A5: Partnering with States to Address Licensing Research Questions

Wednesday, June 28, 2023

10:45 a.m. – 12:00 p.m. | *Gallery II*

1. Descriptive Information

A5: Partnering with States to Address Licensing Research Questions	Presenters
States and territories are responsible for regulating facilities that care for young children, including issuing licenses. Thus, licensing affects a broad array of child care and early education providers. In 2019, the Office of Planning, Research and Evaluation launched a 5-year research project—The Role of Licensing in Early Care and Education (TRLECE)—to synthesize the literature, provide a conceptual framework, analyze existing data, and collect new data to address gaps and priorities. Session presenters will briefly provide an overview of the project and then discuss TRLECE's analysis of licensing administrative data, including licensing data gathered using a web-scraping technique. The team's work highlights the importance of partnering with state leaders when conducting licensing research. A state child care licensing leader also will provide his perspectives on working with researchers on licensing research. The session will address questions such as the following: What are the possibilities and challenges in using a web-scraping technique on licensing or consumer education websites to gather licensing data for research? How can researchers partner with state leaders? What are the possibilities and challenges in using licensing administrative data to address similar research questions across multiple states? What kinds of research questions can be addressed with licensing data?	**Brenda Miranda,** Child Trends **Tracy Gebhart,** Child Trends **Diane Early,** Child Trends **Nina Johnson,** ICF **Dawn Jeffrey,** Arkansas Division of Child Care and Early Childhood Education **Simon Bolivar,** Child Care Licensing, Utah Department of Health and Human Services **Scribe** **Hannah Reutter** **Number of Attendees:** 10

2. Documents/Presentations Shared (Please list any electronic documents, PowerPoint presentations, or web links used during the session.) *Collect presenter PowerPoints or other documents on the flash drive provided.*

- CCEEPRC 2023 TRLECE slides-final

3. Brief Summary of Presentations

Summary of Presentation #1: Overview of Project and Session - Brenda Miranda, Child Trends

- Child care licensing establishes regulations that must be met to legally operate a child care program.
- Child care licensing also monitors and enforces those regulations.
- Purpose of TRLECE Research Project
  - Identify and address gaps in our knowledge of how key licensing features are related to quality and outcomes
  - Collect new data to help address the gaps in our knowledge
  - Develop resources to support states and territories in using data and research to strengthen their licensing systems
- Conceptual Framework for CCEE Licensing

- o This is the conceptual framework for the child care and early education licensing system we developed through the TRLECE licensing research project. We know there's a lot to take in. Although each state and territory has its own licensing system, we think each of the elements in the framework are present in each state/territory licensing system, though the specifics may vary.
  - o CCEE licensing is informed by constituents and interacts with partners and other entities to affect outcomes for providers and families so that, ultimately, children thrive. The framework also acknowledges that licensing systems exist within the context of the state and community.
  - o In reviewing the existing literature on licensing, we saw that much of the research focused on the regulations themselves or provided high-level descriptions about some licensing functions, like monitoring. We are analyzing licensing administrative data and conducting nationwide surveys to help answer questions like, "What are providers' perceptions of the burden and value of licensing?" "Is there evidence of inequities in licensing?" and "What are the roles and responsibilities of licensing staff?" We're focusing today's presentation on our research using licensing administrative data.
- TRLECE Research Using State Administrative Data
  - o We're doing two different types of research partnerships with states that rely on administrative data. The first one focuses on using publicly posted licensing data. We hope this will not only be helpful in better understanding licensing but also be helpful to other researchers who might be interested in using publicly posted licensing data.
  - o This first set of administrative data analysis focused on questions like: Do licensing violations vary by a) community poverty level? b) provider language? C) provider acceptance of subsidies? D) percent of community population that identifies as Black or Hispanic? The questions we can answer are limited by what data are publicly available. We're using licensing violations data and data from the American Community Survey that has data about community characteristics.
  - o Selected four states to scrape publicly posted licensing data to address questions about licensing, such as:
    - Do programs that accept childcare subsidies tend to have more (or fewer) licensing violations?
    - Do programs in communities with higher poverty levels tend to have more (or fewer) licensing violations?
    - Do programs in communities with more individuals who identify as Black tend to have more (or fewer) licensing violations?
    - Do programs in communities with more individuals who identify as Hispanic or Latino tend to have more (or fewer) licensing violations?
  - o Working with three states to use their administrative data (e.g., licensing, quality rating and improvement systems or QRIS) to address questions of mutual interest, such as:
    - Do licensing violations vary based on the percentage of children receiving subsidy in the program?
    - What is the relationship between licensing violations and a program's QRIS rating?
  - o For the second set of analysis, we are working with three states. State licensing leaders helped develop the research questions, and our plan is to examine the same questions across multiple states. We plan to use a range of administrative data—from licensing, subsidy, and QRIS.
- Surveys About Licensing
  - o We are currently conducting three nationwide surveys to better understand people's experiences with and perceptions of child care licensing.
    - 1. The first survey is with licensing administrators and is aimed at understanding how licensing systems function across states/territories and administrators' perceptions of their systems.
    - 2. The second survey is with front-line licensing staff who monitor licensed providers, and is aimed at understanding their characteristics, perceptions of their roles, and job challenges.
    - 3. The third survey is with licensed child care providers—both centers and family child care homes--and is aimed at understanding their experiences with child care licensing and ideas about what's working well and what could be improved.

Summary of Presentation #2: Project Approach to Engaging States and Others - Nina Johnson, ICF
- Purpose of Engagement
  - o Enhance project activities and improve the rigor and relevance of our work
  - o Inform project decisions, shape priority research questions, develop study procedures, interpret findings, and enhance the utilization of findings
- Approach to Engagement

- o The project has a pool of experts – which continues to grow—that we call our Technical Expert Panel (TEP).
  - o We engage subgroups to provide feedback on specific activities or products.
- Activities
  - o The project has or plans to have 43 different engagement activities to gather various types of feedback.
  - o Feedback on product development: TEP members asked to review and provide input on project documents via email. Document review is typically limited to a small number (2-3) of individuals who can provide targeted feedback on a specific project activity (e.g., report, brief, data collection instrument). All public-facing documents will be reviewed by at least one TEP member.
  - o Feedback about instruments: experts asked to provide written input on draft survey instruments or discuss their input in a small group
  - o Feedback about research processes and analyses: experts asked to share feedback via small group discussions on processes for collecting or analyzing information
  - o State research partnerships: States with which the project works to collect/analyze administrative data
- Expert Engagement
  - o The project has engaged 108 people to provide feedback.
  - o The technical experts in this project include individuals who are directly impacted by licensing systems, have substantive expertise in child care and early education (CCEE) licensing policy and practice, or have relevant research expertise.
  - o **Licensing representatives**: individuals who currently work or formerly worked within state licensing agencies and have experience with the development or implementation of licensing regulations, policies, and practices (e.g., regulation, monitoring, caseloads, staff development)
  - o **Researchers:** Individuals who have experience conducting research related to licensing or CCEE systems and can help guide research decisions about study design and data analysis.
  - o **Provider representatives**: providers or individuals who work directly with providers and understand the experiences of providers participating in the licensing system (e.g., entry into system, maintenance of compliance, quality improvement, enforcement, experience of being monitored, navigating multiple sets of standards and monitoring);
  - o **CCEE system representatives:** individuals who work in state or local CCEE agencies outside of licensing and understand the licensing system's position in and interaction with the larger CCEE system;
  - o **Other experts**: Includes state staff who work with licensing data and people with specific knowledge related to facilities needs assessments

Summary of Presentation #3: Lessons Learned from Web-scraping Licensing Data - Tracy Gebhart, Child Trends
- Web Scraping
  - o Web scraping is the process of extracting data from a website and exporting the information into usable format.
  - o Often, this process is done by sending an automated tool, often called a bot, to move through the website to extract specified information.
  - o This automated process allows users to extract large amounts of information from websites much faster than manually copying and pasting the information.
  - Ethical Web Scraping
    - o Though varying in enforceability, ethical users should consider these protections and terms of use <u>before</u> web scraping.
      - ▪ Just because a website exists, doesn't mean one can, or should, engage in web scraping. Sending a bot through a website thousands of times may tax the server, making it unavailable for other users to get the information they need. Additionally, nefarious actors may use the information scraped from a website for personal gain. Regulations regarding web scraping are still in development. However, there are four protections web administrators might employ to protect the information on their website: CAPTCHA (Completely Automated Public Turing tests to tell Computers and Humans Apart), Robots.txt (identifies information that cannot be accessed by bots), Terms of Service, and Website Permissions. An ethical researcher should review and adhere to these protections before engaging in the web scraping process. This recommendation <u>does not</u> constitute legal advice, but instead is offered as a best practice for ethical web scraping.
  - Scraping CCEE Licensing Data

- o 2016 CCDF Final Rule requires states to have a consumer education website with specific information, including licensing monitoring and inspection results
  - o There is variation in these consumer education websites:
  - o CCEE provider types, licensing rules, and visit types/frequencies.
  - o Format (e.g., online databases, summative tables of unpublished data, or scans of physical monitoring forms).
- Lessons Learned in Web Scraping
  - o <u>Start by investigating website structure.</u> Some website structures are easier to scrape than others. Additionally, scraping is limited to the data embedded on the website. Even if you are certain a state collects each licensing violation, but their website only displays a total count of violations, no code can make data appear.
  - o One valuable approach we used to help refine our research questions and select candidates for web scraping was to do a two-phase scan of websites. In the first scan, we observed which data were present as well as the permissions for scraping the data from the website. In the second, we assessed how complicated the scraping process would be. This told us what information we could collect and how timely that process would be before we dove into the work.
  - o <u>Anticipate changes to website or licensing standards.</u> A website may change organization or permissions, which could nullify web scraping code. In the time research is being conducted, there also may be changes to the licensing, like how COIVD-19 precautions changed the nature of licensing visits. Being prepared to adapt your code or adjust research inquiries to respond to differences in data over time is valuable.
  - o <u>Expect changes to CCEE facilities</u>**.** Ensure data cleaners are aware that CCEE facilities can, and do, have duplicative names and that there can be inconsistent names and addresses across data sources (e.g., Child Time v. Child Time LLC; 1412 Main Street vs. Across from the Main Street Wal-Mart). CCEE facilities may close, change ownership, or change location during the period of data collection. Be prepared to encounter changes to the website and the underlying data. With each departure, consider how these changes may introduce bias into the analysis.
  - o <u>Consider data reliability (as with all data sources).</u> Different data sources may populate the website (e.g., licensing reports, QRIS registration, provider reports), some fields may be inaccurate (e.g., language).
  - o <u>Partner for increased interpretability.</u> Data and findings may be difficult to interpret without a state-based licensing expert.
- Successes in Web Scraping
  - o <u>Reduced burden on data custodians.</u> Though we value state partnerships, web scraping offered the opportunity for data collection with reduced burden on licensing administrators and their staff to collate and share the data.
  - o <u>Access to rich data.</u> Web scraping provided access to rich data already published and used by state licensing administrators and consumers alike. In this project, we were also able to connect information scraped from other websites.
  - o <u>Automated processes.</u> The actual web scraping was conducted via bots, allowing for time-efficient automated data collection with minimal user input during the scraping process.
  - o <u>Value-ad.</u> We are in the process of sharing findings with state licensing administrators. When web scraping, sharing results or processes with data custodians could (1) provide opportunities for website improvement by identifying instances where scraped data are inaccurate, missing, or inconsistent and (2) provide insight on trends that may not be immediately apparent in website or database format.

Summary of Presentation #4: Working with States to Analyze Licensing Administrative Data - Diane Early, Child Trends
- Our Work with States
  - o Aimed to answer similar, policy-relevant research questions across 3-4 states.
  - o Outlined questions we thought could be addressed using administrative licensing data
  - o Approached four states, based on our understanding of their data.
  - o Three agreed to partner with us.
  - o Each state required several calls with administrators and data staff to:
    - Understand their licensing system
    - Gain needed permissions
    - Finalize decisions about which questions we would address through our research and analysis.

- Create and execute an MOU
- Request and receive the needed data
- Understand the data they provided
- Research Questions Co-Created with States
  - Predicting licensing violations from:
    - Percent of children receiving subsidies
    - Community poverty level
    - Community race/ethnicity
    - Urban/non-urban
    - QRIS rating
  - Standard set of co-variates in each model
  - Separate models for centers and family child care
  - Separate models predicting total violations and high-risk/severe violations
- Working with Administrative Licensing Data
  - Understanding the data required multiple conversations.
  - Examples:
    - Different terms for violations (e.g., noncompliance), with different meanings within/across states
    - Number of visits in a year may be related to past compliance
    - Documenting every rule inspected vs only those violated
- Answering Similar Questions Across States
  - Cross state analysis decisions:
    - Which licensing visits to include in the dataset
    - Timeframe
    - Ensuring variables were comparable across states
- Opportunities and Challenges
  - Child care licensing is a foundation of the CCEE system, but has received little research attention
  - States have a wealth of licensing administrative data and many are eager to support analysis
  - However, the data are often messy and easily misunderstood
  - Partnering with states to understand their data is critical

Summary of Presentation #5: State Perspectives on Licensing Research - Simon Bolivar, Utah Licensing, and Dawn Jeffrey, Arkansas Licensing
- State perspectives on research
  - We manage a lot of data, and it will be important to know what it is we want to answer and what researchers want to know.
  - When you are in the middle of it, you don't know what questions you should be asking.
  - Wish state licensing teams used the same language. It is hard to understand sometimes when people use different terms It would be helpful to define terms used in licensing.
  - Suggest that researchers develop simple, short reports that leave readers hungry for more. These short reports should reference longer reports of solid research as back-up materials for readers who are interested in learning more. One-pagers and infographics are great.

4. **Brief Summary of Discussion**

   Q&A:
   - Q: When states are collecting info about licensing and different areas of the states, do they involve Tribal entities?
     - A: It varies by state whether it is the same system or if there is a different system for Tribal licensing than state licensing.
     - A: When we began this research, we decided to focus on states. There is a lot to learn about state child care licensing as well as licensing with Tribal communities. We hope that future research could focus on licensing within Tribal communities.
     - A: We incorporate tribes into our system depending on the MOU (from Utah)
   - Q: What program do you use for webscraping?

- o A: We used Python, which is free. The syntax has a learning curve, but we built the code from scratch. It is a steep learning curve, but it is free. We also connected with other researchers in the field to help us think about what we should keep in mind before we started.
- Q: Were your outcomes only licensing violations?
  - o A: Yes. We wanted to try to find questions that can be asked across multiple states. Licensing violations is the thing that every state has in common (though their regulations may vary).
- Q: Did you reconcile the data scraped with administrative data?
  - o A: We did a check of other publications to see what should have been in our data and what we actually got. We also checked with licensing agencies to ask if the data that we got matched with what they were expecting.
- Q: Can you talk about what kinds of things licensing staff read?
  - o A: Graphics are generally preferred. Some simple graphics are helpful. If there is language that I used, it should be a short paragraph accompanied by a picture. If their attention is not captured in the first 3 seconds, they aren't going to read it.
  - o A: People want to know what the kids are doing.
- Q: Are there any specific topics you'd like to learn more about?
  - o A: I would like to see where we are a year from now
  - o A: I would like to see what research says about interaction with children. What triggers or prevents barriers in the early years of education?
  - o A: How best to balance enforcement and TA?
- Q: Do you have regulations in licensing for things like mental health?
  - o A: We are creating rules because there is a lot of unknowns around mental health.
- Q: Have you developed or come across anything that would show the elements of licensing rules across states?
  - o A: The National Database of Child Care Licensing Regulations includes information about child care licensing regulations across states.

5. **Summary of Key issues raised** (facilitators are encouraged to spend the last 3-5 minutes of sessions summarizing the key issues raised during the session; bullets below are prompts for capturing the kinds of issues we're looking for)

- *Licensing administrative data can be used to address interesting, important research questions.*
- *States have a wealth of licensing administrative data, though the data may be messy and difficult to understand.*
- *Research using licensing administrative data is best done in close partnership with states.*