Office's ability to consolidate data from multiple agencies.*

The Office is already combining information from CDE and CDSS to more fully document the experiences of individuals. The primary challenge will be implementing definitions that differ from data produced by CDE and CDSS. For example, there may be some confusion about long-term English learners because state and federal statutes differ in the number of years a student should be classified as an English learner--some say 5 and some say 6, while others do not set a timeframe. There will also be differences based on how CDE and CDSS identify newcomer students.

Cost
Startup Costs: *Costs to begin collection including both direct costs and associated staff time for the Office and its data providers.*

Costs for the Office would focus on identifying how to calculate new metrics and validating the approach with data providers and experts. The staff would also need to spend time developing clarifying language regarding why data shown in the C2C analytical data set does not align with similar information produced by CDE and CDSS. Time spent on these tasks would need to be balanced by reducing time spent on other aspects of managing the C2C analytical data set.

The Office notes that provisioning LTEL and ARTEL data as defined by CDE's definitions would create an increased workload for both the CDE and the C2C. Additionally, it may increase the cost to the P20W system to ingest new data, re-ingest old data, and potentially update all the reports and tables that have been built off the previously submitted files. CDE requested that the C2C office estimate a cost for this work which should include the estimated staff FTE hours to update the file specifications document and update the system to ingest new data. After this, the CDE will calculate the cost for the work of creating flags across all years of data, recompiling and submitting these data, and working

with the C2C team to update the file specifications document as part of determining feasibility.

Ongoing Costs: *Costs necessary to maintain collection of the stated data elements for the Office and data providers.*

Once the new metrics are calculated and validated, ongoing costs should not be significant unless there were changes in the data collection process or elements of the data providers.

Compliance

Legal Requirements: *Compliance with privacy laws, intellectual property rights, and any other relevant regulations.*

All of the data in the C2C P20W dataset are subject to relevant state and federal privacy laws.

Scope: *Whether the proposal is consistent with the scope of work that is described in the Cradle-to-Career Act.*

The [initial legislation](#) prioritized providing information on the "impact of early education on student success and achievement as a student progresses through education segments and the workforce." Providing information on English language fluency in pre-K and K-12 would be consistent with the [dashboards](#) that were proposed to the legislature, including on early education (which lists outcomes like whether students are likely to be classified as English language learners once in K-12) and on college and career readiness (which proposes disaggregating outcomes based on whether a student was a childhood English language learner). However, specific data points that should be included in the analytical data set are not spelled out in the Cradle-to-Career Act.

In 2020-21, during the planning process for C2C, the working group discussed whether information on immigration status should be included in the P20W analytical data set. There was a conscious choice to not include this data point. For example, the [Research Agenda Subcommittee](#) elected to not distinguish between students who applied for federal financial aid and those who applied

for state financial aid for undocumented students who were brought to California as children.

Neutrality of the Office: *Whether the proposal might jeopardize the Office's neutral stance.*

Given that data points regarding English language learner status were part of the original purview of C2C, providing additional information on this subject is not likely to impact its neutrality.

Suppression and Regulatory Feasibility: *The impact of compliance policies on access to data based on the Office's data suppression policy and other regulatory concerns such as reidentification risks.*

Suppression protocols implemented by the Office should be sufficient to protect the privacy of individuals in public-facing tools and exceed suppression protocols used by both CDE and CDSS. However, given the sensitivity of data on immigration status, there is likely to be extra scrutiny on whether people who are flagged as English language learners could be reidentified.

Stewardship and Participation Agreement: *Concerns regarding the participation agreement and stewardship on the proposed data element and its applications.*

Given the sensitivity of data elements pertaining to immigration status, data providers often constrain the information shared within their agencies and so may choose to not provide this information to other state entities like the Office.