

DEFINITIONS MATTER! HOW ONE STATE IS ATTEMPTING TO STANDARDIZE DATA DEFINITIONS

Dawn Thomas¹, Kevin Dolan¹, & Keith Hollenkamp²

Illinois Early Childhood Asset Map, University of Illinois Urbana-Champaign (USA)

¹*Ph.D*

²*MS*

Abstract

As data collection and analyses proliferate our lives and raise expectations for what data can tell us, there are increasing demands for integrating data from several sources to get a wholistic view of the landscape of an area (e.g., a state), a population (e.g., young children), and/or field (e.g., early childhood education). This quest for data is particularly prominent in recent efforts to create more equitable educational systems. However, these initiatives are often hindered by different ways of collecting and defining data elements, such as age and race and ethnicity. The key to addressing this problem is identifying conflicting data definitions from different programs and agencies and working with all stakeholders involved to agree on common data collection and definitions.

Keywords: *Data definitions, longitudinal data systems, early childhood.*

1. Introduction

The Illinois Early Childhood Asset Map (IECAM) was created in 2006 to provide comprehensive early childhood data and maps to local and state agencies and other stakeholders to improve outcomes for Illinois children. As part of this role, IECAM works closely with early childcare education and care (ECEC) services are managed by multiple different government agencies in Illinois, the data collection and storage of these data are siloed away from each other. By having disparate data, analyses done on these data encounter problems with “comparing apples to oranges”: data that should be representing the same indicator not being defined in the same way. This is especially true with specific definitions on indicators such as age, race/ethnicity, and homelessness. Through all of our data work in Illinois, IECAM is trying to identify many of these definitional disagreements and standardize agreed-upon definitions for use across all Illinois agencies and ECEC providers.

In addition, IECAM also has been collaborating with the American Institute of Research (AIR) to identify and provide suggestions to fix the more common differences in definitions in ECE data work.

From the research, some very common disagreements included:

- What words do we use to describe an ECEC entity?
- Is it possible to have an accurate and standardized definition of age ranges?
- How do we define a child’s race and ethnicity so it’s standardized across multiple agencies?
- What is considered a “rural” area in Illinois?

A larger, systemic problem also came to light: How do we ensure an accurate use of terminology across the early childhood spectrum and maintain a consistent use of unique identifiers for each child across the whole system?

2. What words do we use to describe an ECEC entity?

Problem: A very broad topic that kept coming up when comparing data definitions at different agencies was the very general terms used when presenting information about ECEC providers in reports and data systems. Do you describe an ECEC provider by its funding source (federal department of education) or which agency is administering it (state board of education)? Or do you describe it by only its specific program model or by the age of children served? None of these are inherently incorrect, but there is inconsistency in current systems that causes confusion.

Across documentation and data systems, there is also disagreement on what words are used for the physical ECEC entity. According to the AIR report on data definitions, “terms like ‘facility,’ ‘site,’ and ‘program’ are used often interchangeably, yet the meaning of them may vary greatly” (D’Souza et al., 2022).

A clear definition on what constitutes a center-based program and a home-based program is also lacking because there are situations where center-based care providers may provide services inside the home.

Proposed Solution: Consistency in how ECEC entities are described is the major factor here. Making sure all data are inclusive of funding stream, age range, and program type from the beginning would help. This would make combining and analyzing data from different systems much easier and more accurate, especially when trying to pinpoint programs that affect a specific age range or group.

Clear definitions must also be representative of a community's needs and its work force. It should be very clear whether the data is referring to an actual care site or an administrative building. Adding an indicator to provider-level data that distinguishes between these different types of locations would go a long way.

A lack of a definition of what constitutes a center-based care provider is missing at the moment. An agreed upon definition from program providers, including what types of services that center-based care providers offer (including any home-based services) should be included in any data system.

3. Is it possible to have an accurate and standardized definition of age ranges?

Problem: The most consistent confusion was around age ranges for ECEC programs. Many programs in Illinois and the United States are for children of specific age ranges, such as 0–2-year-olds, or 3–5-year-olds. Or sometimes specific federal and state grants will target different age ranges such as birth to 5 or birth to 3. When looking at documentation for these programs and talking with the many different people involved, a lack of consistency in what ages these ranges actually include became apparent. The problem appeared to come from uncertainty about the upper limits of the ranges and whether they are consistent. When someone says that a program is for children “birth to 3,” does that include 3-year-olds? Or are they no longer eligible on their third birthday? Do newborn programs serve only children after birth or pregnant people as well? Using the word “to” in an age range seemed to spark much of the confusion, as different parties interpreted that as both inclusive and noninclusive of the higher age figure.

Proposed Solution: Clear definitions on what age ranges a program is designed for is a must, and those definitions must be agreed upon and used by all relevant parties. Deliberate language that avoids misinterpretation is also incredibly important. Avoiding the word “to” in an age range and instead adopting language like “children age 5 and under” indicates inclusivity. Being more specific by using months instead of years also can add clarity by narrowing the upper figure to a single month rather than an entire year. This can be done by changing “0–3 year olds” to something like “between birth and under 35 months.” As opposed to age 2 and under, the 35 months also provides a more specific reminder of the critical time a child may transfer out of one program and into another, such as the transition from early intervention to early childhood special education.

4. How do we define a child’s race and ethnicity so it’s standardized across multiple agencies?

Problem: Equity and inclusivity in ECEC programs are major concerns in Illinois. Data is a major component of making sure this is done accurately and fairly. Currently there are no standards at the state level that codify how race and ethnicity data are collected, which inherently is not a bad thing, but can cause problems when trying to get an accurate picture of data across programs and agencies. Sometimes this is even a problem within the same ECEC program. One particular program in Illinois houses its data in two separate, distinct databases. One database has 50 values to indicate race and ethnicity and the other has 10, all for the same ECEC program. This may be a case of over specificity, as was reported by AIR: “A clear set of race/ethnicity categories were used in documents, but they aggregated many people of color, undermining representation of diversity.” (D’Souza et al., 2022)

Proposed Solution: IECAM’s suggestion was to use the race and ethnicity definitions used by the U.S. Census Bureau. This would offer across-the-board consistency in collecting information on race/ethnicity. It would also make it easier for datasets from different agencies to be merged on race/ethnicity.

5. What is considered a “rural” area in Illinois?

Problem: Currently, rural areas in Illinois are defined using counties as the principal geographic unit of measure. There are a few definitions on what makes a county “rural.” According to the U.S Census Bureau, rural is defined as a county not part of a metropolitan statistical area (MSA) or a county that is part of a metropolitan statistical area but has a population fewer than 60,000. Another definition used is from the National Center for Health Statistics, which is a bit more granular. It classifies counties as *large central metro*, *large fringe metro*, *medium metro*, *small metro*, *micropolitan*, and *non-core*. From IECAM’s experience providing early childhood demographics for Illinois, we know that using counties for this offers consistency, but it also runs the risk of overlooking diversity and the specific needs of communities within counties. For example, IECAM’s home county has a population of a little more than 200,000, but most of its geographic area would be considered rural. The county’s population is concentrated in two adjacent cities in the middle of the county.

Proposed Solution: Using a smaller geography to determine rural status. IECAM provides population estimates at the zip code level, which could be used for more granular classifications.

6. How do we ensure an accurate use of terminology across the early childhood spectrum and maintain a consistent use of unique identifiers for each child across the whole system?

The overarching, structural problem that kept appearing is that there is not agreement across ECEC agencies and providers. As Illinois moves toward the creation of a new longitudinal data system (ILDS) that will house data across multiple agencies, a standardized set of data definitions is needed more than ever. The goal of the ILDS is to be able to track the entirety of a child’s educational journey in the state. To do this accurately and equitably, all identifiers and indicators used in the data must be consistent and reliable across all data systems. Without clear data definitions, this task becomes much more difficult. Ideally, a set of mandated rules from the Illinois government would be implemented for all agencies to follow. Until then, all we can do is offer researched evidence and suggestions on best practices.

Given our position as a major influence on ECEC data use in Illinois, we also must lead by example. On our website, we have extensive definitions pages for each type of data we provide and make sure to share those as widely as possible. IECAM is also a key developer in the ILDS system, so we are providing our expertise and definitional suggestions at the source.

7. Conclusion

Data usage and analyses are incredibly important when trying to make ECEC programs equitable and accessible to everybody that qualifies for them. In Illinois, many of the programs provided in early childhood are managed by different government agencies with their own data collection and governing standards.

Because of IECAM’s work as a major voice in ECEC data in Illinois, we’ve experienced many of these definitional inconsistencies firsthand and know how they can hinder quality analyses on programs that affect similar populations. To lower the risk of errors in data analyses, especially when that data is used in governmental decision making, ECEC data in Illinois needs to be clearly defined and those definitions need to be consistently used by all parties.

References

D’Souza, S., Howard, E. C., García-Arena, P., Le, V., & Caton, K. (2022, January). *Illinois preschool development grant birth-five needs assessment: Data recommendations report* (ISBE contract 21-586SBE-FINAN-B-21257). Arlington, VA: American Institutes for Research.