MAXIMIZING PILOT PHASE MEASURES TO INFORM QUALITY

IMPROVEMENT: USING A SEQUENTIAL MIXED-METHODS DESIGN WITH

INTERRUPTED TIME SERIES TO EXAMINE FEASIBILITY, UPTAKE, AND

DRIVERS OF AN EVIDENCE-BASED PRACTICE IN PART C/EARLY

INTERVENTION SYSTEMS

by

SONDRA MARIE STEGENGA

A DISSERTATION

Presented to the Department of Special Education and Clinical Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2019

DISSERTATION APPROVAL PAGE

Student: Sondra Marie Stegenga

Title: Maximizing Pilot Phase Measures to Inform Quality Improvement: Using a Sequential Mixed-Methods Design With Interrupted Time Series to Examine Feasibility, Uptake, and Drivers of an Evidence-Based Practice in Part C/Early Intervention Systems

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Special Education and Clinical Sciences by:

Jane Squires	Chairperson
John Seeley	Core Member
Daniel Anders	Core Member
Kent McIntosh	Core Member
Philip Fisher	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019.

DISSERTATION ABSTRACT

Sondra Marie Stegenga

Doctor of Philosophy

Department of Special Education and Clinical Sciences

June 2019

Title: Maximizing Pilot Phase Measures to Inform Quality Improvement: Using a Sequential Mixed-Methods Design With Interrupted Time Series to Examine Feasibility, Uptake, and Drivers of an Evidence-Based Practice in Part C/Early Intervention Systems

Reviews of published research indicate only a 33% median success rate of complex systems change initiatives. This lack of uptake is becoming known as implementation failure. Social emotional interventions with infants and young children and their caregivers also have a history of high attrition rates, putting efforts focused on this area at high risk of implementation failure. This failure rate is especially pertinent because 28 of 55 states and jurisdictions providing Part C early intervention services are focusing improvement efforts on social emotional development as part of the State Systemic Improvement Planning processes. Although policy-mandated quality improvement is occurring in local, state, and federal systems, published research is lacking on quality improvement measures and implementation-focused outcomes within this population. The pilot phase of implementation holds immense potential for identifying quality improvement needs before scale-up, ultimately resulting in large potential savings of money and time by avoiding implementation failure. However, focused objective measures are often missing or underutilized within these cycles. The goal of this study was to examine factors with potential impact on implementation and uptake of an evidence-based practice (EBP) in a quality improvement project. The study

focus was on the feasibility, acceptability, and appropriateness of a policy-related EBP implemented in a Part C early intervention system. Potential determinants (barriers and supports), uptake rates, and driver influence relative to the uptake were analyzed. The efficacy of these elements for improving measurement accuracy and for understanding quality improvement needs in the pilot phase of initial implementation was evaluated. Two specific research questions were addressed: "What are the feasibility, acceptability, and appropriateness of the use of an EBP (i.e., electronic social emotional assessment) used in Part C/early intervention pilot sites?" and "What is the relationship between the identified key drivers and uptake rates for the use of electronic social emotional assessments?" Key drivers identified were initial training, leadership coaching, and implementation completion rates.

CURRICULUM VITAE

NAME OF AUTHOR: Sondra Marie Stegenga

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Grand Valley State University, Allendale, Michigan

DEGREES AWARDED:

Doctor of Philosophy, Special Education, 2019, University of Oregon
Master of Education, Leadership, 2015, Grand Valley State University
Master of Science, Occupational Therapy, 2002, Grand Valley State University
Bachelor of Science, Behavioral Sciences, 2000, Grand Valley State University

AREAS OF SPECIAL INTEREST:

Early social emotional assessment and intervention
Interdisciplinary and interagency collaboration
Implementation science and policy
Data science and methods

PROFESSIONAL EXPERIENCE:

Research Assistant, University of Oregon Prevention Science Program 2018–
   present

Consultant, Raviant Consulting, Eugene, Oregon, 2018–present

Supervisor, University of Oregon Special Education Program, 2015–2018

Coordinator, Ottawa Area Intermediate School District Infant Development
   Program, Zeeland, Michigan, 2011–2015

Occupational Therapist, Ottawa Area Intermediate School District Infant
   Development Program, Allendale, Michigan, 2005–2011

Occupational Therapist, National Heritage Academies Schools, Grand Rapids,
   Michigan, 2003–2005

Occupational Therapist, Mary Free Bed Rehabilitation Hospital, Grand Rapids,
   Michigan, 2002–2007

GRANTS, AWARDS, AND HONORS:

Bricker Dissertation Award, 2019

Louis and Sarah G. Kaplan Scholarship in Educational Leadership, 2018

Travel grant, Cascadia R—Conference on Statistical Programming in R, 2018

Travel grant, American Educational Researchers Association, 2018

Kaplan Scholarship in Early Intervention, 2017, 2018

Florence Wolford Scholarship in Education for promising educational leaders, 2017

Hill M. Walker Scholarship, 2016

University of Oregon Alumni Fund Scholarship, 2016

Honorable Mention, Dean's Award for Outstanding Master's Thesis, Grand Valley State University, 2015

Distinguished Volunteer, Division for Early Childhood, Council for Exceptional Children, 2015

Distinguished Athlete Award, U.S. Marines, 1995

PUBLICATIONS:

Green, A., Maynard, D., & Stegenga, S. (2018). Common misconceptions of suspension: ideas and alternatives for school leaders. *Psychology in the Schools, 55*(4), 419–428.

Kinavey Wennerstrom, E., Stegenga, S., & Lund, E. (2018). Coping with secondary traumatic stress: Self-care strategies for early childhood professionals. In C. Corr and D. Miller (Eds.), *Maltreatment and toxic stress* (Young Exceptional Children Monograph 17) (pp. 32–38). Los Angeles, CA: Division for Early Childhood of the Council for Exceptional Children.

Stegenga, S., Berry, M., & Fisher, P. (2018). Using video coaching to support parents and promote serve and return interactions. In J. McCollum, R. M. Santos, and J. Weglarz-Ward (Eds.), *Enhancing children's access to responsive interactions* (DEC Recommended Practices Monograph Series 5) (pp. 17–26). Los Angeles, CA: Division for Early Childhood of the Council for Exceptional Children.

Gothberg, J. G., Stegenga, S. M., & Cate, D. (2017). Rethinking the paradigm: Comprehensive transition from birth to adulthood. In A. L. Ellis (Ed.), *Transitioning children with disabilities: from early childhood through adulthood.* Boston, MA: Sense Publishers.

ACKNOWLEDGMENTS

It is with immense gratitude that I reflect upon my journey in the doctoral program and the writing of this dissertation. First, I would like to thank all my state agency partners who provided innovation and collaboration throughout this process with special thanks to Laura and Kelly for all your time, ideas, and amazing work you are doing to support children and families and improve systems. I also would like to thank all the innovative leaders and dedicated providers serving infants and toddlers with disabilities in Part C early intervention systems that participated in my study. In addition, I would like to express my gratitude for my Bricker Dissertation Award. I am very grateful for the funds to conduct my research but even more so, the time and mentoring of Dr. Bricker. Next, I would like to express sincere gratitude to my dissertation committee. Without each one of you this would not have been possible. Lisa, thank you for all your hours mentoring me on implementation and measurement. Phil, thank you for your innovation and time problem solving needs and systems supports to increase capacity in systems and improve outcomes. Kent, thank you for the opportunities to collaborate and learn from you as part of your development teams impacting early childhood practices and systems and your teaching on single case research design. Daniel, thank you for all your time mentoring and teaching me in statistics and programming in R. Without all your time and help walking through my messy syntax over the years I never would have made it—and now I love coding!! John, thank you for all your teaching, mentoring, and collaborating on implementation science and research designs in this complex area. Your time and knowledge in this area have been invaluable and have impacted my frames of reference. And to Jane, my advisor, I am forever grateful for all your teaching, mentoring, collaboration, and innovation you have shared with me related to

early intervention, curriculum, screening, and supports. Thank you for all your time not only with my dissertation but my growth as a researcher and person. I literally would not be here without you and could not be more grateful! I also would like to thank Erin Kinavey Wennerstrom and Sloan Storie for your expertise on early intervention systems and time in the qualitative coding process; you rock! I also would like to thank June Gothberg for your time and mentoring throughout my doctoral journey and your expertise on mixed methods. Your mentoring and friendship have been valuable beyond what I can express. I also would like to thank my cohort, Christine, Kelly, and Buket for all your laughter and friendship throughout the coursework that led me to this dissertation. I could not imagine this journey without each one of you. Last, but definitely not least, an immense thank you to my husband, parents, siblings, and close friends who have loved me and believed in me even when I did not always believe in myself. I cannot imagine where I would be in life if I did not have you on this journey with me. Thank you and with so much love and gratitude to you all!

Dedicated to all the children, families, providers, and outstanding leaders in IDEA Part C
early intervention systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

With research reviews indicating a mere 33% median success rate of complex

change initiatives (Smith, 2002) and less than 50% of evidence-based practices ever

making it to full scale use (Balas & Boren, 2000), major funding organizations such as

the Gates Foundation, Institute for Educational Sciences (IES), and National Institutes of

Health (NIH) are becoming increasingly interested in understanding the factors that

contribute to or can prevent this lack of uptake due to the immense waste of money and

time associated with implementation failure (e.g. Bryk, Gomez, Grunow, & LeMahieu,

2015; Jago & Sebire, 2012). Large amounts of money and time are spent on policy-

mandated implementation and scale-up efforts in education and early intervention

systems. These efforts are often underfunded, making it especially imperative to examine

key leverage points within the implementation process that hold potential to improve the

success of scale-up and full implementation.

One such leverage point is the small-scale pilot phase during initial

implementation (Bryk et al., 2015; Jago & Sebire, 2012). Identifying and remedying

potential barriers to full-scale implementation through integration of quality improvement

changes at this point when costs are minimized offers the opportunity to reduce the

potential for implementation failure. However, to make these critical changes requires an

understanding of what determinants (supports and barriers) and drivers are ultimately

contributing to or preventing uptake. In a recent systematic review of quality

improvement research using Plan–Do–Study–Act (PDSA) cycles, the authors identified

many pilot-phase implementations that make minimal to zero quality improvements and

1

often lack use of intentional, objective, and/or validated measures (Taylor et al., 2014). Many measurement challenges occur within the pilot phase of initial implementation due to the small and underpowered nature of these pilot programs and the lack of consistent data sources or identified focus. Leon, Davis, and Kraemer (2011) clearly stated that the focus of pilot studies should be feasibility and quality improvement. Additionally, an overarching publication bias against pilot studies exists due to a lack of understanding of the potential value and the purpose and scope of these studies (Jago & Sebire, 2012). This lack of publication has likely led to both decreased research available on measures and process within this critical phase and likely replication of comparable pilot studies "which ultimately increases the cost of research which is of course the opposite of the funding agencies intentions when calling for pilot feasibility studies" (Jago & Sebire, 2012, p. 548). Although gains are being made in this area with the emergence of some studies on the quality improvement within the pilot phase of initial implementation in early childhood systems (e.g., Metz, Bartley, Ball, Wilson, Naoom, & Redmond, 2015; Schindler, Fisher, & Shonkoff, 2017), there still is an overall dearth of research in this area, with no known published research focused on Part C early intervention systems serving infants and toddlers with disabilities at the time of this study. In addition, there are no known studies harnessing the use of an interrupted time series (ITS) design within this population. Overall, this points to a large gap in the research on the pilot phase of initial implementation in spite of the immense potential value and leverage for making improvements prior to going to full scale (Leon et al., 2011). Therefore, it is critical to begin examining the pilot phase of initial implementation to intentionally measure and

identify what contributes to feasibility and uptake and begin leveraging this high-value point in the implementation process.

To work toward advancing the methodology and understanding of uptake of interventions within the pilot phase of initial implementation in early intervention systems, the present study was focused on two key issues. First, I examined the feasibility, acceptability, and appropriateness of a policy-related implementation of an evidence-based practice (EBP) within a state system. Feasibility, acceptability, and appropriateness are commonly accepted reasons for pilot studies and one of the most commonly published areas of pilot study research (Aarons, Ehrhart, Farahnak, & Hurlburt, 2015). However, feasibility and acceptability tend to rely heavily on self-report measures, which often are not validated (Weiner, Lewis, Stanick, Powell, Dorsey, Clary, Boynton, & Halko, 2017). Therefore, building upon this, I utilized a validated measure for feasibility, acceptability, and appropriateness. In addition, I examined two objective measures related to feasibility and acceptability: (a) uptake rates of the EBP related to key drivers identified by the state agency (initial training and follow-up coaching), and (b) stages of implementation completion of the site. I then gathered lived experiences of providers and program leaders related to implementing the EBP within the complex environments of early intervention systems.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter provides an overview of published literature on three main areas

relevant to implementation and scale-up of EBPs and interventions in Part C/early

intervention settings: (a) history of and reasoning behind implementation science,

including the importance of early phases of implementation related to uptake and

implementation studies within early intervention and early childhood special education

systems; (b) pilot phase research in quality improvement cycles; and (c) measurement

issues and innovation in research design in implementation and quality improvement

within complex social settings such as Part C/early intervention.

Implementation Science

EBPs have the potential to improve the lives of countless individuals (Cook &

Odom, 2013). Yet, less than 50% of evidence-based interventions make it to full

implementation (Balas, & Boren, 2000), thereby missing the opportunity to improve

services and impact lives. Understanding what contributes to successful implementation,

scale-up, and sustainability is imperative to ensuring EBPs actually reach intended

populations. This measured study of factors that contribute to actual use and uptake of

EBPs in real-world systems is the focus of implementation science (Bauer, Damschroder,

Hagedorn, Smith, & Kilbourne, 2015). Studies in implementation science focus on a

range of outcomes, including feasibility and acceptability of the intervention, potential

determinants (supports and barriers) to implementation and uptake, and the use of

specified implementation strategies (Powell et al., 2015), also known as implementation

drivers (Metz, 2012; National Implementation Research Network, Active Implementation

Hub [NIRN], 2018). When used appropriately, implementation strategies can facilitate the successful uptake of EBPs and/or innovations (Proctor, Powell, McMillen, 2013). In recent years, national funding organizations in the United States have begun emphasizing the critical importance of understanding how implementation strategies work (Powell et al., 2015). Recent studies in implementation science have been focused on the measured effect of particular implementation strategies (Smith & Polaha, 2017), ultimately beginning to drill down to what strategies work, for whom they work, and under what conditions. In a recent analysis of published implementation strategies, 73 discrete commonly used effective implementation strategies were found, including clinical supervision; recruiting designating, and training leaders; providing ongoing consultation; mandating change; developing materials; and a combination of such strategies (Powell et al., 2015). The importance of clearly naming, defining, and specifying implementation strategies to accurately and optimally measure and match strategies to barriers and outcomes has been identified in recent research (Proctor et al., 2013).

Although great gains have been made in the last 20 years related to understanding the science of implementation, there remain vast gaps in the research related to implementation science and Part C early intervention systems. Specifically, much of the implementation research has occurred within the field of health, although leading researchers in education, including early intervention and early childhood special education, have recently called attention to the need to address these gaps (e.g., Cook & Odom, 2013; Dunst, Trivette, & Raab, 2013; Halle, Metz, & Martinez-Beck, 2013; Hemmeter & Conroy, 2018; Metz et al., 2015; Odom, 2009).

In addition to a lack of specification of implementation strategies and use of objective measures, there is a lack of focus on the critical timing of when optimization and measured changes should occur within the stages of implementation. With such few practices ever making it past readiness or initial implementation (Balas & Boren, 2000), it becomes critical to identify the determinants and appropriately matched implementation strategies prior to scale-up to optimize outcomes (Powell et al., 2017). Although a few studies have recently emerged in the published literature examining the pilot phase of initial implementation (Metz et al., 2015; Schindler et al., 2017), there remains an overall dearth of literature in this area. Specifically, there are no known studies examining implementation strategies (i.e., drivers) and determinants within a policy-related quality improvement effort in the pilot phase of initial implementation in Part C early intervention.

Pilot Phase Research in Quality Improvement

Although pilot studies have historically faced challenges in publication due to their lack of power and transferability of findings (Leon, Davis, Kraemer, 2011), this bias has begun to diminish because of the potential impact of quality improvement within this phase (Bryk et al., 2015). In recent years, the implementation literature has begun to identify the importance of pilots related to outcomes such as feasibility, acceptability, and uptake (Leon, Davis, Kraemer, 2011). The need for small-scale trials prior to large-scale implementation has been emphasized in recent work examining implementation failure. Bryk et al. (2015) recommended "learning by doing" prior to attempting any large-scale implementation to learn valuable information, avoid potential implementation failure, and ultimately save immense cost and time. Ultimately, pilot projects hold the potential to be

highly impactful as part of the feasibility study to identify the effectiveness of potential determinants and drivers of implementation that could either lead to failure or facilitate full uptake prior to scaling up. However, these pilot projects must take an intentionally chosen, clearly defined, and measured approach to examining these factors (Proctor et al., 2013). This lack of measured approaches to quality improvement in small-scale trials was recently identified in a review of the literature on PDSA cycles (Taylor et al., 2014). However, the small number of participants within pilot studies present measurement challenges that necessitate innovation to improve understanding of determinants and drivers of uptake.

Measurement Challenges and Opportunities for Innovation

Measurement issues and considerations in Part C early intervention programs have been a focus of recent scholarly discussion. Some research has questioned the validity and reliability of current outcome measures utilized in this population of children (Rosenberg, Elbaum, Rosenberg, Kellar-Guenther, & McManus, 2017), which has brought attention to the need for researchers to begin to explore new methods for examining change. Attention to methods for examining change related to rapid cycle quality improvement efforts is especially needed because the aforementioned measures, even if reliable and valid, are summative (as intended by design) and cannot readily capture the rapid cycle changes within initial implementation and quality improvement efforts in early intervention systems. Thus, researchers need new ways to measure change within these quality improvement efforts. These issues are compounded by the overarching difficulty with measurement in real-world systems and implementation research as a whole (e.g., Fixsen, Panzano, Naoom, & Blasé, 2008; Ogden et al., 2012).

In spite of the importance of the pilot phase of implementation for avoiding implementation failure and subsequent associated time and money wastes, recent reviews of the literature identify an overall lack of consistency or precision with measures during PDSA quality improvement cycles (Taylor et al., 2014). There is also a lack of feasibility proof-of-concept studies (Pai et al., 2017), particularly within pilot phases. This lack of precision in measurement of real-world implementation and outcomes in early childhood systems was recently highlighted by major professional organizations (i.e., Fisher, 2018). Advances in technology are bringing a new world of big data, which hold immense potential for new and more sources of data for measurement and process improvements in early intervention systems (Stegenga, Munger, Squires, & Anderson, 2018). These technology advances combined with advances in the field of implementation methods and research make this an optimal time to examine both the changes within quality improvement and feasibility studies and the measures used. Explorations should include opportunities in research design, methods, and data types and uses that are currently under- or unrepresented in the published literature in Part C of early intervention programs, such as triangulation of data sources, big data, and ITS.

Triangulation of Data Sources and Methods

Triangulation is well known as a method for increasing credibility of findings in qualitative research and theory development (e.g., Carter, Bryant-Lukosius, DiCenso, Blythe, & Neville, 2014) and for improving the validity of findings in program evaluation using mixed methods (e.g., Creswell & Plano Clark, 2017; Greene & McClintock, 1985; Saldana, 2014). In addition, implementation science literature includes a number of recent calls for use of mixed methods, particularly when it comes to understanding the

implementation of EBPs (Aarons, Hurlburt, & Horwitz, 2011; Landsverk, Brown, Chamberlain, Palinkas, & Horwitz, 2012; Palinkas et al., 2011; Proctor et al., 2009; Saldana, 2014). Yet, there are no known published studies employing triangulation of methods and data sources in program evaluation of Part C early intervention systems. Across all program evaluation studies citing triangulation of measures, no known published study has used a measure to identify whether the use of a triangulated set of modern program quality improvement measures is feasible, acceptable, and appropriate for use. This validation is important because as with interventions, when the measures used are not feasible, acceptable, and appropriate for the targeted systems these systems will not be fully implemented or utilized. Therefore, it is important to examine the perceived utility of this approach along with its potential for improved understanding of needed quality improvements and for strengthening validity in the limited-duration pilot trials.

Big Data

The amount of data being generated from a variety of sources has increased immensely, with the majority of data having been gathered in only the last four years (McAfee and Brynjolfsson, 2012). This explosion in information has led to what is known as *big data*. Although definitions vary, the overarching consensus is that the term big data is "used to describe situations where the magnitude and complexity of the data demonstrate such an immense scale that it requires special considerations for storage and analysis due to any one or combination of the 3V's": volume, velocity, and variety (Stegenga et al., 2018). Big data encompasses both found data (Connelly, Playford, Gayle, & Dibben, 2016) and data created as a result of large research and administrative

data systems (Baro, Degoul, Beuscart, & Chazard, 2015). This rapid advancement is ultimately due to the ability to cost effectively gather and store large amounts of data. Big data holds potential for use and advancement of measures in a number of areas, but one particular benefit may be its use in small-scale trials. Typically, as for pilot studies, low numbers create power issues. However, as big data collection becomes more common in early intervention systems and by harnessing the large numbers of time point data gathered organically as part of program processes, the potential for use of new analyses emerges, such as ITS or regression discontinuity designs. However, in spite of this potential for impact in early childhood systems (Mead, & Mitchel, 2016; Paschall & Epstein, 2018), the use of big data in Part C early intervention research remains largely underrepresented in the published literature (Stegenga et al., 2018).

ITS in Education and Early Intervention

ITS models have gained attention for their potential suitability both within the health fields (Bernal, Cummins, & Gasparrini, 2017) and more recently within educational research (Hallberg, Williams, Swanlund, & Eno, 2018). Because of the complexities of real-world research in many health and educational settings, traditional randomized controlled trials often are not feasible. ITS design offers many strengths as a quasi-experimental design in these instances. ITS is one of the stronger quasi-experimental designs (Cook, Campbell, Shadish, 2002), and it allows for the examination and use of aggregated data, which are often the most accessible type of data to examine within state and local systems (Hallberg et al., 2018). Routine data are most appropriate for ITS studies (Bernal et al., 2017), again pointing toward high potential for use in state and local systems. Because of the visual layout of the time series models, ITS studies also

can facilitate communication of results to consumers and community partners, which is key for any policy-related change or community partnerships in state and local systems. However, no published studies employing this design have been found within early intervention settings serving infants and toddlers with disabilities, one of our youngest and most vulnerable educational populations. The present study is the first known application of ITS design within the complex special population of interdisciplinary home visiting services for infants and toddlers with disabilities and their families.

Theoretical Frameworks

Frameworks can help guide the selection of implementation strategies and process and are an important part of designing an implementation-focused research study (Powell et al., 2015). Yet, "researchers often fail to explicitly refer to guiding conceptual frameworks" (Powell et al., 2015, p. 178). Therefore, two theoretical frameworks were employed in this study: (a) the Active Implementation Frameworks (AIF) (Metz et al., 2015; NIRN, 2018) and (b) the IDEAS Impact Framework developed recently through the Frontiers of Innovation at the Harvard Center on the Developing Child (Center on the Developing Child, Harvard University, 2018). These frameworks were chosen for several reasons. The AIF was chosen because of its match to the research questions and because it is the underlying framework for the current state systemic improvement planning (SSIP) process work (Early Childhood Technical Assistance Center, 2018), from which this research stems. AIF also has been used as the foundation for other recent early childhood scale-up work examining initial implementation because of its many practical and fitting resources (Metz et al., 2015). The IDEAS Impact Framework also drives research questions and theory. Specifically, the IDEAS Impact Framework was recently

developed to begin addressing some of the measurement and implementation challenges

within early childhood systems, particularly within the initial implementation quality

improvement pilot phases (Schindler et al., 2017). Therefore, its ideals and foundational

processes were used to guide the specific, measurable, and innovative measures and

methods within the pilot phase of initial implementation. The IDEAS Impact Framework

also is used as a lens for discussion related to future research within early intervention

and childhood systems.

Theory of Change

In addition to the research questions and hypotheses being guided by conceptual

frameworks, this study is also anchored in a theory of change. This theory of change was

developed in accordance with the premises of the IDEAS Impact Framework, meaning

that its elements are both measurable and collaboratively developed with community

partners to ensure greatest impact on quality improvement efforts in the rapid cycle

iteration process (Schindler et al., 2017). This intensive engagement of community

partners in the research and quality improvement process has been termed community-

based participatory research. This type of participatory research aims to bridge the

research practice gap through true community engagement and partnership at all stages of

the research process (Wallerstein & Duran, 2006). Therefore, given these principles, the

theory of change was initially developed by the SSIP team and then further refined

through the collaborative, ongoing monthly phone meetings between this researcher and

key stakeholders on the state leadership team. Stakeholders identified the desire to

improve child social emotional outcomes as per the State Identified Measurable Result

as part of the SSIP process. The theory of change focused on improving child social

emotional outcomes is guided by the following assumptions, which were established

through the iterative process: (a) use of evidence-based assessment and/or screening

enables practitioners to identify specific and measurable areas of need (Squires, Bricker,

Heo, & Twombly, 2001); (b) early screening, assessment, and identification of young

children with social emotional needs is foundational for providing individualized,

targeted, and effective interventions and supports (Squires et al., 2001); (c) fidelitous use

of evidence-based interventions, when based on accurate assessment and matched to the

identified need, leads to improved outcomes (Cook & Odom, 2013); (d) early screening

and assessment is considered an EBP worthy of measurement and attention to

implementation and uptake due to its link to improved outcomes; and (e) the study of

implementation factors such as drivers and determinants related to feasibility is especially

important when it comes to interventions that are known for having high attrition rates

and thus are at increased risk of implementation failure, such as early social emotional

interventions (Baggett et al., 2010). A model for the overarching theory of change for

improving child social emotional outcomes is shown in Figure 1.
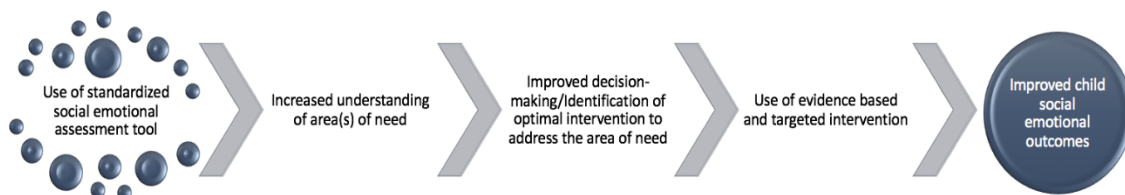
Figure 1. Theory of change for improved child outcomes in social emotional
development.

Overall, given the needs and gaps identified in this review of the literature, the

theory of change, and the results of the community-based participatory research process

identifying a need to understand and measure Step 1 of the theory of change, that is, the

use of a standardized social emotional assessment tool, this study had three specific aims:

(a) understand feasibility, acceptability, and appropriateness of the pilot phase initial

implementation of an EBP within early intervention/Part C systems; (b) examine usage

rates of the EBP relative to key drivers and determinants (supports and barriers) as part of

the theory-building process for quality improvement; and (c) understand the lived

experiences of providers and administrators relative to the use of the social emotional

assessment, including key drivers, within the complex home visiting systems of Part C

early intervention.

Research Questions

To address these aims, the following research questions were proposed.

Feasibility/Acceptability/Appropriateness

1. What are the feasibility, acceptability, and appropriateness of the use of an EBP

   (i.e., the electronic version of the Devereux Early Childhood Assessment [e-

   DECA]) in Part C/early intervention pilot sites?

   a. What are the feasibility, acceptability, and appropriateness ratings for use

      of an electronic early social emotional assessment (the e-DECA) within

      Part C/early intervention pilot service areas?

   b. What are the potential determinants (supports and barriers) related to

      feasibility, acceptability, and appropriateness of use of an electronic early

      social emotional assessment within the pilot service areas?

Impact: Driver Influence and Uptake

2. What is the relationship between the identified key drivers and uptake rates of

   electronic social emotional assessment use? Key drivers identified were (a) initial

   training of full staff, including providers and administrators; (b) follow-up

leadership coaching regarding the social emotional assessment use with focus on systems-level supports; and (c) Stages of Implementation Completion (SIC) scores.

    a. Training and Coaching

        i. What are the usage rate slope trends prior to the training and coaching interventions (e.g., slight positive, slight negative, consistent, or tapering off, which indicates lack of sustainment of the usage)?

        ii. On average, does level and/or trend of the EBP usage increase compared with baseline after (a) initial training and (b) leadership coaching on the EBP use?

        iii. What are the potential determinants (supports and barriers) related to training and leadership coaching on electronic social emotional assessment usage rates of the EBP (i.e., what else do leaders need in this process)?

    b. Implementation Completion

        i. What are the levels of implementation completion for each of the pilot service areas?

        ii. What are the potential determinants (supports and barriers) of service areas completing more stages of implementation?

Hypotheses

Two main hypotheses are related to these research questions. First, electronic social emotional assessment use is moderate to highly feasible, acceptable, and

appropriate on average. Second, a positive relationship exists between key drivers and

uptake rates of the electronic social emotional assessment. These hypotheses were

developed in collaboration with the state partners as part of the research practice

partnership guiding this work.

CHAPTER III

METHODS

A combination of measures and analytic procedures were used to answer the

research questions. Context and setting of the research, research participants, sequence of

systems change efforts, and related timelines assisted in developing these measures and

procedures. Related assurances such as data sharing agreements and data

storage/confidentiality are also described. The research plan was submitted to the

University of Oregon Internal Review Board (IRB)/research compliance services and

determined exempt (protocol 12962919.008).

Context and Setting

The study focuses on the implementation and scale-up of an EBP (the e-DECA)

in service areas providing Individuals with Disabilities Education Act (IDEA) Part C

early intervention services and supports to infants and toddlers with disabilities and their

families in a Midwestern state. The pilot implementation described and examined herein

are part of a federally mandated quality improvement effort occurring in Part C systems

throughout the United States known as the SSIP process (Department of Education,

Office of Special Education Programs, GRADS360° [GRADS360], 2018). In its SSIP

process, this state had targeted systematic electronically based social emotional

assessment as a major goal for its systems improvement related to improving social

emotional outcomes for infants and toddlers with disabilities (see Figure 1). Because of

the critical importance of social emotional screening and assessment (Squires et al.,

2001), a universal approach to screening and assessment was planned. Contrary to

previous efforts using pencil-and-paper tests, the state chose an electronic social

17

emotional assessment measure and targeted its use by IDEA Part C early intervention providers in the state. The pilot, involving four IDEA Part C service areas, is the first step toward scale-up and use of the assessment. Once pilot service areas were identified and formally agreed to pilot participation, the state collaborated with the service areas to schedule formal training, which was conducted by a trainer with expertise in infant mental health who was certified for the assessment. Service areas had access to the electronic system for assessment 4 to 8 months prior to the training dates based on when the state assessment license was purchased. Trainings for the sites occurred in November 2016, January 2017, February 2017, and March 2017.

Participants

Participants were two groups of professionals in the Part C/early intervention programs serving infants and toddlers with disabilities ages birth to 3 years. These two groups were composed of (a) interdisciplinary early intervention providers on implementation teams in the Part C early intervention programs including but not limited to occupational therapists, physical therapists, speech language pathologists, teachers, social workers, and early intervention developmental specialists and (b) local Part C program coordinators of the pilot and beta sites. The results for 77 survey participants were included in the final data analysis after cleaning and screening (see "Analytic Plan" for further information on inclusion criteria). All survey participants had the option to receive a $10 electronic gift card to either Starbucks or Amazon for participation by self-selecting the gift card option at the end of the survey. Informed consent was embedded into the beginning of the survey as per the IRB protocol. At the end of the survey, participants had the option to volunteer for possible participation in a follow up

interview. From these survey participants, 10 interview participants (four administrators and six providers) were selected to ensure representation from all service areas after these participants had self-identified as interested in participating in interviews. Purposeful random sampling of participants in implementation-focused qualitative and/or mixed-methods research is a type of sampling useful for increasing the credibility of findings, such as in this study, given the research questions and focus (Palinkas et al., 2015). All interview participants completed consent forms prior to participating in the interviews. All four pilot service areas were represented in both the quantitative and qualitative data. The 10 interviews with the four administrators and six providers were conducted via phone in winter 2019. All interview participants received a $25 electronic gift card to either Starbucks or Amazon according to the study protocol.

Measures

The following measures were chosen because of their applicability to the research questions and participant groups and their psychometric properties, potential for prediction, and/or ability to provide highly relevant descriptive information for needed theory development and understanding. Clinton-McHarg et al. (2016) discussed the importance of utilizing measures that are valid, relevant, and feasible within implementation and quality improvement research to ensure that change can be detected and accurately understood. The measures are categorized according to a recent framework for measures and metrics in feasibility and proof of concept studies (Pai et al., 2017).

Demographic Information

Demographic information was gathered to provide a foundational understanding of child characteristics, provider characteristics, and organization/service area characteristics being examined in the study. Aggregated demographic data of children served in pilot sites were obtained from the publicly available state early intervention website. Demographics reported included age distributions, percentage served in early intervention and special education, race/ethnicity, and sex. In addition, demographic data for providers in the pilot sites were obtained as part of the data collection process on feasibility and included sex, race, and number of years of experience in early intervention.

Feasibility, Acceptability, and Appropriateness Measures

Feasibility, acceptability, and appropriateness are regularly cited implementation outcomes (Peters, Adam, Alonge, Agyepong, & Tran, 2013), which fit well for pilot phase initial implementation studies (Leon et al., 2011). The Feasibility of Implementation Measure (FIM), Acceptability of Implementation Measure (AIM), and Appropriateness of Implementation Measure (IAM) by Weiner et al. (2017) were chosen for their psychometric property strengths, fit to research questions, succinct approach, and free availability for use. Use of these measures follows the recommendations of Pai et al. (2017) in their systematic review of feasibility measures in which they identified the importance of utilizing measures of feasibility and acceptability that clearly define constructs. According to Weiner et al. (2017), feasibility is defined as "the extent to which a new treatment, or an innovation, can be successfully used or carried out within an agency or setting" (p. 2). Acceptability is defined as "the perception among

implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory" (p.2). Appropriateness is defined as "the perceived fit, relevance, or compatibility of the innovation or evidence-based practice for a given practice setting, provider or consumer" (p. 2). The FIM, AIM, and IAM combined constituted a 12-question self-report measure. Follow-up semistructured interviews with questions developed through a phenomenological lens according to process by Bevan (2014) also were used to gather lived experiences and potential determinants related to the use of the EBP. The unit of analysis of feasibility, acceptability, and appropriateness was individual/provider, and these data were aggregated to examine mean levels across all pilot sites. Survey respondents received a $10 electronic gift card of their choice to either Starbucks or Amazon.

Impact Measures: Uptake and Driver Influence

Impact is defined as the total change in desired outcomes as attributable to a particular program "using the best methodology available" (Pai et al., 2017). This study used a variety of measures to examine uptake as well as the potential influence of drivers on uptake. The unit of analysis was Part C service areas in the target state. As mentioned in the literature review, it is imperative that implementation drivers be clearly defined to optimize measurement and replication (Proctor et al., 2013).

Uptake. Uptake was measured through use of existing electronic data on frequency of electronic social emotional assessment use through detection of new cases of assessment use after initial training and after follow-up coaching. Detection of new cases is listed in the newly developed framework by Pai et al. (2017) as a suggested impact measure.

Driver influence on uptake. Specific drivers that were measured related to potential influences on uptake were (a) initial training including site-specific coaching on the use of the electronic social emotional assessment measured through short comparative ITS (CITS), (b) leadership coaching measured through short CITS, and (c) levels of implementation completion measured through the Universal SIC.

Short CITS. The influences of initial training and leadership coaching on uptake rates were measured through use of an ITS design retrospectively examining state level data aggregated by service area and month. Initial training was defined as the 1-day training on the site-specific use of the EBP, which was an electronic social emotional assessment (i.e., the e-DECA) with full staff (both providers and administrators) in Part C early intervention as documented and provided by the state agency. Leadership coaching was defined as the follow-up coaching session delivered by an expert on the EBP/use of the electronic social emotional assessment and tailored to meet the individualized needs of the service areas, as occurred in the target sites. The coaching was 1 hour or less long, was delivered via phone or in person, and entailed answering questions and providing follow-up resources related to systems-level supports for the service area Part C coordinators. The definition of leadership coaching was established during the monthly research practice partnership meetings and refined with follow-up clarification with the expert/coach. The dates of the initial training and leadership coaching were examined relative to assessment usage levels and trends documented in the electronic system by service area.

SIC. The Universal SIC is a focused measure that is observation based and measures attainment of milestones foundational to successful implementation of EBPs

and/or innovations in systems (Chamberlain, Brown, & Saldana, 2011). The SIC has been used in a number of early childhood and home health visiting programs, with predictive capacity related to implementation success (Saldana, Chamberlain, Wang, & Brown, 2012), therefore holding high potential for suitability and applicability in Part C early intervention systems. The proportions of preimplementation and implementation activities achieved were calculated for each service area, because these activities have been found to be predictive of implementation success in similar complex systems (Brown et al., 2014; Saldana et al., 2012). The number of stages completed was also calculated for each site according to criteria proposed by Chamberlain et al. (2011). See Appendix E for a copy of the universal SIC.

Semistructured interviews. Data on determinants (supports and barriers) and driver influence on uptake were gathered through qualitative semistructured interviews of Part C providers and Part C local program coordinators via phone. Targeted one-on-one interviews with providers and local Part C program leaders were conducted to identify determinants (barriers and/or supports) related to usage rates and impact (or lack thereof) of the initial training, leadership coaching, and SIC. A brief review of aggregated findings from previous qualitative interviews on use of the e-DECA occurred before question development to ensure questions were not redundant of prior state work. An approach by Bevan (2014) was utilized to develop questions through a phenomenological lens to gather providers' lived experiences of EBP use, including understanding of potential determinants that could impact implementation completion. The SIC was used to guide questions with administrators in the second half of the interview session. Because of my expertise in early intervention and prior experience in qualitative

interview processes, I conducted all the interviews. Prior to conducting the interviews, best practice interviewing techniques and legitimization techniques were reviewed, including bracketing and bias reduction (Onwuegbuzie & Leech, 2007).

<div align="center">**Data Sources and Storage**</div>

Both existing data and new data were used to answer the research questions. Existing data were provided by the state and/or obtained from public state data sites in accordance with the signed and documented state-level data sharing agreements (see Appendix A). All data were stored via secure server (e.g. Microsoft OneDrive) to meet data sharing agreement requirements. All data shared from the state were deidentified and aggregated for confidentiality. All new data were gathered confidentially through a secure survey platform (Qualtrics) and stored in Microsoft OneDrive or via a secure USB drive. See Table 1 for research questions, participants, measures, mode of data collection, and proposed analyses.

Analytic Plan

The analytic plan requires intentionality about analyses in accordance with measures, participants, theory, and research questions and requires attention to sequence of the research design.

Design and Sequence

Recent research emphases the importance of mixed-methods research when examining implementation of EBPs and/or innovations in real-world settings due to the increased ability for "understanding and overcoming barriers to implementation" (Palinkas et al., p. 44). Best practices in mixed-methods design necessitate both intentional reasoning for the use of mixed methods dictated by theory and research

Table 1

Targeted Research Questions, Participants, Measures, Mode of Data Collection, and Proposed Analyses

Targeted research question	Main research focus area	Participant group(s)	Measures	Mode of data collection	Analysis
What are the feasibility, acceptability, and appropriateness ratings for use of an electronic early social emotional assessment (e-DECA) within Part C/early intervention pilot service areas?	Feasibility, acceptability, appropriateness	Providers, Part C local program coordinators/administrators	Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM)	Survey	Descriptive statistics
What are the potential determinants (supports and barriers) related to feasibility, acceptability, and appropriateness of use of an electronic early social emotional assessment within the pilot service areas?	Feasibility, acceptability, appropriateness	Providers, Part C local program coordinators/administrators	Qualitative questions: (a) review of aggregated qualitative data results from the state SSIP process, (b) open-ended survey questions, (c) qualitative interview questions	(a) Review of prior qualitative data collected, (b) survey, (c) interview via phone or in person	Phenomenology combined with rapid review process for analysis

Table 1 (continued)

Targeted Research Questions, Participants, Measures, Mode of Data Collection, and Proposed Analyses

What are the usage rates of the electronic social emotional assessment within each of the pilot service areas?	Feasibility, acceptability, appropriateness	Providers	Provider usage rates recorded within the e-DECA system	Existing data deidentified, aggregated data by pilot site from the e-DECA system provided by state data sharing agreement	Descriptive statistics
What are the usage rate slope trends before coaching intervention (e.g., slight positive, slight negative, consistent, or tapering off, indicating lack of sustained usage)	Drivers and uptake: leadership coaching	Providers	Provider usage rates recorded within the e-DECA system	Existing data deidentified, aggregated data by pilot site from the e-DECA system provided by state contact	(a) Visual analysis of frequency plotted by time, (b) segmented regression
On average, do level and/or trend of electronic social emotional assessment usage increase compared with baseline after leadership coaching on assessment use?	Drivers and uptake: leadership coaching	Providers	Provider usage rates recorded within the e-DECA system	Existing data, aggregated data by pilot site from the e-DECA system provided by state contact	(a) Visual analysis of interrupted time series (ITS) plot, (b) segmented regression with interruption based on initial training, and follow-up leadership coaching

Table 1 (continued)

Targeted Research Questions, Participants, Measures, Mode of Data Collection, and Proposed Analyses

What are the potential determinants (supports and barriers) related to leadership coaching on electronic social emotional assessment usage rates? (What else do leaders need in this process?)	Drivers and uptake: leadership coaching	Part C local program coordinators/administrators	Qualitative questions	(a) Review of existing summarized qualitative data, (b) follow-up questions based on whether the site saw a bump with the leadership coaching in the ITS	Phenomenology combined with rapid review process for analysis
What are the levels of implementation completion for each of the pilot service areas?	Drivers and uptake: implementation completion	Part C local program coordinators/administrators	Stages of Implementation Completion (SIC) and Stages of Implementation Analysis (SIA)	Interview	Descriptive statistics
What are the potential determinants (supports and barriers) of service areas completing more of the SIC components?	Drivers and uptake: implementation completion	Part C local program coordinators/administrators	Qualitative questions	Semistructured interviews	Phenomenology combined with rapid review process for analysis

questions and require intentionality about sequencing and timing of the methods
(Creswell & Plano Clark, 2017). I used a sequential mixed-methods design for added
validity, credibility, and depth of understanding (Creswell & Clark, 2017).

Review and Analysis of Existing Data

The first step of the research design involved examining existing state data
sources to gather an overall picture of demographic characteristics and assessment usage.
Existing data examined included qualitative reports on potential use of the electronic
social emotional assessment, demographics of service areas, SSIP process documents,
provider survey data on training related to the electronic social emotional assessment, and
usage data of the EBP from the electronic assessment data system. Electronic social
emotional assessment data were utilized to identify current usage rates of the EBP. Next,
I used a short CITS design to examine whether state-identified key drivers, initial
training, and leadership coaching on use of the EBP demonstrated a significant change in
level or trend of the usage rates.

Concurrent to the secondary data review and CITS analysis, foundational
information about feasibility, acceptability, and appropriateness of the usage of the
electronic social emotional assessment was gathered. Data were collected through use of
validated self-report measures that assessed both providers and program leaders on
feasibility, acceptability, and appropriateness of the EBP in the four pilot sites and
through customized measures related to team structures. All data were gathered through a
secure on-line electronic survey platform (Qualtrics).

Overall, all of the sources of data were analyzed to identify uptake rates and
potential driver influences on uptake. Once preliminary findings on trends and impact

were identified, follow-up qualitative data were gathered and analyzed to identify potential determinants (supports or barriers) to the uptake. In addition, data on stages of implementation completion were gathered through use of the SIC as part of the interview process to optimize accuracy of reporting and consistency of definitions across participants because the Universal SIC is not yet validated for this population (L. Saldana, personal communication, November 9, 2018). Specific quantitative and qualitative analytic procedures are described below.

Quantitative Analysis

Quantitative data were analyzed using R (R Core Team, 2017). Plots of the time series were produced using R with the *ggplot2* (Wickham, 2016). R is an optimal platform for data analysis and data visualization, particularly for ITS, because of its ability to perform necessary data transformations and data cleaning to prepare for and visualize the data, open source accessibility, reproducibility features, and range of packages available that allow for customization and optimal match to needs and data. (All codes for producing the plots will be available upon publication of this study.) Specific quantitative analyses included descriptive statistics and short CITS through use of a multiple baseline design.

Descriptive statistics and demographics. Descriptive statistics of children served in the pilot and beta service areas were pulled from the publicly available state website in accordance with the data sharing agreement. Demographic information for providers and administrators was gathered as part of the feasibility, acceptability, appropriateness, and team structure data gathering process through the Qualtrics Survey platform.

29

Feasibility, acceptability, appropriateness. The surveys on feasibility,

acceptability, and appropriateness were excluded when (a) the participant did not

complete the feasibility, acceptability, and appropriateness nine-question rating survey,

and/or (b) the survey was questionable as to origin/presence of bots as identified by

having an extremely short response time (less than 2 minutes) or having missing or

duplicate IP address information or duplicate responses, etc. In recent years, the impact of

bots on paid surveys has become an issue and should be considered when conducting

cleaning and screening procedures (Shanahan, 2018). The cleaning and screening

procedure resulted in 77 included surveys of the 84 surveys initiated. Means and response

ranges were calculated for each of the three validated measures.

Short CITS. The data for the multiple baseline short CITS included four pilot

sites with 33 months of aggregated total usage counts at monthly time points for a total of

132 observations (4 pilot sites × 33 monthly aggregated counts). A total of 501

assessments were completed and recorded in the electronic database during the 33

months of observations. Each of the four pilot sites had four to eight time points before

the intervention (baseline), 10 to 12 time points in the posttraining/precoaching segment,

and 12 time points in the postcoaching segment. These numbers exceed recommended

numbers for short CITS use with educational data, which are at least three to six time

points per pre- and posttreatment period (Hallberg et al., 2018). Power also increases in

ITS models with added design features, such as when a comparison group is used.

Multiple baseline design was identified as being the optimal design among a comparison

group (Gebski, Ellingson, Edwards, Jernigan, & Kleinbaum, 2011; Shadish, Cook, &

Campbell, 2002), such as used in this study. Parameters were added according to current

research on CITS designs to strengthen the design (e.g., added time points, multiple baseline design). This mixed-methods design was incorporated for added credibility and triangulation of results. Moderate effect size was assumed in this model based on prior training and coaching literature in early childhood settings (Neuman & Cunningham, 2009). Although there have been simulations for power in larger scale traditional ITS designs (e.g., Hawley, Ali, Berencsi, Judge, & Prieto-Alhambra, 2019), there has not been a formalized power calculation for short CITS designs incorporating design elements to add strength (e.g., multiple baseline design). Therefore, power was not formally calculated for this study design but was determined to be sufficient for detecting the effect of level change according to recent CITS articles employing a similar design (Hallberg et al., 2018). Overall, in spite of the best practice recommendations for improving design and power, upon which this analysis was founded and completed, power in CITS models is still being explored (Hawley et al., 2019). See "Discussion" and "Limitations" for further information on power in CITS designs.

I modeled the pretreatment trend of pilot sites to aid in determining whether a parameter should be added to the model to account for any preintervention trends because "CITS designs must contend with uncertainties about the pretreatment functional form of the outcome" (St. Clair, Hallberg, & Cook, 2016, p. 272). Upon inspection, baseline trends were consistent (St. Clair, Hallberg, & Cook, 2016; See Figure 1). Comparisons between the model with the added parameter and the model without revealed that these models were nearly identical. For the model containing the added interaction term, TIME × IMPACT_PHASE, the Akaike information criterion (AIC) value was 217.30. For the model without the interaction term, the AIC value was 213.73. The variance inflation

factors (VIF) were acceptable for all variables with GVIF^[1/(2 × df)] values of 2.67, 2.73, and 2.17 for IMPACT_PHASE, TIME, and PILOT_SITE, respectively (O'Brien, 2007). Therefore, given all these considerations, the most parsimonious model, that without the added interaction parameter to account for baseline trends, was used (Vandekerckhove, Matzke, & Wagenmakers, 2015).

A segmented multiple regression model was used to address the a priori hypotheses and answer research questions related to the impact of initial training and follow-up leadership coaching at each of the four sites as part of a multiple baseline short CITS design. The model was adapted from a model used by Wang, Walter, Grzebieta, and Oliver (2013). The segmented regression of assessment usage counts is as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 \ (\text{TIME}) + \beta_2 \ (\text{IMPACT\_PHASE}) + \beta_3 \ (\text{PILOT\_SITE}) + \beta_4$$
$$(\text{PILOT\_SITE} \times \text{IMPACT\_PHASE}) + e$$

where $\log(\mu_i)$ is the log mean of assessments completed per month; TIME is a continuous variable representing the 33 monthly data points in the time series; IMPACT_PHASE is the phase of impact in the model (baseline, post-training, and post-coaching), which were dummy coded with baseline as the reference group; PILOT_SITE represents each pilot site in the study (which were also dummy coded, with Pilot Site 1 serving as the reference group); and PILOT_SITE × IMPACT_PHASE is the interaction term describing whether the impact of the implementation drivers of initial training and follow-up coaching depend upon the PILOT_SITE (see Figure 3).

The use of comparison groups is a relatively new within short ITS designs but increases power through an added control for time varying confounders (e.g., Bernal,

Cummins, & Gasparrini, 2018). The pilot sites in this study were chosen by the state officials in 2016 toward the beginning of their systemic improvement planning process based on their geographic representation throughout the state. The pilot sites were determined to be appropriate comparisons for use in this model because they were similar enough to account for time varying confounders, such as temporal trends, statewide conferences, and universal dissemination efforts. The pilot sites met criteria outlined by Bernal et al. (2018) for both location- and characteristic-based controls. All pilot sites were in the same state, reducing susceptibility to temporal confounders, and all sites served infants and toddlers with disabilities with the same eligibility and program criteria. Each pilot site was coded for the different time points for the initial trainings as part of the multiple baseline design. Training at the four pilot sites occurred in November 2016, January 2017, February 2017, and March 2017. Follow-up coaching occurred on the same date in March 2018 for all sites in a small group aimed at problem solving systems needs and supports related to use of the assessment.

After specifying the model, it was important to adjust for using count data as the dependent variable. Count data are common in health and education research (e.g., Lee, Han, Fulp, & Giuliano, 2012; Ives, 2015). However, count data can cause skew; hence, the model would not meet the ordinary least squares (OLS) regression assumption of multivariate normality. When this assumption is not met, incorrect assumptions can be made about the significance of findings (Ives, 2015). Although some debate has appeared in the literature about using log transformations versus other types of regression (e.g., Poisson or negative binomial) with count data (Ver Hoef & Boveng, 2007), recent research indicates that log transformations are valid and more interpretable with certain

types of count data (Ives, 2015). Therefore, the AIC was used in model comparisons to determine the best model fit between log-transformed and negative binomial regression models. AIC weights are recommended when comparing competing models, and lower AIC values represent a better model fit (Burnham & Anderson, 2004). The log-transformed regression model had a lower AIC value in the model comparisons/best fit (AIC = 213.73) than did the negative binomial regression model (AIC = 311.86). The log-transformed regression model also had the best fit with the model assumptions (e.g., QQ plot for multivariate normality and multicollinearity tests/VIF). The Durbin-Watson statistic was used to check for autocorrelation as recommended in the literature on ITS (Hallberg et al., 2018; Hawley et al., 2019; Wagner, Soumerai, Zhang, & Ross-Degnan, 2002). Model predictions were also used to examine the data (see Figure 2).
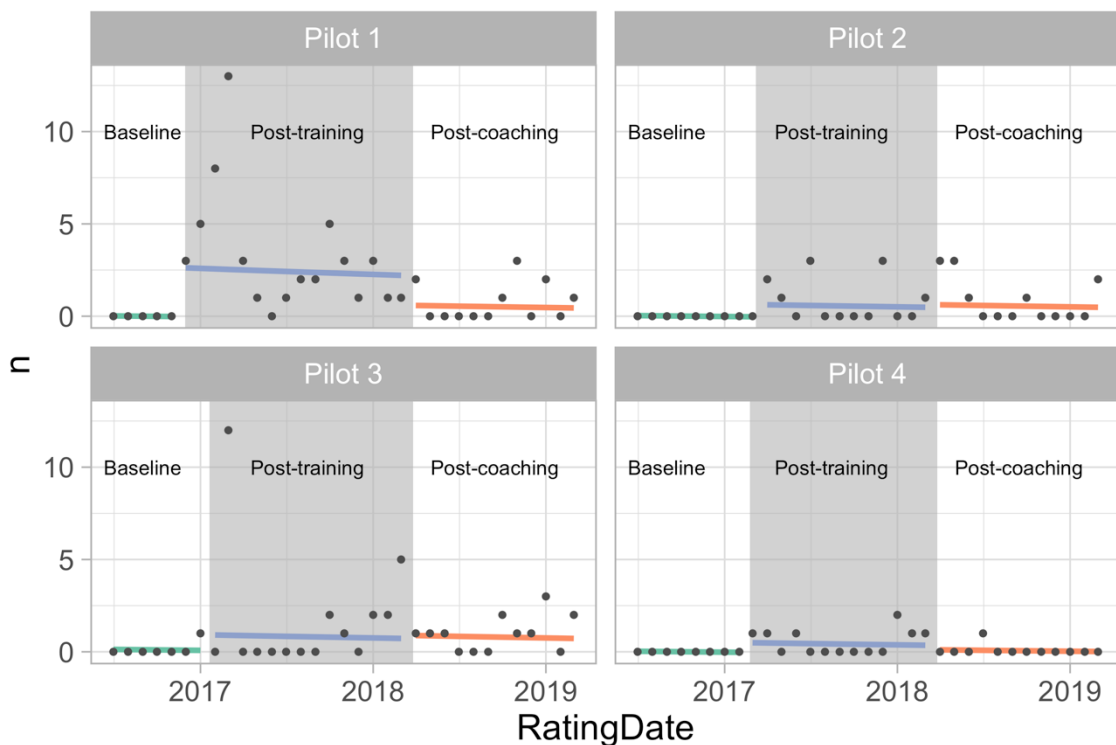
Figure 2. Segmented regression model predictions by phase (baseline, post-training, post-coaching).

Autocorrelation is serial correlation of the residuals, which violates the

assumption of conditional independence (Granger & Newbold, 1974; Rottman, 2016).

"Failing to correct for autocorrelation may lead to underestimated standard errors and

overestimated significance of the effects of an intervention" (Wagner et al., 2002, p. 305).

A low Durbin-Watson value (typically 1.5 or less, depending on the sample size) is used

to identify model misspecification, which can have various causes including the omission

of relevant variables, the inclusion of irrelevant variables, or autocorrelated residuals

(Granger & Newbold, 1974). Therefore, use of a combination of other model fitting

parameters and techniques, such as the AIC, combined with the Durbin-Watson statistic

can help parse out possible autocorrelation (Burnham & Anderson, 2004). The Durbin-

Watson statistic is often used to examine first-order autocorrelation common in time

series and has been recommended as part of procedures in recent articles about the use of

ITS in health and education (Bernal et al., 2017; Hallberg et al., 2018). The Durbin-

Watson value for the model was moderately low and significant (0.86, $p < .01$),

indicating the potential of correlation of error terms over time. However, the fitted-by-

residuals plot did not show significant patterns of correlation. Given that more traditional

time series designs (e.g., auto-regressive integrated moving average) are not feasible with

shorter designs (Hallberg et al., 2018), caution was used in interpreting the statistical

significance and slopes given the potential for either under- or overrepresentation of this

term when autocorrelation is present (Hallberg et al., 2018).

Major threats to validity in ITS designs include history, selection, and

instrumentation (Hallberg et al., 2018; Shadish et al., 2002). History was accounted for

by a multiple baseline comparison design including four pilot sites in the same state.

35

Selection was accounted for by the interruptions because they occurred during the academic year when there would likely be little to no staff turnover, given early intervention and school-based hiring practices that have staff turnover in summer or fall with the new school year based on contract hiring practices. Instrumentation was accounted for by using data from an electronic system that remained consistent in instrumentation and format during the time of the data collection.

In addition to the aforementioned statistical analyses, the impact graphs (see Figure 3) were analyzed visually by two reviewers, the primary researcher and one other member of the research team, who had background training and experience in single-case visual analysis. This visual analysis was used as a second form of data for triangulation to add validity and credibility to the interpretation of the findings due to the noted limitations of short ITS designs. Criteria for visual analysis were rigorous and based on those used in single-case design for visual analysis, including a stable trend in baseline, at least three data points in each phase, and identification of the specific point in time when the intervention plausibly would have an effect. There also must be at least three demonstrations of effect. The current models met all of these criteria. Next, to visually "assess the effects within SCDs, six outcome-measure features were used to examine within- and between-phase data patterns: (a) level, (b) trend, (c) variability, (d) immediacy of the effect, (e) overlap, and (f) consistency of data patterns across similar phases" (Kratochwill et al., 2013, p. 31). All visual models were then given a rating of strong, moderate, or no evidence of a functional relation between the predictor (training or coaching) and the outcome (assessment usage rates). The unit of analysis was the Part C service area. Overall, although these criteria were being used for the visual analysis,

my study was not intended to be construed as a single-case design. To meet criteria from the What Works Clearinghouse for a single-case design, the design must meet three criteria: (a) the independent variable must be manipulated by the researcher versus be a "naturally occurring event," (b) the outcome variable must be measured over time by more than one trained assessor, and (c) there must be agreement of at least .8 by two or more of the trained observers recording the observations of the dependent variable (Kratochwill et al., 2013). Although the features of my study may or may not have met the criteria of a single-case research study, the intent of using the single-case criteria was only for rigor in the visual analysis.

Qualitative Analysis

Qualitative analysis holds potential to "elucidate the complexity of human behavior, facilitate the development and implementation of impactful programs and interventions, and give voice to the lived experiences of inequity" (Raskind et al, 2018, p. 1). Yet, a recent review of research related to implementation and health practices revealed frequent inconsistencies and unclear reporting of qualitative methods (Raskind et al., 2018). Therefore, it is important to have an intentionally outlined and defined process for qualitative analysis to ensure an optimal and ethical approach to analyzing qualitative data. Overall, qualitative processes were guided by the work and recommendations for best practices in qualitative research of Creswell and Poth (2018) and qualitative research guidelines from implementation science (Department of Health and Human Services, National Institutes of Health [DHHS], 2017; Raskind et al., 2018). A rapid review process was used as the underlying foundation for all coding because this study was designed to utilize methods that hold potential for use in real-world systems

improvement. Rapid review processes are often utilized related to policy issues, such as this study, due to the need for expediency and mandated timelines (DHHS, 2017) because "delays may render the findings out of date, reducing their applicability and relevance" (Taylor, Henshall, Kenyon, Litchfield, & Greenfield, 2018, p, 2). Rapid review processes have acceptable validity and quality compared with more traditional methods in situations of quality improvement where time is constrained (Taylor et al., 2018). Being timely with information is sometimes more important than depth in these situations due to the balance of meeting the timelines and relevance goals (Taylor et al., 2018).

Qualitative interviews in this study were recorded by the investigator, transcribed by a professional transcriptionist, and checked for accuracy by the researcher. Atlas TI qualitative software version 8.4.0 was used for accuracy and transparency with coding and document management. Both inductive and deductive coding processes were utilized with double coding for all transcripts by three reviewers with training in qualitative coding and formalized training in early intervention and implementation science. Reviewers 1 and 2 each had over 20 years of experience in a range of health and education settings, including service provision and administration roles in early intervention. Reviewer 3 had 5 years of experience in a variety of early childhood settings as a provider and researcher. All reviewers had formalized training in early intervention and implementation science theory. Reviewers were intentionally chosen based on their foundational understanding of theory and to balance biases and provide analyst triangulation (Raskind et al., 2018). All reviewers were trained in the coding process and provided written directions and definitions of codes along with the transcript files. Deductive codes were based on the research questions and conceptual frameworks

from implementation science that were foundational to this study (AIF and IDEAS Impact Framework). Specific codes were determinants (supports and barriers), usable innovation, implementation stages, implementation drivers, teaming, improvement cycles, and needs/ideas/innovations. As part of the rapid review process, supports and barriers were grouped according to research question by Reviewer 1. Two groups were identified: (a) feasibility, acceptability, and appropriateness determinants; and (b) determinants related to impact, that is, drivers and uptake. Groupings were then confirmed for accuracy by Reviewers 2 and 3, and any discrepancies were settled through consensus. Inductive themes within these research question groupings were allowed to emerge and were coded for increased understanding of the specific determinants related to the research questions. A phenomenological approach was utilized as the overarching lens for inductive code development by allowing latent meaning to emerge to highlight the lived experiences of providers and leaders related to implementation of EBPs in early intervention settings relative to the research questions. The inductive codes emerged during the primary deductive coding process and were captured through the memos and audit trail process of reviewers. This approach allowed for increased accuracy and best practices related to code development and audit trails and was expedient because inductive codes emerged at the same time as initial deductive coding, eliminating the need for a secondary independent process. Any discrepancies between reviewers identified in the coding process went through a reconciliation process to meet consensus, which included meetings and discussions. A process to use an outside expert was planned as part of the reconciliation process but was not needed because consensus was reached through the Stage 1 reconciliation process. Member checking (checking with field

experts, including the state research partners), audit trail with memos about code development by all reviewers, and triangulation of methods (mixed methods) and data sources in the study were used to strengthen the validity of findings (Creswell & Poth, 2018; Onwuegbuzie & Leech, 2007). The negative case technique was employed to identify statements contrary to coded findings but may have been missed in the initial coding process due to scope or personal bias. Negative case is a recommended technique for improving the credibility and validity of findings (Raskind et al., 2018).

Overall, in alignment with current best practice guidelines, the qualitative inquiry in this study was intended as a form of triangulation to strengthen validity and deepen understanding of the larger phenomena of lived experiences related to uptake of EBPs in complex family-based systems of early intervention. Given the mandated timelines in policy-directed systems change work, rapid review processes are inherent to this type of work. Recent guidelines on qualitative research in implementation science reinforce this approach: "qualitative data analysis in implementation science tends to be targeted to specific a-priori research questions, is usually more rapid, and is often integrated with quantitative analyses" (DHHS, 2017, p. 4).

CHAPTER IV

RESULTS

Results and discussion are reported using the SQUIRE guidelines for best practice

reporting in quality improvement research (Goodman et al., 2016).

Demographics

Four pilot service areas providing services to infants and toddlers with disabilities

and their families under IDEA Part C early intervention were included in the study. All

service areas were located in the same Midwestern state and participated as part of the

SSIP process for program improvement. The pilot sites were chosen for the project in

2016 by state administrators based on their geographic representation throughout the

state. As part of the statewide training efforts related to assessment use, a total of 149

professionals, 12 administrators and 137 providers, were formally trained in the use of

the electronic social emotional assessment from October 2016 to March 2017. In a fall

2018 snapshot count, 11,025 children were being served in Part C throughout the state,

with pilot sites serving 100 to 900 children at that time. Exact numbers served by each

pilot site are intentionally not provided to preserve confidentiality.

Research Questions

The following a priori research questions were addressed in the study.

Feasibility/Acceptability/Appropriateness

1. What are the feasibility, acceptability, and appropriateness of the use of an EBP

   (e.g., an electronic social emotional assessment, the e-DECA) in Part C/early

   intervention pilot sites?

a. What are the feasibility, acceptability, and appropriateness ratings for use of an electronic early social emotional assessment (the e-DECA) within Part C/early intervention service areas?

b. What are the potential determinants (supports and barriers) related to feasibility, acceptability, and appropriateness of use of an electronic early social emotional assessment within the service areas?

Impact: Driver Influence and Uptake

2. What is the relationship between the identified key drivers and uptake rates of the electronic social emotional assessment use? Key drivers identified were (a) initial training of full staff, including providers and administrators; (b) follow-up leadership coaching regarding the social emotional assessment use with focus on systems-level supports; and (c) SIC scores.

   a. Training and Coaching

      i. What are the usage rate slope trends prior to the training and coaching interventions (e.g., slight positive, slight negative, consistent, or tapering off, which indicates lack of sustainment of the usage)?

      ii. On average, does level and/or trend of the EBP usage increase compared with baseline after (a) initial training and (b) leadership coaching on the EBP use?

      iii. What are the potential determinants (supports and barriers) related to training and leadership coaching on electronic social emotional

assessment usage rates of the EBP (i.e., what else do leaders need in this process)?

b. Implementation Completion

 i. What are the levels of implementation completion for each of the pilot service areas?

 ii. What are the potential determinants (supports and barriers) of service areas completing more stages of implementation?

Feasibility, Acceptability, Appropriateness

Feasibility, acceptability, and appropriateness information was derived from the survey using the FIM, AIM, and IAM and from qualitative interviews.

Feasibility, Acceptability, Appropriateness Ratings

The 77 respondents that were included in the final analysis after screening and data cleaning resulted in 74 to 76 individuals responding to each of the 12 items for the FIM, AIM, and IAM. Participants did have the option to skip any items in the survey they did not feel comfortable answering; hence, I am unable to determine whether questions were missed due to error or intentional skipping. However, there was a very high completion rate for all who attempted to answer all the questions: 96% of participants fully completed the survey, and 2.5% partially completed. The FIM, AIM, and IAM were scored on a scale of 1 to 5, where 1 is the highest level of disagreement and 5 is the highest level of agreement or positive response. All measures had mean scores in the moderate range; means were 4.08, 4.00, and 4.03 for the FIM, AIM, and IAM, respectively. These numbers were supported the original hypothesis that the electronic social emotional assessment is rated as at least moderately feasible, acceptable, and

43

appropriate for use by Part C early intervention providers and leaders. However, answers to individual items within the sections ranged from 2 to 5, indicating variance in the feelings of feasibility, acceptability, and appropriateness.

Determinants to Feasibility, Acceptability, Appropriateness

Qualitative interviews focused on a range of questions related to providers' and administrators' lived experiences relative to the use of the electronic social emotional assessment. interviews included open questions that allowed themes to emerge and specific questions aimed at issues of feasibility, acceptability, and appropriateness.

Feasibility. Feasibility is defined as "the extent to which a new treatment, or an innovation, can be successfully used or carried out within an agency or setting" (Weiner et al., 2017, p. 2). Supports related to feasibility of the electronic social emotional assessment use included affordable and quick to use because of its on-line access. Barriers to feasibility centered around difficulty with efficiency, lacking time and technology, inefficiencies and user layout, and lacking the tools for efficiency and/or inefficient processes.

Feasibility supports. Affordability and accessibility were mentioned as making it more feasible for use in early intervention.

"What I like best is, you know, the affordability and the accessibility." —Admin 1.

Several interviewees discussed the feasibility of the assessment for use because of the efficiency and on-line access.

I like how it's really quick. —Provider 4

I love the usability and the fact that it's online … because I think that just kind of cuts down on like, you know, the having to transfer it and score it differently, etc., etc. —Admin 2

44

I mostly use my iPad for it. Um, every now and then if I have to leave a copy of it for a family member if I don't see them, I'll print it out, but for the most part I'm doing it on my iPad. – Provider 5

I like that it's online. —Provider 2

I like that it's electronic so that it's easy to track and we can go through and have multiple, um, I can fill it out, a daycare provider can fill it out, the family can fill it out, um, and it's just all right there to be able to be compared and looked at, at any time. —Provider 5

Feasibility barriers. References to difficulties related to efficiency or needing

more time were noted throughout many of the interviews.

We still are, you know, trying to streamline our assessment tools, you know, so I think finding a really good one for social/emotional development is difficult, and I do really like the e-DECA. It's just I think from a practical standpoint, you know, because it's not assessing other areas of development, um, you know, we try to use, and we're having to, you know, utilize like, we use the DC-2 now, um, and that has a social/emotional section, and you have to complete the entire document. So, I just, you know, from like I said, from a productivity and practicality standpoint, you know, pulling out in another evaluation unfortunately isn't always the easiest. —Provider 6

I think that if we were given more time and, uh, you know, more time to do because—just as along with all other services I think around the country, people are stressing and feeling they are asked to do more with less time. —Provider 5

I guess just more hours in my day. If you could create the 27-hour day that would be fantastic. —Provider 2

Then we go ahead and put it into the computer system. I think, um, it's time-consuming. We keep being told that it should take about 5 minutes, it should take about 5 minutes. It usually ends up taking more like 20 to 25. —Provider 3

I simply, I don't have the time. —Provider 1

So, they see it as potentially being used but their, um, barrier is time. —Admin 3

I would say that's the biggest because it's often an additional so there's not often times that's all we're giving unless it's for, um, practitioner purposes. So, in assessment we're typically doing another tool as well. —Admin 2

45

In addition, themes related to needing the assessment to be more efficient or built into routines versus being an add-on were noted by a number of providers and administrators.

> I would say it does not come up as far as being utilized that frequently, um, during our teaming meetings. —Provider 6

> In all of our families it's not just one area we're working on, we're working on all those areas, and so a whole visit I talk about feeding, and then all of a suddenly we talk about sleep or we talk about this, and all of a sudden I've done six visits, and I completely forgot to, you know, bring out the assessment tool because we've just gotten so wrapped up in other things. —Provider 2

> It's just another thing that you have to add on to—that's why I don't use it on every child. I simply, I don't have the time. —Provider 1

> The e-DECA, um, has been a challenge to put into practice for us. —Admin 3

> Faithfully maybe 20%, um, and I think that's really just because it's just another thing, it's not because they don't want to do it. I think it's just, oh yeah, that's right, I can use that for eligibility. It's like it's not ingrained as part of their routine yet. —Admin 2 [regarding how many staff are using the assessment]

Themes related to technical difficulties and/or barriers impacting feasibility emerged. These quotes are related to usability and technology.

> Um, it's not terribly user friendly. I think if I, you know, if we were using it more consistently.... —Provider 6

> It's electronic and we have some team members who … don't know how to use computers as well. —Provider 4

> I mean, we're definitely able to navigate it and get what we want, but there are sometimes staff—it's like, oh, I know there's something like this that's part of there. Like, where is that again? —Provider 4

> I'm not very computer savvy, so just putting it out there, I really am not good at all on the computer so it takes me a little bit to navigate it, but it's not like it's challenging comparatively with my, like, electronic documentation for my power school or whatever, you know. —Provider 2

> I think if my coworkers could be here they say they have a hard time logging in. —Provider 2

46

It's not super user friendly for somebody who doesn't have a lot of time … if it was a little, um, more streamlined and user friendly where staff didn't have to do so many different screens and that to put the information in and get the results they want and all that, I think that it might be, um, more acceptable to use. — Admin 3

In a multidisciplinary team if you're wanting me to consult on a kid I'd like to be able to just pull their file review type of thing, and sometimes that's tricky because you can't really look as a group, you can only search by provider. — Admin 2

When it first came out may have been, um, a little more innovative, but it's a little bit clunky. —Admin 1

Several interviewees noted a lack of tools necessary for efficiency (e.g., an iPad) to be able to complete the assessment in the home and/or a less efficient process of filling out the assessment on paper and needing to reenter the data into the computer back at the office after the home visit.

I don't like entering it online from the paper version because it is so tiny and I find, I mean, this sounds kind of ridiculous and tedious, but I find I get off a lot on trying to enter it because I'm okay, well where's that radio button, where's that one? Um, so that's one thing that just from a user perspective is difficult. — Provider 1

The way we have to do it is we have to print off a copy of it, take it to the home with us or give it to the parent at a point, and then go back and input all the data. —Admin 3

I have the parent fill it in on the paper and then I, um, so between their session, you know, so they leave me, and then before their next session I complete it online and then I, um, print off any appropriate handouts that go with that. — Provider 5

We don't have iPads or Hotspots, so that would be lovely, but no, I take out the paper version and do more of kind of an interview with a family and then come back and put it into the computer at my desk. —Provider 1

The problem is we don't have access to mobile devices to take them out on the road, and so that's another barrier for us. —Admin 3

Acceptability. Acceptability is defined as "the perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable,

palatable, or satisfactory" (Weiner et al., 2017, p. 2). Acceptability supports and barriers themes both centered around seeing the importance, or lack thereof, of social emotional development and assessment. In addition, issues related to the strategies were both a support and a barrier to acceptability.

Acceptability supports. Statements related to the overall importance of addressing social emotional development and using it to qualify children for services were noted.

> I think ongoing social/emotional development is, you know, obviously is super important. —Provider 6

> They use it for eligibility. They use it for, um, programing. They see the value in it. They see it as something that is, um, very helpful because we get a lot of kids for Part C that, um, well, it's just because of, um, you know, behavior or it's because of, like some of the foster kids with all their placements, um, that they're showing, exhibiting some concerns with their behavior, some things like that. — Admin 3

> We were allowed to qualify him for, you know, under ECDD and social/emotional and that qualified him." —Provider 2 [regarding what the provider liked best about the assessment]

> I think that it's important work and important work for parents especially, because I don't think it is one of those areas of development that we, um, traditionally kind of look at. —Admin 4

> I do think with this e-DECA and with us being the pilot I think it has brought it to our attention that, or not brought it to our attention, but it makes us remember that the social/emotional development is kind of the foundation for all of other learning and all of their success in later life. —Provider 5

> When you work with infants and their families social/emotional, um, development of the infant as well as the family is just a huge piece of the puzzle. —Provider 4

> I don't think anybody would say that there is not, that it is not, you know, this area of development [social emotional] is not a priority because we know for sure that it is. —Provider 3

In addition, statements related to acceptability from having strategies to use as follow-up from the assessment were noted.

I love how the assessment's set up so that when it says, you know, you have an area of need in, um, you know, attachment or you have an area of need in you know, whatever it is, then it will give you the strategies, you know, I think it like, I'm probably saying it wrong but it like, gives you those strategies like, within the classroom or within the home for the, you know, about the child's age to address that area of need. So that's kind of where I pull those from to give some family-friendly ideas. —Provider 4

I've really liked the fact that you can access strategies. —Admin 4

I like the fact that there's resources connected to it. —Admin 3

Acceptability barriers. There also were statements that social emotional

development is not as important a focus area as are other areas of development.

I work with a lot of infants, um, itty bitties out of the NICU, and so the social/emotional hasn't entirely taken, you know, the families are more focused on their getting them to grow and develop and you know, um, gain weight, if you so to say. —Provider 4

You have people who say they understand it, they get it, it appears so on surveys and that, but when you get down to the nitty gritty there's still not that true understanding of social/emotional and the importance of it. —Admin 3

They didn't know that we could really serve kids with just a social/emotional delay. If that makes sense? They thought that they had to have bigger and broader developmental needs. —Admin 2

I think just basically it's still in its infancy and there's still, um, skepticism on it. —Admin 3 [regarding using the electronic social emotional assessment]

In addition, providers noted frustration regarding using the tool but also difficulty

because of lack of appropriate supports to refer to after conducting the assessment.

Because we address it, talk about it, and then send them out there to almost nobody. —Provider 3

You kind of see this whole topic of social/emotional development coming, um, over from all agencies, so clearly it's a concern, and I think with, um, mental health, how it is today that there's less and less supports for mental health, that maybe that's a reason." —Provider 5 [regarding why there is not more use of the assessment]

Some acceptability barriers were related to the strategies, with many interviewees noting that they would like to see improvements in this area because it is one of the largest areas that promotes the acceptability and use of the tool.

> I think folks were a little bit confused, mostly about what to do with the information, so some were giving it pretty faithfully, but as a whole group they weren't seeing the benefit or understanding how to pull apart the strategies. —Admin 2

> The only other thing that I just don't love about it is I really don't—although I think it gives really good information, um, I feel like the, um, like the you know, ideas or the suggestions and strategies, I just don't feel like they're super comprehensive or detailed. —Provider 6

> I will go get other handouts from, um, things that I already have, which is I think more useful for myself and for my families than to stop and take the time to go through that feedback, because I haven't found any time where I'm like, wow, that was some great information. —Provider 5

> I mean, a lot of them are common sense. —Provider 2 [regarding the e-DECA strategies]

> The strategies seem a little generic. —Provider 1

> So, I guess my downfall with the e-DECA is I've had to really search out strategies on my own because I found the strategies were limiting. They were like, well yeah, everybody knows you need to read to your child, but what does that look like? —Provider 1

> So, that's, you know, that was some comment that I had heard is that the resources were just too basic and too general. —Admin 3

> What staff have said to me about the e-DECA is that they, and this is something I hear, I've heard for years, that they wished there were more interventions, that they wished there was, um, more—it was more robust. —Admin 1

Appropriateness. Appropriateness is defined as "the perceived fit, relevance, or compatibility of the innovation or evidence-based practice for a given practice setting, provider or consumer" (Weiner et al., 2017, p. 2). Supports identified related to appropriateness included that the assessment was family friendly and breaks down descriptions and milestones for ease of use appropriate for IDEA Part C services. Barriers

related to appropriateness included providers feeling the assessment did not provide added information beyond what they already knew and was not compatible when working with families in group settings.

Appropriateness supports. Interviewees referenced the fit related to use with families and caregivers. In addition, several references were made to appropriateness for use in early intervention because the assessment breaks down social emotional development and milestones into usable descriptions and is easy to understand because it is comprehensive and objective.

Statements related to appropriateness first centered around the assessment being family friendly and conducive for use with caregivers.

I think it's, you know, family friendly. —Provider 6

I do like that it gives me, you know, a pretty good report of in terms of like, the breakdown is helpful for me in talking with parents. —Provider 2

I do like that it factors in, um, a difference in scoring, you know, based on the reporter" —Provider 6 [regarding how different scores/norms are determined based on whether the caregiver or service provider is answering the questions]

Some statements were related to how the assessment provided a more comprehensive and objective approach to social emotional development and helped break down milestones and provided an understanding of development for use in early intervention settings.

I think that it has been good to bring us back to thinking, um, okay this is really where we need to start. —Provider 5

Up until this point we've never really, um, looked at it [social emotional development] in terms of the actual developmental milestones, the age ranges, the, um, specific behaviors, and again, you know, ways to describe and what to call that. So, I think that the e-DECA has done that. —Provider 2

The e-DECA is I think, um, it's an excellent tool for somebody who is new to early childhood … I think it's an excellent tool for that, um, because it gives somebody that's new to it a lot of information. And a lot of, even for me I think it gives some talking points to work with families. —Provider 5

I think it takes away that judgmental piece. It's more objective in what it's looking at. —Admin 1

I love how comprehensive it is. —Provider 4

I think probably the best thing about it is that it is, um, it does definitely direct your attention to specifically social/emotional development … Um, and it gives you, um, you know, vocabulary and descriptions of that. —Provider 2

Appropriateness barriers. Some providers did not feel the assessment provided additional or needed information.

I know that, um, that myself and the staff that I work with, we're all pretty seasoned teachers, we've been doing this for well over 20 to 25 years, that it's not giving me anything new that I don't already have access to. —Provider 5

In addition, providers doing service provision in group settings noted a lack of fit/appropriateness in those settings.

A lot of times we're working in groups and so sometimes it's difficult to, when you're running a group, to sit and take time to talk with a family individually about, um, going through the e-DECA, which becomes a little bit more confidential. —Provider 5

I think also the fact that we are not fully in the natural environment makes it difficult because, um, you know, the—like for instance, the suggestions for home and the suggestions for school, we are a school but the references are to, you know, like a child who is at day care or, um, you know, in some kind of a center as their natural environment, not so much what's happening here. So, there's a little bit of a disconnect just because of the way we provide services. —Provider 3

Impact: Driver Influence and Uptake

The segmented regression model was significant ($R^2 = .36$, $F(12, 119) = 5.51$, $p < .01$), accounting for 36% of the variance. A significant positive increase in mean

assessment usage numbers was noted for all pilot sites after the initial training ($p < .05$)

(see Table 2 for estimates).

Table 2

Regression Results – Estimates for Monthly Number of Assessments Completed by

Intervention and Site

Coefficients	*Estimate*	*SE*	*p*
Intercept	0.016	0.232	0.946
Post-training	1.310	0.296	0.000***
Post-coaching	0.609	0.418	0.148
Pilot 2	0.016	0.288	0.956
Pilot 3	0.107	0.302	0.724
Pilot 4	0.012	0.294	0.968
Time	-0.008	0.013	0.538
Post-training by Pilot 2	-0.789	0.348	0.025*
Post-training by Pilot 3	-0.730	0.356	0.042*
Post-training by Pilot 4	-0.876	0.351	0.014*
Post-coaching by Pilot 2	0.008	0.357	0.982
Post-coaching by Pilot 3	0.066	0.368	0.857
Post-coaching by Pilot 4	-0.368	0.362	0.311

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. Estimates based on a log(n+1) transformation.

However, mean assessment usage numbers per month after coaching were not

significantly different from numbers before the follow-up coaching ($p > .05$). Mean rates

were calculated for each of the pilot sites to further examine impact after initial training

and after follow-up coaching by exponentiating the estimates for interpretation. Often,

the incidence rate ratios or odds ratios can be used for interpretation with types regression (Peng, Lee, & Ingersoll, 2002). However, because specific count data were used in this study standard counts were most meaningful and hence transformed for numeric interpretation, which has been noted with some cases of categorical count variables.

Baseline Usage Rates and Slope Trends

Baseline is defined as the time period from when the pilot sites had access to the electronic assessment until the date of the initial full staff training on the electronic assessment. Mean baseline rates were parallel and linear; all pilot sites had a mean baseline count rate of close to zero assessments per month before training with monthly mean rates of 0.02, 0.03, 0.13, and 0.03 for pilot sites 1, 2, 3, and 4, respectively.

Training and Coaching Impact

Impact of initial training and follow-up leadership coaching on assessment usage rates were also examined in the model.

Initial training impact. Initial training, represented by the interruption between Phases 1 and 2 on Figure 3, had a significant positive impact ($p < .05$) on all pilot site usage rates compared with baseline in all areas. After training, Pilot Site 1 had a mean of 2.77 assessments, Pilot Site 2 had a mean of 0.71 assessments per month, Pilot Site 3 had a mean of 0.81 assessments per month, and Pilot Site 4 had a mean of 0.56 assessments per month. Visual analysis revealed evidence of immediacy of effect with a noted change in level at all sites after training and overall consistency across Phases 1 and 2. However, an immediate drop in usage rates after the initial training month at two of the four sites was noted. This drop confounded the requirements of overlap and variability, ultimately resulting in a determination of moderate evidence of a functional relation between initial

training and usage rates (Kratochwill et al., 2013). See Figure 3 for the segmented regression impact phases.

Follow-up coaching impact. Follow-up leadership coaching, as represented by the interruption between Phases 2 and 3 in the model, did not have a significant impact across any of the sites ($p > .05$). After coaching, Pilot Site 1 had a mean of 0.87 assessments per month, Pilot Site 2 had a mean of 0.89 assessments per month, Pilot site 3 had a mean of 1.01 assessments per month, and Pilot Site 4 had a mean of 0.29 assessments per month. The visual analysis revealed the same findings: no immediacy of effect, no overall change in level or trend, and confounding overlap at all pilot sites, resulting in a rating of no evidence of a functional relation (Kratochwill et al., 2013) (see Figure 3).
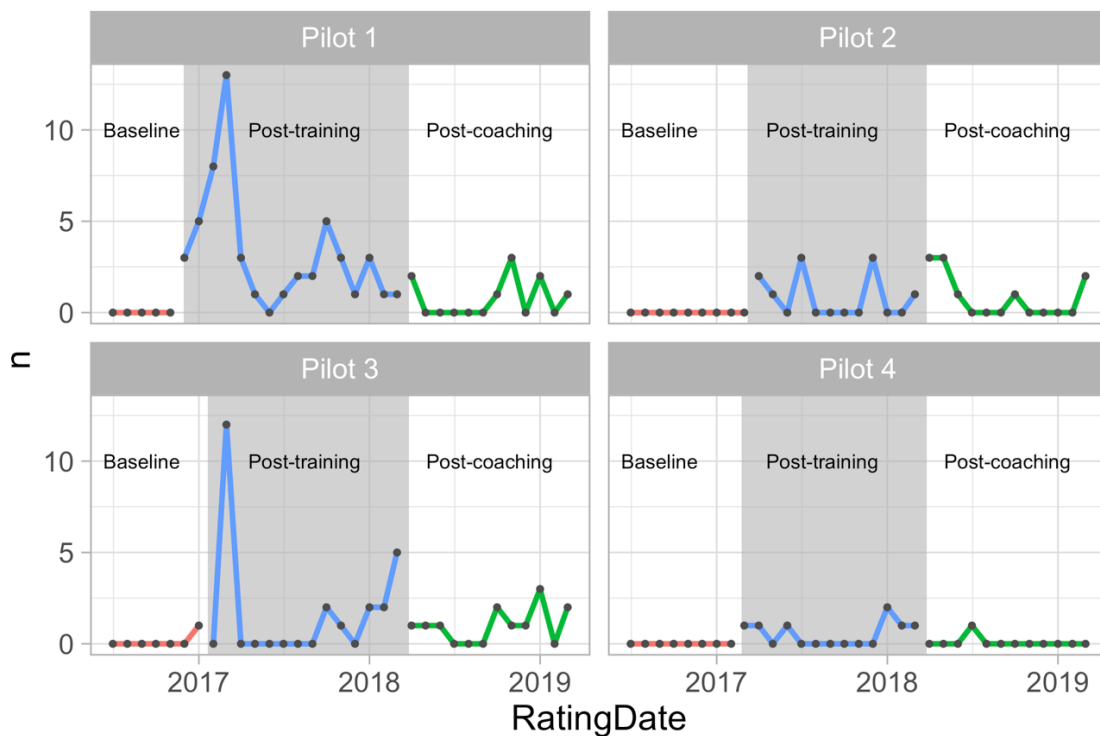

Figure 3. Multiple baseline driver impacts of initial training and follow-up coaching by pilot site.

Qualitative Findings: Driver Influence and Uptake

Qualitative findings relative to driver influence and uptake included themes related to determinants (supports and barriers) relative to coaching and training and an overall theme of difficulty with sustainability and waning usage.

Determinants: uptake. Interviews revealed a number of quotes related to usage patterns of the electronic social emotional assessment, with greater usage initially and quotes related to a more recent lack of use of the assessment. Particularly, there were statements that guidelines for use in the beginning were well defined and resulted in use in the beginning after the initial training.

> Our area was using it pretty highly in the beginning. I don't know if that continues to be the case. And I would take responsibility for that because I feel like I haven't revisited the e-DECA this year. —Admin 1

> It was actually an annual goal of mine and a few of my coworkers to use the e-DECA with um, I think 5 families. —Provider 2

> Everybody on staff was supposed to find three children or two children that we needed to do the assessment with, do the documentation, give the information, and then we kind of reconvened in a professional development or a staff meeting to talk about what um, our feelings about it, how it went and, you know, what the pros and the cons of using the e-DECA were. —Provider 3 [regarding process/guidelines for use]

However, there were many statements that more recently there is not as much usage.

> I have not done—I haven't even done a e-DECA this school year. —Provider 5

> I haven't used it within, you know, since 2019 which sounds crazy, but, um, I think I had a hard time kind of finding the strategies, like, one of the tabs. —Provider 2

> Quite honestly, I don't know that that is, if that's still the directive because we've also had some kind of major changes in the way we run the program, um, over the past year and it sort of went to the wayside. —Provider 3

> So, we're struggling with the e-DECA. —Admin 4

Determinants: driver influence. Themes of potential determinants (supports and barriers) related to the initial driver influence of training and leadership coaching on electronic social emotional assessment usage rates were identified through the qualitative interviews.

Training and coaching supports. Statements related to practices supporting the drivers of coaching and training were noted in the interviews. Statements referenced having access to a team lead or expert coach and the comprehensive nature of recent trainings.

Several interviewees referenced having a designated team lead and/or expert coach as support for the use of the electronic social emotional assessment.

> So, we do definitely have a point person, but the majority of the staff has been trained in that. —Provider 6

> Periodically we've, um, you know, we've checked in with, um, we've had great support from [expert coach], um, whenever we came up with a, you know, across something that, you know, we weren't sure or there was a glitch. So, we've had great support from her with the system. –Admin 3

> First, they gave us an introduction to it through, um, conference calls, and we had meetings at the state. —Admin 3

> We do monthly, like, calls and then quarterly visits as a face-to-face. – Admin 2

There were also references to the comprehensive nature of recent trainings on social emotional development.

> I think our county and our program has done, you know, a really good job in recent years, um, with doing a lot of, um, trainings on social/emotional development so not always related to the e-DECA but, um, in general I think they've done a, you know, pretty good job. —Provider 6

> I think that the e-DECA has done a wonderful job of implementing trainings on them to make sure that it was as easy as possible to use and as user friendly as possible, and I've been very impressed with the follow-up and them wanting feedback to try to improve it as much as possible. I think it has been very impressive. —Provider 5

Training and coaching needs/barriers. Themes emerged related to the need for further training, need for further coaching and/or more time with an expert coach, and a need for further training on specified strategies. Multiple interviewees noted that one of the largest barriers to further training was "always time" (Admin 4 and Provider 1). The need for follow-up coaching was referenced both on the provider and administrator levels:

> Staff is trained in e-DECA, but I just feel like they haven't had that follow up. — Admin 1

> My wish would have been and is, and I tell them all the time, that [expert coach] is able to provide the same level of support that she provides for infant mental health. —Admin 1

> We do actually, and we have, but again, we haven't met since the summer I don't think. —Provider 3 [regarding whether there is a point person on site for questions about the e-DECA use]

> I have asked our leadership team at the state for more support from a content expert, um, just as a leader. —Admin 1

> References were made to a need for reflective supervision and general coaching in

addition to coaching specific to the electronic social emotional assessment.

> I feel like we have to have the infant mental health model of supporting our staff so then they can support our families so then they can support their babies. —Admin 1

> In my career, I have had a professional coach for, um, probably 3 years. Um, at different times I had a coach and I'm telling you, that was so helpful. Even though the person was not a content expert, um, in our field, uh, it was very, very helpful in staying on track and getting support and, uh, just not losing my way. —Admin 1

> There were also statements related to a need for further training related to the

specific strategies and interventions, continued training on the overall importance of social emotional development and assessment, and training on implementation science and how to make systems changes.

I still feel that piece is missing of let's get down, not that we need to provide infant mental health services, but let's get down to the nit and gritty like, what could a home visit look like? What do you do when a child walks up and hits you? Like, let's talk about specific things, strategies. What do you say? What do you do? —Provider 1

I think folks were a little bit confused, mostly about what to do with the information … they weren't seeing the benefit or understanding how to pull apart the strategies. —Admin 2

There still needs to be more of that training of the importance of it in early intervention and in families, and, you know, how that all fits in the, you know, the big picture. —Admin 3

In addition to needing more specified training, barriers related to staff turnover

and how to handle ongoing training emerged.

Of course some of the staffing patterns have changed, so some of those that were providing services then, um, were no longer with us, and then we added some new people, and so continuing of training of that has been somewhat interesting. —Provider 1

Overall, participants stated that the training and coaching were of benefit but in

most instances could not recall what the training and coaching had entailed. Clear

guidelines and manualization are common components related to fidelity and have been

noted to be core components to achieving higher levels of implementation completion

(Saldana et al., 2012).

Implementation Completion

Information used to calculate the implementation completion scores were

obtained through use of the Universal SIC in the second half of the administrator

qualitative interviews. Specific SIC scores and themes of determinants related to

implementation completion were identified through the coding process.

SIC scores. SIC stages completed and proportion of activities were determined

according to criteria by Chamberlain et al. (2011). Stages completed indicates the last

stage at which at least one activity was completed, and proportion of activities completed

is the total activities completed divided by the number possible. Overall, all pilot sites

demonstrated a stage completion score of 6: "services and consultation begin"

(Chamberlain et al., 2011). Proportion of activities scores in each of the stages were

similar across most sights and ranged from 0% to 100% completion (see Table 3).

Table 3

SIC Proportion Scores by Site

Stage	Completion (%)			
	Pilot Site 1	Pilot Site 2	Pilot Site 3	Pilot Site 4
1. Engagement	100	100	100	100
2. Consideration of feasibility	75	75	75	75
3. Readiness planning	60	60	60	60
4. Staff hired and intro training	100	100	100	100
5. Fidelity monitoring process	25	25	25	25
6. Services and consultation begin	75	75	50	50
7. Model fidelity and staff competence and adherence tracked	0	0	0	0
8. Competency	0	0	0	0

Note. All completion rates are as of February 2019.

Determinants to SIC. Qualitative interviews revealed several themes related to

supports and barriers impacting implementation completion. Supports to completion

included having access to expert coaches/state level supports and pilot teams/early

adopters at the local sites. Regarding barriers to stages of implementation, themes were

related to lack of leadership power and the need for formalized/written usage criteria, and

overall themes were related to lack of sustainment.

Expert coach/supports. Several statements emerged related to having access to a

coach or point person to ask questions and receive coaching on the use of the electronic

assessment.

I mean, we do have access to, you know, [expert coach] and then we had access as a state-appointed kind of point person to also work with us on this project. — Admin 4

I do have a small group of people who are, um, the pilot team for social/emotional. —Admin 3

I took, like, a core group of the early adopters to a training that ... was doing, um, with [expert coach] and sort of learned a little bit about what we weren't doing or not understanding, and part of that was because I learned it with them. I wasn't, like, a train the trainer kind of thing so I wasn't able to see, like, what they weren't utilizing it for. —Admin 2

Barriers to implementation completion. Several barriers to implementation

completion were noted, including lack of leader power/lack of ability to make systems-

level decisions, lack of written criteria for when and with whom to use the assessment

(fidelity), and a need for program leader training in systems change/implementation

science.

Lack of authority for decision making. Several administrators provided

statements related to the difficulty in making change because they do not have the

authority to make systems-level decisions as the coordinator. In addition, there were

statements related to the overall effects leadership.

I'm not the decision maker. I have administration that I have to report to who makes the decisions even though I am the coordinator. —Admin 3

So, part of the challenge … is that I'm not everybody's boss and I/we're decentralized and that, um, that's gonna be changing next year. —Admin 2

We also have had change in leadership here." —Admin 1 [regarding barriers to implementation and use of e-DECA]

Need for formalized guidelines/criteria on assessment usage. Next there was a

noted lack of fidelity criteria/written guidelines with all service area leaders and

providers. Consistency was also lacking within the criteria provided verbally. Some

references were made to a need for formalized expectations because this is the impetus

for prioritizing usage when time is short.

Regarding the criteria for usage, both providers and administrators shared

common strands that there were not written guidelines/criteria.

We don't have written guidelines yet. —Admin 3

The only guidelines that I give them is we are really supposed to do it on all of our foster care kids. —Admin 2

No, we don't really have any current guidelines of when we're using it. —Provider 3

So, they kind of left it up to the coordinators to figure out based on what our individual needs and strengths and weaknesses were to figure out how this would work best for us. So, some went in different directions. We all kind of went in different directions I think. —Admin 3

Providers noted confusion related to this lack of criteria at times, often resulting in

decreased use of the assessment because it was not a mandate and there were not clear

criteria.

So, we'd have different directives each year —Provider 5

We are using them for all kiddos in foster care, um, but aside from that I don't think we have anything other than just the recommendation of it's here, this is what we have available use, use it if there are any concerns in social/emotional areas. —Provider 4

If it's something that I'm not mandated to do then I just feel like the, it's, um, it's better just to, to shoot through the hip and go through things as we see them. —Provider 5

Having to get the annual done and that takes priority because that's a state, you know, federal guideline that I can't break. —Provider 2

We've had some encouragement to use it a little more, especially with all of the, um, you know, COSF. —Provider 6

However, both administrators and providers noted the importance of having guidelines and had provided verbal guidelines, with some administrators working toward more formalized definitions and guidance for use.

> If I really built the time where we did a check in about it, so more accountability around it. —Admin 2

> I think it's also us having clear guidelines as our service area. When do we want to use it, when do we not want to use it, or should we just use it for everybody? We have to, I think, just kind of put some parameters about when it would be the most useful for us, if that makes sense. —Admin 4

> We thought maybe if, um, staff did one it would then snowball into, you know, seeing the value in doing more, you know, with more families. So that didn't happen. —Admin 3

> [Expert coach] brought this out, and I thought this was wonderful of, you know, not just focusing—using [the electronic assessment] not to just focus on kids that have maybe a deficit or, you know, a concern or a weakness but let's do it on everybody so you can also celebrate those kids that have really good solid social/emotional. —Admin 3

> Just having it just be a standard part of every evaluation would probably be really helpful. —Provider 4

Systems change coaching and support. An overarching theme was noted related to support needed in understanding how to make systems changes.

> It's one of the things that's kind of missing from the pilot, is that practical how do, you know, how do you implement things. —Admin 1

> I mean, they have talked about implementation science but I'm like, okay sister, I just want a plan, you know. I don't want a bunch of charts. I just want a real plan with what I'm doing. —Admin 1

CHAPTER V

DISCUSSION

The results and use of a sequential mixed-methods design to examine the implementation of an EBP as part of a policy-mandated initiative within the complex interdisciplinary system of early intervention/Part C of IDEA serving infants and toddlers with disabilities have been presented. I examined the use of a real-world application of an ITS as part of the sequential mixed-methods design for added validity and understanding of determinants to uptake of the assessment. Given that less than 50% of interventions make it to full scale, growing attention to the concept of implementation failure has been noted in the literature (Bryk et al., 2015). Implementation failure is important because there is immense time and money waste associated when interventions fail to reach scale. In systems where there is already a shortage of funding, such as IDEA Part C early intervention systems (Grant, 2005; IDEA Infant and Toddler Coordinators Association, 2018), it is important to begin harnessing the pilot phase of initial implementation to make quality improvements prior to scale-up and thus improve uptake rates. This requires innovation and improvement in the use of objective measures with subsequent targeted quality improvements. In this study, a systems implementation of a new practice—conducting an electronic social emotional assessment—was analyzed prior to scale-up. Four service areas in a Midwestern state were part of the pilot phase, and those participants were interviewed in this study to examine the feasibility and uptake of the EBP. The following discussion includes interpretation of results, limitations, suggestions for future research, and concluding remarks according to SQUIRE guidelines for reporting mixed-methods research (Goodman et al., 2016).

Interpretation

This study employed the use of mixed methods both for increasing understanding of phenomena and for the use of triangulation of data and findings for added validity, which are noted rationale for using mixed methods research (Clark & Ivankova, 2016), particularly when addressing implementation-related issues (Aarons et al., 2011; Landsverk et al., 2012; Palinkas et al., 2011; Proctor et al., 2009). Therefore, the results are triangulated and discussed as a whole according to each of the research questions. I discuss feasibility, acceptability, and appropriateness findings, the impact of driver influence (training and coaching) and uptake of the electronic social emotional assessment, and the SIC of service areas implementing the electronic social emotional assessment. Next, I discuss the big picture of these three areas combined relative to overall outcomes and future planning and recommendations. I also discuss the overall potential for use of this triangulated mixed-methods approach in program evaluation and pilot implementation and process improvement.

Feasibility, Acceptability, Appropriateness

As mentioned throughout this work, feasibility, acceptability, and appropriateness are three of the most common focus areas in pilot studies (Leon et al., 2011). Yet, rarely are validated measures used as part of this process (Taylor et al., 2018). Therefore, this study employed the first known use of a validated feasibility, acceptability, and appropriateness measure in this setting. Results supported the hypothesis; the use of the electronic social emotional assessment in early intervention settings as evaluated was moderate to highly rated as feasible, acceptable, and appropriate. This conclusion was supported by the qualitative findings from interviews with providers and administrators

who provided many positive statements in each of these areas. However, there also were a number of statements that emerged, particularly in the area of feasibility and appropriateness, that indicated other factors impacting implementation. For example, in spite of an average rating of moderate to highly feasible on the validated self-report survey, a number of providers and administrators expressed that they did not have time to complete the electronic assessment. This attitude was mirrored in the range of the scores on the validated measures (from 2 to 5) assessing feasibility, acceptability, and appropriateness. For example, although average scores were in the moderate range, 24% of the respondents (18 of 76) felt that the social emotional measure was not an appropriate match for use in early intervention and 25% of respondents (19 of 76) felt the assessment was not easy to use. These statements aligned with providers who also noted in the interviews that they did not have access to the necessary technology (e.g., iPads or mobile devices) to complete the tool electronically while in the client's home. Instead, many providers were filling out the assessment on paper while on a home visit and then returning to the office and entering the results into a computer, essentially requiring the providers to do the assessment twice. This added time for entering combined with comments about the electronic system not feeling user friendly and requiring extra time to navigate could lead to the assessment being less feasible in actuality, particularly when providers already expressed feeling a lack of time to complete necessary job duties. This lack of efficiency in processes combined with lack of time to complete all required duties often put providers in a situation where they would feel they were choosing between using the electronic social emotional assessment and completing federally and locally mandated paperwork. Because the electronic social emotional assessment was not a

requirement, it is not surprising that many providers were prioritizing other duties over using this assessment, even though most providers noted the importance of early social emotional development. However, these difficulties with efficiency were in contrast with a provider who reported using an iPad to directly enter and complete the assessment while in the home. This provider reported only statements about efficiency and likeability of the assessment and also reported regular usage, which differed from the experience of other users. Hence, the theory generated from these findings is that adequate tools and efficient processes may be linked to increased feelings of feasibility and usage. (See "Future Research" for further discussion.) In businesses focused on technology-related products, user-centered design and user experience are a large focus. Yet, rarely is this focus noted in educational product research. Usability and user experiences contribute to use and sustainability, according to industry studies (Forlizzi & Battarbee, 2004). Therefore, the user experience should be a consideration when promoting use and scale with electronic systems in education. This untapped area of development could have significant impact on systems use of electronic and technologically based practices and interventions.

Overall, feasibility, acceptability, and appropriateness scores were high. This result was confirmed in the qualitative interviews in which providers and administrators provided statements that supported feelings of feasibility, acceptability, and appropriateness concerning the use of this assessment in Part C early intervention systems. However, in spite these scores, overall steep rates of drop-off and low flat-lined rates of usage were noted in the time series. These findings indicate that reports of feasibility, acceptability, and appropriateness should be used cautiously when these are

the only measures determining scale-up. This issue is of concern because most pilot studies focus on use of self-report measures and rarely include more objective measures of uptake and impact (Pai et al., 2017; Taylor et al., 2018). In essence, many programs are deciding to scale up based on the self-report of feasibility, acceptability, and appropriateness alone. This is not to say that these types of measures and self-report studies are not valuable; we do know the results are correlated with outcomes and provide valuable theory forming part of the foundation for understanding readiness and future use (Saldana et al., 2012). However, further conversation is needed about when and how we might consider use of the feasibility measures and consider pairing them with additional data sources, such as qualitative interviews and ITS designs, for a more robust source of information on which to base going to scale.

Impact: Driver Influence and Uptake

A key part of this study was to examine the impact of particular drivers on the uptake of usage of an early social emotional assessment in selected service areas by Part C providers. The drivers were identified and defined during the ongoing monthly partnership meetings between the researcher and core team members at the state agency. Drivers included initial training and follow-up leadership coaching on systems-level supports and were based upon efforts initiated at the state level through their SSIP process. Findings indicated a significant change in usage rates after the initial training. However, steep drop-off in usage was also noted after the initial training. This finding aligns with prior concerns raised in the literature that training alone does not result in sustainment or long-term usage of an intervention or EBP (Cook & Odom, 2013). This finding also aligns with the qualitative data indicating that even though the initial

68

guidelines provided were verbal and not written, they were very clear. The initial

directions provided in the training and/or initial messaging were reported by several

interviewees as requiring them to assess three to five children in their caseload that would

be benefitted by or appropriate for the assessment. One provider described the criteria by

saying,

> Everybody on staff was supposed to find three children or two children that we
> needed to do the assessment with, do the documentation, give the information,
> and then we kind of reconvened in a professional development or a staff meeting
> to talk about what, um, our feelings about it, how it went and, you know, what the
> pros and the cons of using the e-DECA were. —Provider 3

However, once the initial criteria were met, providers noted the assessment had to take

second place to other more formalized mandates and caseload demands. This resulted in

the large initial bump in usage numbers followed by immediate drop-off. "Our area was

using it pretty highly in the beginning. I don't know if that continues to be the case"

(Admin 1). This was confirmed by the data showing that rates dropped off to an average

of less than one assessment per month in some pilot sites in the most recent phase. This

was echoed in the qualitative data, where several providers stated they have not

completed an electronic social emotional assessment in 2019 with one provider stating, "I

have not done—I haven't even done a e-DECA this school year" (Provider 5).

In spite of these areas of potential focus, overall the participants at the pilot sites

have continued to use the assessment, with no service areas withdrawing from use as

often happens with implementation of EBPs (Bryk et al., 2015). This continued, albeit

decreasing, use may be due to additional efforts and supports of the state with ongoing

messaging related to early social emotional development, ongoing access to a specialized

coach, a Part C pilot site leader community of practice webinars, and statewide early

intervention training and technical assistance systems in place, which were mentioned by many providers in the study.

Regarding the impact of follow-up coaching, the visual analysis revealed that the follow-up leadership coaching related to systems-level supports appeared to be optimally timed at a point when usage rates had dropped to one of the lowest levels. However, the statistical analysis and/or formal visual analysis did not reveal a significant change in usage rates after the follow-up coaching.

Caution should be used in the interpretation of these findings given the small sample size and potential issues related to autocorrelation. The analysis may not have the statistical power to detect smaller effects, so ultimately the coaching may have had an impact, just a small one. However, both the formal visual analysis and the qualitative results mirrored the statistical finding that administrators, in spite of stating that they valued the expert coaching, ultimately felt they needed more regular coaching to be successful and continue to improve usage of the assessment given all the other demands and requirements in their work. One administrator stated, "My wish would have been and is, and I tell them all the time, that [expert coach] is able to provide the same level of support that she provides for infant mental health" (Admin 1). Administrators also noted that receiving coaching on implementation science theory and systems change supports would be a benefit. "It's one of the things that's kind of missing from the pilot, is that practical how do, you know, how do you implement things" (Admin 1). This issue is mirrored in recent research. Lyon et al. (2018) discussed the need for providing supports and education to leaders on the "how to" of using implementation science frameworks and approaches.

Overall, the results of the time series and qualitative follow-up on driver influence and uptake revealed that many researched-based practices were put into place in efforts to impact uptake of the EBP. However, many of the known predictors of sustainment and full-scale implementation, such as having written guidelines for usage, consistent coaching that is initiated by the coach on regular intervals versus at will by the administrator, and having a written plan (Saldana et al., 2012), were missing. These missing components were identified through use of the SIC to guide qualitative interview questions with administrators as part of the interview process. These pieces were missing in spite of implementation frameworks being foundational to all the SSIP work, indicating a need for further assessment and supports in this area. Research on implementation stages and predictors of sustainment indicate that we now know key components that predict outcomes. Therefore, greater use of tools that identify these components predictive of implementation throughout all scale-up processes is suggested. This suggestion is in alignment with recent recommendations of Raskind et al. (2018) to use implementation frameworks to guide questions and analysis in qualitative processes examining implementation-related issues. Without use of the SIC during the interviews, many components indicative of successful scale-up would not have been identified through the traditional qualitative information gathering on determinants and lived experience, as noted by examining data from the first half of the interview prior to questions related to the SIC. Overall, these findings emphasize the point made by Raskind et al. (2018) that qualitative studies in implementation-focused research should use implementation frameworks to guide question development and analysis.

Big Picture

Overall this study emphasizes the potential impact of using more objective measures and new types of analysis as part of mixed-methods research in quality improvement efforts occurring in the pilot phase of initial implementation. With new data sources emerging daily, there is increasing potential to leverage these for use in early intervention systems. If we begin to intentionally incorporate more electronic data sources in our data collection processes, the door to conducting more rigorous analyses upon which to base systems improvement and scaling decisions is opened. According to Shadish et al. (2002), to examine causal inference in nonrandomized studies it "requires more data than in non-randomized studies" (p. 160) because it ultimately allows stronger models with improved ability to test assumptions. Wearable technology, mobile devices, and electronic curricula, communications, and resources all hold potential for gathering data that would be highly valuable for program planning and improvement without adding large time or effort burdens to already stressed systems (Stegenga et al., 2018). It is time we begin thinking beyond traditional measures and data collection and innovating our processes, measures, and data in early intervention systems.

Recommendations

These results lead to recommendations related to findings for quality improvement and next steps to work toward increasing the uptake of the assessment usage. I recommend making the following changes as part of a PDSA cycle prior to moving to scale-up to determine whether implementing the following truly impacts usage rates. Specifically, I recommend considering use of the findings in this study to implement changes as part of a PDSA cycle in the pilot sites prior to scaling up to

72

determine whether the identified components contribute to increased usage rates. PDSA cycles are key to making measured improvements and increasing implementation outcomes (Fixsen et al., 2008). Specific recommendations to consider implementing as part of the PDSA cycle prior to scale-up include:

- Establish clear written guidelines with descriptive explanations for use both locally and at the state level. Fidelity and written guidelines are not only known predictors of successful implementation (Saldana et al., 2012) but also indicative of higher levels of collaboration and shared understanding (Cross, Dickmann, Newman-Gonchar, & Fagan, 2009).

- Manualize trainings so they can be replicated over time and with fidelity for future trainings with new staff and scale-up sites. Manualizing training and interventions is foundational to measuring and understanding what aspects contribute to uptake, or lack thereof, of EBPs (Center on the Developing Child, Harvard University, 2018).

- Continue to provide ongoing messaging and training about the overall importance of social emotional development and the role of early intervention providers in supporting this area of development. Although many participants stated that supporting social emotional development was important and a key role of early intervention, this feeling was not universal. Social emotional development is a key indicator in early intervention systems through the child outcome summary process and is a focus of the SSIP process new indicator (C11) with the majority of states focusing on social emotional development (GRADS360, 2018). Therefore, it is important to continue work in this area.

- Consider focusing additional training and resources on specified strategy use linked to findings of the electronic social emotional assessment. The use of a curriculum aimed toward improving social emotional development in infants and toddlers might also be considered because the use of curriculum in educational and early childhood settings is linked to higher quality programs and improved outcomes (Camilli, Vargas, Ryan, & Barnett, 2010; Frede, 1995; Macy, Bricker, & Squires, 2005). It is also recommended that trainings be manualized for further replication and future study in understanding "what works" in the training as practices continue to scale up (Center on the Developing Child, Harvard University, 2016).

- Implement training and supports related to leader use and understanding of implementation science foundational principles for evidence-based systems change. The importance and impact of the immediate program leader, such as the Part C program coordinators in this study, in supporting implementation has come to the surface in recent research (Lyon et al., 2018). However, many leaders feel they lack formal training or supports in this area, such as identified in this study. Therefore, promoting intentional usage and support of leaders specific to implementation could be an added driver for promoting successful uptake and long-term use of EBPs in early intervention. A process for examining leader impact and understanding of implementation, the Implementation Leadership Scale, was recently validated within school settings (Lyon et al., 2018).

- Increase the use of implementation frameworks and tools, such as the SIC, for understanding and guiding the implementation of EBPs in early intervention systems change efforts, such as the SSIP process. Although this implementation will require

added innovation and research in this area, such as the adaptation and validation of tools such as the SIC within early intervention settings, this study demonstrates its promise for use in identifying key areas that contribute to uptake and sustainability of the EBP.

- Consider a user-centered design study of the electronic assessment for improved efficiency, access, and ultimately a better user experience. User-centered design is known to impact usage rates and sustainability in technology industries (Zviran, Glezer, & Avni, 2006). Barriers identified related to the usability of the technological aspects of the assessment (e.g., users stating it felt "klunky" or was "not intuitive" and often felt that it took extra time due to difficulty navigating the system) are indicators that improvement in this area may link to improved use.

- Improve funding to IDEA Part C early intervention systems to allow for adequate services and necessary tools for efficiency, such as mobile devices for efficiently and accurately completely assessments. Lack of funding is a known issue in early intervention systems impacting service delivery throughout the United States (IDEA Infant and Toddler Coordinators Association, 2018). Without funds necessary to provide adequate services and supports, providers will remain in a state of scarcity, feeling they must choose between simply meeting mandates, as discussed in the results, and being able to employ truly EBPs, such as the use of the assessment in this study.

- Continue to work toward increasing and mechanizing data collection for improved measurement opportunities sensitive to detecting systems change. More data that are collected on a regular basis open the door for use of stronger designs such as the ITS

design presented here or other strong quasi-experimental designs such as regression discontinuity (Shadish et al., 2002). This approach may mean little to no added burden on providers and systems when passive data and technology are used for increased ease of data collection (Stegenga et al., 2018).

- Continue research focused on drivers impacting uptake and sustainability of EBPs to truly identify "what works, for whom, and under what conditions" (Center on the Developing Child, Harvard University, 2016). Use of factorial designs that can start to pull apart these factors, such as the Sequential Multiple Assignment Randomized Trial or the Multiphase Optimization Strategy designs for randomizing strategies as part of state scale-up efforts, may increase understanding of the mechanisms and strategies contributing to uptake and sustainability in future scale-up (Collins, Murphy, & Strecher, 2007).

Limitations

In spite of findings related to critical elements impacting usage rates, this study has limitations. In spite of a decent response rate as reported in the results, the feasibility, acceptability, and appropriateness survey data that were collected did not constitute a large enough sample to achieve optimal power due to the inherent nature of pilot studies, which have a limited pool for recruitment and data. This small sample size limits types of analyses. Although power was assumed sufficient for the ITS, a power calculator has not yet been developed for CITS designs, and this issue is still being researched and debated in the literature (Hawley et al., 2019). In addition, mild autocorrelation was noted. This is common in CITS, and in ITS designs in general, and is why triangulation of data sources (qualitative and visual analysis) was added for increased credibility and validity. Many

components along with specialized design features (such as the multiple baseline) were added to strengthen the design and power and ultimately add to the validity and credibility of interpretation of results. Another limitation is this analysis employed a more complex analysis with the ITS, use of log transformed data due to use of count data. This complicated interpretation of coefficients and the model. Because the aim of this study was to begin examining measures that may be of use in everyday systems, complex models and advanced knowledge for interpretation may ultimately impact usability. The ITS multiple baseline design could have been stronger with more time points between the interruption of interventions. For example, if the initial training had been spread at least 3 months apart for each site, more time points could have been assessed. However, given that this study was an analysis of secondary data in a real-world setting and not part of a research design, the multiple baseline design was considered an added benefit, providing strength to the design, and hence was added to the model versus merely pooling the pilot site data together. However, analysis could be optimized if this study were replicated or if practices were scaled up in the future.

In addition to the statistical limitations, the qualitative processes had some limitations. Foundationally, a rapid review (e.g., DHHS, 2017; Taylor et al., 2018) was utilized. Although this approach is commonly recommended and utilized in implementation science and mixed-methods studies (DHHS, 2017; Raskind et al., 2018), it does not provide the depth of information that a traditional qualitative study may entail (Taylor et al., 2018). However, as is noted in the implementation science literature, this lack of depth is a trade-off because in many policy and systems-related studies too much time taken in the research and analysis may render the findings no longer valid or useful

due to the mandated timelines (Taylor et al., 2018). Overall, in spite of these limitations, I feel this mixed-methods approach, which included new objective measures in addition to traditional self-report and qualitative measures, provided a more comprehensive foundation for determining scale-up and program improvement than would have been available from use of a single data source.

Future Research

This study revealed several areas in need of additional research. More research is needed on the identification of drivers, strategies, and mechanisms that contribute to the uptake of EBPs in early intervention settings. Methods in pilot research also need to be explored.

Identification of Drivers, Strategies, and Mechanisms of Uptake

Given the potential for identifying critical features necessary for implementation and scale-up with use of the SIC, future study on the adaptation and validation of this tool in early intervention systems is recommended. This tool would be the first of its kind validated for use within early intervention settings and ultimately holds potential to provide a more objective measure for identifying needs related to scale-up and implementation within early intervention systems. The identification of critical drivers linked to improved implementation outcomes may also strengthen statistical models and analyses, ultimately allowing for better prediction and understanding of the interplay of factors contributing to scale-up. In this vein, teaming has been identified in the literature as a key driver to implementation (NIRN, 2018) and is a recommended practice within early intervention and early childhood research (Division for Early Childhood, Council for Exceptional Children, 2014). Therefore, examining teaming structures and networks

in relation to dissemination of training and coaching and their relationships to implementation outcomes would be a next logical step in the research of drivers impacting uptake of EBPs. Examination of the use of a leadership-based training focused on implementation, such as the tools and frameworks developed by Lyon et al. (2018), would be beneficial given the requests in this study by leaders for further understanding of systems change and implementation science.

Methods Development in Pilot Research

Two main areas of issues related to methods development emerged in this study. A formal examination should be conducted on whether the use of the triangulated set of measures in this study would be considered feasible, acceptable, and appropriate to state leaders and whether they feel it provides additional information that adds to their ability to make decisions related to scale-up. In addition, further development and examination of the quality of information obtained through the qualitative processes described in this study using implementation frameworks combined with the traditional methodology of phenomenology would be beneficial to understanding and improving work in this area, because this is the first known application of a formalized coding process developed from implementation frameworks. A study examining the use of a rapid review process founded on theoretical implementation frameworks and guided by traditional qualitative frameworks such as phenomenology for question development and analysis would be helpful for future mixed-methods and qualitative policy implementation projects due to the short timelines and specific needs of the projects and research.

In addition to qualitative methods development, a need for methods research related to quantitative analysis emerged. Given the potential for a variety of functional

forms in each segment of the segmented regression and the unique distributions created

by real-world implementation efforts combined with seasonality impacts, techniques such

as the Generalized Additive Model (GAM) hold potential for future exploration and

applicability (D. Anderson, personal communication, April 15, 2019). GAM has many

benefits; it is still based on a formula similar to a traditional linear model and hence

demonstrates potential for higher levels of translatability and understanding compared

with other statistical techniques. However, GAM uses a function in place of the beta

predictors to more optimally represent the functional form than is possible with more

traditional linear models. GAM has been used in air quality models with some caution as

to potential inconsistencies in effect sizes related to the type of smoothing procedure used

in the modeling (Steib, Judek, & Burnette, 2003). However, GAM has yet to be explored

related to short ITS and educational and/or early childhood data. Therefore, although it

demonstrates promise, GAM should be studied with a critical eye across a variety of data

types commonly used in time series (e.g., count data, percentages, and z scores) due to

past issues (Steib et al., 2003). GAM also should be compared with other regression

techniques typically used with short ITS such as OLS, Poisson, and negative binomial

models, with development of consistent smoothing procedures to ensure accuracy of

effects estimates (Steib et al., 2003). Another beneficial examination would be whether

the use of GAM could improve model fit–related issues with functional form in the

segmented approach and specifically whether it impacts autocorrelation issues. If

improvement were noted, it would increase the number of applications in real-world

systems where data often do not follow regular structures and merely transforming the

data has not always been beneficial for creating an accurate model.

Overall, CITS has immense potential, even with its limitations, for adding increased objectivity to decision making in systems, particularly in pilot phases of initial implementation. However, triangulation of measures to ensure validity and credibility and the possible use of visual analysis pending model criteria should also be considered in short time series designs. Further research should be conducted related to the triangulation of methods and whether this process contributes to improved decision making and ultimately better outcomes in scale-up compared with use of traditional measures or approaches in quality improvement in early intervention systems. Further research also is needed on package development in open-source software platforms, such as R. Packages with built-in guidance for decision-making criteria related to model specification (e.g., prompts for model diagnostics) may make time series designs more accessible and affordable to a range of analysts and decision makers conducting and leading systems-level work. New measures that are available but not understandable and accessible from an affordability standpoint (e.g., software cost or training) will not have the most reach and impact. R is an open-source freely available platform with many free resources (e.g., books, trainings, and guidance documents).

Conclusion

Data and processes that inform decisions of when and how to scale EBPs warrant further attention and innovation. We know from years of research that there is a lack of interventions that actually make it to scale in spite of years of effort and millions of dollars spent on training and supports, leading to what is becoming known as implementation failure (Bryk et al., 2015). New sources of data are available due to rapid advances in technology with data storage, analysis, and data gathering, and this is a

perfect time to begin rethinking our methods and measures to better inform quality improvement and make beneficial and truly impactful targeted changes during the pilot phase of implementation of new practices. We need to begin exploring a triangulated approach and move beyond relying mainly on self-report measures in this phase. My study provides preliminary support for this further exploration and potential for use. If we can begin making these shifts and preempt implementation failure by basing decisions on more objective triangulated measures, we have the potential to impact even more lives. If the interventions and EBPs are not actually reaching the intended consumers, individuals are missing out on potentially life-changing interventions and supports. This issue becomes especially important when considering some of youngest and most vulnerable populations, such as the infants and toddlers with disabilities in this study. Although the window of opportunity is not closed with regard to development, we do know that it becomes more difficult to make changes as the child grows older and their brains develop specialized pathways (Center on the Developing Child, Harvard University, 2007). Therefore, the early years, when a child is receiving early intervention services, is the most opportune time for interventions and supports to improve outcomes for children and families that ultimately impact both current life quality and actual brain architecture and future outcomes, *as long as the interventions actually reach the child.* Hence, it is not just a consideration but an ethical imperative that we continue this conversation and innovation in the area of measurement during the pilot phase of initial implementation efforts. If we truly want to move the field of early intervention forward and begin changing lives and impacting outcomes for all, we need to begin with a measured and objective approach to implementation and systems change.

82

APPENDIX A

PARTNERSHIP AGREEMENT

STATE OF MICHIGAN
DEPARTMENT OF EDUCATION
LANSING

RICK SNYDER
GOVERNOR

SHEILA A. ALLES
INTERIM STATE SUPERINTENDENT

MEMORANDUM

DATE: June 29, 2018

TO: Sondra Stegenga, University of Oregon, Special Education – Early Intervention Program

FROM: Reneé DeMars-Johnson, Director
Office of Great Start/Early Childhood Development and Family Education

SUBJECT: Letter of Support for Collaboration

The Michigan Department of Education, Office of Great Start/Early Childhood Development and Family Education (MDE) fully supports collaborating with you as you examine readiness and potential determinants of successful implementation of innovative and evidence-based practices in the Part C of the Individuals with Disabilities Education Act programs, known here as *Early On®*. We look forward to using the findings of this collaborative research to inform scale up of the practices being implemented as part of the Michigan Part C State Systemic Improvement Plan (SSIP).

This collaboration will include, but is not limited to, the following:
- A commitment to collaborate through regularly scheduled meetings with MDE staff,
- An introduction to pilot sites and participation in the SSIP Implementation Committee, and the State Coordination and Evaluation Committee as needed/available, and
- The completion of a Data Sharing Agreement to ensure data is shared and protected in a form and manner that is acceptable to both parties.

We look forward to this collaboration and the opportunity to improve the implementation of early intervention practices based on the findings from this research. Please contact me if you have any questions or wish to discuss this further. I can be reached via email at DeMars-JohnsonR@michigan.gov or by telephone at 517-241-0162.

STATE BOARD OF EDUCATION

CASANDRA E. ULBRICH – CO-PRESIDENT • RICHARD ZEILE – CO-PRESIDENT
MICHELLE FECTEAU – SECRETARY • TOM MCMILLIN – TREASURER
NIKKI SNYDER – NASBE DELEGATE • PAMELA PUGH
LUPE RAMOS-MONTIGNY • EILEEN LAPPIN WEISER

608 WEST ALLEGAN STREET • P.O. BOX 30008 • LANSING, MICHIGAN 48909
www.michigan.gov/mde • 833-633-5788

APPENDIX B

ORIGINAL DATA SHARING AGREEMENT

Data Sharing Agreement

By and Among Sondra Stegenga and the Michigan Department of Education's Office of Great Start

This Data Sharing Agreement ("DSA") is an agreement entered into by Sondra Stegenga and the Michigan Department of Education's ("MDE")-Office of Great Start ("OGS") for the purpose of sharing data between the parties in a manner consistent with the Family Educational Rights and Privacy Act of 1974 ("FERPA").

PURPOSE

The MDE Office of Great Start is partnering with Sondra Stegenga to examine levels of readiness, determinants of practice (e.g., needs, barriers, or facilitators) and potential implementation strategies for scale up of the e-DECA assessment in alignment with the current State Systemic Improvement Plan (SSIP) for Part C/Early Intervention in Michigan. State and service organizations are tasked to scale up supports related to the state systemic improvement plan (SSIP) related to the new Part C indicator, C11. Yet, we know from recent research that implementing and scaling up evidence-based practices in complex systems is immensely difficult. This difficulty of implementing evidence-based practices is known as the research to practice gap and has been noted as a significant concern in most organizations including early intervention and early childhood systems (e.g. Odom, 2009). This has created a call for the use of implementation science to bridge the gap between research and practice in early childhood systems (e.g. Dunst, 2013). Yet little research exists examining implementation readiness, specific strategies, and scale up processes specific to the complex interdisciplinary and family-based systems of Part C of IDEA serving infants and toddlers with disabilities. This is particularly concerning since social emotional interventions, which are the focus of 37 of the 56 states and jurisdictions providing IDEA Part C early intervention, are known to have increased rates of attrition (e.g. Baggett et al., 2010); ultimately pointing to an even greater need to examine readiness and implementation strategies to ensure uptake. This, coupled with recent research findings that over 50% of interventions do not make it past the pre-implementation stage to full implementation and sustainability (Saldana, 2014), demonstrates an imperative need for implementation focused studies to optimize outcomes for young children and families in Part C systems.

As a true partnership, the goals, focus, and data included in this research have been developed in collaboration with the Michigan SSIP team representatives, in alignment with current SSIP timelines and work.

This DSA is for the purpose of evaluating the effectiveness of e-DECA assessment in pilot and beta Intermediate School Districts (ISD) as part of the current State Systemic Improvement Plan (SSIP). Sondra Stegenga wishes to evaluate and identify levels of readiness, determines of practice, and potential implementation strategies for scale up of the e-DECA child assessment in alignment with the current Michigan SSIP for Part C/Early Intervention. The final evaluation will be provided in a report to OGS for application and inclusion in the annual SSIP reporting for Part C.

FERPA contains important exceptions that allow sharing of confidential data without prior consent. This data request falls under the category of:

Audit and evaluation for the purpose of:
    X   Auditing or evaluating a Federal or State-supported education program; or
    ☐   Enforcing or complying with Federal legal requirements related to those programs

Studies for the purpose of:
    ☐   Developing, validating, or administering predictive tests
    ☐   Administering student aid programs
    ☐   Improving instruction

APPENDIX C

DATA SHARING AGREEMENT AMENDMENT

Amendment
Data Sharing Agreement
MDEOGS-Stegenga2018

By and Among Sondra Stegenga and Michigan's Department of Education's Office of Great Start

This Amendment is made and entered into by Sondra Stegenga and Michigan's Department of Education's ("MDE")-Office of Great Start ("OGS") and is effective on the date indicated below. This Amendment amends and supplements the Data Sharing Agreement ("DSA") between Sondra Stegenga and OGS dated July 30, 2018.

Amendments: The Data Sharing Agreement is amended as follows:

Additional Data Deliverables:

Data Element Crosswalk			
Data Requested	**Public, Confidential, or Not Public/Not Confidential Data Source**	**Years Requested**	**Data Provision**
Early On SSIP Evaluation Report and aggregate data	Not Public/Not Confidential	2017-2018 through 2018-2019	MPHI will provide the *Early On* SSIP Evaluation Report and aggregate data to MDE. MDE will then provide the report and aggregate data to Sondra Stegenga.
Early On SSIP Pilot and Beta Site Evaluation Reports and aggregate data	Not Public/Not Confidential	2013-2014 through 2018-2019	WSU will provide the SSIP Pilot and Beta Site Evaluation Reports and aggregate data to MDE. MDE will then provide the reports and aggregate data to Sondra Stegenga.

Term: Unless earlier terminated as provided herein, the term of this Amendment shall commence upon the execution hereof and shall continue until **December 31, 2019**.

Continuance: Except as expressively amended in this Amendment, the terms of the Data Sharing Agreement remain in effect unchanged. If any inconsistency exists between this Amendment and the Data Sharing Agreement, the provisions in this Amendment shall control.

APPENDIX D

MEASURES

Part C Coordinator and Provider Measures

Part C Coordinator

1.   Feasibility (FIM)
2.   Accessibility (AIM)
3.   Acceptability (IAM)
4.   Qualitative Semi-Structured Interviews
5.   Stages of Implementation Completion (SIC)

Provider

1.   Feasibility (FIM)
2.   Accessibility (AIM)
3.   Acceptability (IAM)
4.   Qualitative Semi-Structured Interviews

Acceptability of Intervention Measure (AIM), Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM)

PROVIDERS & PROGRAM COORDINATORS

GENERAL INSTRUCTIONS: These measures could be used independently or together. The IAM items could be modified to specify a referent organization, situation, or population (e.g., my clients). Please check and report the psychometric properties with each use or modification.

	Completely disagree	Disagree	Neither agree nor disagree	Agree	Completely agree
1. The e-DECA assessment tool meets my approval.	①	②	③	④	⑤
2. The e-DECA assessment tool is appealing to me.	①	②	③	④	⑤
3. I like the e-DECA assessment tool.	①	②	③	④	⑤
4. I welcome using the e-DECA assessment tool.	①	②	③	④	⑤

Intervention Appropriateness Measure (IAM)

	Completely disagree	Disagree	Neither agree nor disagree	Agree	Completely agree
1. The e-DECA assessment tool seems fitting.	①	②	③	④	⑤
2. The e-DECA assessment tool seems suitable.	①	②	③	④	⑤
3. The e-DECA assessment tool seems applicable.	①	②	③	④	⑤
4. The e-DECA assessment tool seems like a good match.	①	②	③	④	⑤

Feasibility of Intervention Measure (FIM)

	Completely disagree	Disagree	Neither agree nor disagree	Agree	Completely agree
1. The e-DECA assessment tool seems implementable.	①	②	③	④	⑤

2. The e-DECA assessment tool seems possible.	①	②	③	④	⑤
3. The e-DECA assessment tool seems doable.	①	②	③	④	⑤
4. The e-DECA assessment tool seems easy to use.	①	②	③	④	⑤

Pragmatic Qualities:

- Readability tested by substituting "This EBP" for "Insert Intervention." Flesch reading ease score (and grade level) is 95.15 (5th grade) for AIM, 99.60 (5th grade) for IAM, and 94.17 (5th grade) for FIM.
- No specialized training is needed to administer, score, or interpret the measures.
- Cut-off scores for interpretation not yet available; however, higher scores indicate greater acceptability, appropriateness, or feasibility.
- Norms not yet available.
- Scales can be created for each measure by averaging responses. Scale values range from 1 to 5. No items need to be reverse coded. Good measurement practice: assess structural validity to confirm the unidimensionality of each measure and calculate alpha coefficient to ascertain reliability.
- There is no cost to use these measures.
- Time to complete: less than 5 minutes per measure.

From: Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M.H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, *12*(1), 108.

Provider Semi-Structured Interviews
30 minutes

1. Tell me about your role in assessing and providing intervention related to early social emotional development.
2. Tell me about your role and experience in using the e-DECA.
3. What do you like best about the e-DECA?
4. What do you like least about the e-DECA?
5. What, if anything, prevents you from you using the e-DECA?
6. What would help you to use the e-DECA more?
7. What other supports do you need to be able to provide intervention and assessment related to early social emotional development?
8. Does your service area have written guidelines for when and with whom to use the e-DECA? If so, what are the guidelines?
9. If you do not have written guidelines, what is the criteria for conducting/using the e-DECA/how do providers decide when to use the e-DECA?
10. Is there anything else you would like to share either about providing services and/or assessment related to early social emotional development?

Part C Coordinator Semi-Structured Interviews and/or Focus Groups
30 minutes

1. Tell me about your role providing program services and supports related to early social emotional development (e.g. ensuring referrals, choosing assessments, etc.)
2. Tell me about your experience with the e-DECA.
3. What do you like best about the e-DECA?
4. What do you like least about the e-DECA?
5. What, if anything, prevents your providers/service area from you using the e-DECA?
6. What would help your service area to use the e-DECA more?
7. What other supports do you need for your service area to be able to provide intervention and assessment related to early social emotional development?
8. Does your service area have written guidelines for when and with whom to use the e-DECA? If so, what are the guidelines?
9. If you do not have written guidelines, what is the criteria for conducting/using the e-DECA/how do providers decide when to use the e-DECA?
10. Is there anything else you would like to share either about providing services and/or assessment related to early social emotional development?

Universal Stages of Implementation Completion (SIC)
Sample Extract Included with Permission of Author, Dr. Lisa Saldana

Stages of Implementation Completion (SIC) – Variable Descriptions

Stage 1 – Engagement	
Date site is informed/learns services/program available	• Agency or System Identifies that a Program is available for Scale-up. • Site might have proactively sought out information • Information might have been purposefully disseminated to site
Date of interest indicated	• Site reaches out to purveyor or developer group requesting initial information • Site reaches out to purveyor or developer group express a desire to assess program fit
Date agreed to consider implementation	• Site notifies purveyor or developer that they want to move forward with a potential implementation plan • Site chooses the EBP from a list of practices available to implement • Site agrees to talk with other relevant parties within the system or organization to determine if they would support adoption
Date initial cost information sent	• Site expresses high enough interest that the purveyor or developer provides information regarding the costing structure for implementation • General cost and resource information is provided to the site (not necessarily specific to site structure).

Stage 2 – Consideration of Feasibility	
Date of 1st site planning contact	• Date of first discussion to describe the implementation process and expectations in detail • Date of first discussion where implementation is outlined including negotiation to fit implementation plan within the parameters of the site's rollout
Date Stakeholder meeting #1	• Date of first meeting with leadership and key members involved in the implementation process • Meeting is most often in person, but can also occur via videoconference or teleconference • Concrete information is provided to key members of site's initiative and expectations are clearly defined • Key steps necessary to achieve positive outcomes are described
Date Feasibility Questionnaire completed	• Documentation of feasibility is sometimes recorded by the site and sometimes by the purveyor • Regardless, a dialogue occurs to address if it is feasible for site to implement the EBP using the typical implementation strategy • Concrete expectations (e.g., regarding population served, flexible scheduling, collaboration with psychiatrist) are outlined and the value of specific needs clarified.
Date liaison/Program Champion representative identified to purveyor	• Identification of the site's employee or team member responsible for taking the lead on the implementation efforts with the purveyor.

Stage 3 – Readiness Planning	
Date of cost calculator / funding plan review	• Site and Purveyor look over program cost projections • Site is provided with estimates for program costs and calculations are reviewed with purveyor specific to site
Date of staff sequence, timeline, hire plan review	• Job titles, FTE and roles are discussed for the varying program positions. • Purveyor provides a staffing timeline to make sure roles are filled in an efficient manner; e.g. therapist hired prior but close to training.
Date of recruitment review	• Reviewing recruitment of non FTE positions essential to the implementation; e.g. foster parents, skills coaches… • Might involve preparing pamphlets, advertising, attending community gatherings… • The date should be the start of this process as it will continue and evolve over the entire implementation.
Date of referral criteria review	• Establishing the source of the target population of the implementation efforts. • Might involve preparing pamphlets, advertising, establishing locations to present on the intervention… • The date should be the start of this process as it will continue and evolve over the entire implementation.

REFERENCES CITED

Aarons, G. A., Ehrhart, M. G., Farahnak, L. R., & Hurlburt, M. S. (2015). Leadership and organizational change for implementation (LOCI): A randomized mixed method pilot study of a leadership and organization development intervention for evidence-based practice implementation. *Implementation Science, 10*(1), 11.

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in child welfare. *Administration and Policy in Mental Health and Mental Health Services Research, 38,* 4–23.

Baggett, K. M., Davis, B., Feil, E. G., Sheeber, L. L., Landry, S. H., Carta, J. J., & Leve, C. (2010). Technologies for expanding the reach of evidence-based interventions: Preliminary results for promoting social-emotional development in early childhood. *Topics in Early Childhood Special Education, 29*(4), 226–238.

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. In J. Bemmel & A. T. McCray (Eds.), *Yearbook of medical informatics 2000: Patient-centered systems* (pp. 65–70). Stuttgart, Germany: Schattauer Verlagsgesellschaft.

Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a literature-driven definition of big data in healthcare. *BioMed research international*, *2015*.

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., & Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychology, 3*(1), 32.

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology, 46,* 348–355.

Bernal, J. L., Cummins, S., & Gasparrini, A. (2018). The use of controls in interrupted time series studies of public health interventions. *International Journal of Epidemiology, 47,* 2082–2093.

Bevan, M. T. (2014). A method of phenomenological interviewing. *Qualitative Health Research, 24,* 136–144.

Brown, C. H., Chamberlain, P., Saldana, L., Padgett, C., Wang, W., & Cruden, G. (2014). Evaluation of two implementation strategies in 51 child county public service systems in two states: Results of a cluster randomized head-to-head implementation trial. *Implementation Science, 9,* 134. doi:10.1186/s13012-014-0134-8

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261–304.

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.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. J. (2014). The use of triangulation in qualitative research. *Oncology Nursing Forum, 41,* 545–547.

Center on the Developing Child, Harvard University. (2007). *In brief, the science of early childhood development*. Retrieved from https://46y5eh11fhgw3ve3ytpwxt9r-wpengine.netdna-ssl.com/wp-content/uploads/2007/03/InBrief-The-Science-of-Early-Childhood-Development2.pdf

Center on the Developing Child, Harvard University. (2016). *From best practices to breakthrough impacts: A science-based approach to building a more promising future for young children and families*. Retrieved from https://developingchild.harvard.edu/resources/from-best-practices-to-breakthrough-impacts/

Center on the Developing Child, Harvard University. (2018). *Our innovation approach: IDEAS Impact Framework.* Retrieved from https://developingchild.harvard.edu/innovation-application/innovation-approach/

Chamberlain, P., Brown, C. H., & Saldana, L. (2011). Observational measure of implementation progress in community based settings: The Stages of Implementation Completion (SIC). *Implementation Science, 6,* 116. doi:10.1186/1748-5908-6-116

Clinton-McHarg, T., Yoong, S. L., Tzelepis, F., Regan, T., Fielding, A., Skelton, E., ... Wolfenden, L. (2016). Psychometric properties of implementation measures for public health and community settings and mapping of constructs against the Consolidated Framework for Implementation Research: A systematic review. *Implementation Science, 11,* 148. doi:10.1186/s13012-016-0512-5

Collins, L. M., Murphy, S. A., & Strecher, V. (2007). The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): New methods for more potent eHealth interventions. *American Journal of Preventive Medicine, 32*(5), S112–S118.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1-12.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*(2), 135–144.

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Thousand Oaks, CA: Sage.

Cross, J., Dickmann, E., Newman-Gonchar, R., & Fagan, J. (2009). Using mixed-method design and network analysis to measure development of interagency collaboration. *American Journal of Evaluation, 30*(3), 310–329.

Department of Education, Office of Special Education Programs, GRADS360° (2018). *2018 Part C FFY 2016 SPP/APR indicator analysis booklet.* Retrieved from https://osep.grads360.org/#communities/pdc/documents/17332

Department of Health and Human Services, National Institutes of Health. (2017). *Qualitative methods in implementation science.* Retrieved from https://cancercontrol.cancer.gov/IS/docs/NCI-DCCPS-ImplementationScience-WhitePaper.pdf

Division for Early Childhood, Council for Exceptional Children. (2014). DEC recommended practices in early intervention/early childhood special education 2014. Retrieved from https://www.dec-sped.org/dec-recommended-practices

Dunst, C. J., Trivette, C. M., & Raab, M. (2013). An implementation science framework for conceptualizing and operationalizing fidelity in early childhood intervention studies. *Journal of Early Intervention, 35*(2), 85–101.

Early Childhood Technical Assistance Center. (2018, October). *SSIP Phase II: Tools and resources.* Retrieved from http://ectacenter.org/topics/ssip/ssip_phase2_resources.asp

Fisher, P. (2018, October). *Translating scientific knowledge regarding healthy development under conditions of adversity.* Keynote presented at the Zero to Three Annual International Conference, Denver, CO.

Fixsen, D., Panzano, P., Naoom, S., & Blasé, K. (2008). *Measures of implementation components of the National Implementation Research Network frameworks.* Chapel Hill, NC: Authors.

Forlizzi, J., & Battarbee, K. (2004, August). Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 261–268). New York, NY: ACM.

Frede, E. C. (1995). The role of program quality in producing early childhood program benefits. *The Future of Children, 5*(3), 115–132.

Gebski, V., Ellingson, K., Edwards, J., Jernigan, J., & Kleinbaum, D. (2012). Modelling interrupted time series to evaluate prevention and control of infection in healthcare. *Epidemiology & Infection, 140,* 2131–2141.

Goodman, D., Ogrinc, G., Davies, L., Baker, G. R., Barnsteiner, J., Foster, T. C., ... Thor, J. (2016). Explanation and elaboration of the SQUIRE (Standards for Quality Improvement Reporting Excellence) Guidelines, v. 2.0: Examples of SQUIRE elements in the healthcare improvement literature. *BMJ Quality and Safety* 25. doi:10.1136/bmjqs-2015-004480

Granger, C. W., & Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics, 2,* 111–120.

Grant, R. (2005). State strategies to contain costs in the early intervention program: Policy and evidence. *Topics in Early Childhood Special Education, 25*(4), 243–250.

Greene, J., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. *Evaluation Review, 9,* 523–545.

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short comparative interrupted time series using aggregate school-level data in education research. *Educational Researcher.* doi:10.3102/0013189X18769302

Halle, T., Metz, A., & Martinez-Beck, I. (Eds.). (2013). *Applying implementation science in early childhood programs and systems*. Baltimore, MD: Brookes.

Hawley, S., Ali, M. S., Berencsi, K., Judge, A., & Prieto-Alhambra, D. (2019). Sample size and power considerations for ordinary least squares interrupted time series analysis: A simulation study. *Clinical Epidemiology, 11,* 197–205.

Hemmeter, M. L., & Conroy, M. A. (2018). Advancement of evidence-based programs for young children with social and emotional learning difficulties. *School Mental Health, 10,* 199–201.

IDEA Infant and Toddler Coordinators Association. (2018). *State challenges report.* Retrieved from http://www.ideainfanttoddler.org/pdf/2018-ITCA-State-Challenges-Report.pdf

Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution, 6,* 828–835.

Jago, R., & Sebire, S. (2012). Publishing pilot and feasibility evaluations of behavioural interventions: Implications for preventive medicine. *Preventative Medicine, 55,* 548–549.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Landsverk, J., Brown, H., Chamberlain, P., Palinkas, L. A., & Horwitz, S. M. (2012). Design and analysis in dissemination and implementation research. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Translating science to practice* (pp. 225–260). New York, NY: Oxford University Press.

Lee, J. H., Han, G., Fulp, W. J., & Giuliano, A. R. (2012). Analysis of overdispersed count data: Application to the Human Papillomavirus Infection in Men (HIM) study. *Epidemiology & Infection, 140,* 1087–1094.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research, 45,* 626–629.

Lyon, A. R., Cook, C. R., Brown, E. C., Locke, J., Davis, C., Ehrhart, M., Aarons, G. (2018). Assessing organizational implementation context in the education sector: Confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. *Implementation Science, 13.* doi:10.1186/s13012-017-0705-6

Macy, M. G., Bricker, D. D., & Squires, J. K. (2005). Validity and reliability of a curriculum-based assessment approach to determine eligibility for Part C services. *Journal of Early Intervention, 28,* 1–16.

McAfee, A., & Brynjolfsson, E. (2012, October). Big data: The management revolution. *Harvard Business Review.* Retrieved from http://hbr.org/2012/10/big-data-themanagement-revolution/ar

Mead, S., & Mitchel, A. L. (2016). *Moneyball for Head Start: Using data, evidence, and evaluation to improve outcomes for children and families.* Sudbury, MA: Bellwether Education Partners.

Metz, A., Bartley, L., Ball, H., Wilson, D., Naoom, S., Redmond, P. (2015). Active implementation frameworks for successful service delivery: Catawba county child wellbeing project. *Research on Social Work Practice, 25*(4), 415-422.

National Implementation Research Network, Active Implementation Hub. (2018, October). *Module 1.* Retrieved from: https://implementation.fpg.unc.edu/module-1

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46,* 532–566.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41,* 673–690.

Ogden, T., Bjørnebekk, G., Kjøbli, J., Patras, J., Christiansen, T., Taraldsen, K., & Tollefsen, N. (2012). Measurement of implementation components ten years after a nationwide introduction of empirically supported programs—a pilot study. *Implementation Science, 7,* 49.

Olivier, J., Wang, J. J., Walter, S., & Grzebieta, R. (2014). Anti-helmet arguments: lies, damned lies and flawed statistics. *Journal of the Australasian College of Road Safety*, *25*(4), 10.

Onwuegbuzie, A. J., & Leech, N. L. (2007). Validity and qualitative research: An oxymoron? *Quality & Quantity, 41,* 233–249.

Pai, N. P., Chiavegatti, T., Vijh, R., Karatzas, N., Daher, J., Smallwood, M., ... Engel, N. (2017). Measures and metrics for feasibility of proof-of-concept studies with human immunodeficiency virus rapid point-of-care technologies: The evidence and the framework. *Point of Care, 16*(4), 141–150.

Palinkas, L. A., Aarons, G. A., Horwitz, S., Chamberlain, P., Hurlburt, M., & Landsverk, J. (2011). Mixed method designs in implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 38,* 44–53.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42,* 533–544.

Paschall, K., & Epstein, D. (2018, October). Using data to improve the well-being of infants and toddlers and their families [Blog]. Retrieved from https://www.naco.org/blog/using-data-improve-well-being-infants-toddlers-and-their-families

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research, 96*(1), 3–14.

Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: What it is and how to do it. *BMJ, 347*. https://doi.org/10.1136/bmj.f6753.

Plano Clark, V., & Ivankova, N. (2016). Mixed Method Research: A Los Angeles.

Powell, B. J., Beidas, R. S., Lewis, C. C., Aarons, G. A., McMillen, J. C., Proctor, E. K., & Mandell, D. S. (2017). Methods to improve the selection and tailoring of implementation strategies. *Journal of Behavioral Health Services & Research, 44*(2), 177–194.

Powell, B. J., Waltz, T. J., Chinman, M. J., Damschroder, L. J., Smith, J. L., Matthieu, M. M., ... Kirchner, J. E. (2015). A refined compilation of implementation strategies: Results from the Expert Recommendations for Implementing Change (ERIC) project. *Implementation Science, 10,* 21.

Proctor, E. K., Landsverk, J., Aarons, G., Chambers, D., Glisson, C., & Mittman, C. (2009). Implementation research in mental health services: An emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research, 36,* 24–34.

Proctor, E. K., Powell, B. J., & McMillen, J. C. (2013). Implementation strategies: Recommendations for specifying and reporting. *Implementation Science, 8,* 139.

R Core Team (2017). R: A language and environment for statistical computing. Retrieved from: https://www.r-project.org/

Raskind, I. G., Shelton, R. C., Comeau, D. L., Cooper, H. L., Griffith, D. M., & Kegler, M. C. (2018). A review of qualitative data analysis practices in health education and health behavior research. *Health Education & Behavior, 46,* 32–39.

Rosenburg, S., Elbaum, B., Rosenburg, C., Kellar-Guenthar, Y., McManus, B. (2017). From flawed design to misleading information: US Department of Education's early intervention childhood outcomes evaluation. *American Journal of Evaluation, 39*(3), 350-363.

Rottman, B. M. (2016). Searching for the best cause: Roles of mechanism beliefs, autocorrelation, and exploitation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1233–1256.

Saldana, L. (2014). The stages of implementation completion for evidence-based practice: Protocol for a mixed methods study. *Implementation Science, 9,* 43.

Saldana, L., Chamberlain, P., Wang, W., & Brown, C. H. (2012). Predicting program start-up using the Stages of Implementation measure. *Administration and Policy in Mental Health and Mental Health Services Research, 39,* 419–425.

Schindler, H. S., Fisher, P. A., & Shonkoff, J. P. (2017). From innovation to impact at scale: Lessons learned from a cluster of research–community partnerships. *Child Development, 88,* 1435–1446.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Shanahan, T. (2018, March 22). Are you paying bots to take your online survey? [blog post]. Retrieved from https://www.forsmarshgroup.com/knowledge/news-blog/posts/2018/march/are-you-paying-bots-to-take-your-online-survey/

Smith, J. D., & Polaha, J. (2017). Using implementation science to guide the integration of evidence-based family interventions into primary care. *Families, Systems, & Health, 35,* 125–135.

Smith, M. E. (2002). Success rates for different types of organizational change. *Performance Improvement, 41,* 26–33.

Squires, J., Bricker, D., Heo, K., & Twombly, E. (2001). Identification of social-emotional problems in young children using a parent-completed screening measure. *Early Childhood Research Quarterly, 16,* 405–419.

St. Clair, T., Hallberg, K., & Cook, T. D. (2016). The validity and precision of the comparative interrupted time-series design: Three within-study comparisons. *Journal of Educational and Behavioral Statistics, 41,* 269–299.

Stegenga, S. M., Munger, K., Squires, J., & Anderson, D. (2018, September 14). A mixed methods systematic scoping review of the use of big data in early intervention research: Ethical and practical implications. *OSF Preprints.* https://doi.org/10.31219/osf.io/5jg2p

Stieb, D. M., Judek, S., & Burnett, R. T. (2003). Meta-analysis of time-series studies of air pollution and mortality: update in relation to the use of generalized additive models. *Journal of the Air & Waste Management Association*, *53*(3), 258-261.

Taylor, B., Henshall, C., Kenyon, S., Litchfield, I., & Greenfield, S. (2018). Can rapid approaches to qualitative analysis deliver timely, valid findings to clinical leaders? A mixed methods study comparing rapid and thematic analysis. *BMJ Open, 8*(10), e019993.

Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic review of the application of the plan–do–study–act method to improve quality in healthcare. *BMJ Quality and Safety, 23*(4), 290–298.

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The Oxford handbook of computational and mathematical psychology* (chap. 14). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199957996.013.14

Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology, 88,* 2766–2772.

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27,* 299–309.

Wallerstein, N. B., & Duran, B. (2006). Using community-based participatory research to address health disparities. *Health Promotion Practice, 7,* 312–323.

Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science, 12,* 108.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag: New York.

Zviran, M., Glezer, C., & Avni, I. (2006). User satisfaction from commercial Web sites: The effect of design and use. *Information & Management, 43,* 157–178.