

**Development and Preliminary Content Validation of ENGAGE as an Assessment of
Preschool Interactions and Active Engagement**

LeAnne D. Johnson, Maria L. Hugh, Andrea L.B. Ford, Danielle Dupuis, Kelsey A. Young, and

Brenna Rudolph

Department of Educational Psychology

University of Minnesota

Author's Note

LeAnne Johnson, Ph.D.  <https://orcid.org/0000-0003-0597-638X>

Maria Hugh, Ph.D.  <https://orcid.org/0000-0002-5696-4170>

Andrea Ford, Ph.D.  <https://orcid.org/0000-0003-0213-526X>

Danielle Dupuis, Ph.D.  <https://orcid.org/0000-0002-4960-1988>

Kelsey Young, Ph.D.  <https://orcid.org/0000-0001-7418-4138>

This work was supported by the National Center for Special Education Research, Institute of Education Sciences (Grant #R324A170032) and the National Center for Leadership in Intensive Intervention, Office of Special Education Programs (Grant #H325H140001). The statements do not necessarily reflect the views of our funders and all errors and omissions are our own.

We would like to thank the numerous educators, children, and their families that participated in this work. Without their support, this work would not have been possible. We would also like to thank Krystle Aman, Drake Bauer, and many additional members of our research team for their contributions to this work. For more information, please contact LeAnne Johnson (leannej@umn.edu).

Development and Preliminary Content Validation of ENGAGE as an Assessment of Preschool
Interactions and Active Engagement

Abstract

ENGAGE is in development as a web-based observation system with core features we believe will facilitate its use as a scalable assessment-to-action coaching and instructional support system in preschool classrooms. ENGAGE assesses adult interaction practices and children's active engagement such that classroom teams receive data to inform their intentional design and delivery of embedded instruction that can be made more effective for children's developmental needs through differentiation and intensification. For this study, we describe important precursors to evaluating other psychometric properties within an argument-based approach to validation. We used iterative cycles to gather evidence to refine and validate the content of our measurement targets (i.e., adult interaction practices and child active engagement) as well as constructs (i.e., groupings of practices by theorized mechanisms). Following two iterative rounds using online questionnaires, responses from participants representing intended users of ENGAGE generally demonstrated consistency with our definitions and conceptualizations, with more varied responses for child engagement. We discuss our procedures and results in terms of reducing measurement error that may be attributable to measurement targets in advance of future examinations of observation conditions (e.g., observer training, session duration) and situational variance (e.g., instructional pedagogy, adult-child ratios) that are often influential within observation-based assessments.

Key words: engagement, interaction, embedded learning, observation, preschool

To create meaningful learning opportunities for young children with developmental delays or disabilities in early childhood classrooms, early childhood educators must be intentional and systematic in their selection and use of developmentally appropriate practices. This selection and use requires both a foundation of high-quality practices that are effective for all children and approaches to intensifying practice use to meet the needs of specific children (Shepley & Grisham-Brown, 2019). The foundation of effective universal practices includes thoughtfully arranging the environment, using preventative practices, and being responsive within interactions (Hemmeter et al., 2007). From that foundation, intensifying intervention for specific children involves attending to the ways that children's instructional needs (i.e., visual models, dense repetitions, prompting) and modes of responding (i.e., verbal, visual, gestural, written) can be leveraged to promote children's generalized use of skills. More specifically, intensified practice use by adults should intentionally create opportunities for children to acquire and demonstrate skills across: (a) different instructional routines (e.g., large group vs. exploratory play), (b) different criteria for success (e.g., imitating a hand raise in response to a question vs. imitating a series of movements with a song), and (c) different stimuli eliciting behaviors (e.g., a visual schedule vs. verbal direction; e.g., Conroy et al., 2014; MacSuga-Gage & Simonsen, 2015). When the need for differentiated and intensified practice use intersects with the diversity of the instructional context and classroom teams' professional roles in early childhood settings (Johnson, 2017), the degree to which teams are prepared to effectively promote opportunities for children with developmental delays or disabilities to master needed skills requires attention.

To foster more effective use of practices across adults and instructional routines, early childhood classroom teams may benefit from valid and reliable assessment of their own practice

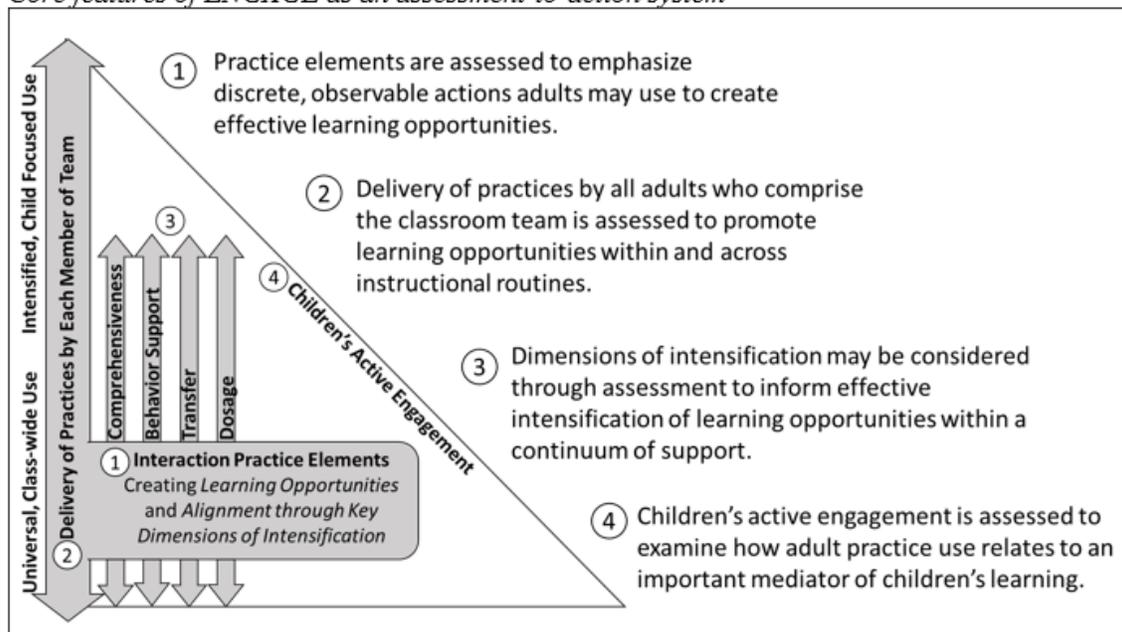
use and its impact on children's engagement and skill use. This kind of assessment may not only draw a team's attention to what research-supported practices are being used or not, but perhaps more importantly, also to how their use of those practices may facilitate or hinder children's engagement in critical, differentiated, and intensified learning opportunities. Although several global classroom quality measures exist that broadly assess practice use (e.g., Classroom Assessment Scoring System [CLASS]; Pianta et al., 2006), there is still a need for an observation-based classroom assessment that facilitates monitoring of how classroom teams use specific research-supported practices relative to children's engagement that demonstrates skill use. This type of dual assessment of adult and child performance within one system may contribute to improved impact on children's developmental outcomes through better alignment across universal practice use, differentiated use within approaches to intensifying intervention, and children's active engagement that demonstrates use of skills. To address this need and support alignment, our team is currently developing ENGAGE, a web-based, direct observation assessment system that includes several core features necessary for supporting the move from assessment to action.

Core Features of an Assessment-to-Action Observation System

Figure 1 provides our representation of four core features of the ENGAGE system that we conceptualize as necessary to producing actionable information based on data gathered using the assessment system. The first core feature is the assessment of individual practice elements (PEs) that create discrete learning opportunities within and across routines (Carta & Greenwood, 1985). This feature involves assessing adults' use of observable interaction behaviors that promote active engagement and facilitate learning, rather than broad collections of intervention strategies (e.g., social skills training, incidental teaching). Like definitions posed by others (e.g.,

Chorpita & Deleiden, 2009; Sutherland et al., 2019), we define PEs as the individual observable actions associated with an evidence-based program that educators deliver, in isolation or in combination, to achieve desired student outcomes. Examples of PEs in early childhood include praise, opportunities to respond, and comments that follow-in on a child’s attentional focus (DEC/NAEYC, 2009). When research supported PEs are incorporated into the classroom interactions between adults and children, the form those interactions take create the mechanism through which learning opportunities may or may not be effective. A mechanism is defined as the process through which a PE operates to affect desired outcomes (Lewis et al., 2018). PEs used by adults to provide explicit opportunities that support children to demonstrate skills in ways that are frequent, responsive, and differentiated are important aspects of an effective learning process (DEC/NAEYC, 2009).

Figure 1
Core features of ENGAGE as an assessment-to-action system



The second core feature necessary to providing actionable information is assessment that includes the performance of each adult who comprises the classroom instructional team to make

decisions about supportive actions needed to improve children's outcomes (Carta & Greenwood, 1985). By including all classroom team members (e.g., related service providers, assistants, special educators), this assessment approach facilitates examinations of the degree to which specific adults are using practices within and across different instructional routines as part of a system of support that aligns foundational universal practice use with differentiated and intensified practice use. As such, the team may also consider how children's outcomes may be impacted relative to not only what practices are present, but how the frequency and type of practices used differentiate and intensify learning opportunities.

Explicit consideration for how teams intensify learning opportunities is the third core feature of our assessment-to-action system. Key dimensions for intensifying instruction and intervention (Fuchs et al., 2017) may function as determinants of the effect learning opportunities may or may not have on children's outcomes. Determinants are factors, often malleable, that enable or hinder whether practice use achieves desired intervention effects or outcomes (Lewis et al., 2018). The ability of key determinants to help explain why practice use may or may not have had an effect underscores their importance in ENGAGE. Therefore, ENGAGE is being developed to facilitate examinations of practice use relative to four key determinants: (a) dosage of opportunities created for children to display skills and receive feedback, (b) transfer of skills to other formats and contexts, (c) comprehensiveness of the explicit instructional principles within the intervention, and (d) behavioral supports fostering self-regulation and minimizing nonproductive behavior (Fuchs et al., 2017).

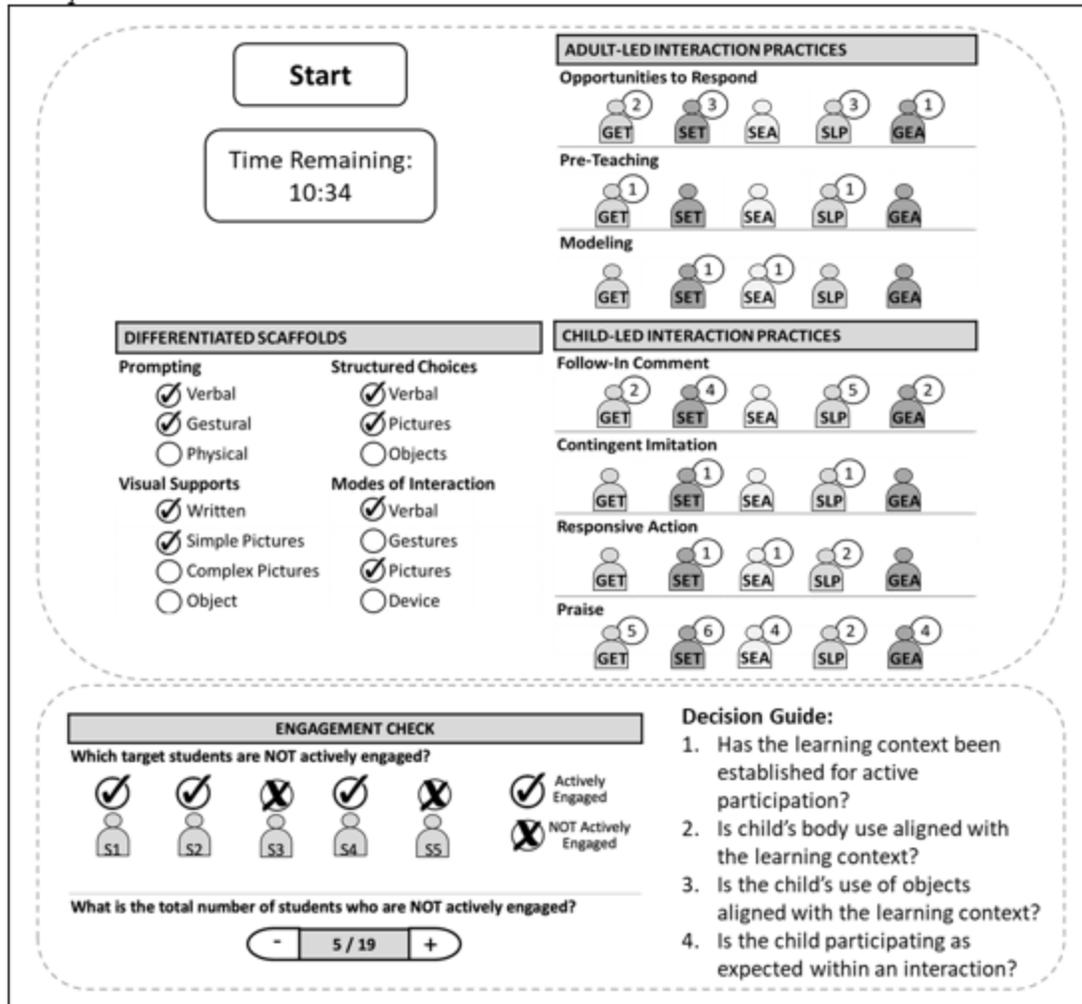
A fourth and final core feature is the direct assessment of children's active engagement. The dual assessment of adult and child behavior within ENGAGE is intended to provide classroom teams with a means for monitoring their practice use relative to an important mediator

of children's learning—active engagement (e.g., Vitiello & Williford, 2016). This real-time data on child engagement supports the team in making empirically informed decisions to adapt, differentiate, and/or intensify practice use within and across routines when necessary.

Developing an Observational Assessment of Preschool Learning Opportunities: ENGAGE

ENGAGE is a web-based direct observation assessment system being designed for practitioner and researcher use within professional development systems to assess how educators within preschool classrooms deliver intentional learning opportunities in service of promoting active engagement needed to improve children's outcomes. To offer specific information about adult use of interaction PEs and child active engagement, ENGAGE complements, not replaces, global measures of classroom quality (e.g., CLASS; Pianta et al., 2006). Observation sessions with ENGAGE use feasible time-sampling approaches (Lane & Ledford, 2014) of approximately 15-minute sessions in up to three different types of preschool instructional contexts—adult-directed large group, adult-directed small group, and child-directed exploratory play. Within an observation session, observers use ENGAGE to monitor *adult and child behaviors*. Figure 2 provides examples of the ENGAGE observation screens. The top panel provides an example of the main observation screen used to monitor adult's use of key interaction PEs through event recording of practice use as observed in real-time with children. The bottom panel of Figure 2 provides an example of how child behavior is monitored for specific focal children as well as the whole class using momentary time-sampling (Wood et al., 2016). While testing of ENGAGE is underway to determine the optimal sampling of engagement within an observation session, current capabilities allow observers to select an interval of one to five minutes to determine how frequently to assess engagement within an observation session.

Figure 2
Example ENGAGE observation screens



Note. GET = general education teacher, SET = special education teacher, SEA = special education assistant, SLP = speech language pathologist, GEA = general education assistant, S#[1,2,3,4,5] = focal student

Types of Adult Interaction Practice Elements Monitored with ENGAGE

Informed by the literature (e.g., Epstein, 2007; DEC/NAEYC, 2009; Pianta & Hamre, 2009), the research team identified and grouped adult interaction PEs into three types based on specific mechanisms through which each PE is expected to influence children’s learning (see Table 1 for definitions and measurement approach). The first type, *adult-led interaction PEs*, includes specific practices that are antecedents to a child’s behavior that promote more frequent and varied active child responses as an important mechanism for producing change in children’s

outcomes (DEC/NAEYC, 2009; Epstein, 2007). The second type, *child-led interaction PEs*, include practices that occur in response to a child’s attentional focus to build on a child’s interest and/or in the moment behavior as the mechanism for facilitating improved outcomes (DEC/NAEYC, 2009). The final grouping, *differentiated scaffolds*, includes clusters of practices (e.g., prompting, visual supports) that, when multiple PEs within a cluster (e.g., verbal promoting and gestural prompting) are delivered by adults, create differentiated ways of supporting a variety of children in engaging in learning through active participation (Barton & Smith, 2015; Burchinal, 2018; DEC/NAEYC, 2009). Therefore, we conceptualize the use of multiple PEs to provide differentiated scaffolding, or levels of support, as the mechanism for producing changes in children’s outcomes. In the early stages of the development process for ENGAGE, both the form and conceptual groupings of the adult interaction PEs into types were reviewed and refined based on feedback from nationally recognized experts in preschool learning environments.

Table 1*Definition and Measurement of Adult Practice Elements by Practice Type*

	Definition	Measurement
Adult-Led	Modeling: An adult secures attention and then demonstrates a skill	Frequency
	Pre-Teaching: An adult explains upcoming activities or expectations	Frequency
	Opportunities to Respond: An adult uses a question, statement, or gesture to seek a observable response from a child/children related to instruction	Frequency
Child-Led	Contingent Imitation: After a child/children has/have performed an action, an adult imitates motor actions of that child/children	Frequency
	Follow-in Comment: Given a child’s observable attention or interactions, an adult makes a related comment that expects no child response	Frequency
	Praise: Given a child/children’s behavior, an adult provides approval through positive feedback	Frequency
	Responsive Action: Following a child’s/children’s verbal or non-verbal communicative attempt, the adult responds with an action	Frequency
Differentiated Scaffolds	Modes of Interaction: Varying ways for children to communicate, including (a) verbal, (b) gestures, (c) pictures, and (d) device	Dichotomous for (a) – (d)
	Prompting: Varying types of supports to promote participation and skill use, including (a) verbal, (b) gestural, and (c) physical	Dichotomous for (a) - (c)
	Structured Choices: Varying ways of helping children select from at least two options, including (a) verbal, (b) picture, and (c) object	Dichotomous for (a) - (c)
	Visual Supports: Varying visual information, including (a) written, (b) complex pictures, (c) simple pictures, and (d) objects	Dichotomous for (a) - (d)

Children's Active Engagement

Children's active engagement is the behavior of interest given that engagement is an important mediator of learning and, given our definition, includes important demonstrations of children's skill use (e.g., using objects in ways aligned to the activity, demonstrating motor responses, participating in an interaction). Given the intended use of ENGAGE to support more effective embedded learning opportunities for all children, though especially for those with delays or disabilities, the definition and measurement within ENGAGE emphasizes aspects of active engagement rather than passive engagement. Passive engagement is often described as the child appropriately participating in the environment without manipulating objects or verbalizing, including behaviors such as watching others (e.g., Kemp et al., 2013). In contrast, active engagement is broadly defined as manipulating materials or vocalizing appropriately given the activity or context, for example, counting aloud or rolling play dough (Kemp et al., 2013). To minimize error that may be attributed to observers needing to make inferences about the universe of child behaviors that may represent active engagement, a decision-guide was developed for determining if a child is displaying active engagement or not during at each interval. The main aspects of the decision guide are shown in the lower panel of Figure 2. The research team developed each aspect of the decision guide with definitions that allowed for variability in the form of a child's behavior; based on either the child's developmental skills or aspects of the instructional routine. With this guide, observers determine if a child is not actively engaged rather than if they are actively engaged thereby, in effect, ensuring that any and all observable forms of active engagement showing developmentally appropriate skill use are captured. In the early stages of development for ENGAGE, like with the adult interaction PEs, expert researchers and stakeholders reviewed and provided feedback on the engagement observation decision guide

and measurement approach that contributed to refinements in advance of further testing.

Measurement Approaches Aligned to Intended Use and Guiding Validation

The adult interaction PEs included in ENGAGE (see Table 1) are shared by many evidence-based programs and practices along continuums of support (e.g., Shepley & Grisham-Brown, 2019), creating greater opportunities to leverage PEs as mechanisms of change. To promote more intentional creation of learning opportunities for children through delivery of adult interaction PEs guided by key determinants (e.g., dosage, transfer, etc.), the measurement approaches included within ENGAGE needed to be different than traditional classroom assessment systems. Our measurement model shifts away from traditional rating schemes and assessment of a singular lead educator (e.g., Pianta & Hamre, 2009) to assess the use of PEs by all adults across instructional routines. ENGAGE relies on event recording of occurrence and frequency of practice use by *each* adult who is part of the routine delivery of instruction in the classroom and observation. This approach to measurement creates an assessment system in which researchers and educators may examine data summaries with consideration of key determinants relative to the specific mechanisms for learning (i.e., adult-led, child-led, and differentiated support). For example, ENGAGE data may be used to examine the: (a) ways adults with different professional roles (e.g., teacher, occupational therapist) contribute variance to the instruction and learning opportunities through variable use of PEs that may facilitate or hinder determinants (e.g., transfer); (b) ways delivery of different types of PEs vary by instructional routine, professional role, type of program, adult-child ratio, and observation occasion to optimize dosage and comprehensiveness; and (c) effects and outcomes different interaction PEs may produce when each determinant is part of an intensified intervention plan to meet individual children's developmental needs. With these potential data uses in mind, attending to how

observers and classroom teams may interpret targets of measurement and resulting scores is important to achieving our goal of ENGAGE functioning as an assessment to action system within continuums of support (Kane, 2013). Combined with the goal of creating a scalable system, minimizing measurement error associated with the targets of measurement is an important precursor within our argument-based approach to validation before evaluating other psychometric properties such as stability, criterion-related validity, and inter-observer reliability (Kane, 2013).

A Need for Alternative Approaches to Gathering Content and Construct Validity Evidence

A systematic, argument-based approach (Kane, 2013) to the selection and validation of the content within the current version of ENGAGE involved: (a) systematic reviews of the literature, (b) iterative development and testing of definitions applied to a variety of classrooms, (c) expert and stakeholder review, and (d) iterative cycles gathering evidence to validate the content and operationalized definitions. Manuscripts detailing the literature supporting selection of the adult interaction PEs and conceptualization of child engagement, as well as the observer training manual, are available from the first author. This paper focuses on describing the iterative cycles used to gather evidence that helped to refine and validate the content (i.e., adult interaction PEs and child active engagement) as well as constructs (i.e., groupings of PEs by mechanism) as important precursors to evaluating other psychometric properties.

In traditional approaches to assessment development, researchers often evaluate individual items to determine the degree to which each item assesses the intended assessment targets and then evaluate groupings of items to determine if those groupings represent the intended theoretically-derived construct. Within ENGAGE, the individual items are the adult and child behaviors that are assessed using event recording of simple occurrence/non-occurrence or

frequency of occurrence (see Table 1). Given the intended use of ENGAGE, validating that the operational definitions of each item clearly and consistently lead observers to identify the intended target behavior when it is observed is a salient and important precursor to subsequent examinations of other types of validity and reliability evidence needed to support inferences and use. Therefore, we needed to develop an approach that facilitated evaluation of the degree to which our brief definitions for each item facilitated accurate identification of behaviors across multiple exemplars without needing extensive training. Similarly, validating that the individual items within the assessment system could be arranged to represent meaningful groupings is another important aspect of forming a validity argument that supports inferences that are consistent with the intended use of the assessment (Kline, 2005). In traditional approaches, how related items are as a measure of a construct is often evaluated using a variety of statistical approaches (e.g., confirmatory factor analysis, Kline, 2005). In contrast, for ENGAGE, the adult interaction PEs are not grouped because they purport to assess the same construct, but rather, they leverage the same theorized mechanism supporting learning. Therefore, it is not expected that individual interaction PEs will be correlated, which makes traditional approaches to establishing the factor structure and construct validity of an instrument untenable for ENGAGE. As such, we needed to develop a new approach to establish the underlying factor structure of ENGAGE as a precursor to developing scoring and reporting components that support the inferences researchers and practitioners using the system will make.

Research Questions

Given the measurement approaches of ENGAGE, traditional approaches to validating the performance of specific items as part of an assessment system were insufficient. Yet, when ENGAGE is taken to scale, it remains necessary to validate the measurement approach and

reduce measurement error attributed to variability in how users make inferences about the presence of observable behaviors relative to operationalized definitions of PEs, groupings of PEs by type, and active engagement. Our team used an iterative process to understand the extent to which the performance of ENGAGE intended users was consistent with the defined adult and child behaviors proposed by the research team for observation-based assessment within ENGAGE as they: (a) identify adult-led and child-led interaction PEs, (b) identify differentiated scaffold PEs, (c) categorize PEs into practice types, and (d) identify a child as actively engaged.

Method

Our team iteratively developed and deployed two rounds of questionnaires to answer each part of the research question, yielding eight questionnaires. Round 1 represented initial definitions of adult-led and child-led PEs, differentiated scaffolds, practice types, and active engagement. The results of Round 1 informed changes to definitions that were then included in the Round 2 questionnaires. Round 2 represented refinements with a goal of demonstrating clarity of definitions and establishing preliminary validity evidence.

Recruitment Procedures

For Round 1, our team recruited intended users (i.e., researchers and practitioners) in-person and online using three outlets: (a) an in-person exhibit booth at a national early intervention professional organization conference, (b) an email to a statewide professional development listserv for early childhood practitioners, and (c) an email to participants from the ongoing field test of ENGAGE. After revising definitions based on the findings from Round 1, we recruited a new sample of participants for Round 2 through the same professional development listserv and participants in the ongoing field test. We also recruited through two additional online outlets: (a) an email to the national early intervention professional association

membership list and (b) web-based postings to subgroups of a national professional organization.

Administration Procedures

For Round 1, two sets of administration procedures were used that varied by the recruitment method. For in-person recruitment (i.e., exhibit booth), to create an even distribution of participation across tasks, a member of the research team selected which task the participant would complete. We provided participants with a tablet that contained the questionnaire and headphones for listening to videos within the task. For online recruitment (i.e., listserv and field-test participants), individuals were provided with a link to the project website that included all four surveys. We instructed these participants to complete as many tasks as they chose, in any order, over a two-week period. The combination of approaches led to an unequal distribution of participants across tasks (i.e., some tasks had more participants than others).

Round 2 was entirely web-based recruitment. Participants were provided with a link to access the questionnaires. Upon clicking the link and consenting to participate, participants were randomly assigned to a set of two of the four tasks. We chose this approach to (a) reduce the load on participants and (b) facilitate an equal distribution of participants across tasks.

Participants

Table 2 includes an overview of the professional roles and work settings for participants for all questionnaires when provided. For Round 1, questionnaires received responses from 23 to 53 participants depending on the task. Of those providing demographic information, the majority were female ($M = 93\%$, *range* 88 – 96%) and White ($M = 92\%$, *range* 74 – 100%) with a master's degree or higher ($M = 69\%$, *range* 60 - 76%). Of the four tasks, respondents to the Practice Types (PT) task included the most racial diversity with four identifying as Bi-Racial, two as Asian, and one as Black. For Round 2, we grouped the tasks such that 42 participants

randomly received the PT task and the Differentiated Scaffolds (DS) task while 40 participants randomly received the Adult-Led PEs (AL), Child-Led PEs (CL), and Active Engagement (AE) tasks. Most participants in Round 2 provided demographic information. That majority were female ($M = 99\%$, *range* 98 – 100%) and White ($M = 89\%$, *range* 86 – 93%) with a masters degree or higher ($M = 87\%$, *range* 86 – 88%). Of the two tasks, respondents to the PT/DS questionnaire included the most racial diversity with four identifying as Black and one as Bi-Racial.

Table 2
Professional Demographics for Individuals that Completed Each Questionnaire

Demographic Variable	Round 1				Round 2	
	AL & CL	DS	PT	AE	PT/DS	AL & CL/AE
Total Number of Responses	26	23	53	27	42	40
Professional Role						
Behavior Consultant or Coach	3	1	7	1	2	2
General Education Teacher	3	3	6	3	2	1
Research/Educator in Higher Education	0	0	4	1	7	8
Special Education Teacher	7	10	13	10	8	4
Speech-Language Pathologist	2	2	2	2	16	18
Other	10	5	18	8	7	7
Current Work Setting(s)^a						
General Education Classroom	4	3	10	3	6	6
Inclusive Classroom	15	16	26	16	25	28
Institute of Higher Education	1	0	5	2	13	15
Self-Contained Classroom	5	9	12	9	8	11
Other	15	9	26	12	24	26

Note. AL & CL = Adult-led and Child-Led Practice Elements Task, DS = Differentiated Scaffolds Task, PT = Practice Types Task, AE = Active Engagement Task.

^a More than one work setting could be selected, so totals will not add to the total number of responses.

Instruments

For all tasks, participants completed an online questionnaire in Qualtrics™ (2015), which

included an introduction to the task, a consent form, the task itself (described below), and an opportunity to provide demographic information. Within each questionnaire, participants were presented with written definitions, and videos that were carefully selected and edited to highlight the focal behavior (adult interaction PEs, child engagement). We trimmed videos to highlight just one behavior whenever possible and embedded arrows to point to the focal child or adult. Qualtrics™ randomized the order of all items and response options across participants and tasks.

For the *Identifying Adult-led and Child-led PEs* task, participants were first provided with an overview of the eight AL and CL PEs. For each item, participants were presented an array of four adult interaction PEs and their definitions. After reviewing the definitions, they were asked to watch a short video clip. When the clip was finished, they were provided with the same list of four adult interaction PEs and asked to identify which of the four they saw the focal adult use. In Round 1, there were a total of nine items (7 practices shown one time; 1 practice—praise—was mistakenly shown two times). In Round 2, there were a total of seven items, one for each practice, as FI-OTR and OTR were collapsed into one category.

For the *Identifying Differentiated Scaffolds PEs* task, participants were first provided with an overview of the four clusters of practices within DS (e.g., visual supports, prompting) and the specific adult interaction PEs within those sets. For each item, participants were provided with an array of four DS PEs (e.g., mode of interaction – gesture, prompting – gestural, etc.) and their definitions. After reviewing the definitions, participants watched a short video clip. When the clip was finished, they were provided with the same list of the four DS PEs and asked to identify which one of the four they saw the adult use. In Round 1 and 2, there were a total of 14 PEs displayed across respondents with no modifications between rounds.

For the *Categorizing PEs by Type* task, participants were shown definitions of three

practice types (i.e., AL, CL, & DS), 8 adult-led and child-led PEs, and the four DS clusters. We then asked participants to complete a closed sort to categorize each PE into one of the three practice types. No information was provided regarding how many PEs should be placed in each category. The Round 2 surveys were identical in format with two changes. Because follow-in OTR was not consistently distinguished from OTR, follow-in OTR was removed for a total of 11 PEs. In addition, definitions for a three PEs (i.e., modeling, pre-teaching, and responsive action) were updated in response to the findings from Round 1.

For the *Identifying Active Engagement* task, we provided participants with a brief orientation to the ENGAGE definition of active engagement and two activities. In the first activity, we provided a definition of a specific element of active engagement (i.e., body use, material use, participation in an interaction) within the decision guide. Participants then watched a short video clip and were asked to decide whether the child in the clip was actively engaged with respect to that element. For each element, participants completed two items with corresponding videos which were randomly selected from a set of three. In the second activity, we provided a definition of active engagement that described how to apply the decision guide to consider all of the elements in determining whether a child was actively engaged. We provided a brief video example of how and when to look for each element to determine if a child is actively engaged. After viewing the video example, participants were asked to apply the decision-making guide to a random selection of three (of five) example videos and determine whether the child was actively engaged. In Round 2, we made two modifications: (a) we added a video explanation that more clearly delineated the difference between active and passive engagement, and (b) we refined the definition of full physical prompting described with the decision guide.

Results

Identifying Individual Adult-led and Child-led Interaction PEs

Table 3 provides the results of both Round 1 and Round 2. In Round 1, except for opportunity to respond (OTR), approximately 70% or more of the responses, averaged across individual items for each PE, were consistent with our proposed definitions. For OTR, averaged across both items, only 34.6% of respondents correctly identified OTRs. In reviewing the responses to these items, 39% ($n=5$) of responses in item 1 and 46% ($n=6$) responses in item 2 selected follow-in opportunity to respond (FI-OTR), suggesting that participants were not consistently distinguishing when an OTR was adult-led or child-led. Given this difficulty and that FI-OTR overlapped in content with OTR, our team removed FI-OTR as a PE and refined the definition of OTR. We did not make any other changes to definitions for the six other PEs.

Table 3*Mean Percent of Responses Consistent with the Proposed Measurement Model*

Elements by Task	Round 1		Round 2	
	<i>N</i>	Mean (SD)	<i>N</i>	Mean (SD)
Identifying Adult-Led and Child-Led PEs^a				
Modeling	11-15	84.2% (3.4%)	20	85.0% (14.1%)
Pre-Teaching	10-16	95.0% (7.1%)	20	90.0% (7.1%)
Opportunity to Respond	13	34.6% (5.4%)	19-21	90.2% (6.4%)
Contingent Imitation	10-16	100.0% (0.0%)	20-21	87.6% (10.8%)
Follow-in Comment	9-17	94.1% (8.3%)	19-20	92.1% (11.2%)
Responsive Action	11-15	69.7% (4.3%)	20	62.5% (3.5%)
Praise	26	82.7% (2.7%)	20-21	95.0% (7.1%)
Identifying Differentiated Scaffolds^a				
Modes of Interaction	6-17	80.1% (10.9%)	20-23	80.8% (10.9%)
Prompting	7-15	91.3% (3.7%)	20-23	94.7% (1.3%)
Structured Choices	7-15	85.5% (20.3%)	21-22	81.7% (19.2%)
Visual Supports	6-16	85.2% (14.1%)	20-22	80.9% (9.1%)
Identifying Active Engagement^b				
Element - Body Use	16-19	53.3% (45.7%)	20-31	89.0% (7.8%)
Element - Material Use	18	83.3% (29.5%)	29-30	85.5% (22.2%)
Element - Participate in an Interaction	16-18	92.6% (12.8%)	29-30	73.1% (29.5%)
Applying the Decision Guide	15	94.7% (7.8%)	24-27	57.9% (35.7%)

Note. *N* = Sample size range due to unbalanced randomization; SD = Standard deviation; PE = Practice Element.

^a This task two items per PE.

^b This task had three items per element and five items for applying the decision guide.

In Round 2 and for six of the seven PEs, 85% (e.g., modeling) to 95% (e.g., follow-in comment) of responses were consistent with our proposed definition. For one PE, responsive action, only 60% (item 1, *n*=12) and 65% (item 2, *n*=13) of responses were consistent with our proposed definition. For these questions, a brief OTR also occurred within the video, which was the option selected by those that did not select responsive action. While our team worked to ensure that only one PE was represented in each video and given that PEs often occur in quick

succession, there was no video available with a responsive action that did not also include an educator using an OTR. Though OTR should not have been included as an option in the array of response options to test the responsive action PE, this does at least provide evidence that this group of participants could consistently identify each of the AL and CL PEs.

Identifying Differentiated Scaffolds

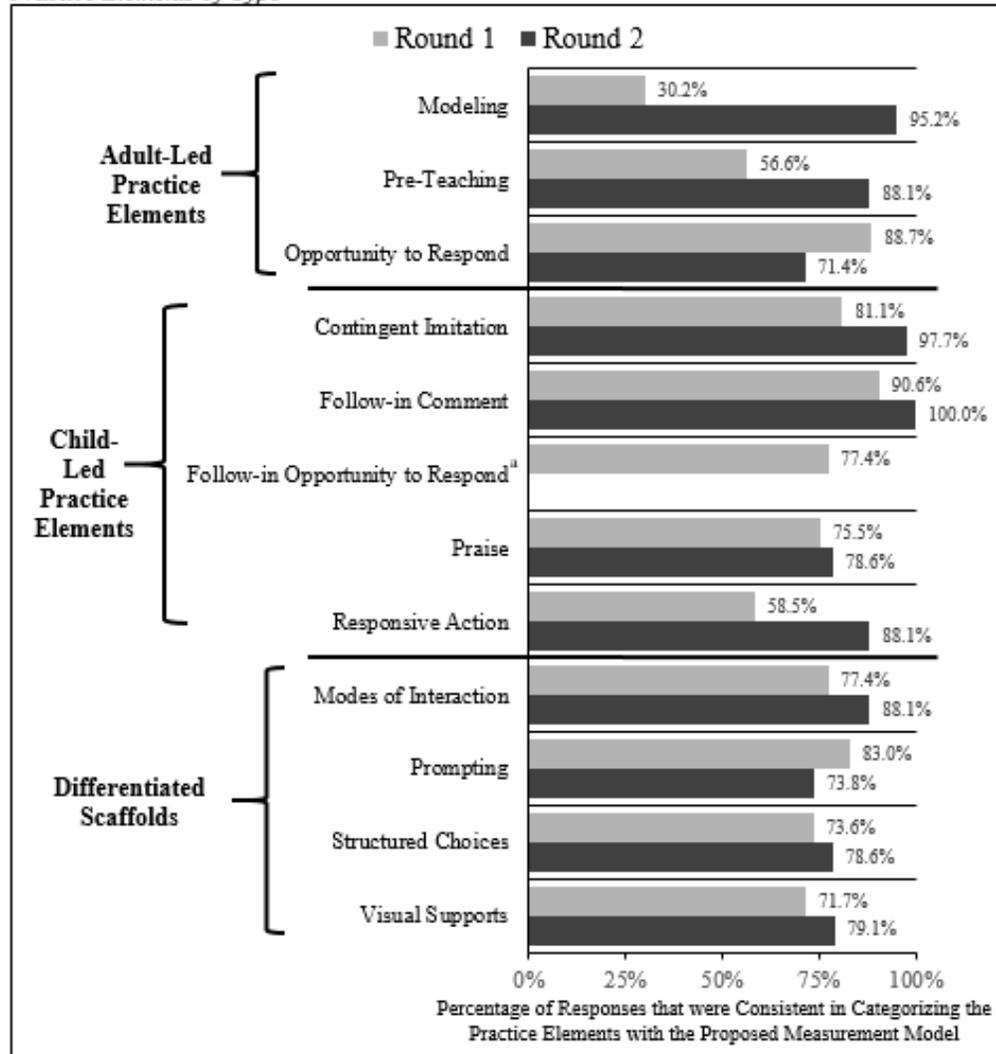
Table 3 provides the means for Round 1 and Round 2 for each set of PEs comprising DS. Across the four sets of PEs, a mean of 80.1% to 91.3% of responses, averaged across individual items within each set, were consistent with our proposed definitions. Within each set, our randomization process yielded an uneven distributions of responses for specific PEs comprising a set. Therefore, we secured a larger sample for Round 2, with sample sizes of 20 to 21 per item across the four sets of DS practices. Again, across the four sets of PEs, a mean of 80.8% to 94.7% of responses that were consistent with our proposed definitions. Like the previous tasks, the results suggest that this group of participants could consistently identify each of the PEs within the four sets DS practices.

Categorizing Practice Types

Figure 3 displays the results to the Round 1 and Round 2. In Round 1, for nine out of 12 PEs, 70% ($n=38$) or more of responses were sorted in a manner consistent with our proposed categorization. For three PEs, less than 60% of responses from participants were consistent with the researcher defined categories: (a) responsive action ($n=31$), (b) pre-teaching ($n=30$), and (c) modeling ($n=16$). Given the low degree of consistency, we refined the definitions of these three practices for Round 2. In Round 2, we presented 11 PEs (FI-OTR removed given PE identification task) and all 11 PEs had at least 70% of responses that were consistent with our proposed categorization. This result suggested that this group of individuals could distinguish

between the practice types and categorize PEs into AL, CL, and DS in ways that were consistent with the researchers’ conceptualizations.

Figure 3
Percentage of Participants from Round 1 and Round 2 who Consistently Categorized the Practice Elements by Type



^a For Round 2, Follow-in Opportunity to Respond was removed as a practice element following the Identifying Adult-Led and Child-Led Practice Elements Round 1 Task.

Identifying Children’s Active Engagement

Table 3 provides the results of the Round 1 and Round 2. For the first activity in Round 1 (i.e., identifying a child as actively engaged or not based on a singular aspect of the decision guide) and averaged across all items within an aspect, more than 80% of responses were

consistent in with our proposed determination of actively engaged or not. For the second activity in Round 1 (i.e., applying the decision guide within 5 items), the results were mixed. In reviewing responses by item, only 20% ($n=3$) for item 2, 47% ($n=7$) for item 3, and 0% of responses for item 5 were consistent. Across these items, participants often indicated a child was actively engaged if they were passively engaged (e.g., sitting and looking at a teacher while the teacher read a book). Given that ENGAGE is designed to capture active engagement, we recognized this response pattern as a need to operationalize our definition of active engagement more clearly. Therefore, in Round 2, we determined it was necessary for us to include an additional video that more clearly delineated (a) the need for the context to be set for active engagement and (b) that passive engagement did not fit our criteria for engagement.

For the first activity in Round 2, results were, on average, 8% lower than Round 1 across all items. For the second activity in Round 2, the results continued to be mixed across all five questions. There were improvements in answering item 2 (44%, $n=11$) and item 5 (25%, $n=6$) but a decrease in consistent responding in item 3 (23%, $n=7$).

Discussion and Implications

In this article, we described one-step of our iterative development process of ENGAGE, a web-based observational system designed to capture specific aspects of learning opportunities provided in preschool classrooms with the intent to support all children's learning, though especially those children with disabilities and delays. Features of this system combine with the intended use to create a need for us to develop novel approaches to validating definitions and conceptual groupings of measurement targets (i.e., adult interaction PEs and child active engagement) that are foundational to the way data produced by the system will be used to make inferences and guide actions of instructional teams (Kane, 2013). Therefore, within our

argument-based approach to validation, we used iterative cycles of tasks within multiple brief questionnaires. The tasks created opportunities for us to propose and refine our definitions to validate that intended users are consistently able to recognize the interaction PEs and children's active engagement assessed within the system. By reducing measurement error that may be attributed to the definitions of the measurement targets within the observation system, we expect that subsequent examinations of reliability and validity relative to use of ENGAGE will contribute to more precise evaluations of other conditions of administration (e.g., observer training, session duration) and scoring (e.g., stability, reliability; Kane 2013).

The results of this study support the validity of our specification and conceptualization of key interaction PEs that adults may use in preschool classrooms to create learning opportunities for children. This type of psychometric testing is an important component of preparing an observational assessment to be taken to scale for the intended use (Pianta & Hamre, 2009). Therefore, findings of this study provide preliminary validity evidence that educators can identify the varying forms of interaction PEs when present within the typical conditions of a preschool classroom provide an essential foundation for subsequent evaluations given the ability to reduce variability that may be attributed to the observers, a strength over global rating systems (Pianta & Hamre, 2009). Without this type of evaluation, future examinations of educators' ability to reliably use the system may incorrectly attribute lower reliability to the conditions of the observation (e.g., type of classroom, number of adults present, type of instructional routine) rather than the way the targets of measurement were specified for observers.

The measurement approach within ENGAGE also facilitates assessment that may be more sensitive to individual adult and situational variance than global ratings (Pianta & Harne, 2009), which is essential to interpretations and use that help to optimize ongoing and embedded

learning opportunities for individual children. Combining discrete and directly observed practice use with our approach to grouping practices by type and sampling across instructional routines reduces the inferences classroom teams need to make when exploring how to use ENGAGE to inform instructional decisions (Kane, 2013). In addition to data about the use of specific interaction PEs, the validation of our groupings of PEs into three practice types (adult-led, child-led, and differentiate scaffolding) based on theorized mechanisms through which learning may be facilitated may contribute to more effective instructional decisions based on ENGAGE data. By making learning mechanisms more explicit and displaying use of specific interaction PEs by different adults across instructional routines, classroom teams may begin to make instructional decisions about use of interaction PEs in more intentional ways to facilitate active engagement and improved outcomes through differentiated and intensified delivery of learning opportunities.

Validating our approach to specifying children's active engagement and the associated decision guide provided mixed results. Though educators in this study were able to identify discrete aspects of children's engagement (e.g., body use, materials use, participation in an interaction) in ways that were relatively consistent, their performance utilizing the full decision guide to decide whether a child was actively engaged was less consistent with the research team. There are multiple possible explanations for this, all of which warrant further examination within future testing. For example, the construct of engagement is widely recognized as a complex construct that may encompass an array of observable and unobservable behaviors relative to the theoretical paradigm underscoring engagement as a measurement target within different use cases (Fredreicks et al., 2004). As such, it is quite possible that educators were drawing on different paradigms when asked to make an overall determination rather assess one specific aspect. It is also possible that the decision guide is too complex for educators to be fluent in its

use based on the very brief introduction provided to them as part of this study. In which case, there is a need to evaluate if more robust training procedures mitigate this issue.

Limitations and Future Directions

There are limitations to this study that also highlight opportunities for future work. The persistence of lower degrees of consistency with the research team and some variability in responses, despite refined definitions, may be a result of videos including use of more than one practice. Though PEs are defined as discrete behaviors that do not overlap in their definitional form, use of certain PEs do often overlap in delivery in naturally occurring classroom interactions. For example, in a classroom, a structured choice may be offered in a way that invites a response using a gestural mode of interaction, representing two important PEs that would both be recorded in real-world use of ENGAGE. Nevertheless, for our purposes, oversight of this type of dynamic occurrence of multiple PEs and failure to properly control the array of response options available to participants may explain some of the inconsistencies and variability within our study. We also acknowledge that though we intentionally sought participants who represent intended user groups, our recruitment methods likely contributed to a racially and culturally homogeneous sample. Future work to examine how data are used to establish performance expectations and inform instructional decision-making will need to attend to how race and culture factor into observer recordings of behaviors as well as how classroom teams use those data. We also acknowledge that the nature of the tasks and varied contexts for participation may also contribute to varied and somewhat mixed findings. Future work is needed to examine how the provision of information and practice opportunities within a formal training process may contribute to greater consistency beyond refinements to definitions. Despite these limitations, we believe this work holds promise as one aspect of preparing a scalable observation system that

leverages the variability of educators' situational interactions (Pianta & Harme, 2009) as a means for enhancing embedded learning opportunities for children who are most in need of intentional, differentiated, and intensified instruction.

References

- Barton, E. E., & Smith, B. J. (2015). Advancing high-quality preschool inclusion. *Topics in Early Childhood Special Education, 35*(2), 69–78.
<https://doi.org/10.1177/0271121415583048>
- Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives, 12*(1), 3–9. <https://doi.org/10.1111/cdep.12260>
- Carta, J. J., & Greenwood, C. R. (1985). Eco-behavioral assessment: A methodology for expanding the evaluation of early intervention programs. *Topics in Early Childhood Special Education, 5*(2), 88-104.
- Chorpita, B.F. & Daleiden, E.L. (2009). Mapping evidence-based treatments for children and adolescents: Application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of Consulting and Clinical Psychology, 77*(3), 566-579.
<https://doi.org/10.1037/a0014565>
- Conroy, M. A., Sutherland, K. S., Vo, A. K., Carr, S., & Ogston, P. L. (2014). Early childhood teachers' use of effective instructional practices and the collateral effects on young children's behavior. *Journal of Positive Behavior Interventions, 16*(2), 81–92.
<https://doi.org/10.1177/1098300713478666>
- DEC/NAEYC. (2009). *Early childhood inclusion: A joint position statement of the Division for Early Childhood (DEC) and the National Association for the Education of Young Children (NAEYC)*. <https://doi.org/10.1177/1096250609347736>

- Epstein, A. S. (2007). *The intentional teacher: Choosing the best strategies for young children's learning*. National Association for the Education of Young Children.
- Fredricks, J., A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.
- Fuchs, L. S., Fuchs, D., & Malone, A. S. (2017). The taxonomy of intervention intensity. *TEACHING Exceptional Children*, 50, 35-43.
<https://doi.org/10.1177/0040059917703962>
- Hemmeter, M. L., Fox, L., Jack, S., & Broyles, L. (2007). Program-wide model of positive behavior support in early childhood settings. *Journal of Early Intervention*, 29, 337–355.
<https://doi.org/10.1177/105381510702900405>
- Johnson, L. D. (2017). Scaling the pyramid model across complex systems providing early care for preschoolers: Exploring how models for decision making may enhance implementation science. *Early Education and Development*, 28(7), 822-838.
<https://doi.org/10.1080/10409289.2017.1286205>
- Kane, M.T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50, 1-73. <https://doi.org/10.1111/jedm.12000>
- Kemp, C., Kishida, Y., Carter, M., & Sweller, N. (2013). The effect of activity type on the engagement and interaction of young children with disabilities in inclusive childcare settings. *Early Childhood Research Quarterly*, 28(1), 134–143.
<https://doi.org/10.1016/j.ecresq.2012.03.003>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
- Lane, J. D., & Ledford, J. R. (2014). Using interval-based systems to measure behavior in early

- childhood special education and early intervention. *Topics in Early Childhood Special Education*, 34(2), 83–93. <https://doi.org/10.1177/0271121414524063>
- Lewis, C.C., Klasnja, P., Powell, B.J., Lyon, A.R., Tuzzio, L., Jones, S., Walsh-Bailey, C., & Weiner, B. (2018). From classification to causality: Advancing understanding of mechanisms of change in implementation science. *Frontiers in Public Health*, 6(136), 1-6. <https://doi.org/10.3389/fpubh.2018.00136>
- MacSuga-Gage, A. S., & Simonsen, B. (2015). Examining the effects of teacher-directed opportunities to respond on student outcomes: A systematic review of the literature. *Education and Treatment of Children*, 38(2), 211-239. <https://doi.org/10.1353/etc.2015.0009>.
- Pianta, R.C. & Hamre, B.K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. <https://doi.org/10.3102/0013189X09332374>
- Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2006). *Classroom assessment scoring system: Preschool (pre-K) manual*. Brookes Publishing.
- Shepley, C. & Grisham-Brown, J. (2019). Multi-tiered systems of support for preschool-aged children: A review and meta-analysis. *Early Childhood Research Quarterly*, 47, 296-308. <https://doi.org/10.1016/j.ecresq.2019.01.004>
- Sutherland, K. S., Conroy, M. A., McLeod, B. D., Kunemund, R., & McKnight, K. (2019). Common practice elements for improving social, emotional, and behavioral outcomes of young elementary school students. *Journal of Emotional and Behavioral Disorders*, 27(2), 76–85. <https://doi.org/10.1177/1063426618784009>
- Vitiello, V., & Williford, A. P. (2016). Relations between social skills and language and literacy

outcomes among disruptive preschoolers: Task engagement as a mediator. *Early*

Childhood Research Quarterly, 36, 136-144. <https://doi.org/10.1016/j.ecresq.2015.12.011>

Wood, B. K., Hojnoski, R. L., Laracy, S. D., & Olson, C. L. (2016). Comparison of

observational methods and their relation to ratings of engagement in young children.

Topics in Early Childhood Special Education, 35(4), 211–222.

<https://doi.org/10.1177/0271121414565911>