



# Implementation Drivers as Practical Measures of Data-Driven Decision-Making: An Initial Validation Study in Early Childhood Programs

Jared L. Barton<sup>1</sup> · Becci A. Akin<sup>1</sup>

Received: 3 January 2022 / Accepted: 18 April 2022 / Published online: 12 May 2022  
© The Author(s) 2022

## Abstract

Leveraging data to demonstrate program effectiveness, inform decision making, and support program implementation is an ongoing need for social and human service organizations, and is especially true in early childhood service settings. Unfortunately, early childhood service organizations often lack capacity and processes for harnessing data to these ends. While existing literature suggests the *Active Implementation Drivers Framework (AIF Drivers)* provides a theoretical basis for data-driven decision-making (DDDM), there are no practical applications or measurement tools which support an understanding of readiness or capacity for DDDM in early childhood settings. This study sought to address this gap through the development and initial validation of the *Data-Driven Decision-Making Questionnaire (DDDM-Q)* based on the nine core factors in the AIF Drivers. The study piloted the 54-item questionnaire with 173 early childhood program administrators. Findings from this study suggest using the AIF Drivers as a theoretical basis for examining DDDM supports three of five categories of validity evidence proposed by Goodwin (2002), including (1) evidence based on test content, (2) evidence based on internal structure, and (3) evidence based on relationships to other variables. This study may inform future research seeking to develop theoretically based instruments, particularly as it pertains to expanding use of the AIF Drivers. Practice-wise, the study findings could enhance and complement early childhood programs as well as other social and human service implementations by presenting the DDDM-Q as a platform for understanding organizational readiness for DDDM and identifying strengths as well as areas for improvement.

**Keywords** Implementation drivers · Data-driven decision-making · Active Implementation Frameworks · Instrument development · Validation study

## Background and Objectives

As social and human service providers across the globe amass large quantities of data, calls to use data to support decision making and implementation are growing. While challenges associated with data use are relevant across all human service sectors (Gillingham, 2019), early childhood programs are experiencing dramatically increasing demands to uptake research-based decisions, implement evidence-informed practices, and continuously use data to support decision making. These expectations are clear and

pervasive. Those engaged in these conversations cross multiple practice and policy levels from direct service providers and organizational administrators working to demonstrate program efficacy and impact (Snow & Van Hemel, 2008; Stein et al., 2013; Yazejian & Bryant, 2013) to local jurisdictions investing in early childhood integrated data systems (ECIDS) (Coffey et al., 2020) and national and global efforts expanding evidence-based policy initiatives (Haskins, 2018; Haskins & Margolis, 2014; Yoshikawa & Kabay, 2015). A notable example of this in the United States is the Maternal, Infant, and Early Childhood Home Visiting (MIECHV) program. MIECHV is a federal initiative that expanded state implementations of evidence-based home visiting and requires the collection of key benchmark data to demonstrate program performance and accountability (Barton, 2016; Haskins, 2018). Taken together, it is perhaps not surprising that data and its use in decision making are essential

✉ Jared L. Barton  
jaredlee@ku.edu

<sup>1</sup> University of Kansas School of Social Welfare, 1545 Lilac Lane, Lawrence, KS 66045, USA

for early childhood program accountability and continuous quality improvement (Sirinides & Coffey, 2018; Yazejian & Bryant, 2013; Zweig et al., 2015).

Data-driven decision-making (DDDM) as an intentional process has been identified as a hallmark of successful implementation and improved outcomes (Fixsen et al., 2009). Despite widespread technological advancements leading to vast quantities of data, human service organizations and public agencies can be slow to incorporate data to inform decision making (Coulton et al., 2015; Sirinides & Coffey, 2018). Two major obstacles and gaps may inhibit the use of data to support decision making for early childhood programs. First, a limited amount of research exists to provide guidance on best practices for DDDM (Yazejian & Bryant, 2013). Yazejian and Bryant point out that without this guidance, early childhood programs may not understand what to do with data they do collect or may be overly focused on individual-level data elements, failing to address other contexts and systems impacting children and families (Yazejian & Bryant, 2013). Second, early childhood programs experience limited capacity to utilize and learn from data (Mandinach, 2012; Sirinides & Coffey, 2018; Yazejian & Bryant, 2013). As technical capacity for collecting and organizing data has increased, so too has the need for data literacy capacities that enable organizations to turn data into actionable strategies to inform practice with individual children and families as well as support administrative decisions (Little et al., 2019; Mandinach, 2012; Sirinides & Coffey, 2018). At an operative level, early childhood leaders have reported difficulty engaging in reflective practices to support decision making, leading researchers to suggest the need for future work that identifies alternative pathways to support effective decision making (Aubrey et al., 2012). A specific example of this challenge for early childhood programs is the data needed to fully understand families' and children's experiences and how those experiences influence outcomes are often housed in multiple data systems (Zweig et al., 2015). Efforts to link these distinct data systems require human and financial resources dedicated to data preparation, analysis, and evaluation that many early childhood programs lack.

### Active Implementation Drivers

While extant literature contains little evidence about what factors may lead to readiness for DDDM in early childhood programs, existing literature supports the usefulness of the Active Implementation Drivers Framework (referred to henceforth as *AIF Drivers*) as a promising guiding framework for early child programs (Metz & Bartley, 2012). *AIF Drivers* include nine critical components and structures commonly found in successful practice and program implementations (Fixsen et al., 2005, 2009, 2019). *AIF Drivers*

interact in a way that allows for weaknesses in one driver to be compensated by strengths of another (Fixsen et al., 2005, 2009), and when integrated, drivers will support the uptake of practice innovations as well as fidelity and sustainability of programs (Metz & Bartley, 2012).

The nine *AIF Drivers* fall into three categories: competency drivers, organization drivers, and leadership drivers (Fixsen et al., 2009). *Competency Drivers* are processes that develop, improve, and sustain the capacity of program staff to use innovations and implement programs (Fixsen et al., 2015; Metz & Albers, 2014; NIRN, n.d.). The four competency drivers include staff selection, training, coaching, and performance assessment. *Organization Drivers* are key supports organizations put in place to create a welcoming environment for program staff to implement programs and use innovations. These drivers provide backbone support for the accessibility of effective competency drivers and availability for data to support improvement (Metz & Bartley, 2012; NIRN, n.d.). The three organization drivers include facilitative administration, systems intervention, and decision support data systems. *Leadership Drivers* involve targeting and developing appropriate strategies to maintain and support program implementation and innovation as challenges emerge (Fixsen et al., 2015; NIRN, n.d.). The two drivers related to leadership include technical leadership and adaptive leadership, both of which are required for successful implementation. Technical leadership guides implementations when greater certainty exists around presenting implementation challenges and corresponding solutions (Bertram et al., 2015). As such, technical challenges respond well to more established methods and action plans known to produce the desired outcome such as staff selection and performance assessment (Bertram et al., 2015; Fixsen et al., 2015). Adaptive leadership, on the other hand, is necessary when less is known about the presenting challenges and their corresponding solutions (Bertram et al., 2015). Adaptive leaders are often drivers and champions early on in change efforts and are needed as guides for coaching, facilitative administration, and systems-level interventions (Bertram et al., 2015; Fixsen et al., 2015).

### Implementation Drivers and Data Use

The importance of DDDM is prevalent across the *AIF Drivers*. Fixsen and colleagues (2015) stressed the importance of reliable and valid measures of *AIF Drivers* and assessed best practices for these measures including specific recommendations related to data use across implementation drivers. Using data across all competency drivers ensures ongoing improvement in practices and processes involving staff selection, training, coaching, and performance assessment. As an example, Fixsen et al. (2015) recommend collecting knowledge and skill-based pre- and post-test results during

staff trainings on delivery and implementation of evidence-based interventions or innovations. Results from these training pre- and post-tests provide decision support for other competency drivers in the forms of ‘feedback’ and ‘feed-forward’ mechanisms. As a feedback mechanism, organizations can use these data to inform how well they assess the potential of new staff recruits throughout the selection and hiring process. Training data should also be used as a feed-forward mechanism guiding coaches and supervisors in their continuous efforts to develop staff (Fixsen et al., 2015). Performance assessment data can support an organization’s understanding of outcome achievement and offer insight into the long-term impact of the organization’s staff selection, training, and coaching processes (Fixsen et al., 2015).

Organization drivers provide staff at practice, administrative, and leadership levels the necessary structural supports to use data as a decision support (Fixsen et al., 2015). A critical driver includes an accessible and comprehensive decision support data system that includes data and measurement of relevant short- and long-term outcomes as well as data for assessing performance and fidelity (Kaye et al., 2012). Decision support data systems need to be robust to facilitate data usage throughout multiple levels of an organization’s hierarchy and across competency and leadership drivers. Specifically, a decision support data system should be robust enough to support staff selection, training, and coaching of direct service providers in addition to informing leadership decisions related to program and policy (Kaye et al., 2012). Even with a robust decision support data system, additional organization drivers are necessary to ensure data are used effectively to support DDDM. Facilitative administrators, for instance, may serve as key drivers of data use internally within organizations, modeling how to use data to identify improvement opportunities, champion outcomes, and support practitioners (Fixsen et al., 2015). The systems intervention driver supports sustainable contexts for organizations to deliver programs over time (Fixsen et al., 2015). Organizations must establish accountability and credibility for when external financial and political forces may threaten their sustainability. Organizations that use data effectively to support decision making, improve their work, and demonstrate their impact may be better equipped to handle such threats (Coulton et al., 2015; Stoesz, 2014).

While competency and organization drivers focus on process, structural, and contextual factors needed for DDDM, leadership drivers identify key leaders who influence the uptake of new practice approaches, innovations, and programs (Fixsen et al., 2015). Leaders often directly engage in processes related to competency drivers including staff selection, training, and performance assessment as well as decisions about organization drivers regarding decision support data systems, internal facilitative administration, and how the organization presents itself to external systems.

DDDM for technical leadership includes data to support policy, procedure, and staffing decisions. Meanwhile, DDDM for adaptive leadership involves seeking multiple sources of data to guide decision making, understand organizational effectiveness, and establish a culture that values open lines of communication with practitioners (Fixsen et al., 2015). Taken together, organizational leaders may strongly influence how data are used to support decision making.

The AIF Drivers surfaced in two empirical studies for their usefulness for encouraging the use of data to support child welfare practice as well. In the first study, organizational supports (e.g., supervision, facilitative administration, and decision support data systems) influenced front-line child welfare worker data use and skills more than the influence of their own perceptions of their skills (Collins-Camargo et al., 2011). In a second study, Collins-Camargo and Garstka (2014) found child welfare agencies that installed outcomes-oriented team structures and reinforced outcome achievement through routine supervision and coaching were more likely to use data to support decision making and implement evidence-informed practices. While these studies provide encouraging insights on the influence of some implementation drivers for DDDM in child welfare agencies, further research is needed to build more knowledge of DDDM and understand what factors contribute to DDDM in other contexts, including early childhood programs.

## Current Study

The current study aimed to develop and test a pragmatic measure of DDDM that would support the effective implementation of evidence-informed programs in social and human service organizations. We used the AIF Drivers as the main conceptual framework for developing the measure by considering items for each of the nine drivers. The study involved piloting and assessing initial validation of the *Data-Driven Decision-Making Questionnaire (DDDM-Q)* to establish whether the instrument measured what it intended to measure with a sample of administrators involved in implementing evidence-based early childhood programs. The framework used to evaluate validity for this study was based on Goodwin’s (2002) recommendation to consider the accumulation of evidence across five categories: (1) evidence based on test content, (2) evidence based on response processes, (3) evidence based on internal structure, (4) evidence based on relationships to other variables, and (5) evidence based on consequences of testing. Goodwin’s (2002) guidance emphasizes a movement toward a unitary conceptualization of validity based on these five evidence categories rather than a focus on different types of validity (i.e., content, criterion, and construct). Presented in this article are the results of this initial validation.

## Methods

### Participants

This study collected data from 173 early childhood program administrators across six states in the central region of the United States. For the purposes of this study, early childhood programs included those serving pregnant women and families with children birth to five years old on outcomes related to maternal and child health, child development, parental support, and early childhood education and school readiness. To identify individuals in the target population, we contacted officials across six state agencies who oversaw public funding for early childhood initiatives in their respective states. We requested permission to contact program administrators as well as lists of program administrator contacts in each state. After verifying the active work status and email account of each contact, 545 early childhood administrators were sent email invitations to complete the online questionnaire. Prior to data collection, the study received approval from the university's institutional review board (IRB). Participation was voluntary and anonymous, and each participant received a statement of informed consent before completing the questionnaire. The final response rate was 32% with 173 program administrators completing the DDDM-Q.

### Data Collection Procedure

Program administrators received an email invitation that provided information regarding the study and a URL link to complete the DDDM-Q via a secure, web-based survey platform called Research Electronic Data Capture (REDCap). REDCap is designed specifically to support rapid development of data collection solutions for clinical and translational research (Harris et al., 2009). The researcher provided weekly follow-up email reminders to program administrators for two additional weeks after the initial email invitation. Upon accessing the URL, respondents reviewed the statement of informed consent before completing the DDDM-Q where respondents were asked to respond to each statement. No identifying information was tracked in this system and responses were maintained on a secure university server.

### Measures

#### Data-Driven Decision-Making Questionnaire

Development of the DDDM-Q occurred iteratively over four months following recommendations for scale construction from Carpenter (2018) and Hinkin et al. (1997).

We approached initial item development deductively with each item written corresponding intentionally to data use and DDDM within one of the nine AIF Drivers (see Hinkin et al., 1997). Each item took the form of a statement with a five-point Likert-style response scale (1 = *Strongly disagree*, 2 = *Disagree*, 3 = *Neither agree nor disagree*, 4 = *Agree*, and 5 = *Strongly agree*). Upon completion of this initial assessment, we engaged a panel of seven experts to review the instrument, provide feedback, and assess the conceptual consistency of the questionnaire's content and adequacy. The seven panelists included two university faculty members familiar with the methodological design of the study as well as Active Implementation Frameworks, two applied researchers with knowledge in early childhood program research and instrument development, two state-level leaders of statewide early childhood programs, and one early childhood program administrator in charge of overseeing organizational data and performance operations.

We refined the DDDM-Q over the course of three iterations based on panelists' feedback. Based on feedback received during the first iteration, we added more detailed instructions for respondents and provided definitions for AIF Driver concepts that panelists felt may not be commonly known by respondents (e.g., definitions for the terms "coach" and "coaching"). During the second iteration, we launched an electronic version of the questionnaire on REDCap (Harris et al., 2019). Panelists accessed the revised instrument through REDCap and provided additional feedback regarding the clarity of revised items and if items would be clear to respondent as well as their suggestions for removal or addition of items and the time it took to complete the questionnaire online. From this feedback, we removed seven items that panelists deemed redundant or unrelated to AIF Drivers. At their suggestion, we added two items related to DDDM and AIF Drivers as well as items to collect demographic and program characteristics about study respondents. With the third and final iteration, we compiled the final instrument and shared it with the two faculty member panelists to ensure coherent and complete integration of panelists' feedback. Lastly, we updated REDCap with the final version of the DDDM-Q. The final questionnaire contained 54 items with each item matched to one of the AIF Drivers to form nine distinct subscales. Supplement 1 contains all 54 items within each of the nine corresponding subscales.

#### Participant Demographics and Program Characteristics

In addition to responding to questions about DDDM, participants were also asked to respond to 15 items on their demographics and program characteristics. Demographic items included age in years, race (American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, or Other), ethnicity



(Hispanic or Latino, Not Hispanic or Latino), highest education level (Less than high school education, High school graduate or Graduate Equivalency Degree [GED], Some college, Bachelor's degree, Master's degree or higher), and number of years worked in early childhood work programming. Program characteristics included the type of early childhood program (Preschool or other center-based, Home visiting or other home-based, Maternal-child health, Parenting groups, or Other), number of families served by a program in a given year, number of staff employed in the program, family populations targeted (Low income families, Parents under 21 years old, Households with a history of substance abuse, Households with a history of child abuse or neglect, Non-English speaking families, or Other), and the program's target outcomes (Prevention of child abuse and neglect, Child development, School readiness and kindergarten readiness, Maternal health, Infant and child health, Family economic self-sufficiency).

### Analytic Approach

The study utilized a three-step analytic approach. First, descriptive analysis of each item on the DDDM-Q was used to determine how early childhood program administrators rate their organization's readiness for DDDM. Descriptive analyses including frequencies, means, and standard deviations for each item were observed and analyzed using IBM-SPSS version 25 (IBM Corporation, 2017).

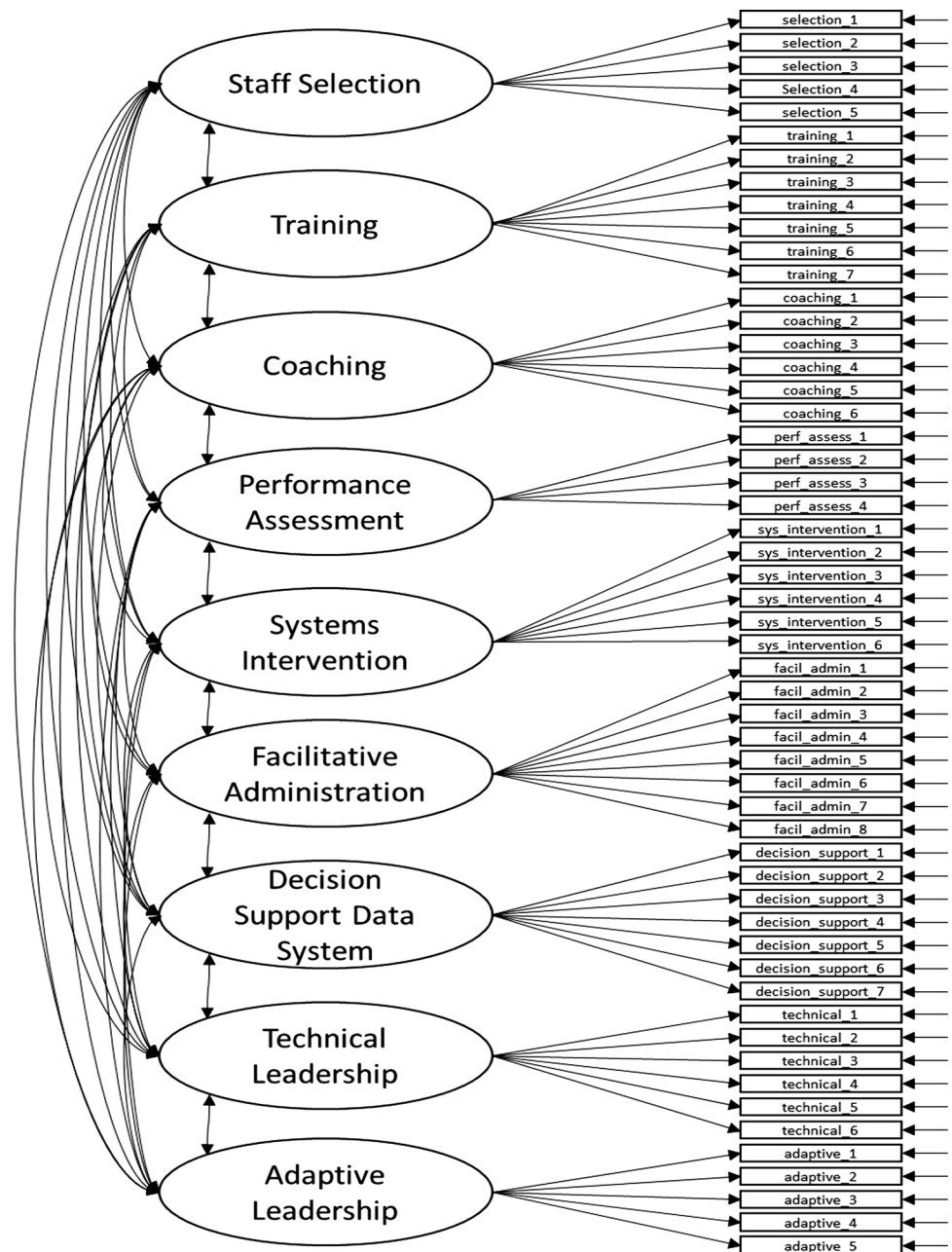
Second, a confirmatory factor analysis (CFA) was performed to evaluate the goodness-of-fit of nine AIF Drivers as an underlying factor structure for the DDDM-Q. The 9-factor model included each of the AIF Drivers (i.e., selection, training, coaching, performance assessment, systems intervention, facilitative administration, decision support data systems, adaptive leadership, and technical leadership) as an individual factor. A path-based depiction of the 9-factor model is shown in Fig. 1. Four goodness-of-fit indices suggested by Brown (2015) and Lewis (2017) were used to evaluate the model's goodness-of-fit, including the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and Chi Square fitness indices as well as statistical significance ( $p$  values). Guiding the evaluation of goodness-of-fit in this study is Schreiber et al.'s (2006) recommended cut-offs for accepting model fitness (i.e.,  $CFI \geq .95$ ;  $TLI \geq .96$ ;  $RMSEA < .06$  with confidence intervals). The CFA was performed in R Version 3.5.3 (R Core Team, 2019) utilizing the lavaan package to test the model's fit (Rosseel, 2012). Furthermore, because item responses in the DDDM-Q are ordinal (i.e., Likert-style), the CFA utilized a robust weighted least squares estimator (WLMSMV) as recommended by Brown (2015). An analysis of missing data using Little's test in R determined 26 missing patterns to be missing

completely at random (MCAR),  $\chi^2(813) = 709.81$ ,  $p = .996$ . As such, pairwise deletion functions within R's lavaan package were applied (Brown, 2015; Peugh & Enders, 2004; Rosseel, 2012). Finally, Cronbach's alphas ( $\alpha$ ) were calculated in SPSS version 25 (IBM Corporation, 2017) to measure internal consistency of the 9-factor model.

The third analytic approach included a bivariate analysis of 72 potential relationships between eight demographic and program characteristic variables and the scale scores of each of the nine factors of the DDDM-Q. The nine score variables were computed variables that summed items corresponding to the nine driver-level factors examined. The eight covariates included three categorical demographic variables: race, (0 = white, 1 = non-white), ethnicity (0 = non-Hispanic/non-Latino, 1 = Hispanic/Latino), and education (0 = Less than 4-year college degree, 1 = Bachelor's degree, 2 = Master's degree or higher). Demographic covariates also included two continuous variables: age in years and early childhood work experience in years. The three categorical program characteristic variables included number of families served annually (recoded as 1 = 60 or fewer families served, 2 = 61 to 199 number of families served, 3 = 200 or more families served), number of staff in program (recoded as 1 = 5 or fewer staff, 2 = 6 to 14 staff, 3 = 15 or more staff), and program type (recoded as 0 = Other program, 1 = Home visiting program).

Bivariate analyses were conducted within IBM-SPSS version 25 (IBM Corporation, 2017). Normality tests confirmed the normal distribution of variables so appropriate parametric tests were used based on the categorical or numerical status of the covariates (Norman, 2010; Rubin & Babbie, 1993). One-way analysis of variance (ANOVA) tests were performed for relationships between categorical covariates and the scale scores. One relationship tested—program type and the performance assessment score—required a robust non-parametric test because it violated the Levene's test for the assumption of homogeneity of variances. Accordingly, a Welch-ANOVA test was used to test this relationship (Moder, 2010). For relationships between scores and continuous covariates, Pearson correlation coefficients ( $r$ ) with two-tailed significance tests were computed. Given the MCAR assumption, missing data for bivariate analyses were handled on a case-by-case basis using pairwise deletion functions in SPSS (IBM Corporation, 2017; Peugh & Enders, 2004). Because we tested a total of 72 potential relationships, statistical significance of each relationship was evaluated based on an adjusted alpha level using the Bonferroni correction method (McDonald, 2014). As such we tested significance of each relationship with an alpha criterion of .0006. Effect sizes for the Welch-ANOVA test were determined based on the omega-squared value ( $\omega^2$ ) (Skidmore & Thompson, 2013). Effect sizes for all other categorical variables were determined based on the eta-squared values ( $\eta^2$ ),

Fig. 1 Nine-factor path model



while effect sizes for continuous variables were determined based on Pearson correlation coefficients ( $r$ ) (Lakens, 2013).

## Results

### Participant Characteristics

Nearly all participants self-identified as female (95.3%). The vast majority of participants (89%) were White, while 6% identified as Black or African American, 3% as American Indian or Alaska Native, and less than 1% as either Asian, Native Hawaiian or Other Pacific Islander, or other race.

For ethnicity, nearly 9% of participants identified as Hispanic or Latino. The vast majority of program administrators reported having at least a 4-year college degree with 49% obtaining a Bachelor's Degree and 41% obtaining a Master's Degree or higher. Participants' mean age was 46.6 years old ( $SD = 11.2$ ), and their mean years of experience working in early childhood programs was 15.2 years ( $SD = 8.9$ ).

About two-thirds (67%) of participants reported working as administrators in home visiting or other home-based early childhood programs. Of the remaining participants, 16% worked in preschools or other center-based programs, 3.5% worked in parenting groups, 3.5% worked in maternal-child health programs, and 9% reported working in some other

type of program. Large majorities of program administrators reported their programs targeting families with a variety of social, health, and risk disparities including low-income families (96%), young parents under 21 years of age (83%), households with histories of substance abuse (76%) or child abuse or neglect (76%), and families speaking languages other than English (65%). Thirty-one respondents (18%) reported serving other target populations. Target outcomes of the programs reported by participants at the child level included prevention of child abuse and neglect (84%), child development (90%), school and/or kindergarten readiness (82%), and infant and child health (78%). About two-thirds of participants reported target outcomes at the parent or family level including improving maternal health (67%) and supporting family economic self-sufficiency (62%). The median number of families served was 100 with an interquartile range of 239.5 (Q1 = 53, Q3 = 292.5). The mean number of program staff was 24.1 with a standard deviation of 40.8.

## Descriptive Results

Of the final 54 items on the DDDM-Q, three items were reversed scored to ensure higher scores would reflect or imply higher use of data. These three items included selection\_3 (“Our program relies primarily on the gut feelings and opinions of the hiring team to hire new staff”), perf\_assess\_4 (“Program supervisors use their own discretion to evaluate the performance of individual service providers.”), and decision\_support\_3 (“Even though we put data into our system, we cannot get data out in meaningful reports.”). Supplement 1 contains the item label and descriptive statistics for each item on the DDDM-Q. Mean responses to the 54 individual DDDM-Q items ranged from 2.68 to 4.49 with the range of score possibilities spanning 1.0 to 5.0. In general, descriptive statistics for each item demonstrate respondent agreement as 47 of the 54 items mean scores were 3.5 or higher, and five items were between 3.0 and 3.5.

Only two items’ means were slightly below 3.0 implying some disagreement with the item.

Because subscales varied in their number of items, we calculated a mean per item index score to standardize comparison and analysis across subscales. We computed the mean per item index score by summing the scores for every item in a given subscale and then dividing the summative score by the number of items in the subscale. For the staff selection subscale, for example, we summed the five responses to each item in the staff selection subscale and divided by five. The range of possible mean per item index scores is 1.0 at the lowest to 5.0 at the highest. All subscales had a mean per item index score above 3.5, indicating general respondent agreement overall. The systems intervention subscale had the highest mean per item score ( $M=4.23$ ,  $SD=.52$ ), while the training subscale had the lowest mean per item score ( $M=3.34$ ,  $SD=.64$ ). Descriptive statistics including means and standard deviations of these mean per item index scores as well as number of items per subscale are located in Table 1.

## Confirmatory Factor Analysis

The model fit indices examined suggested the data fit the proposed 9-factor model well based on Schreiber et al.’s (2006) recommended model fitness criteria (CFI = .98; TLI = .97; RMSEA = .03, 90% CI .021–.036). The ratio of the chi square statistic relative to the degrees of freedom ( $\chi^2/df = 1.14$ ,  $\chi^2(1341) = 1534.65$ ,  $p < .001$ ) also suggested overall goodness-of-fit for the 9-factor model following recommendations from Alavi et al. (2020).

## Internal Consistency

Table 1 provides Cronbach’s alphas ( $\alpha$ ) for the nine factors analyzed as an indicator of internal consistency. Alpha levels for eight of the nine factors ranged from .73 to .90, indicating that the majority of factors have levels of internal

**Table 1** Means and estimates of internal consistency for the DDDM-Q ( $N=173$ )

Domain	Subscale	# of items	Mean per item index score	SD	Cronbach’s Alpha (Internal consistency)
Competency driver	Staff selection	5	3.55	.63	.670
	Training	7	3.34	.64	.779
	Coaching and supervision	6	4.08	.58	.850
	Performance assessment	4	3.57	.66	.732
Organization driver	Systems intervention	6	4.23	.52	.787
	Facilitative administration	8	4.07	.54	.863
	Decision support data systems	7	3.62	.63	.795
Leadership driver	Technical leadership	6	3.86	.65	.865
	Adaptive leadership	5	3.90	.69	.904

consistency that range from acceptable to excellent (Nunnally, 1978). The factor that did not meet these levels—staff selection—fell slightly below the suggested threshold at  $\alpha = .67$ . Given its proximity to the acceptable threshold in combination with the results of the CFA, this level of internal consistency may be sufficient given the practical context of DDDM in early childhood service implementations of this study (Taber, 2018).

### Associations Between DDDM-Q Subscales, Participant Demographics, and Program Characteristics

Bivariate analyses found generally small effect sizes and no statistically significant relationships at the Bonferroni-adjusted  $p < .0006$  level between eight participant demographics and program characteristics and the 9-factor subscales of the DDDM-Q. Table 2 shows the associations between demographic characteristics, program characteristics, and DDDM-Q factor scores, including the effect sizes and statistical significance results of these bivariate analyses.

### Discussion

Even though research supports applying core implementation drivers to inform DDDM and existing literature calls on early childhood programs to build capacity for DDDM, there was a notable lack of a suitable measure of capacity and readiness for DDDM among early childhood programs or human service organizations more generally (Albers et al., 2017; Barwick et al., 2011; Fearing et al., 2014; Fischer et al., 2014; Graff et al., 2010; Kimber et al., 2012; McCrae et al., 2014; Metz et al., 2015; Salverson et al., 2015). The current study aimed to fill this gap by developing, piloting, and conducting the initial validation of the Data-Driven Decision-Making Questionnaire (DDDM-Q). The DDDM-Q was theoretically grounded in the AIF Drivers framework and piloted with a sample of early childhood program administrators. The study informs implementation research and practice in three key ways: (1) it provided a theoretically driven model for assessing individual administrators’ perceptions of data use; (2) it established initial evidence of the underlying factor structure and internal consistency of the AIF Drivers for measuring DDDM; and (3) it found no evidence that participant demographics or program characteristics were related to mean scores on DDDM-Q.

While no single study will establish complete validity of a measure, we evaluated the initial validity of the DDDM-Q based on the accumulation of evidence across five validity evidence categories (Goodwin, 2002). This study produced evidence supporting the validity of the DDDM-Q in three of these categories, including (1) evidence based on test

**Table 2** Measures of association between demographic characteristics, program characteristics, and DDDM-Q factor scores

	Effect test	9-factor score correlations ( $r$ ), Eta-squared values ( $\eta^2$ ), or Omega-squared values ( $\omega^2$ )								
		Selection and hiring	Training	Coaching and supervision	Performance assessment	Systems intervention	Facilitative admin	Decision support data system	Technical leadership	Adaptive leadership
<i>Categorical variables</i>										
Race	$\eta^2$	.000	.000	.017	.001	.009	.010	.008	.017	.002
Ethnicity	$\eta^2$	.003	.014	.022	.023	.016	.042	.006	.020	.016
Education	$\eta^2$	.001	.009	.012	.015	.001	.000	.013	.013	.008
# of families served	$\eta^2$	.008	.025	.004	.010	.013	.037	.008	.050	.013
# of staff	$\eta^2$	.002	.015	.019	.022	.013	.009	.001	.027	.012
Program type	$\eta^2$ or $\omega^2$	.000	.012	.016	.039 <sup>ab</sup>	.021	.000	.031	.001	.000
<i>Continuous variables</i>										
Age in years	$r$	.150	.096	.080	.033	-.147	.179	.039	.117	.169
Experience in years	$r$	.076	.127	.182	.157	-.010	.196	.159	.168	.136

\*Statistical significance based a Bonferroni corrected alpha,  $p < .0006$

<sup>a</sup>Relationship tested using a Welch-ANOVA

<sup>b</sup>Effect size value based on Omega-Squared ( $\omega^2$ )



content, (2) evidence based on the internal structure, and (3) evidence based on the relationships with other variables.

### Validity Evidence Based on Test Content

Validity evidence based on test content examines the extent to which the actual content of a measurement instrument relates to the content domain (Goodwin, 2002). For DDDM-Q, it was imperative that the content of the questionnaire directly related to data use in early childhood programs. To ensure this, a panel of experts were selected to review, provide feedback, and support the iterative development of the questionnaire. Selection of the panel representatives involved recruiting experts with relevant backgrounds, expertise, and an understanding of the practical implications for piloting the instrument with a real sample. Additionally, selection of panel representatives involved recruiting panelists who could draw from broad and overlapping knowledge bases including those familiar implementing evidence-based practices in early childhood, those with training in instrument development and validation, and those familiar with the implementation science frameworks used to guide survey development. These experts provided critical feedback about DDDM-Q's understandability, usefulness, length, and completeness. We used the feedback to revise the questionnaire and define commonly misunderstood terms (i.e., coach and coaching).

Existing literature on validity standards suggests this approach produced reasonable evidence of DDDM-Q's validity based on its test content. As a starting point, the initial generation of items for the questionnaire was consistent with recommendations to base item development on relevant theory (Holmbeck & Devine, 2009). Our approach to developing the DDDM-Q also followed existing recommendations to refine the questionnaire iteratively with reviews from a panel of experts (Goodwin, 2002; Holmbeck & Devine, 2009). Finally, we aligned our approach to selecting panelists with the *Standards for Educational and Psychological Testing*, including panelists with a wide range of practical, theoretical, and empirical expertise (Goodwin, 2002; Holmbeck, 2009). This process provided a powerful foundation for a strategic instrument development process, vetted the potential usefulness and shortcomings of the questionnaire with others, and led to a critical evaluation of the questionnaire's content.

### Validity Evidence Based on Internal Structure

The study also produced evidence to support validity based on the internal structure of DDDM-Q through a robust statistical analysis of the internal factor structure of the questionnaire. CFA is often cited as the recommended analytic method to generate this form of evidence (Goodwin, 2002;

Holmbeck & Devine, 2009). The CFA in this study revealed a goodness-of-fit for the 9-factor model, indicating a match between the items on the questionnaire and the AIF Drivers. In addition to the CFA, an analysis of Cronbach's alphas provided statistical evidence of the internal consistency of the nine factors. These findings are consistent with both theoretical assumptions and previous empirical research on measure development (Ogden et al., 2012). Taken together, these statistical analyses provide evidence based on DDDM-Q's internal structure to support its validity and its application of AIF Drivers its underlying factor structure for measuring DDDM with early childhood program administrators.

### Validity Evidence Based on Relationships to Other Variables

Finally, we examined the relationships between external variables and responses to the DDDM-Q. As Goodwin (2002) contended, this type of validity evidence is critical to understand how other variables affect instrument scores in expected and unexpected ways. Our series of bivariate analyses presented an opportunity to assess if evidence based on relationships to other variables existed through an examination of 72 relationships between eight demographic and program characteristics and the nine subscale scores on the questionnaire. These analyses, however, found no statistically and no practically significant relationships to suggest DDDM-Q subscales are related to the studied demographic and program characteristic variables.

This finding provides encouraging evidence supporting DDDM-Q's validity. Even though some demographic and program characteristic variables can change over time, these characteristics are more likely to remain fixed and make it difficult to understand their effect on changes in outcomes or behaviors over time (LeCroy, 2019). In the context of this study, if these variables were related to DDDM, it would be difficult to design data-related trainings, interventions, and decision-making processes based on these fixed variables. This finding, combined with the results of the CFA, suggests that AIF Drivers serve as a more practical framework for such activities. Consider, for example, what an organization might do to improve its practices related to data use. It seems logical, fair, and ethical that an organization would improve DDDM by adjusting its training and coaching practices related to data than it would be for an organization to hire based on fixed demographics such as race or age.

### Implications for Practice and Research

The development of the DDDM-Q provides the field with a feasible and practical assessment for understanding program and organizational readiness for DDDM. Because the questionnaire is structured logically around the AIF Drivers,

it may provide a simple platform and outline for identifying key areas of strength as well as areas for improvement. Frequent use of DDDM-Q over time may support awareness of strengths and opportunities at various points and help organizations understand how their needs related to DDDM readiness may change or remain constant. The use of the DDDM-Q in practice is, in and of itself, an exercise in data-driven decision-making that could lead to additional pathways for innovation and program improvement. Lyons et al. (2018) recognized the same implication with regard to the development of strategic implementation measures in the education field. Future research can expand upon the use of the DDDM-Q by examining longitudinally its use with both quantitative and qualitative data analyses.

Three key lessons learned from the development of DDDM-Q may provide insight to support future research efforts. First, this study provides empirical support for using AIF Drivers as a guide and general structure for developing research instruments. Indeed, while previous work has highlighted the AIF Drivers as useful for guiding a supporting program implementation (Fixsen et al., 2019), this study expands the utility of the framework into other domains of inquiry. Second, the study provides a model for applying Goodwin's (2002) validity assessment framework to initiate validation efforts of new measures. This model includes assessing the merit of accumulating multiple forms of evidence to support validity overall. Finally, this study underscores the importance of engaging a diverse panel of expert representatives to vet an instrument prior to its use and tailor it to the target respondent population. Such an approach reinforces the value of community-based participatory actions that are needed in research to ensure field relevance and practice uptake (Wallerstein & Duran, 2010).

## Limitations

A few general limitations of this study should be considered in the context of interpreting its findings and implications. First, the study piloted DDDM-Q at one point in time and lacks data to support its usefulness over time. Implementation, however, often occurs over the course of multiple years (Fixsen et al., 2019; Ogden et al., 2012), so future research with DDDM-Q should attempt to collect data longitudinally, which may support better identification of trends and patterns related to DDDM changes throughout program implementation and accumulate evidence of DDDM-Q's test-retest reliability. Second, the study's sample is limited in that it focused exclusively on program administrators, relied on non-probability sampling techniques to engage them, and lacked demographic diversity particularly when it comes to race and gender identity. As such, while program administrators were a logical starting population given the variety of their roles and responsibilities, the study is

limited in its ability to generalize to other key participant groups. Finally, while the response rate for this study is consistent with other response rates for web-based survey research (Kelfve et al., 2020; Schouten et al., 2009), we cannot be certain there was no non-response bias. Future studies should recruit larger, more diverse samples, consider incentives to increase response rates, and engage other respondent groups including direct service providers, executive leaders, and families receiving services.

In addition to these general limitations, a few limitations related to the validity evidence also emerged. While this study produced evidence in three of the five evidence categories, it did not produce evidence based on response processes nor did it produce evidence based on the consequences of testing (Goodwin, 2002). To accumulate evidence based on response processes, future research may include interviews of respondents to gain an understanding of how they interpreted the questionnaire's content and whether their interpretations aligned with the questionnaire's intent. Finally, to accumulate evidence based on the consequences of testing, future studies could examine the extent to which use of the DDDM-Q helped organizations achieve the anticipated benefits of using the questionnaire, which may include identifying areas to improve decision-making processes resulting from data use.

## Conclusion

The present study reported on the development, piloting, and initial validation of the Data-Driven Decision-Making Questionnaire (DDDM-Q) with early childhood program administrators. While this study is the first effort to use and validate this questionnaire, its findings offer foundational evidence of DDDM-Q's validity, including evidence based on test content, internal structure, and relationships with other variables. As such, the DDDM-Q grounded in established Active Implementation Frameworks can serve as an empirical tool to guide future research, assess organizational readiness and capacity for DDDM, and inform and improve practice. This could prove especially helpful to early childhood service organizations, and other social and human service organizations more broadly, as they encounter increasing requirements to use data to support decision-making processes and successful implementations.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s43477-022-00044-5>.

**Acknowledgements** The authors would like to acknowledge and thank all those who contributed their time and expertise to this study. These individuals included those on the panel of experts (Dr. Jacklyn Biggs, Janet Horras, Dr. Jessica-Sprague Jones, Dr. Amy Mendenhall, and LaShawn Williams) and the early childhood program administrators

who participated in this study. Without each of their contributions, the study would not have been possible.

**Author Contributions** Both authors contributed to the study conception and design. Participant recruitment, material preparation, data collection, and analysis were performed by JB. The first draft of the manuscript was written by JB. BA wrote a revised version of the manuscript. Both authors read and approved the final manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. L. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing*, 76(9), 2209–2211. <https://doi.org/10.1111/jan.14399>
- Albers, B., Mildon, R., Lyon, A. R., & Shlonsky, A. (2017). Implementation frameworks in child, youth and family services—Results from a scoping review. *Children and Youth Services Review*, 81, 101–116. <https://doi.org/10.1016/j.childyouth.2017.07.003>
- Aubrey, C., Godfrey, R., & Harris, A. (2012). How do they manage? An investigation of early childhood leadership. *Education Management, Administration, and Leadership*, 41(1), 5–29. <https://doi.org/10.1177/1741143212462702>
- Barton, J. L. (2016). Federal investments in evidence-based early childhood home visiting: A multiple streams analysis. *Poverty & Public Policy*, 8(3), 248–262. <https://doi.org/10.1002/pop4.142>
- Barwick, M., Kimber, M., & Fearing, G. (2011). Shifting sands: A case study of process change in scaling up for evidence based practice. *International Journal of Knowledge, Culture and Change Management*, 10(8), 97–114.
- Bertram, R. M., Blase, K. A., & Fixsen, D. L. (2015). Improving programs and outcomes: Implementation frameworks and organization change. *Research on Social Work Practice*, 25(4), 477–487. <https://doi.org/10.1177/10497315154537687>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.
- Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12(1), 25–44. <https://doi.org/10.1080/19312458.2017.1396583>
- Coffey, M., Zamora, C., & Nguyen, J. (2020). Using data to support a comprehensive system of early learning and care in California. SRI International.
- Collins-Camargo, C., & Garstka, T. (2014). Promoting outcome achievement in child welfare: Predictors of evidence-informed practice. *Journal of Evidence-Based Social Work*, 11(5), 423–436. <https://doi.org/10.1080/15433714.2012.759465>
- Collins-Camargo, C., Sullivan, D., & Murphy, A. (2011). Use of data to assess performance and promote outcome achievement by public and private child welfare agency staff. *Children and Youth Services Review*, 33(2), 330–339. <https://doi.org/10.1016/j.childyouth.2010.09.016>
- Coulton, C. J., Goerge, R., Putnam-Hornstein, E., & de Haan, B. (2015). Harnessing big data for social good: A grand challenge for social work (pp. 1–20). *American Academy of Social Work and Social Welfare*.
- Fearing, G., Barwick, M., & Kimber, M. (2014). Clinical transformation: Manager's perspectives on implementation of evidence-based practice. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(4), 455–468. <https://doi.org/10.1007/s10488-013-0481-9>
- Fischer, R., Anthony, B., & Dorman, R. (2014, December). *Big data in early childhood: Using integrated data to guide impact*. Presented at the 29th Annual Zero to Three National Training Institute, Ft. Lauderdale, FL.
- Fixsen, D., Blase, K., Naoom, S., Metz, A., Louison, L., & Ward, C. (2015). *Implementation drivers: Assessing best practices*. University of North Carolina at Chapel Hill.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531–540. <https://doi.org/10.1177/1049731509335549>
- Fixsen, D. L., Blase, K., & Van Dyke, M. K. (2019). *Implementation practice & science*. Active Implementation Research Network.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., & Friedman, R. M. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Gillingham, P. (2019). Decision support systems, social justice and algorithmic accountability in social work: A new challenge. *Practice*, 31(4), 277–290. <https://doi.org/10.1080/09503153.2019.1575954>
- Goodwin, L. D. (2002). Changing conceptualizations of measurement validity: An update on the new standards. *Journal of Nursing Education*, 41(3), 100–106. <https://doi.org/10.3928/0148-4834-20020301-05>
- Graff, C. A., Springer, P., Bitar, G. W., Gee, R., & Arredondo, R. (2010). A purveyor team's experience: Lessons learned from implementing a behavioral health care program in primary care settings. *Families, Systems, & Health*, 28(4), 356. <https://doi.org/10.1037/a0021839>
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCAP): A meta-driven methodology and workflow process for providing translational research informatics. *Journal of Biomedical Informatics*, 42(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>
- Haskins, R. (2018). Evidence-based policy: The movement, the goals, the issues, the promise. *The Annals of the American Academy of Political Social Science*, 678(1), 8–37. <https://doi.org/10.1177/0002716218770642>
- Haskins, R., & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy*. Brookings Institution Press.
- Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1), 100–120. <https://doi.org/10.1177/109634809702100108>
- Holmbeck, G. N., & Devine, K. A. (2009). Editorial: An author's checklist for measure development and validation manuscripts. *Journal of Pediatric Psychology*, 34(7), 691–696. <https://doi.org/10.1093/jpepsy/jsp046>
- IBM Corporation. (Released 2017). IBM SPSS Statistics for Windows, Version 25.0. IBM Corp.
- Kaye, S., DePanfilis, D., Bright, C. L., & Fisher, C. (2012). Applying implementation drivers to child welfare systems change:



- Examples from the field. *Journal of Public Child Welfare*, 6(4), 512–530. <https://doi.org/10.1080/1554872.2012.701841>
- Kelfve, S., Kivi, M., Johansson, B., & Lindwall, M. (2020). Going web or staying paper? The use of web-surveys among older people. *BMC Medical Research Methodology*, 20(1), 1–12. <https://doi.org/10.1186/s12874-020-01138-0>
- Kimber, M., Barwick, M., & Fearing, G. (2012). Becoming an evidence-based service provider: Staff perceptions and experiences of organizational change. *The Journal of Behavioral Health Services & Research*, 39(3), 314–332. <https://doi.org/10.1007/s11414-012-9276-0>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lewis, T. F. (2017). Evidence regarding the internal structure: Confirmatory factor analysis. *Measurement and Evaluation in Counseling and Development*, 50(4), 239–247. <https://doi.org/10.1080/07481756.2017.1336929>
- Little, M., Vogel, L. C., Merrill, B., & Sadler, J. (2019). Data-driven decision making in early education: evidence from North Carolina's Pre-K program. *Archivos Analíticos de Políticas Educativas = Education Policy Analysis Archives*, 27(1), 17, 1–27.
- Lyons, A. R., Cook, C. R., Brown, E. C., Locke, J., Davis, C., Ehrhart, M., & Aarons, G. A. (2018). Assessing organizational implementation context in the education sector: Confirmatory factor analysis of measures of implementation leadership, climate, and citizenship. *Implementation Science*, 13(1), 5. <https://doi.org/10.1186/s13012-017-0705-6>
- Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85. <https://doi.org/10.1080/00461520.2012.667064>
- McCrae, J. S., Scannapieco, M., Leake, R., Potter, C. C., & Menefee, D. (2014). Who's on board? Child welfare worker reports of buy-in and readiness for organizational change. *Children and Youth Services Review*, 37, 28–35. <https://doi.org/10.1016/j.childyouth.2013.12.001>
- McDonald, J. H. (2014). *Handbook of biological statistics* (3rd ed.). Sparky House Publishing.
- Metz, A., & Albers, B. (2014). What does it take? How federal initiatives can support the implementation of evidence-based programs to improve outcomes for adolescents. *Journal of Adolescent Health*, 54(3), S92–S96. <https://doi.org/10.1016/j.jadohealth.2013.11.025>
- Metz, A., & Bartley, L. (2012). Active implementation frameworks for program success. *Zero to Three*, 32(4), 11–18.
- 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. <https://doi.org/10.1177/1049731514543667>
- Moder, K. (2010). Alternatives to F-Test in One Way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling*, 52(4), 343–353.
- National Implementation Research Network. (n.d.). *Active implementation hub*. University of North Carolina Frank Porter Graham child development institute. <https://nirn.fpg.unc.edu/ai-hub>
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advancements in Health Science Education*, 15, 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- 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(1), 49. <https://doi.org/10.1186/1748.5908.7.49>
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525–556. <https://doi.org/10.3102/00346543074004525>
- R Core Team. (2019). R: A language and environment for statistical computing. R Project for Statistical Computing. <https://www.R-project.org/>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rubin, A., & Babbie, E. (1993). *Research methods for social work* (2nd ed.). Brooks/Cole Publishing Company.
- Salverson, M., Bromfield, L., Kirika, C., Simmons, J., Murphy, T., & Turnell, A. (2015). Changing the way we do child protection: The implementation of Signs of Safety® within the Western Australia Department for Child Protection and Family Support. *Children and Youth Services Review*, 48, 126–139. <https://doi.org/10.1016/j.childyouth.2014.11.011>
- Schouten, B., Cobben, F., & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, 35(1), 101–113.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Sirinides, P., & Coffey, M. (2018). Leveraging early childhood data for better decision making. *State Education Standard*, 18(1), 35–38.
- Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior Research Methods*, 45(2), 536–546. <https://doi.org/10.3758/s13428-012-0257-2>
- Snow, C. E., & Van Hemel, S. B. (2008). *Early childhood assessment: Why, what, and how*. The National Academies Press.
- Stein, A., Freel, K., Hanson, A. T., Pacchiano, D., & Eiland-Williford, B. (2013). The Educare Chicago research-program partnership and follow-up study: Using data on program graduates to enhance quality improvement efforts. *Early Education & Development*, 24(1), 19–41. <https://doi.org/10.1080/10409289.2013.739542>
- Stoesz, D. (2014). Evidence-based policy: Reorganizing social services through accountable care organizations and social impact bonds. *Research on Social Work Practice*, 24(2), 181–185. <https://doi.org/10.1177/1049731513500827>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research on Science Education*, 48, 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Wallerstein, N., & Duran, B. (2010). Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. *American Journal of Public Health*, 100(Suppl 1), S40–S46. <https://doi.org/10.2105/AJPH.2009.184036>
- Yazejian, N., & Bryant, D. (2013). Embedded, collaborative, comprehensive: One model of data utilization. *Early Education and Development*, 24(1), 68–70. <https://doi.org/10.1080/10409289.2013.736128>
- Yoshikawa, H., & Kabay, S. (2015). *The evidence base on early childhood care and education in global contexts*. Background paper for the Education for All Global Monitoring Report 2015. UNESCO.
- Zweig, J., Irwin, C. W., Kook, J. F., & Cox, J. (2015). Data collection and use in early childhood education programs: Evidence from the northeast region. REL 2015-084. *Regional Educational Laboratory Northeast & Islands*.