SECTION 3, CHAPTER 12

EQUITY AS A PERSPECTIVE FOR IMPLEMENTATION RESEARCH IN THE EARLY CHILDHOOD FIELD

Milagros Nores, Ph.D., National Institute for Early Education Research

A data collector in a U.S. preschool classroom observed a teacher call security because she perceived a child as being disrespectful and difficult. The preschooler was observed being removed from the classroom. This occurred during a standard observation of classroom quality in one of our research projects. Standard research practices with respect to processes in early childhood may end with the classroom being given a high "negative discipline" score. Because of the limitations of standard protocols, unanswered questions remain when looking at the data. Was the child black? A boy? Hispanic? All three? To the extent that research on processes inquires more deeply into these questions, it may more fully account for how programs operate and are implemented and shed light on the biases that are reproduced in early childhood systems.

This anecdote is one of many in the research that demonstrates how the measures we use and the protocols we enact provide only a limited view of the issues and problems embedded in the implementation of policies and practices in early childhood. This chapter therefore delves into the question of equity and why equity matters in early childhood education and development (ECED) programs. It also explores the central role of research in deciphering how and when ECED programs do in fact contribute to equity (or not), and, more specifically, how equity can be embedded in evaluation designs.

Equity is "the absence of systematic and potentially remediable differences in one or more aspects ... between groups of people characterized socially, geographically, or demographically" (Starfield, 2007, p. 483). Inequities may be rooted in discrimination due to gender, disability, race/ethnicity, language, minority status, or religion; structural poverty; geographic isolation; weak governance; and cultural norms (Bamberger & Segone, 2011). Critical race theory—which contends that research and discussion of social inequity, and school inequity in particular, should consider race and racism—has been central to strengthening the ECED field's conceptualization of inequities (Ladson-Billings, 2004).

A vision of increasing equity inspired the growth of ECED programs that reduce disparities, readiness gaps, and inequities at the starting gate, and equalizing the playing field at kindergarten entry—goals that are part of the mission of many preschool programs across the country.[1] This vision and mission derive from years of research on how preschool programs may affect not only middle-class children but also disadvantaged, special needs, and dual language children, among others (Yoshikawa, et. al, 2013).

---

[1] For example, Head Start states "that every child, regardless of circumstances at birth, has the ability to succeed in life" (https://www.nhsa.org/about-us/mission-vision-history). The Abbot preschool program implementation guidelines state that "intensive, high-quality preschool programs can close much of the early achievement gap for lower income children" (https://www.nj.gov/education/ece/guide/impguidelines.pdf). The Seattle preschool program includes a "commitment to early learning as the foundation for future academic success and a strategy for closing opportunity gaps" (https://www.seattleschools.org/cms/One.aspx?portalId=627&pageId=33661301).

But not all programs are created equal (Yoshikawa et al., 2013; Camilli, Vargas, Ryan, & Barnett, 2010). Research on program quality and processes and on implementation has helped us understand why some programs work and some do not, and why some work for some children and not others—information that is crucial to an equity-based evaluation (Bamberger & Segone, 2011). Research can not only help bring to light what works in the early years but can also document how programs contribute to increasing equity (or reducing inequity) and at what point in the education process they do so. That is, it can help us understand the effectiveness, efficiency, relevance, impact, and sustainability of ECED programs with respect to equity goals.

However, research on what occurs in preschools classrooms, teacher practices, interactions, the effectiveness of programs or preschool curricula, and ultimately, their effect on children cannot be separated from the biases and inequities that children and families may experience in the education process and the social structures in which schools and individuals are embedded. Biases and racism are present as early as preschool and kindergarten, whether it be in teachers' perceptions of Black children's behavior (Ladson-Billings, 2011; Yates & Marcelo, 2014), in perceptions of Black girls as less innocent and more adult-like, a perception known as adultification (Epstein, Blake, & Gonzalez, 2017), or in children's own perceptions of race (Farago, Sanders, & Gaias, 2015). More recently, research on preschool expulsion has also shown how implicit biases in preschool may also be determining disciplinary behavior early on (Mitchell, Fonseca, & LaFave, 2016).[2] To the extent that we care about equity, research should, when feasible, measure the degree to which processes and programs in early childhood reduce or exacerbate inequities and what exactly in the program's design or its implementation is contributing to these results.

Yet we cannot escape the fact that research itself—and the measures, researchers, observers, interviewers and other agents of research—may introduce biases of its own to any evaluation process. And if questions pertaining to equity are not asked, then equity is not assessed at all.

> **We cannot escape the fact that research itself—and the measures, researchers, observers, interviewers and other agents of research—may introduce biases of its own to any evaluation process.**

All of this matters in terms of research validity (American Evaluation Association, 2011; Kirkhart, 2010, 2013). Kirkhart defines multicultural validity as the "accuracy or trustworthiness of understandings and judgments, actions, and consequences, across multiple, intersecting dimensions of cultural diversity" (2010, p. 401). She argues that validity is enhanced when attention to cultural diversity and reflection on cultural biases helps guide the choices of epistemologies, methods, and procedures. She further argues (2005) that validity is threatened when culture is ignored or diversity stereotyped.

---

[2] Research on implicit biases and behavior expectations of teachers reveals that preschool teachers are more likely to expect challenging behaviors from black children and, in particular, black boys (Gilliam, Maupin, Reyes, Accavitti, & Shic, 2016). The authors define "implicit bias" as the "automatic and unconscious stereotypes that drive people to behave and make decisions in certain ways" (p. 3).

Equity in research implies capturing the extent to which programs, policies, and interventions reduce or increase inequities, validly defining inequities in relation to the context and the disadvantages that are present, and integrating the concept of equity into all components of research, from the questions asked to the analysis and interpretation stage. In sum, understanding equity means being able to answer questions that attend to equity concerns. Who are the less advantaged, and how does this evaluation capture their experience with ECED policies and programs?

EARLY CHILDHOOD PROGRAM EFFECTS AND EQUITY

Research on early childhood has provided quite robust evidence regarding the importance of preschool and has still more to contribute in terms of structure, curriculum, program features, and leadership, among other aspects (Bowne et. al., 2017). Research on quality preschool programs has shown that small- and large-scale public programs can have long-term and substantial effects on children's developmental trajectories (Camilli et al., 2010, McCoy et al., 2017, Yoshikawa et al., 2013). Research also shows that while all children can benefit significantly, children from low-income backgrounds (Gormley, Gayer, & Phillips, 2008; Weiland & Yoshikawa, 2013), children with special needs (Phillips & Meloy, 2012; USHHS, 2010; Weiland, 2016), dual language children (Barnett et al., 2007; Bloom & Weiland, 2015; Bumgarner & Brooks-Gunn, 2015; Dickinson & Porche, 2011; Goldenberg, 2012; Puma et al., 2010; Slavin, Madden, Calderón, Chamberlain, & Hennessy, 2011; Wilson Dickinson, & Rowe, 2013), and children from a racial or ethnic minority background (Gormley, Gayer, & Phillips, 2008; Weiland & Yoshikawa, 2013) may benefit as much or more than others.

For example, studies of universal preschool programs in Boston (Weiland & Yoshikawa, 2013) and Tulsa (Gormley, Gayer, & Phillips, 2008) have found positive effects on children's math and reading achievement scores (among others) at kindergarten entry. These effects were larger for low-income, African American, and Hispanic children. Figure 1 (based on Friedman-Krauss, Barnett, & Nores, 2016, p. 11) shows average effects across these two programs reported in months of learning. Using these averages, Friedman-Krauss et al. (2016) have estimated that on average, universal programs of the same quality could reduce gaps in math skills for African Americans by 45% and for Hispanics by 78% and eliminate reading gaps for both these groups of children. While individual state population compositions and readiness gaps differ, with some of them exhibiting large percentages of white low-income or native low-income children, these projections have nationwide implications. A meta-analysis that covers 23 early education programs from the perspective of gender equity (Magnuson et. al, 2016) finds that effects are generally similar for boys and girls. Differences are observed mostly across middle childhood, when the programs seem to have a greater impact on boys with respect to grade retention and special education placement.
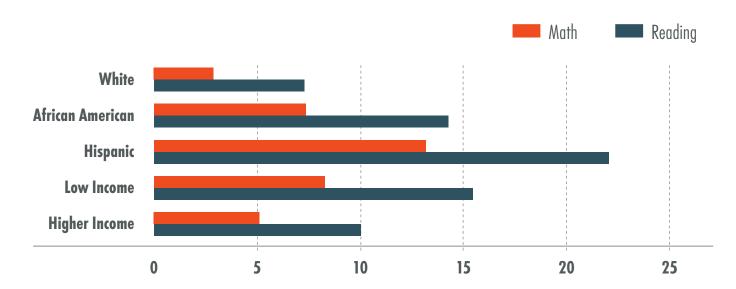
Figure 1. Average positive effects across two universal preschool programs in months of learning.

Notes: The low-income category includes children with household incomes at or below 200% of the federal poverty guidelines, while the higher-income category includes children with household incomes above 200% of the federal poverty guidelines. Source: Friedman-Krauss et al., 2016, p. 11.

HOW DO WE DEFINE EQUITY IN RESEARCH?

Equity-focused implementation research can be understood as "analyzing the impact of internal and external processes, as well as foundational assumptions and interpersonal engagement, on marginalized and under-served individuals and communities" (Spark Policy Institute, 2014) within the process of implementation research, that is, within the process of inquiring how programs, policies, and individual practices are enacted in real-world settings (Halle, 2020). Equity, therefore, is a perspective a researcher brings to the research process that calls for understanding the "complexity and multidimensionality of context, culture and power as fundamental elements to be addressed in evaluation" (Dean-Coffey, Casey, & Caldwell, 2014, p. 84). Ultimately, the goal of equity in research is to ensure that research components capture whether a program is working toward reducing inequities and is validly defining these inequities in relation to the context and populations at hand and that evaluations of processes and programs are not introducing biases that reduce the chances of understanding whether the program works and, if it does, for whom.

A similar and highly interconnected concept (or evaluation paradigm) that has gained traction as a mechanism with an equity perspective is cultural competence, which involves understanding the unique and defining characteristics of different populations with which researchers engage (Harvard Clinical and Translational Science Center, 2010). The culturally competent researcher values diversity, understands the dynamics of the differences among subpopulations,

and has the capacity to adapt to diversity (Shiu-Thornton, 2003).[3] An analogous concept is cultural responsiveness, which is defined as "a theoretical, conceptual and inherently political position that includes the centrality of, and attunement to, culture in the theory and practice of evaluation" (Hood, Hopson, & Kirkhart, 2015, p. 283).

Lastly, intersectional approaches "challenge practices that isolate and prioritize a single social position and emphasize the potential of varied inter-relationships of social identities and interacting social processes in the production of inequities" (Bécares & Priest, 2015, p. 3). From a research perspective, intersectionality means adopting an approach to the subject of study in which multiple marginalizations (by sex, gender, race, ethnicity, income, social class, education, age, sexuality, immigration history, geography, among others) are considered, rather just a single difference. Bauer (2014) proposes that these should be considered in an additive scale (in quantitative studies, this relates to measuring the combined added effect of two characteristics as different from the sum of each individual characteristic alone). Such approaches can further the field's capacity to specifically document inequities in early childhood within intersectional groups—African American boys, African American girls, Native American girls, Hispanic immigrant children, or Muslim immigrant children, for example (Ford & Harawa, 2010). As Bauer (2014) points out, carefully considering intersectional issues can reduce measurement bias, improve construct validity, allow identification of heterogeneity of effects, and avoid the problem of average total effects that do not represent any true group (see also Whitesell, 2017).

Equity, cultural competence and responsiveness, and intersectional approaches all interconnect in central ways in the design, collection, analyses, and interpretation stages of the research work. At their core is an emphasis on understanding the complexity of social and power dynamics and an explicit attempt to recognize, measure, and assess differences, as well as reduce biases (as much as possible) and employ culturally appropriate methods. In essence, as we assess early education programs, we must take into account that these programs take place in various settings and contexts; that they have differential effects on children of different racial, ethnic, language backgrounds, of differing genders, and with differing needs (among other aspects); that children in different types of settings (e.g., urban versus rural) may have different levels of cumulative deprivation; and that all of this is central to understanding (and measuring) differences, effectiveness of processes, interactions, curriculums, and detractors and contributors throughout. At the same time, researchers should minimize any biases introduced by the research itself and strive to comprehend any cultural limitations to its methods, instruments, collection processes, or analyses.

These processes are applicable regardless of the type of research. The discussion in the next section recognizes that this may encompass (but not be limited to) basic science research, clinical or randomized trial research, ethnographic research, mixed methods research, or community-based participatory research, among others.

---

[3] In addition, the American Evaluation Association defines it as a process of learning and relearning, awareness of self and one's cultural position, refraining from assuming a full understanding of stakeholder perspectives, and recognizing dynamics of power (2011, p. 3).

The same is true for process, progress, and summative evaluations. Process evaluations focus in particular on how program or project components interconnect and are being implemented. Equity in this sphere would ensure not only that implementation is being documented but that the methods and measures used for the process apply an equity lens for interpreting progress (Frierson Hood, Hughes, & Thomas, 2010). Progress evaluation focuses on whether progression toward stated goals is taking place. Equity questions that may be put forward include whether the goals respond to different types of individuals and needs and whether there is any indication of equitable progress. Summative evaluations are intended to show a program's effectiveness. The role of equity here is to assess whether gains are inclusive and to situate the results in the contexts and environments necessary to interpret them adequately.

WHY ARE THESE APPROACHES IMPORTANT FOR RESEARCH?

Grounding research in equity-based perspectives, cultural competence, and intersectional approaches enhances it in various ways. Cultural competence heightens effective interactions between researchers and participants in both qualitative and quantitative research. This happens because researchers actively seek to engage with the diverse perspectives and segments of the community, respect the cultures represented, and remain aware of how their own backgrounds and experiences limit or enhance the conduct of research (American Evaluation Association, 2011).

> **Cultural knowledge comes from understanding differences, similarities, and inequities that may be structurally determined.**

More specifically, Papadopoulos and Lees (2001) put forward a model for the development of culturally competent researchers based on cultural awareness, cultural knowledge, and cultural sensitivity. These authors developed their framework in the nursing field, but these concepts can be incorporated into the more general notion of cultural competence. They illuminate how cultural competence can enhance interactions between researchers and participants via awareness, which they define as a process in which researchers reflect "on how their own values, perceptions, behavior, or presence and those of respondents can affect the data they collect" (p. 260). Cultural knowledge comes from understanding differences, similarities, and inequities that may be structurally determined. Cultural sensitivity derives from a true partnership with the agents of research. The authors argue that matching ethnicities of interviewers and participants, for example, encourages the latter, although it does not guarantee it (Frierson, Hood, & Hughes, 2002). Researchers, they add, should also ensure that all research components, including design, data collection, analyses, interpretation, and dissemination, are guided by cultural awareness, knowledge, and sensitivity. Cultural competence is not a series of steps that a researcher carries out apart from an evaluation or research process; rather, it undergirds how that process is carried out (Frierson et al., 2010; Hood, Hopson, & Kirkhart, 2015).

Integrating these concepts into various research components can ensure that racism is challenged, ethnocentricity is considered, and essentialism (blaming culture for results observed for a group) is avoided.

Another strength of foregrounding equity and cultural responsiveness is that it improves communication with racial and ethnic minorities or other groups (for example, language minorities) in research studies. It also produces a more accurate representation of cultural processes and practices because the researcher understands and effectively responds to factors that might influence individuals' participation, whether they be children, families or staff members, such as their history, their circumstances, and current policies that affect them. Kien Lee (2007) provides examples: an evaluator in a Native American community will be much better equipped if she understands the history of oppression, sovereignty struggles, and research misrepresentation that Native Americans have faced (see also LaFrance & Nichols, 2010). Likewise, evaluators working with women need to understand and account for existing gender roles. Similarly, working in settings with large immigrant populations requires understanding immigration policy (see, for example, Allman & Slavin, 2018).

An equity lens also incorporates an adequate representation of groups (Hood, Hopson, & Kirkhart, 2015). This requires purposeful methods for securing consent, sampling, and recruiting. Intersectional or multicultural representations across categories (race, ethnicity, religion, gender, age, language, disability, and socioeconomic background) allow for an understanding of differences and inequities as well as of pathways for inequities (Kirkhart, 2010; Bécares & Priest, 2015). The categorical labels that are most frequently used to represent individual characteristics (race, ethnicity, gender, age, language, or disability) do not capture the whole of human diversity because diversity is also constituted within categories, and it is crucial to understand the intersecting cultural identifications that these combinations represent (Kirkhart, 2010).

When it comes to measuring implementation in ECED programs, Aboud and Prado (2018) suggest that there may be various alternatives depending on the goal of implementation, whether it is piloting a program to determine feasibility or examining a well-developed program, in which the focus would likely be on quality and fidelity, among others. They explain that most ECED programs can be categorized as being delivered to children either directly (e.g., preschool) or indirectly via caregivers (e.g., home visiting). In this context, equity will come into play through the effects of the program on children (e.g., when assessing a pilot), the practices and processes observed by caregivers and teachers, curriculum enactment, enrollment practices, exclusion/inclusion of children/parents, attendance rates of children/home visitors, or expulsion practices, among other things.

COMPONENTS OF RESEARCH

Thomas and McKie (2006) provide examples of how researchers' values, beliefs, and biases can compromise an evaluation process. The questions asked and the questions not asked, what is focused on versus what is minimized, the evaluation approach selected versus the one discarded, the data collected versus the data disregarded, the interpretations made, and how and to whom the results are presented can all undermine an evaluation.

An approach to research that truly incorporates equity requires integrating equity concepts across all these components, from questions asked to interpretation (Hood, Hopson, & Kirkhart, 2015).

THEORETICAL FRAMEWORK AND EVALUATION QUESTIONS

Research and evaluation are grounded in theory: evaluation theories, social science theories, program theories, and theories of change, all of which signify implicit and explicit assumptions about how programs or practices operate and how individuals respond to such programs or practices (American Evaluation Association, 2011). Therefore, as the theoretical framework for research is developed, researchers should explicitly examine the values, beliefs, and approaches embedded in it as well as whether it fits the "evaluated" population. The American Evaluation Association (2011) advocates that researchers thoughtfully consider alternative competing frameworks, assess fit of theory to the context, and pay attention to complex power explanations within systems.

A crucial step in any evaluation is defining the questions to be addressed. The questions and how they are worded are critical to setting the evaluation on the right path. They may address needs and strengths, processes, use of resources, progress toward outcomes, and effectiveness, among other things (Hood, Hopson, & Kirkhart, 2015). Thinking in terms of equity when developing research questions entails considering whether processes are strengthened or hindered by culture, which may point to cultural fitness, on the one hand, or suggest that adaptations are needed, on the other. It also requires understanding the distribution of benefits. For example, is the program benefiting some groups more than others? Is the program reducing initial disparities among individuals? Are research questions addressing differences across and within relevant groups? Are processes reducing inequities? And if the answer to any of these questions is yes, implementation researchers must explain why. For example, are any subgroups with lower rates of absenteeism, and if so, why? Does any group show high teacher turnover and, if so, which teachers and why?

DESIGN AND SAMPLING

Design encompasses the sources and type of data, the individuals from whom evidence will be drawn, the approach (quasi-experimental, experimental, ethnographic, case study, or mixed methods), and the timing, among other aspects. Here equity will define who is represented, whether differences between and within groups can be

assessed, and how much information is collected in processes that will contextualize and identify the sources of differences across groups. Examples of questions researchers can use to guide design and sampling are who is included with this design, who is excluded, and whether the different groups that make up the target population will be well represented.

The degree to which design decisions bear on who is included and who is excluded is a central equity consideration. In quantitative designs, researchers pay close attention to selection bias and its implication for the design, the analytical strategy, and the interpretation. Heckman (1990) defines selection bias as the "distorted representation of a true population as a consequence of a [nonrandom] sampling rule" (p. 201). Distorted selection rules are likely the outcome of self-selection decisions by families, children, teachers, principals, and so forth. And selection rules introduced by the design may also generate selection biases. For example, say we are studying a program that assesses the impact of a specific racial justice curriculum, but only parents who are interested opt in to these classes, while parents of other children just continue in general education classes. The evaluation will then confound program effects with the effects of families or home environments. These parents are particularly motivated by this type of content, which very likely impacts other choices and behaviors in the home and, ultimately, would also impact the outcome of interest. If we understand the selection rules that define who is the target of a program or the intervention focus of a particular study, we can understand who is left at the margins, whether the design can find ways to include them, and to what degree the research is valid and generalizable (Willis & Rosen, 1979; Grimes & Shulz, 2002). Randomization helps to avoid selection bias and create comparable groups at baseline, yet it does not eliminate biases, such as those due to measurement, attrition, or low response rates, from other evaluation components (Torgerson & Torgerson, 2003).

Closely tied to the issue of selection bias are process aspects such as barriers to participation in the intervention or program evaluated, as well as in the study itself. It is important to create design and research strategies that address participation and take into account timing and sampling. Will the researchers be able to distinguish differences across disadvantaged groups with the design and sample size that is proposed? That is, is the statistical power sufficient for quantitative inquiries such as subgroup analyses, and are all groups in fact represented so that the investigation is qualitatively adequate? Sampling also has key implications for coherence and biases in qualitative methods, where researchers need to specify what is included or excluded when it comes to sample size, sampling strategy (random sampling, convenience sampling, stratified sampling, cell sampling, quota sampling, or a single-case selection strategy), and sample source (Robinson, 2014). For example, in both quantitative and qualitative studies, we pay close attention to teachers but rarely include teacher assistants as informants on quality. Yet they often more closely represent the children's culture than do the lead teachers (Figueras-Daniel, 2016). Similarly, the literature often does not follow up on what drives program attrition, and attendance issues and costs of ECED programs are rarely reported (Connolly & Olson, 2012; Logan, Piasta, Justice, Schatschneider, & Petrill, 2011; Greenwood et al., 2018).

INSTRUMENTS

Instruments may themselves introduce biases. The American Evaluation Association (2011) recommends choosing data collection instruments that have been used with the populations of interest and that have shown sensitivity to those populations. This does not guarantee a lack of bias, as there is no perfect instrument. But it does make it more likely that an instrument will effectively capture increases in equities (changes over time and between groups) in the disadvantaged populations of interest. When using standardized instruments, researchers may have to review their weaknesses for particular subgroups in the population of interest. Who does the instrument not measure well? That is, researchers should reflect critically on "what constitutes meaningful, reliable, and valid data" (American Evaluation Association, 2011, p. 9), starting at the planning stage and continuing throughout data collection.

As an example, quantitative evaluations measuring the impact of specific preschool-age interventions and/or preschool programs have many times relied on the Peabody Picture Vocabulary Test (PPVT) (Dunn & Dunn, 1997; Dunn, Dunn, & Dunn, 2007). The PPVT has shown sensitivity in gauging growth in receptive English vocabulary in children identified as African American (e.g., Weiland & Yoshikawa, 2013), Hispanic (e.g., Bloom & Weiland, 2015; Weiland & Yoshikawa, 2013), and dual language learners (e.g., Bloom & Weiland, 2015; Durán, Roseth, & Hoffman, 2010; Slavin et al., 2011) across many evaluations. Despite having shown sensitivity to specific population groups, instruments may have biases that are yet unclear, and the PPVT has been challenged on the basis of limitations in assessing dual language competencies in the early years (Bandel, Atkins-Burnett, Castro, Wulsin, & Putnam, 2012). Further research could help establish measurement invariance for different subgroups. For example, Nores and Barnett (2018) have established that the PPVT-III performs equally well between English and Spanish home language speakers and between boys and girls. Because they lacked a sample with a language difference for the PPVT-IV, the authors could only replicate this process for gender difference, establishing partial measurement invariance between boys and girls for the measure. Similar analyses are needed for most measures used with preschool children and infants.

Including individuals from the population of interest in the processing of vetting instruments that are being piloted would help reduce biases (O'Brien et al., 2006; O'Brien, Harris, Beckman, Reed, & Cook, 2014). This vetting process could take culture, race, ethnicity, and language into account as well (O'Brien et al., 2006; Public Policy Associates, 2015; see Appendix). The same is true when translating or adapting an instrument (Dettlaff & Fong, 2011).

We also have much more to learn about the weak associations between existing measures of classroom quality and children's learning (Burchinal, 2018). Researchers have started to push for more depth or further content specialization in the process measures used in early childhood education to understand quality (Atkins-Burnett, Sprachman, Lopez, Caspe, & Fallin, 2011; Goodson, Layzer, Smith, & Rimdzius, 2004; Zaslow et al., 2016) and to measure program impact on different subgroups of children, such as dual language learners or children with special

needs (Castro, Espinosa, & Páez, 2011; Halle, Vick Whittaker, & Anderson, 2010; Peisner-Feinberg et al., 2014; Soukakou, Winton, West, Sideris, & Rucker, 2014).

Similarly, measures are starting to be developed to further inquire into leadership and climate (e.g., Pacchiano, Klein, & Hawley, 2016; Whitebook & Ryan, 2012) in early childhood education settings. These are still new in the ECED field, and pending further inquiry we do not yet know whether these measures respond to the different types of programs and different populations served.

FIELDWORK

Fieldwork encompasses ethics approvals, recruitment strategies and training of field personnel, management of data collection, consenting procedures, survey and interview protocols and procedures, focus group protocols and procedures, retention policies and strategies, and translation and interpretation services. A lot of culturally responsive work should occur at the fieldwork stage, where one-on-one interactions take place between a research team and partners in the field who are willing to be research subjects and agents.

Cultural competency assessments and frameworks are highly relevant to this stage of work. The Appendix lists questions associated with various frameworks and self-assessments regarding whether assessors require culturally competent training, how to determine criteria for choosing interviewers, and how to create a flexible process that accounts for the needs of individuals or contexts (O'Brien et al., 2006, Public Policy Associates 2015; Whitesell, 2017).

Consent strategies and issues of representation are central to any evaluation. It's critical to use strategies that promote comprehensive participation, including making accommodations for language as necessary (American Evaluation Association, 2011), and to reduce barriers to the participation of groups in the study. This is especially important because active consent already reduces representation of disadvantaged populations in education research (Bergstrom et al., 2009; Flay & Collins, 2005). Accommodations should also extend beyond the consent period, to communication, assessment, survey, interview, and all evaluation activities (American Evaluation Association, 2011); this may necessitate translation or interpretation services.

Retention policies and strategies (including incentives) should reflect the culture and the individuals or children who take part in the study. They should also be effective at reducing the impact of differential attrition of particular subgroups. This will help retain validity and preserve the capacity of the study to answer questions on equity. Research on factors affecting survey response (Edwards et al., 2002; Fan & Yan, 2010), as well as on effective retention strategies for samples (Robinson, Dennison, Wayman, Pronovost, & Needham, 2007) has shown that accounting for these factors—and for demographic differences among leadership, staff, and children—can increase response rates and reduce differential attrition.

METHODS AND ANALYSES

Initial checks at this stage should ensure that attrition and/or survey response has not been differential. That is, the processes used for design and sampling, instruments, and fieldwork should not result in a sample that is more representative of a particular category (by language, race, ethnicity, gender, immigration status, or other identification) than the target population. Did only some teachers answer the surveys? Who attended the focus groups? Who finished the assessments? Who attended the program? The training? Differences between the target group and the final sample need to be clearly reported, both because they may bias results and because they are necessary to interpret analyses.

Central equity questions at this stage include the following. Are there outcomes differences, intended and unintended? Are the data disaggregated along demographic lines so that it is possible to understand programs along lines of race, culture, socioeconomic status, language, and so forth? Were there factors that contributed to disparities (or reduced disparities)? Were there any unintended changes or consequences due to cultural/racial/ethnic considerations? (O'Brien et al., 2006, Public Policy Associates, 2015; see Appendix). The study has to have the statistical power to answer such questions across subgroups or intersections.

INTERPRETATION AND DISSEMINATION

Dissemination and interpretation should be based on all the concepts presented so far. Questions that can be addressed at this stage include the following. Are the main results consistent for all subgroups, or is there evidence of heterogeneous subgroup differences? Are interpretations of subgroup differences contextualized? Are institutional or programmatic factors that contributed to subgroup effects shown? Does the program reduce equity for participants along particular dimensions? Is it neutral? Negative? What factors are contributing to or hindering equity?

Interpretation should reflect the context studied and address whether the feedback based on race, ethnicity, gender, language, or another individual characteristic allows the program and agents of change to engage the system in long-term equitable change (O'Brien et al., 2006). As the Tribal Evaluation Workgroup (2013) puts it, "Evaluation should inform practice, program, and system improvement, providing information to answer questions that local program directors and staff have about how to better serve the children and families in their communities" (p. 23). In addition, assessing social (economic, sociological, political, and cultural) explanations of processes and outcomes, as well as the social institutions and processes that influence the generation and allocation of resources, can further support a comprehensive equity-focused agenda (Östlin et al., 2011).

Efforts such as the CONSORT, STROBE, COREQ, SRQR and SAGER guidelines have strengthened the research field by requiring consistency in reporting on quantitative and qualitative research (Schulz, Altman, & Moher, 2010; Bastuji-Garin et al., 2013; Tong, Sainsbury, & Craig, 2007; O'Brien et al., 2014; Heidar, Babor, De Castro, Tart, & Curno, 2016). Yet most of these do not address equity per se. SAGER focuses on sex and gender in reporting, COREQ addresses possible biases in qualitative designs, and more recently, the CARE guidelines (Yousafzai et al., 2018) have put forward a framework for reporting on implementation research. But even though these guidelines do not directly address equity, they require contextualizing results and thus provide an initial step toward strengthening reporting in implementation studies.

CONCLUSION

In essence, addressing equity in research implies capturing the extent to which programs, policies, and interventions reduce or increase inequities, validly defining inequities in relation to the context and the disadvantages that participants in programs face, and taking care that the research process itself does not introduce biases. All of this is of central importance in the context of current ECED policies that aim to reduce inequities and disadvantages before kindergarten entry.

Addressing equity in this context includes (although is not limited to) going beyond a consideration of individual race, gender, or ethnic associations that is currently the more common approach in the field. Research needs to further examine intersections among different social hierarchies and identities; explore cumulative impacts, levels, pathways, and social (economic, sociological, political, and cultural) explanations; consider the dynamic nature of inequities; and assess social institutions and processes that influence the allocation of resources and its social determinants.

In research, the concept of equity, together with cultural competence, cultural responsiveness, and intersectionality, can permeate all components and phases of research. An equity lens makes the research process more responsive to the equity goals of early childhood education, takes into account existing disadvantages, and leads to processes that make it easier to engage agents and individuals in long-term equity change. Only by understanding what's working, what is not, and why, with the intention of advancing equity across children and families, can research strongly support the development of policies for all of our children.

References

Aboud, F. E., & Prado, E. L. (2018). Measuring the implementation of early childhood development programs. *Annals of the New York Academy of Sciences, 1419*(1), 249-263.

Allman, K. R., & Slavin, R. E. (2018). Immigration in 2018: What is a teacher educator to do? *The Teacher Educator, 53*(3), 236-243.

American Evaluation Association. (2011). *Public statement on cultural competence in evaluation.* Fairhaven, MA: American Evaluation Association. https://www.eval.org/p/cm/ld/fid=92.

Atkins-Burnett, S., Sprachman, S., Lopez, M., Caspe, M., & Fallin, K. (2011). The Language Interaction Snapshot (LISn): A new observational measure for assessing language interactions in linguistically diverse early childhood programs. In C. Howes, J. T. Downer, & R. C. Pianta (eds.), *Dual language learners in the early childhood classroom* (pp. 117-146). Baltimore, MD: Brookes Publishing.

Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations.* New York: UNICEF Evaluation Office.

Bandel, E., Atkins-Burnett, S., Castro, D. C., Wulsin, C. S., & Putnam, M. (2012). *Examining the use of language and literacy assessments with young dual language learners (final report).* Chapel Hill, NC: University of North Carolina, FPG Child Development Institute, CECER-DLL.

Barnett, W. S., Yaroz, D.J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly, 22*(3), 277-293.

Bastuji-Garin, S., Sbidian, E., Gaudy-Marqueste, C., Ferrat, E., Roujeau, J. C., Richard, M. A., Canoui-Poutrine, F. 2013. Impact of STROBE statement publication on quality of observational study reporting: Interrupted time series versus before-after analysis. *PloS One, 8*(8): e64733.

Bauer, G. R. (2014). Incorporating intersectionality theory into population health research methodology: Challenges and the potential to advance health equity. *Social Science and Medicine, 110*(3), 10-17.

Bécares, L., & Priest, N. (2015). Understanding the influence of race/ethnicity, gender, and class on inequalities in academic and non-academic outcomes among eighth-grade students: Findings from an intersectionality approach. *PloS One, 10*(10), e0141363.

Bergstrom, J. P., Partington, S., Murphy, M. K., Galvao, L., Fayram, E., & Cisler, R. A. (2009). Active consent in urban elementary schools: An examination of demographic differences in consent rates. *Evaluation review, 33*(5), 481-496.

Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study.* New York: MDRC. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594430.

Bowne, J. B., Magnuson, K. A., Schindler, H. S., Duncan, G. J., & Yoshikawa, H. (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis, 39*(3), 407-428.

Bumgarner, E., & Brooks-Gunn, J. (2015). The association between early care arrangements, quality, and emergent bilingual Latino American children's math and literacy skills in English. *Early Childhood Research Quarterly, 30*(1), 32-44.

Burchinal, M. (2018). Measuring early care and education quality. Child Development Perspectives, 12(1), 3-9.

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112*(3), 579-620.

Castro, D., Espinosa, L., & Páez, M. (2011). Defining and measuring quality in early childhood practices that promote dual language learners' development and learning. In M. Zaslow, I. Martinez-Beck, K. Tout., & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 257-280). Baltimore, MD: Brookes Publishing.

Connolly, F., & Olson, L. S. (2012). *Early elementary performance and attendance in Baltimore City schools' pre-kindergarten and kindergarten.* Baltimore, MD: Baltimore Education Research Consortium.

Dean-Coffey, J., Casey, J., & Caldwell, L. D. (2014). Raising the bar – Integrating cultural competence and equity: Equitable evaluation. *The Foundation Review, 6*(2), 81-94.

Dettlaff, A. J., & Fong, R. (2011). Conducting culturally competent evaluations of child welfare programs and practices. *Child welfare, 90*(2), 49-68.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development, 82*(3), 870-886.

Dunn, L. M., & Dunn, L. L. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis: Pearson Assessments.

Dunn, L. M., Dunn, L. L., & Dunn, D. M. (1997). *Peabody picture and vocabulary test* (3rd ed.). Circle Pines, MN: American Guidance Services.

Durán, L. K., Roseth, C. J., & Hoffman, P. (2010). An experimental study comparing English-only and transitional bilingual education on Spanish-speaking preschoolers' early literacy development. *Early Childhood Research Quarterly, 25*(2), 207-217.

Edwards, P., Roberts, I., Clarke, M., DiGuiseppi, C., Pratap, S., Wentz, R., & Kwan, I. (2002). Increasing response rates to postal questionnaires: systematic review. *BM, 324*(7347), 1183.

Epstein, R., Blake, J., & Gonzalez, T. (2017). *Girlhood interrupted: The erasure of black girls' childhood.* Washington, DC: Georgetown Law Center on Poverty and Inequality.

Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in human behavior, 26*(2), 132-139.

Farago, F., Sanders, K., & Gaias, L. (2015). Addressing race and racism in early childhood: Challenges and opportunities. In J. Sutterby (Ed.), *Discussions on Sensitive Issues* (pp. 29-66). Bingley, UK: Emerald Group Publishing Limited.

Figueras-Daniel, A. (2016). *Key influences on the quality and outcomes of preschool education for dual language learners: Professional development and bilingual staffing patterns* (unpublished doctoral dissertation). Rutgers University-Graduate School, New Brunswick.

Flay, B. R., & Collins, L. M. (2005). Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *Annals of the American Academy of Political and Social Science, 599*(1), 115-146.

Ford, C. L., & Harawa, N. T. (2010). A new conceptualization of ethnicity for social epidemiologic and health equity research. *Social Science and Medicine, 71*(2), 251-258.

Friedman-Krauss, A., Barnett, W. S., & Nores, M. (2016). *How much can high-quality universal pre-k reduce achievement gaps?* Washington, DC: Center for American Progress.

Frierson, H. T., Hood, S., & Hughes, G. (2002). Strategies that address culturally responsive evaluation. In J. Frechtling (Ed)., *The 2002 user-friendly handbook for project evaluation* (pp. 63-73). Alexandria, VA: National Science Foundation.

Frierson, H. T., Hood, S., Hughes, G. B., & Thomas, V. G. (2010). A guide to conducting culturally responsive evaluations. In National Science Foundation, *The 2010 user-friendly handbook for project evaluation* (pp. 75-96). Alexandria, VA: National Science Foundation.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?* Research Study Brief. New Haven, CT: Yale University, Yale Child Study Center.

Goldenberg, C. (2012). Research on English learner instruction. In M. Calderón (Ed.), *Breaking through: Effective instruction and assessment for reaching English learners* (pp. 39-61). Bloomington, IN: Solution Tree Press.

Goodson, B. D., Layzer, C., Smith, W. C., & Rimdzius, T. (2004). *Observation measures of language and literacy instruction (OMLIT)*. Cambridge, MA: Abt Associates.

Gormley, W., Gayer, T., & Phillips, D. A. (2008). Preschool programs can boost school readiness. *Science,* June 27, 1723-1724.

Gormley, W., Gayer, T., Phillips, D. A., & Dawson, B. (2005). The effects of universal pre-k on cognitive development. *Developmental Psychology, 41*(6), 872-884.

Greenwood, P. W., Karoly, L. A., Everingham, S. S., Houbé, J., Kilburn, M.R., Rydell, C. P, . . . & Chiesa, J. (2018). Estimating the costs and benefits of early childhood interventions: Nurse home visits and the Perry Preschool. In B. Welsh. D. P. Farrington, & L. W. Sherman (Eds.), *Costs and benefits of preventing crime* (pp. 123-148). New York: Routledge.

Grimes, D. A., & Schulz, K. F. (2002). Bias and causal associations in observational research. *The Lancet, 359*(9302), 248-252.

Halle, T. (2020). How implementation science and improvement science can work together to improve early care and education (this volume).

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures* (2nd ed.) Washington, DC: Child Trends.

Harvard Clinical and Translational Science Center (2010). *Cultural competence in research*. Cambridge, MA: Program for Cultural Research, Harvard Catalyst, Program for Faculty Development and Diversity.

Heckman, J. J. (1990). *Selection bias and self-selection.* In J. Eatwell, M. Milgate, & P. Newman (Eds.), Econometrics (pp. 201-224). London: Palgrave Macmillan.

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review, 1*(1), 1-9.

Hood, S., Hopson, R. K., & Kirkhart, K. E. (2015). Culturally responsive evaluation. In K. Newcomer, H. P. Hatry, & J. S. Wholey (eds.), *Handbook of practical program evaluation* (4th ed.) (pp. 281-317). New York: Wiley.

Kirkhart, K. E. (2005). Through a cultural lens: Reflections on validity and theory in evaluation. In S. Hood , R. Hopson , & H. Frierson (Eds.), *The role of culture and cultural context: A mandate for inclusion, the discovery of truth, and understanding in evaluative theory and practice* (pp. 21-39). Greenwich, 36CT: Information Age Publishing.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation, 31*(3), 400-413.

Kirkhart, K. E. (2013, April). Repositioning validity. Paper presented during a plenary panel, Perspectives on Repositioning Culture in Evaluation and Assessment, at CREA Inaugural Conference, Chicago.

Ladson-Billings, G. (2004). Just what is critical race theory and what's it doing in a nice field like education. In G. Ladson-Billings & D. Gillborn (Eds.), *The RoutledgeFalmer reader in multicultural education* (pp. 49-67). London: RoutledgeFalmer.

Ladson-Billings, G. (2011). Boyz to men? Teaching to restore black boys' childhood. *Race Ethnicity and Education, 14*(1), 7–15.

LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation, 23*(2), 13.

Lee, K. (2007). *The importance of culture in evaluation: A practical guide for evaluators*. Denver: The Colorado Trust.

Li, Z., Nores, M., and Barnett, W. S. (2018, April). Measurement invariance of the Peabody Vocabulary Test. Paper presented at the 2018 American Education Research Association Annual Meeting, New York.

Logan, J. A., Piasta, S. B., Justice, L. M., Schatschneider, C., & Petrill, S. (2011). Children's attendance rates and quality of teacher-child interactions in at-risk preschool classrooms: Contribution to children's expressive language growth. *Child and Youth Care Forum, 40*(6), 457-477.

Magnuson, K. A., Kelchen, R., Duncan, G. J., Schindler, H. S., Shager, H., & Yoshikawa, H. (2016). Do the effects of early childhood education programs differ by gender? A meta-analysis. *Early Childhood Research Quarterly, 36*(3), 521-536.

McCoy, D. C., Yoshikawa, H., Ziol-Guest, K. M., Duncan, G. J., Schindler, H. S., Magnuson, K., . . . Shonkoff, J. P. (2017). Impacts of early childhood education on medium-and long-term educational outcomes. *Educational Researcher, 46*(8), 474-487.

Mitchell, S., Fonseca, M., & LaFave, A. (2016). *Suspension and expulsion in preschool development states: Policies and practices*. State Technical Assistance Report. Preschool Development and Expansion Grant Technical Assistance. Herndon, VA: AEM.

O'Brien B. C., Harris I. B., Beckman T. J., Reed D. A. & Cook D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine, 89*(9): 1245-1251.

O'Brien, R. L., Kosoko-Lasaki, O., Cook, C. T., Kissell, J., Peak, F., & Williams, E. H. (2006). Self-assessment of cultural attitudes and competence of clinical investigators to enhance recruitment and participation of minority populations in research. *Journal of the National Medical Association, 98*(5), 674-682.

Östlin, P., Schrecker, T., Sadana, R., Bonnefoy, J., Gilson, L., Hertzman, C., . . . Muntaner, C. (2011). Priorities for research on equity and health: Towards an equity-focused health research agenda. *PloS medicine, 8*(11), e1001115.

Pacchiano, D., Klein, R., and Hawley, M.S. (2016). *Reimagining instructional leadership and organizational conditions for improvement: Applied research transforming early education*. Chicago: The Ounce of Prevention Fund.

Papadopoulos, I., & Lees, S. (2002). Developing culturally competent researchers. *Journal of advanced nursing, 37*(3), 258-264.

Peisner-Feinberg, E., Buysse, V., Fuligni, A., Burchinal, M., Espinosa, L., Halle, T., & Castro, D. C. (2014). Using early care and education quality measures with dual language learners: A review of the research. *Early Childhood Research Quarterly, 29*(4), 786-803.

Phillips, D., & Meloy, E. (2012). High-quality school-based pre-k can boost early learning for children with special needs. *Exceptional Children, 78*(4), 471-90.

Public Policy Associates. (2015). Considerations for conducting evaluation using a culturally responsive and racial equity lens. Lansing, MI: Public Policy Associates. http://publicpolicy.com/wp-content/uploads/2017/04/PPA-Culturally-Responsive-Lens.pdf.

Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., . . . Ciarico, J. (2010). *Head Start impact study: Final report*. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation.

Robinson, K. A., Dennison, C. R., Wayman, D. M., Pronovost, P. J., & Needham, D. M. (2007). Systematic review identifies number of strategies important for retaining study participants. *Journal of clinical epidemiology, 60*(8), 757-e1.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative research in psychology, 11*(1), 25-41.

Schulz, K. F., Altman, D. G., & Moher, D. 2010. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and Pharmacotherapeutics, 1*(2), 100-107.

Shiu-Thornton, S. (2003). Addressing cultural competency in research: Integrating a community-based participatory research approach. *Alcoholism: Clinical and Experimental Research, 27*(8), 1361-1364.

Slavin, R., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis, 33*(1), 47–58.

Soukakou, E. P., Winton, P. J., West, T. A., Sideris, J. H., & Rucker, L. M. (2014). Measuring the quality of inclusive practices: Findings from the inclusive classroom profile pilot. *Journal of Early Intervention, 36*(3), 223-240.

Spark Policy Institute. (2014). Tools for integrating an equity lens. http://tools.sparkpolicy.com/introduction-taking-an-equity-lens.

Starfield, B. (2007). Pathways of influence on equity in health. *Social Science and Medicine, 64*(7), 1355-1362.

Thomas, V. G., & McKie, B. K. (2006). Collecting and utilizing evaluation research for public good and on behalf of African American children. *The Journal of Negro Education, 75*(3), 341-352.

Tong, A., Sainsbury, P., & J. Craig. 2007. Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal of Quality Health Care, 19*(6): 349–357.

Torgerson, D. J., & Torgerson, C. J. (2003). Avoiding bias in randomised controlled trials in educational research. *British Journal of Educational Studies, 51*(1), 36-45.

Tribal Evaluation Workgroup. 2013. *A roadmap for collaborative and effective evaluation in tribal communities.* Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services.

Weiland, C. (2016). Impacts of the Boston prekindergarten program on the school readiness of young children with special needs. *Developmental Psychology, 52*(11), 1763-1776.

Weiland, C., & Yoshikawa, H. (2013). Impacts of a prekindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84*(6), 2112-2130.

Whitebook, M., & Ryan, S. (2012). Supportive environmental quality underlying adult learning (SEQUAL). Berkeley, CA: Center for the Study of Child Care Employment, University of California.

Whitesell, N. R. (2017). *Evidence and equity: Challenges for research design.* OPRE Report #2017-76. Washington, DC: Administration for Children and Families, Office of Planning, Research and Evaluation.

Willis, R. J., & Rosen, S. (1979). Education and self-selection. *Journal of Political Economy, 87*(5, Pt. 2), S7-S36.

Wilson, S. J., Dickinson, D. K., & Rowe, D. W. (2013). Impact of an early reading first program on the language and literacy achievement of children from diverse language backgrounds. *Early Childhood Research Quarterly, 28*(3), 578-592.

Yates, T. M., & Marcelo, A. K. (2014). Through race-colored glasses: Preschoolers' pretend play and teachers' ratings of preschooler adjustment. *Early Childhood Research Quarterly, 29*(1), 1-11.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T., . . . Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* New York: Foundation for Child Development.

Yousafzai, A. K., Aboud, F. E., Nores, M., & Kaur, R. (2018). Reporting guidelines for implementation research on nurturing care interventions designed to promote early childhood development. *Annals of the New York Academy of Sciences, 1419*(1), 26-37.

Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Daneri, P., Green, K., . . . Martinez-Beck, I. (2016). Quality thresholds, features, and dosage in early care and education: introduction and literature review. *Monographs of the Society for Research in Child Development, 81*(2), 7–26.

Appendix: Self-assessments and considerations for research

The following includes a compilation of reflection or self-assessments drawn from various perspectives on cultural competence, cultural congruence, and cultural responsiveness that are organized by context, perspective, program, design and sampling, procedures and analyses, and dissemination. All of these perspectives inform research in different ways and support reflection about all stages of research.

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Context	How do people from this culture typically greet each other?	Have you learned the history of this community and of the evaluand?	Have you researched the cultural behavior and needs of the language population? For example, accommodations for language?	Have you decided whether cultural competency training is needed?
	Whom should I greet first if I am approaching a group of people?	Have you identified the relevant geographic boundaries and characteristics of this context?	Sought clarity on demographics and other characteristics of the local community?	Do you learn about the socioeconomic status, culture, or other aspects of the priority population and accommodate differences?
	How do people from this culture tend to view someone with authority and power?	Have you identified the strengths of this context?		
	What past experiences has the community had with researchers and evaluators?	Have you paid attention to how power is distributed through formal or informal structures?		
	Who are the typical knowledge holders in this culture?			
	What contextual conditions and structural inequities exist in this context?			

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Perspective	What social identities and groups do I belong to? How might these color the lens through which I view the world? What social identities and groups do people who don't know me think I belong to? Who is knowledgeable enough to help me ensure multicultural validity?	Have you considered? the values espoused by the funders of this evaluand? If the program is built on prior empirical research, have you paid attention to participated in the original body of evidence and how culture was addressed?		
Program	Why is the initiative of the program important? What potential impact, both positive and negative, can the evaluation have on the community and beyond? Do I know what policies, procedures, and practices may affect the program's impact? Do I know what policies, procedures, and practices may affect the staff's performance in the evaluation?	What cultural characteristics are most salient in understanding the consumers of this program? Diverse? Homogenous? What cultural characteristics are most salient in understanding the providers of this program? Diverse? Homogeneous? What are the admission criteria? How does it restrict diversity?	Have you sought clarity on eligibility criteria?	

Continues on next page

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Design, Sample	Who is in my sample and what do I need to know about them? What is the best time for me to collect data from them? Who should collect the data to ensure that participants feel comfortable and safe?		Do you routinely involve the priority population in designing some/all evaluation steps? Do you take race/ethnicity into account in designing survey/instrument(s)? Have you considered demographic or underserved populations?	Do you take race/ethnicity into account when designing an instrument? Do you consider demographic differences between leadership, staff, and children? Community context? Underserved populations?
Procedures	Is the location for the interview/activity easily accessible, familiar, and comfortable for the people with whom I will meet? What am I assuming about each group of stakeholders in the evaluation?		Do you find yourself changing the way you speak, and the words you use based on verbal or nonverbal cues from your recipients? Have you determined criteria for identification of interviewers? Have you decided whether interviewers need cultural competency training?	Do you find yourself changing verbal and nonverbal responses (words and tones) in response to who you interview? Do you understand the need to adapt and be flexible in your process to the needs of individuals? Have you determined criteria for interviewers?

Concept	Lee, 2007 Cross-Cultural Competence	Kirkhart, 2010 Cultural Congruence	Public Policy Associates, 2015 Cultural Responsiveness	O'Brien et al., 2006 Cultural Competence
Analyses, Dissemination	Can the average person not steeped in evaluation terminology understand me?		Have you checked for outcomes and differences, intended and unintended?	Do you disaggregate data along demographic lines to understand programs along race, culture, socioeconomic status, and language lines?
	How will the findings be used by the community members, politicians, policymakers, journalists, and special interest groups?		Have you determined who or what is changed/affected?	Do you analyze and interpret outcomes, differences, and intersections?
	Will the findings place a stigma on a certain group or give the group power to access resources and improve their situations?		Have you observed any unintended changes or consequences due to cultural/racial/ethnic considerations?	Do you think about how you can use the type of feedback you receive based on racial, ethnic, or other characteristics of individuals who participate in the system to engage them in long-term equitable change?
	What are the self-serving purposes of the research for the sponsor and the evaluator?		Do you ensure that the program is accessible to the target population?	
			Do you make recommendations that focus on equity?	
			Do you make use of disaggregated data along demographic lines in order to adapt your evaluation processes to the race/culture of recipients?	