Article

Hammill Institute on Disabilities

Topics in Early Childhood Special Education
1–13

DOI: 10.1177/02711214221129237
tecse.sagepub.com
SAGE

Advances in the Technical Adequacy of the Early Cognitive Problem-Solving Indicator Progress Monitoring Measure for Infants and Toddlers

Dale Walker, PhD[1], Jay Buzhardt, PhD[1], Fan Jia, PhD[2], Alana Schnitz, PhD[1], Dwight W. Irvin, PhD[1], and Charles R. Greenwood, PhD[1]

Abstract

Engaging, focusing, and persisting in the completion of tasks are among the skills needed for school success. Tracking whether a child is learning cognitive problem-solving skills is essential in knowing if they are acquiring skills important for development and school readiness; and if not, how they are responding to early intervention. Use of the Early Cognitive Problem-Solving Indicator (EPSI) was documented by data for 2,614 children (6–42 months of age) collected by the early childhood staff from 45 programs. Results indicated that the EPSI was (a) scalable across programs, assessors, and assessment occasions, (b) reliable, (c) sensitive to growth over months of age, (d) comprised a dynamic continuum of skills within and across skills over time, and (e) moderated by children's disability status but not gender or home language. Implications for research and practice are discussed.

Keywords

cognitive, development, problem solving, infancy, toddlers, progress monitoring, assessment

Among some of the most critical skills for school success is the ability to engage, focus on, and persist in the completion of challenging tasks and problems (Keen, 2011; Martin et al., 2013; Zelazo et al., 1997). Parents and professionals identify problem-solving skills as an important general outcome for young children (Priest et al., 2001). Children who can explore, manipulate, and play with toys, attend to a task, try out potential solutions, and persist in solving problems are learning how to learn (Siegler & Alibali, 2005). The development of early problem-solving has implications for children's executive functioning (Brock et al., 2009; Diamond & Lee, 2011; Fitzpatrick & Pagani, 2012), social competence (Landry et al., 2009; Stichter et al., 2016), later academic proficiency and school success (Garon et al., 2008; McWayne et al., 2004).

While early problem-solving skills develop throughout childhood, they emerge in simple forms during the earliest years of life (Babik et al., 2019; Keen, 2011; Willatts, 1999). Considerable individual variation exists, however, in the problem-solving abilities of young children and these skills may be delayed for some children. Early identification of delays in cognitive problem-solving skills leading to effective intervention may help move children toward age-expected outcomes (Colombo, 2004; Institute of Medicine and National Research Council, 2000; Zelazo et al., 1997). Consequently, a focus of early education and early intervention is promotion of children's cognitive and problem-solving skills (Diamond & Lee, 2011; National Research Council, 2001).

Unfortunately, few adequate measures designed to inform intervention decision-making exist even though individual progress monitoring is mandated in legislative and professional statutes. For example, Head Start Program Performance Standards [45 CFR Sec 1304.20(f) (1)] requires programs use of information from assessments of children's language, cognitive, physical, and social-emotional development to guide individual instruction goals and respond to children's strengths and needs (Office of Head Start, 2012; 45 CFR Sec 1308.3(b)(2)(ii)). The Division for Early Childhood (DEC, 2014) Recommended Practices highlights two characteristics of Individual

[1]The University of Kansas, Kansas City, USA
[2]University of California Merced, USA

Corresponding Author:
Dale Walker, Juniper Gardens Children's Project, The University of Kansas, 444 Minnesota Avenue, Suite 300, Kansas City, KS 66101, USA.
Email: walkerd@ku.edu

Growth and Development Indicators (IGDIs): (1) *Practitioners implement systematic ongoing assessment to identify learning targets, plan activities, and monitor the child's progress to revise instruction as needed*; and (2) *Practitioners use assessments with sensitivity to detect child progress and needs*.

The lack of adequate measures exists for several reasons. One is that intervention decision-making measures require important additional technical features beyond traditional psychometrics (reliability, validity) that make them usable by program staff to screen and monitor individual intervention effects. These measures support practitioners' intervention strategies by indicating which children are falling below benchmark performance levels at specific ages and by indicating rate of growth over time (slope) after a change in intervention has been made. For these purposes, measures must be brief to administer, repeatable, and scores interpreted by program staff, and supported by data management tools that make learning, using, and interpreting results program wide feasible. The infant-toddler *Individual Indicators of Growth and Development (IGDIs)* were developed for this purpose (Greenwood et al., 2011) and the benefits of using IGDIs by early intervention providers has been demonstrated through research on practice (Buzhardt et al., 2018; Carta et al., 2015; Kincaid et al., 2020; VanDerHeyden & Snyder, 2006).

Another reason is a result of history where there has been an emphasis on documenting delay or diagnosis at the expense of assessment to inform intervention decision-making (Neisworth & Bagnato, 2004; Walker et al., 2008). Few measures with both high technical standards and practical value exist for early educators to use to inform intervention and program improvement efforts (Buzhardt et al., 2020; Carta et al., 2002; Diamond et al., 2013; Moreno & Klute, 2011).

While some measures are closely tied to progress evaluation only within a curriculum (Lambert, 2020), others are primarily for developmental screening and early intervention eligibility decision making (Limbos & Joyce, 2011). Consequently, early education and intervention programs struggle to find assessment systems to meet their intervention decision-making needs (Bagnato et al., 2014).

Another barrier has been the technical adequacy of infant/toddler measurement. Access and usability of existing measures for screening, informing intervention, and monitoring progress for infants and young children over time is limited (Akers et al., 2016). Compared to assessments of children at older ages, the scores from traditional infant and toddler assessments are often poor predictors of later cognitive functioning (Colombo, 1993; Goodman, 2002). Traditional measures are not sensitive to short-term growth and do not provide information to help early educators detect growth in development in a timely manner (Bagnato et al., 2011; Diamond et al., 2013). Many rely solely on parent report (Charkaluk et al., 2017; Squires & Bricker, 2009) rather than on what children actually show they can do. Consequently, information gathered can rarely be used to reflect progress or help to select and guide intervention implementation (Buzhardt et al., 2018; Walker et al., 2008). Thus, early interventionists struggle to find assessment measures that meet their intervention decision making needs (Bagnato et al., 2011).

Purpose

In this investigation, we sought to advance the validity claims behind the Early Problem-Solving Indicator (EPSI) by examining new evidence likely to strengthen the rationale for using the EPSI for the universal screening and progress monitoring of infants and toddlers. The EPSI is one of four infant-toddler IGDIs developed by this team to inform intervention decision making including planning and individualizing intervention (Carta et al., 2002; Greenwood et al., 2011; Walker et al., 2008). The other infant-toddler IGDIs are the: (1) Early Communication Indicator (ECI; e.g., Greenwood, Carta, et al., 2006), (2) Early Movement Indicator (EMI; e.g., Greenwood et al., 2018), and (3) Early Social Indicator (ESI; Carta et al., 2004).

Compared to traditional assessments or caregiver reports, the infant-toddler IGDIs are direct observations of a child's performance on a few critical skills using valid, repeatable (monthly, quarterly) brief probes (6 minutes) with alternate forms for the purpose of estimating individual growth rates (Fuchs & Deno, 1991; Greenwood et al., 2011). The IGDIs are standardized measures in their administration procedures making scores generalizable across individuals and occasions of measurement. Administration occurs in the context of child play with standard toy items and a familiar adult partner using a standard interaction style, that of following the child's lead in play.

Given the unique purpose, function, and role of IGDIs in practice with very young children (Wilson, 2003, 2005, 2013), we selected Classical Test Theory (CTT; Crocker & Algina, 1986) as an appropriate approach for guiding the development and validation of the Infant-Toddler IGDIs. Key considerations in this decision were: (a) Infants and toddlers freely show their skills during play with a familiar adult; (b) infants and toddlers have few test taking skills, they are unable to sit at a desk or reliably respond to prepared items, such as paper cards or images on digital devices; (c) very young children may be frightened of unfamiliar adults and may shutdown when approached (Brooker et al., 2013); (d) There are a limited number of age-appropriate toys (analogous to "items" used in traditional assessments) that can be used during a brief play session that

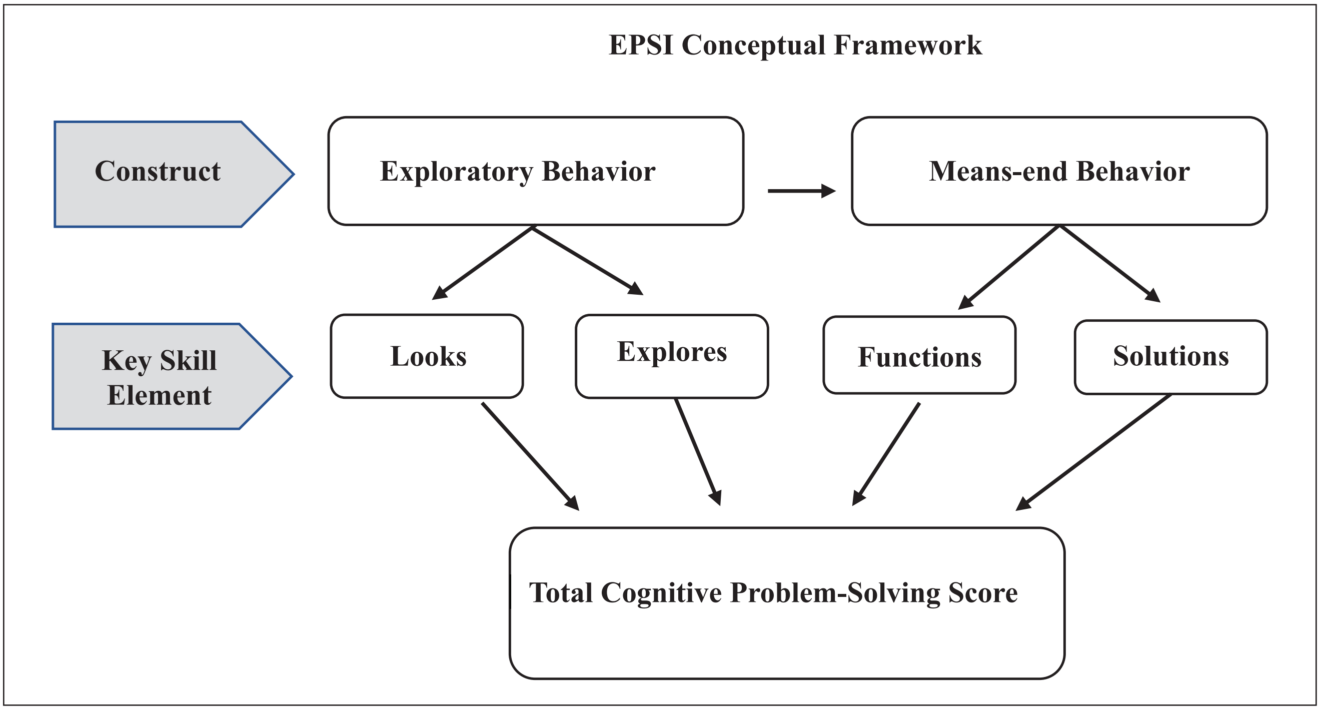

Figure 1. Theoretical construct for early cognitive problem-solving indicator (EPSI).

naturally evoke the skills of interest. Infant-toddler IGDIs are not intended to measure all skills in the targeted domain, instead measuring those that have strong theoretical relationship to the domain, are observable, expected to grow over time, and that practitioners can reliably score with training. Furthermore, scores obtained must be interpretable by practitioners and parents who are using the measures to inform change in their practices (Greenwood et al., 2011). These and other considerations supported use of CTT over other approaches such as Item Response Theory (IRT; e.g. Embretson & Yang, 2006; Jabrayilov et al., 2016) where large numbers of items and children assessed on these items are expected to support interpretation of item level scores.

Briefly, the sequence of development of each infant-toddler IGDI has proceeded as follows: (a) Building out a prototype measure using CTT and testing proof of concept and pilot psychometrics in a small sample of children with the assessments administered by research staff (Carta et al., 2004; Greenwood et al., 2002; Greenwood, Walker, et al., 2006; Luze et al., 2001); (b) improving the measure based on these results and refinement of protocols (Greenwood, Carta, et al., 2006; Greenwood et al., 2013); (c) development of web tools that support program staff efforts to be certified assessors, coders who score performance, and reporters and users of the results with caregivers (Buzhardt et al., 2011, 2022); (d) testing use of the IGDI in a larger number of programs and children supported by website tools for program, and data management (Greenwood et al., 2010, 2018, 2020) , (e) testing sensitivity to intervention effects (Buzhardt et al., 2018, 2020), and (f) psychometric studies including invariance test in large, time-displaced samples (Greenwood et al., 2013, 2020).

Cognitive Problem-Solving Indicator (EPSI)

Development of the EPSI has followed this same iterative cycle. In the first development step, a construct map for the measure was developed based on review of the developmental and early childhood assessment literatures on cognition for infants and toddlers. Supporting this information was a national survey of parents and providers regarding the importance of this early childhood outcome statement, *Child solves problems that require reasoning about objectives, situations, and people* (Priest et al., 2001). The theoretical model for the EPSI illustrated in Figure 1 links the desired outcome to cognitive problem-solving constructs including Visual and Object Exploratory and Means-End behaviors. Aligned to each construct are individual key skill behaviors to be measured, Looking (visual), Exploring (touching or manipulating) addressing visual and object exploratory construct, Function (making a toy perform its intended effect), and Solutions (persistent, sustained behaviors) addresses the means-end construct as measurement indicators (Greenwood, Walker, et al., 2006; Hupp & Abbeduto, 1991; Willatts, 1999). Combined, these key skills form a Total Problem-Solving per minute score.

Toy set items for authentic play situations involving cognitive problem-solving behaviors for infants and toddlers were identified. Criteria used to select items included: (a) toys that had a high likelihood of evoking problem-solving responses (e.g., pressing a lever, stacking cups by size), (b) toys that maintained child interest across the age range based on logical screening and testing, and (c) toys that were widely available making them easy for programs and individuals to purchase (Greenwood, Walker, et al., 2006). Toy items fitting our criteria were pop-up toys, stacking cups, and ball drop tower. Parallel alternate forms (A and B) were selected for each toy.

Infant-Toddler IGDIs, including the EPSI, are designed for use with infants and young children with and without disabilities. Children administered the EPSI are not disadvantaged if they use augmentative or assistive technologies (e.g., positioning support, wheelchair), sign language, picture communication systems, or need accommodations for visual or hearing impairments. The EPSI may be administered to children who are dual language learners as long as the adult play partner speaks the same language as the child to facilitate play. The child's primary home language and any additional languages they use are documented in the Online Data System by program staff.

Based on the concept map and literature review, we developed behavioral definitions for each child key skill. These were Look, Explores, Functions, and Solutions. A Look is recorded when a child orients their body by facing or moving toward a toy or person in front of the child. Looks are only scored when they occur in the absence of other key skills. Explore is recorded when the child touches, moves, puts the toy in their mouth, rubs, shakes, pushes, pulls, bangs the toy. A Function is recorded when the child makes a toy perform a function or creates an effect for which it was designed (e.g., popping up, fitting into, taking out, sorting, and opening). A Solution is recorded when the end-point for the toy was reached (e.g., all cups stacked, all balls retrieved).

EPSI Key Skill scores are in terms of responses per minute metrics based on dividing the number of child key skill responses by the session duration (6 minutes). The EPSI's Total Weighted Problem-Solving Score (TWPS) was a composite of summing the frequency of Looks, Explores, Functions, and Solutions key skill scores and then assigning a "1" to each occurrence of Look and Explore, "2" for each Function, and "3" for each Solution. Weighting enabled the more proficient behaviors to make a larger contribution to the Total EPSI benchmark score. These calculations are made automatically when assessors enter EPSI data into the IGDI Online Data System.

In another step, we evaluated the appropriate length of an individual EPSI session by using a 10-minute standard for comparison of shorter durations. The 6-minute session proved to be the shortest period to provide equivalent results. Thus, the final standard administration protocol used by the assessor during an assessment required that the child be presented with three toy items to play with for 2 minutes each, and the frequency of key skill behaviors were recorded for each 2-minute session (Greenwood, Walker, et al., 2006).

Administration and Scoring of the EPSI

Administration of the EPSI involves conducting three, 2-minute play sessions for a total of 6-minutes with the child and a familiar adult as play partner centered around one of two alternate toy sets that share common stimulus features. Pilot user testing of administration and protocol procedures was conducted in an initial proof of concept study with a small sample of children ($N=30$; Greenwood, Walker, et al., 2006).

Research Questions

Advancing to the next step in EPSI development and evaluation involved a large-scale deployment of the EPSI supported by the website with assessments and data collected by providers to advance what we know about the EPSI. We addressed in this investigation the following validity claims:

RQ1: Was use of the EPSI scalable by trained, certified practitioners in early childhood programs as measured by the volume of individual child data collected across states, programs, children, and occasions?

RQ2: To what extent do patterns of growth in EPSI key skill trajectories and intercorrelations within and across key skills conform to a continuum of changing key skills over time?

RQ3: Does evidence indicate that EPSI Total Weighted Problem-Solving (TWPS) score is reliable?

RQ4: To what degree was the EPSI benchmark TWPS score sensitive to growth over time?

RQ5: To what degree was growth in children's TWPS score moderated by children's characteristics including disability status, home language, and gender?

Method

Participants

Programs. Forty-five programs located across seven states adopted the EPSI for progress monitoring, screening, and/or intervention decision-making participated in this study. Programs represented a variety of early childhood education and early intervention service models including, Early Head Start (EHS), and Part C of the Individuals with Disabilities Act (IDEA, 2004). The distribution of programs that adopted the EPSI and contributed child data were: 20

Table 1. Children's Sociodemographics.

Variable	Infant/toddler sample
Children, *n*	2,614
Children's age (months) at first assessment	
M	24.0
SD	10.7
Min	6
Max	42
Gender (%)	
Male	52.6
Female	47.4
Home language (%)	
English	80.5
Spanish	16.7
Home language not Spanish	2.8
Individual family service plan (%)	
No	88.0
Yes	12.0

programs (44%) EHS and/or Part C, and 25 (56%) EHS + Part C, or Part-C only. These programs reported either providing home-based 11 (24%), center-based 4 (9%), or a combination 30 (67%) of home- and center-based services to children and families.

Assessors. Staff ($N = 175$) from 45 programs who completed EPSI training and certification participated as assessors/coders in conducting EPSI assessments. The mean number of staff conducting assessments was 4 per program, ranging from 1 to 50. The average number of children assessed by staff at each program was 30, ranging from 1 to 186. The roles of staff assessing and coding across the programs were described as early childhood and special educators, interventionists, home visitors, and evaluators.

Children. Child participants were 2614 infants/toddlers (53% boys and 47% girls) between 6 and 42 months of age (Table 1). The mean age of children was 24 months ($SD = 10.7$ months), with a range of 6 to 42 months. Twelve percent ($n = 314$) had an IFSP/IEP. English was the primary home language reported (80%; $n = 2{,}091$), followed by Spanish (17%; $n = 444$).

Measurement and Procedures

In addition to the EPSI scores, each child's profile of personal information was entered into the online data system. These items were determined by the program directors and included children's date of birth, and other demographic variables including gender (Male, Female), home language (native home language, dual language learner), IFSP/IEP status (No, Yes), and Disability Type (e.g., language, cognition, movement, etc.; Buzhardt et al., 2018). Gender, home language, and disability status were examined in this research as potential moderators of EPSI performance. It was possible to disaggregate the IFSP/IEP variable as follows: children (a) with no IFSP/IEP (typically developing), (b) children with an IFSP/IEP, and children with an IFSP/IEP and a documented cognitive concern.

The EPSI, like all Infant-Toddler IGDIs, can be administered in any language as long as the play partner, child, and the individual scoring the assessment speak that language. Assessments can be scored in-person (live) or later by watching a video-recorded session. IGDIs can also be scored using the IGDI Mobile app in lieu of paper scoring forms. After recording the key skills during an assessment session, data are entered into the password protected IGDI Online Data System https://igdi.ku.edu/ in which all data for a particular program are securely managed. Using the online system, data for individual and groups of children may be summarized in reports, including graphical displays of children's performance generated through the system (Figure S5).

IGDI training and scoring certification. Infant-toddler early educators and interventionists learn to administer and score the Infant-Toddler IGDIs, including the EPSI, through a combination of didactic training, online resources, and scoring practice to criterion. To achieve certification for an IGDI, a trainee scores two assessments with at least 85% reliability agreement with a master coding. If agreement is below 85%, they consult with a certified EPSI trainer who provides feedback for improvement until at least 85% reliability is achieved and documented in the IGDI Online Data System. Trainees also demonstrate at least 85% of the administration steps on the EPSI fidelity checklist to be certified to administer. It is recommended that staff conduct annual coding recalibration checks, involving dual coding such that agreement between a reliability assessor and a primary assessor can be checked. This is facilitated by the IGDI Online Data System which allows staff to enter a "reliability" score for an existing assessment, and the data system calculates interrater agreement between the two providers' scores where % agreement = [(agreements/(agreements + disagreements)) × 100].

Statistical Analyses

All EPSI assessments ($N = 5{,}477$) were screened to identify outliers before analyses. Forty-three (0.8%) assessments that were more than +3.0 *SD* above the mean were removed, leaving $N = 5{,}434$ to be included in analyses. The number of assessments collected per child ranged from 1 to 22, with 49% of children having at least two. Simple descriptive statistics, such as mean (*M*) and standard deviation (*SD*) to address the scalability research question (RQ1) are reported. Here, scalability refers to the increased implementation of an evidence-based practice across states, programs, personnel, and children including training needed to promote and support use at scale (Fixen et al., 2005; Greenwood et al., 2018; Milat et al., 2015).

To address the questions about children's EPSI growth trajectories (RQ2 and RQ4), we conducted growth curve analyses (GCA) using the multilevel modeling framework (MLM; Snijders & Bosker, 2012). The GCA fits longitudinal data into a hypothetical growth model and estimates the growth parameters, such as means of intercept, linear slope, and quadratic slope. These parameters define the shape of the growth trajectory that best fits the data. One advantage of using GCA in the MLM framework is that it can appropriately handle unequal spacing of time and unbalanced data (Raudenbush & Bryk, 2002). It is suitable for this study because children entered and left programs at different ages and for different reasons (e.g., moved in or away, aged out, etc.). Another advantage of this approach is that the intercept value can be estimated at any one age at the discretion of the researchers. We estimated the intercept at 36 months of age because it is an endpoint for children exiting IDEA Part C (Infant/Toddler) early intervention services and a transition point for the onset of Part B (Preschool) intervention services.

To address sensitivity to growth in the EPSI Score trajectory (RQ4), we conducted a two-step preliminary analysis to identify the appropriate growth models. In the first step, we compared different empty models (i.e., random intercept models) to determine the extent to which change over time was observed due to children, accessors and programs. As a result, a two-level cross-classified model (i.e., assessment at level 1 nested within children and assessors at level 2, but without the program level) was identified as the best fitting model for the EPSI based on a series chi-squared difference tests ($\Delta\chi^2[1]$) and comparisons of AICs (a.k.a., Akaike's Information Criteria). In the second step, we carried out a chi-squared difference test to identify the correct trajectory shape (linear vs quadratic) for the EPSI. The result showed that a quadratic model better described the trajectory ($\Delta\chi^2[1] = 11.95, p < 0.001$). After the preliminary analysis, we estimated the benchmark growth trajectories representing the mean trajectory as well as the −1.5 and –1.0 and +1.5 *SD* trajectories over the age span. These benchmarks are helpful when deciding whether an individual child's growth is falling within or outside of the range of expectation for their chronological age. Equation (1) illustrates the model used for this trajectory in which the mean intercept (γ_{00}), linear slope (γ_{10}), and quadratic slope (γ_{20}), were parameters of interest. Conducted through R (v4.1.0), variances of the residual terms u_{0j} and v_{0k} reflected the random effects due to children and accessors, respectively.

Level 1

Level 2

$$
\begin{aligned}
y_{i(jk)} &= \beta_{0(jk)} + \beta_{1(jk)} Age_{i(jk)} + \beta_{2(jk)} Age^2_{i(jk)} + e_{i(jk)} \\
\beta_{0(jk)} &= \gamma_{00} + u_{0j} + v_{0k} \\
\beta_{1(jk)} &= \gamma_{10} \\
\beta_{2(jk)} &= \gamma_{20}
\end{aligned} \tag{1}
$$

We then estimated a multivariate growth curve model for the key skills and compared all features to a theoretical continuum of growth in problem-solving proficiency (RQ4). The advantages of a multivariate compared to univariate growth model are more accurate standard errors plus information on the interrelationships between and among the key skill trajectories. After the preliminary steps of examining the correct clustering structure and shape of trajectory for each key skill, we estimated quadratic trajectories of all four key skills in a joint multivariate growth curve model considering random intercepts of children and assessors. The chi-squared differences comparing a linear vs. a quadratic model were 4.657 ($p = .030$) for Looks, 8.322 ($p = .003$) for Explores, 6.341 ($p = .011$) for Functions, and 78.254 ($p < .001$) for Solutions. This joint multivariate growth curve model was estimated using Mplus (v7.4) with Bayes estimator (see Mplus code for RQ4 in Supplemental Material).

To address the moderation question (RQ5), we examined the effect of three child-level moderators (i.e., gender, home language, and IFSP status) by adding them one at a time in the model and estimating the cross-level interactions between them and the growth parameters (equation (2)). The key parameters of interest in this model were those associated with the moderator: the differences in intercept (γ_{01}), linear slope (γ_{11}), and quadratic slope (γ_{21}). This analysis was conducted through R (v4.1.0). We used Cohen's *d* (equation (7) in Feingold [2009]) to estimate effect sizes for moderators.

Level 1

Level 2

$$
\begin{aligned}
y_{i(jk)} &= \beta_{0(jk)} + \beta_{1(jk)} Age_{i(jk)} + \beta_{2(jk)} Age^2_{i(jk)} + e_{i(jk)} \\
\beta_{0(jk)} &= \gamma_{00} + \gamma_{01} Moderator_{(jk)} + u_{0j} + v_{0k} \\
\beta_{1(jk)} &= \gamma_{10} + \gamma_{11} Moderator_{(jk)} \\
\beta_{2(jk)} &= \gamma_{20} + \gamma_{21} Moderator_{(jk)}
\end{aligned} \tag{2}
$$

Results

Was the EPSI Scalable Within and Across Early Childhood Programs (RQ1)?

The potential for scaling up the EPSI was demonstrated given the number and range of early childhood programs whose staff implemented the EPSI. The total number of

Table 2. Number of Early Problem Solving Indicator Assessments by Age and Total *M*, *SD*.

Age at test (months)	EPSI assessments	*M*	*SD*	Age at test (months)	EPSI assessments	*M*	*SD*
6	101	10.9	5.1	25	177	20.0	9.1
7	96	9.7	4.7	26	188	21.8	9.4
8	74	9.3	4.4	27	206	20.9	8.6
9	115	10.4	5.2	28	137	22.0	8.1
10	108	10.1	5.2	29	173	22.6	8.7
11	101	11.1	6.0	30	160	24.7	10.5
12	117	10.4	4.6	31	184	24.3	9.5
13	123	12.5	5.4	32	172	24.9	9.7
14	140	12.5	5.8	33	225	25.4	9.0
15	129	13.2	6.5	34	199	26.4	10.8
16	149	14.8	6.7	35	215	27.7	10.1
17	160	15.3	5.9	36	148	27.1	10.7
18	154	15.3	6.5	37	132	26.4	10.3
19	161	16.7	7.6	38	104	28.3	9.6
20	178	17.5	7.3	39	122	27.2	10.6
21	174	17.9	7.4	40	102	29.0	9.8
22	175	18.3	7.4	41	103	29.1	12.0
23	169	18.2	7.8	42	90	28.7	10.6
24	173	20.0	7.8				
Total					5,434		
M					20.0		
SD					10.3		
Min					0		
Max					68.83		

Note. $n = 2{,}614$ children. Assessments after removal of 43 outliers.

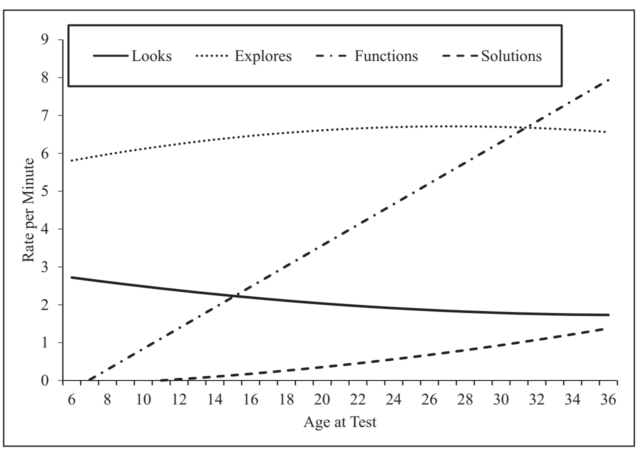

Figure 2. EPSI key skill elements fitted growth trajectories.

children with EPSI data used in these analyses was 2,614 with a total of 5,434 EPSI assessments administered and scored by certified program staff. The number of EPSI assessments at each month of age (6–36 months) ranged from 74 to 225 with the TWPS score growing from a mean of 10.9 (*SD* = 5.1) per min at 6 months to 27.1 (*SD* = 10.7) per minutes at 36 months (Table 2). Children were assessed on a rolling basis at different ages as they entered and exited programs.

Did Patterns of Growth in EPSI Key Skill Trajectories Reflect a Continuum (RQ2)?

A continuum was reflected in several ways. The key skill trajectories were different in their age of onset and the overall trend and change in growth rate over time (Figure 2; Table 3). At 6 months of age, infants were looking at and exploring the EPSI toys. Infants Explored toys at a rate of less than 6 per minute. The linear growth rate of Looks decreased across this age span, as the rate of Functions emerged and accelerated substantially over time reaching a rate close to eight occurrences per minute by 36 months. Solutions emerged slowly beginning at around 11 months of age and continued building slowly to a rate of 1.4 per minute by 36 months.

Ages at onset and visual patterns of growth trajectories suggested a complex continuum of cognitive problem-solving skills overtime. Looks and Explores are observed to be in children's repertoire at 6 months, wherein Functions and Solutions emerge closer to when children become toddlers (Figure 2). The extent to which children only Looked at toys decreased as engagement in Functions or Solutions of

Table 3. EPSI Key Skills Growth Models (Age Centered at 36 Months).

Model		Intercept	Linear slope	Quadratic slope
Looks	Est.	1.731	−0.003	0.001
	Post. *SD*	0.093	0.007	0.000
	CI lower	1.487	−0.015	0.001
	CI upper	1.888	0.011	0.002
Explores	Est.	6.561	−0.035	−0.002
	Post. *SD*	0.192	0.016	0.001
	CI lower	6.221	−0.066	−0.003
	CI upper	6.955	−0.006	−0.001
Functions	Est.	7.936	0.273	0.000
	Post. *SD*	0.168	0.016	0.001
	CI lower	7.599	0.242	−0.001
	CI upper	8.269	0.301	0.001
Solutions	Est.	1.377	0.080	0.001
	Post. *SD*	0.055	0.005	0.000
	CI lower	1.270	0.071	0.001
	CI upper	1.482	0.089	0.001

Note. Non-significant estimates are underscored. CI lower = the lower limit of 95% credible interval; CI upper = the upper limit of 95% credible interval.

toys increased. Correlations between the growth parameters of the four key skills provided additional evidence. Small to medium correlations were found between the intercepts and slopes of Looks with Explores, Functions, and Solutions. Explores had small to medium correlations with Solutions. All growth parameters of Functions were correlated to those of Solutions, the largest between-skill correlation ($r = .59$) was found between Functions and Solution intercepts at 36 months (see Table S5).

Because of these patterns of accelerating and decelerating key skills over the age range, we applied a score weighting convention to the key skills to provide a positively accelerating benchmark indicator of EPSI TWPS. The TWPS is a sum of 1 for each occurrence of Look and Explore, 2 for each Function, and 3 for each Solution. Weighting also allowed for occurrence of more advanced problem-solving behaviors to have a larger contribution to the TWPS benchmark score. These calculations are made automatically when assessors entered raw scores into the IGDI Online Data System.

Were the TWPS Rate Scores Reliable (RQ3)?

Several forms of CTT total score reliability were demonstrated. The Pearson *r* correlations between the three toy types for 2-minute total scores and the TWPS score were 0.84, 0.80, and 0.84, respectively, for pop-up toys, stacking cups, and the ball drops. These correlations are indicative of internal consistency. Two groups equivalence analysis demonstrated that the groups' TWPS grand mean scores' were equivalent and distributionally similar ($M_{Grp1} = 20.00$, $SD = 10.32$, $n = 2665$ vs. $M_{Grp2} = 20.04$, $SD = 10.20$, $n = 2769$). The TWPS mean difference of 0.034 between groups was minor ($d = 0.003$) as was the narrow 95% confidence interval of −0.057 to 0.050. The groups' mean linear slopes were 0.60 and 0.63 per minute per month of age, respectively, with age accounting for 36% versus 31% of the variance in TWPS in each.

Were EPSI TWPS Trajectories Sensitive to Growth Over Time (RQ4)?

The Total EPSI score trajectories were sensitive to growth over time. The growth trajectory for the Total EPSI score was estimated from the quadratic model (equation (1)). On average, children showed a rate of 29.128 ($SE = 0.476$, $p < .001$) total cognitive problem-solving behaviors per minute at 36 months. The pattern of growth in the Total EPSI score was captured by a linear slope of 0.851 ($SE = 0.044$, $p < .001$) and a positive acceleration rate (i.e., quadratic slope) of 0.005 ($SE = 0.001$, $p < .001$). These growth parameters were used to compute benchmark means for each month of age from 6 to 36 months. Scores were computed at 1.5 *SD* above, 1 and 1.5 *SD* below benchmark for purposes of screening and progress monitoring (Figure 3).

Was Growth in the Total EPSI Score Moderated by Child Characteristics (RQ5)?

Differences in children's Total EPSI growth scores by Gender or Home languages, were not significant in intercept, linear slope, or quadratic slope (Table S4). Children with vs. without an IFSP also were not significantly different. When we examined disability types however, we found

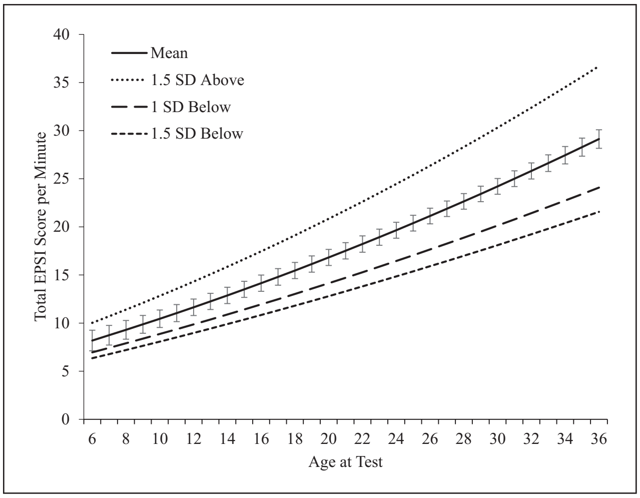

Figure 3. Total EPSI cognitive problem solving fitted mean, standard deviation benchmark trajectories.

that children with a cognitive disability concern for early intervention services were 8.214 per min lower in their Total EPSI score than children without an IFSP at 6 months (SE = 3.125, p = .009). Cohen's d for this difference was −0.844, suggesting a large effect. This difference decreased over time, and became −3.075 (SE = 1.603, p = .055) at 36 months showing a small effect (Cohen's d = −0.316). Differences between children with speech, motor, unspecified disabilities, and those with no disability were non-significant (see Figure S4).

Discussion

Proficiency in problem-solving is an important outcome in early childhood related to promoting cognitive development. Early educators need a valid, practical measurement tool for intervention decision making. We sought to advance the validity claims behind the Early Problem-Solving Indicator (EPSI) by examining new evidence likely to strengthen the rationale for using the EPSI for the universal screening and progress monitoring of infants and toddlers. The present study built on earlier research and development of the EPSI (Greenwood, Walker, et al., 2006), extending the child age range downward from 14 to 6 months of age and increasing the volume of EPSI data by scaling to a large number of programs and interventionists who used the EPSI data that they collected in the education and treatment of infants and young children.

Results indicated that the EPSI was successfully implemented by 45 programs, their staff (N = 175) and children (N = 2,614) using the website and its tools to produce a total of N = 5,434 EPSIs. The EPSIs key skill trajectories within and across the age range did function as a continuum supported by several outcomes. These were mean level, sequencing of and variation in skill onset at 6 months and thereafter, patterns of change within and across key skills, patterns of the correlations among key skills trajectories, and differential mean intercepts at 36 months of age. Additionally, with and across key skills and age in months, patterns of correlation indicated numerous concurrent and predictive interrelationships.

The rate of Explores started highest, growing slightly before reaching a plateau and then gradually decelerating as the rate of Functions overlapped and exceeded the rate of Explores. Similarly, the rate of Looks started higher than either Functions or Solutions, but slowly decreased as the rate of Functions increased followed by the onset of Solutions. Trajectories of the EPSI key skills in this study were consistent with previous reports (Greenwood, Walker, et al., 2006). Growth patterns that emerged for Looks and Explores compared to Functions and Solutions permitted adjustment and updating of the constructs in the conceptual model for the EPSI. Looks and Explores are grouped into a combined construct of Visual and Tactile Exploration behaviors that precede Means-end behaviors of Functions and Solutions (Figure 1).

CTT reliability analyses indicated that the TWPS score was reliable. In the present study, the mean TWPS score followed an increasing rate of progress in cognitive problem-solving accelerating from 6 to 36 months of age. The steepest TWPS trajectory occurred for children whose scores were +1.5 SD above the mean. Children with TWPS score trajectories that fell between −1.0 and −1.5 SD below the mean had slower growth trajectories but still paralleled the shape of the mean TWPS score. Taken together, these distributional benchmarks provide important comparatives for evaluating an individual child's trajectory in reference to the mean across months of age as well as slope forecast regarding future growth.

Moderation analyses indicated that children with an indicated cognitive disability were observed to moderate the TWPS score. These findings are consistent with those for other Infant-Toddler IGDIs including the EMI (Greenwood et al., 2018), ECI (Greenwood et al., 2013), and the ESI (Greenwood et al., 2021) contributing further to the suite of IGDI measures available for measuring skills needed by infants and toddlers for early development and later readiness for school. While differences in performance moderated by child home language have been reported for some IGDIs (Greenwood et al., 2018), home language did not moderate child EPSI performance. Gender was not a significant moderator of TWPS score either. Child disability status was found to be associated with lower, but not statistically significant, TWPS scores.

Limitations

While this sample of programs, caregivers, and children was recruited across seven states from Early Head Start, Part C, and community child care programs, this was a sample of convenience because of the relevance of the EPSI to their program goals and aspirations to serve children and families. The sampling frame employed was not nationally representative. The IGDI data management system used by programs is a practice-based system developed over the years with input from practitioners and did not include in children's profiles information about race, ethnicity, or family SES, programmatic variables describing curricular content, interventions, staffing demographics, treatment fidelity, or intervention dosage. While many of the programs implementing the EPSI had income eligibility requirements (EHS/HS), personal information about family SES was not collected in the IGDI data system. Thus, it was not possible to examine the moderating influences of these variables in this study.

Training administration and certification for scoring are required for those using the Infant-Toddler IGDIs including the EPSI. Collection of interrater reliability, while encouraged and supported through training and the IGDI Online Data System, is not required. An annual recertification calibration process entailing scoring videotaped EPSI administrations previously scored by a certified trainer is recommended. Therefore, while interrater reliability is reported elsewhere (Greenwood, Walker, et al., 2006; Walker & Greenwood, 2010), it is not for this study.

Future Research

This present study of the EPSI advanced what is known about its technical properties and utility by building out a large sample of diverse programs and children, including infants and young children with and without disabilities, as well as, children who are dual language learners. Further work is needed to make the EPSI data archive nationally representative. Next step investigations in the validation of the EPSI include multigroup invariance analysis of psychometric equivalence (Greenwood et al., 2013) and prospective analyses addressing the reliability, criterion, and predictive validity of the EPSI. Other next steps will focus on demonstrating that the EPSI is sensitive to intervention effects for individuals and groups of children. The EPSI was recently included in a pilot demonstration of remote administration of IGDIs in an EHS program during the COVID-19 pandemic (Greenwood et al., 2021). Further investigation of remote application of the EPSI will be explored.

Implications for Practice

Results indicated that the EPSI, in its present form, can be scaled for use across child care, early learning, and intervention programs. Staff who used the EPSI for progress monitoring and screening represented early educators and interventionists in center- and home-based programs infant-toddler programs located in seven U.S. states. The sample included infants and children attending EHS and Part C programs in addition to early learning and child care and intervention staff. We were able to establish comparative benchmarks for a larger number of infants and young children receiving early childhood services than previously reported for the EPSI. The updated technical standards should serve to increase the confidence in using this measure to inform practice based on child-level data to improve interventions that promote children's cognitive problem-solving skills.

Author Note

The opinions expressed herein are those of the authors and do not represent the views of the U.S. Department of Education, Institute for Education Sciences (IES), the National Center for Special Education Research (NCSER), the National Institutes of Health (NIH), or other agencies.

Acknowledgments

We acknowledge the contributions of colleagues Judith J. Carta, Susan Higgins, Christine Muehe, Debra Montagna, and Stacy Johner to this work. We also recognize our research partners in the participating programs, the teachers and home visitors, the participating children and their families, without whom this research would not be possible.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grants from the Institute for Education Sciences (IES), National Center for Special Education Research (NCSER) #R324A150166 U.S. Department of Education to the University of Kansas. Additional support for this work was received from the Kansas Intellectual and Developmental Disabilities Research Center, National Institutes of Health (NIH) #U54HD090216, University of Kansas, Schiefelbusch Institute for Life Span Studies.

ORCID iD

Charles R. Greenwood https://orcid.org/0000-0002-6274-3075

Supplemental Material

Supplementary material for this article is available on the Topics in Early Childhood Special Education website at http://tecse.sagepub.com.

References

Akers, L., Del Grosso, P., Snell, E., Atkins-Burnett, S., Wasik, B. A., Carta, J. J., Boller, K., & Monahan, S. (2016). Tailored teaching: Emerging themes from the literature on teachers;

use of ongoing child assessment to individualize instruction. *Dialog*, *18*(3), 133–150.

Babik, I., Cunha, A. B., Ross, S. M., Logan, S. W., Galloway, J. C., & Lobo, M. A. (2019). Means-end problem solving in infancy: Development, emergence of intentionality, and transfer of knowledge. *Developmental Psychobiology*, *61*, 191–202.

Bagnato, S. J., Goins, D. D., Pretti-Frontczak, K., & Neisworth, J. T. (2014). Authentic assessment as “best practice” for early childhood intervention: National consumer social validity research. *Topics in Early Childhood Special Education*, *34*(2), 116–127. https://doi.org/10.1177/0271121414523652

Bagnato, S. J., McLean, M., Macy, M., & Neisworth, J. T. (2011). Identifying instructional targets for early childhood via authentic assessment: Alignment of professional standards and practice-based evidence. *Journal of Early Intervention*, *33*(4), 243–253. https://doi.org/10.1177/1053815111427565

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., & Grimm, K. J. (2009). The contributions of ‘hot’ and ‘cool’ executive function to children’s academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly*, *24*, 337–349. https://doi.org/10.1016/j.ecresq.2009.06.001

Brooker, R. J., Buss, K. A., Lemery-Chalfant, K., Aksan, N., Davidson, R. J., & Goldsmith, H. H. (2013). The development of stranger fear in infancy and toddlerhood: Normative development, individual differences, antecedents, and outcomes. *Developmental Science*, *16*(6), 864–878. https://doi.org/10.1111/desc.12058

Buzhardt, J., Greenwood, C. R., Jia, F., Walker, D., Schneider, N., Larson, A. L., Valdovinos, M. G., & McConnell, S. R. (2020). Technology to guide data-driven intervention decisions: Effects on language growth of young children at risk for language delay. *Exceptional Children*, *87*(1), 74–91. https://doi.org/10.1177/0014402920938003

Buzhardt, J., Greenwood, C. R., Walker, D., Anderson, R., Howard, W., & Carta, J. J. (2011). Effects of web-based support on Early Head Start home visitors’ use of evidence-based intervention decision making and growth in children’s expressive communication. *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field*, *14*(3), 121–146.

Buzhardt, J., Greenwood, C., Walker, D., Jia, F., Schnitz, A., Higgins, S., Montagna, D., & Muehe, C. (2018). Web-based Support for data-based decision making: Effect of intervention implementation on children’s communication. *Journal of Early Intervention*, *40*(3), 246–267. https://doi.org/10.1177/1053815118788059

Buzhardt, J., Leonard, J., Ai, J., Higgins, S., Greenwood, C., Consolver, K., Walker, D., & Carta, J. (2022). Technology to facilitate progress monitoring of infant-toddler growth and development: Measuring implementation in community-based agencies. *Journal of Special Education Technology*. Advance online publication. https://doi.org/10.1177/01626434221108882

Carta, J. J., Greenwood, C. R., Atwater, J., McConnell, S. R., Goldstein, H., & Kaminski, R. A. (2015). Identifying preschool children for higher tiers of language and early literacy instruction within a response to intervention framework. *Journal of Early Intervention*, *36*, 281–291. https://doi.org/10.1177/1053815115579937

Carta, J. J., Greenwood, C. R., Luze, G. J., Cline, G., & Kuntz, S. (2004). Developing a general outcome measure of growth in social skills for infants and toddlers. *Journal of Early Intervention*, *26*(2), 91–114. https://doi.org/10.1177/105381510402600203

Carta, J. J., Greenwood, C. R., Walker, D., Kaminski, R., Good, R., McConnell, S., & McEvoy, M. (2002). Individual growth and development indicators (IGDIs): Assessment that guides intervention for young children. In M. Ostrosky & E. Horn (Eds.), Assessment: Gathering meaningful information. *Young Exceptional Children Monograph Series* (Vol. 4, pp. 15–28). Sopris West.

Charkaluk, M. L., Rousseau, J., Calderon, J., Bernard, J. Y., Forhan, A., Heude, B., & Kaminski, M. (2017). Ages and Stages Questionnaire at 3 years for predicting IQ at 5–6 years. *Pediatrics*, *39*(4), e20162798. https://doi.org/10.1542/peds.2016-2798

Colombo, J. (1993). *Infant cognition: Predicting later intellectual functioning*. SAGE.

Colombo, J. (2004). Visual attention in infancy: Process and product in early cognitive development. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 329–341). Guilford.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, *333*(6045), 959–964.

Diamond, K. E., Justice, L. M., Siegler, R. S., & Snyder, P. A. (2013). *Synthesis of IES research on early intervention and early childhood education (NCSER 2013-3001)*. http://ies.ed.gov/ncser/pubs/20133001/pdf/20133001.pdf

Division for Early Childhood. (2014). *DEC recommended practices in early intervention/early childhood special education*. http://www.dec-sped.org/recommendedpractices

Embretson, S., & Yang, X. (2006). Item response theory. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 385–409). Lawrence Erlbaum Associates Publishers.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods*, *14*(1), 43–53.

Fitzpatrick, C., & Pagani, L. S. (2012). Toddler working memory skills predict kindergarten school readiness. *Intelligence*, *40*, 205–212. https://doi.org/10.1016/j.intell.2011.11.007

Fixen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models., *Exceptional Children*, *57*, 488–500.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60.

Goodman, J. F. (2002). Infant intelligence: Do we, can we, should we assess it? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment* *of children: Intelligence and achievement* (pp. 183–208). Guilford.

Greenwood, C. R., Buzhardt, J., Walker, D., Jia, F., & Carta, J. J. (2020). Criterion validity of the early communication indicator for infants and toddlers. *Assessment for Effective Intervention*, *45*(4), 298–310. https://doi.org/10.1177/1534508418824154

Greenwood, C. R., Carta, J. J., & McConnell, S. (2011). Advances in measurement for universal screening and individual progress monitoring of young children. *Journal of Early Intervention*, *33*(4), 254–267. https://doi.org/10.1177/1053815111428467

Greenwood, C. R., Carta, J. J., Walker, D., Hughes, K., & Weathers, M. (2006). Preliminary investigations of the application of the Early Communication Indicator (ECI) for infants and toddlers. *Journal of Early Intervention*, *28*(3), 178–196. https://doi.org/10.1177/105381510602800306

Greenwood, C. R., Higgins, S., McKenna, M., Buzhardt, J., Walker, D., Ai, J., Irvin, D., & Grasley-Boy, N. (2021). Remote use of individual growth and development indicators (IGDIs) for infants and toddlers. *Journal Early Intervention*, *44*, 168–189. https://doi.org/10.1177/10538151211057552

Greenwood, C. R., Luze, G. J., Cline, G., Kuntz, S., & Leitschuh, C. (2002). Developing a general outcome measure of growth in movement for infants and toddlers. *Topics in Early Childhood Special Education*, *22*(3), 143–157.

Greenwood, C. R., Walker, D., & Buzhardt, J. (2010). The early communication indicator for infants and toddlers: Early head start growth norms from two states. *Journal of Early Intervention*, *32*(5), 310–334.

Greenwood, C. R., Walker, D., Buzhardt, J., Howard, W., McCune, L., & Anderson, R. (2013). Evidence of a continuum in foundational expressive communication skills. *Early Childhood Research Quarterly*, *28*, 540–554.

Greenwood, C. R., Walker, D., Buzhardt, J., Irvin, D., Schnitz, A. G., & Fan, J. (2018). Update on the EMI for infants and toddlers. *Topics in Early Childhood Special Education*, *38*(2), 105–117. https://doi.org/10.1177/0271121418777290

Greenwood, C. R., Walker, D., Carta, J. J., & Higgins, S. (2006). Developing a general outcome measure of growth in the cognitive abilities of children 1 to 4 years old: The early problem-solving indicator (EPSI). *School Psychology Review*, *35*, 535–551. https://doi.org/10.1080/02796015.2006.12087960

Head Start Performance Standards 45CFR, Sec. 1304

Hupp, S., & Abbeduto, L. (1991). Persistence as an indicator of mastery motivation in young children with cognitive delays. *Journal of Early Intervention*, *15*, 219–225.

Individuals with Disabilities Act (IDEA 2004) (PL. 108-446) (2004).

Institute of Medicine and National Research Council. 2000. *From neurons to neighborhoods: The science of early childhood development*. The National Academies Press. https://doi.org/10.17226/9824

Jabrayilov, R., Emons, W., & Sijtsma, K. (2016). Comparison of classical test theory and item response theory in individual change assessment. *Applied Psychological Measurement*, *40*(8), 559–572. https://doi.org/10.1177/0146621616664046

Keen, R. (2011). The development of problem-solving in young children: A critical cognitive skill. *Annual Review of Psychology*, *62*(1), 1–21.

Kincaid, A., McConnell, S. R., & Wackerle-Hollman, A. K. (2020). Assessing early literacy growth in preschoolers using individual growth and development indicators. *Assessment for Effective Intervention*, *45*(3), 173–183. https://doi.org/10:1177/1534508418799173

Lambert, R. (2020). *Technical manual for the teaching strategies GOLD® assessment (second edition): Birth through third grade*. Center for Educational Measurement and Evaluation, University of North Carolina Charlotte. https://teachingstrategies.com/wp-content/uploads/2020/10/2020-Tech-Manual_GOLD.pdf

Landry, S. H., Smith, K. E., & Swank, P. R. (2009). New directions in evaluating social problem-solving in childhood: Early precursors and links to adolescent social competence. *New Directions for Child and Adolescent Development*, *123*, 51–68.

Limbos, M. M., & Joyce, D. P. (2011). Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *Journal of Developmental and Behavioral Pediatrics*, *32*(7), 499–511.

Luze, J. L., Linebarger, D. L., Greenwood, C. R., Carta, J. J., Walker, D., Leitschuh, C., & Atwater, J. (2001). Developing a general outcome measure of growth in expressive communication of infants and toddlers. *School Psychology Review*, *30*(3), 383–406.

Martin, A., Ryan, R. M., & Brooks-Gunn, J. (2013). Longitudinal associations among interest, persistence, supportive parenting, and achievement in early childhood. *Early Childhood Research Quarterly*, *28*(4), 658–667.

McWayne, C. M., Fantuzzo, J. W., & McDermott, P. A. (2004). Preschool competency in context: An investigation of the unique contribution of child competencies to early academic success. *Developmental Psychology*, *40*(4), 633.

Milat, A. J., Bauman, A., & Redman, S. (2015). Narrative review of models and success factors for scaling up public health interventions. *Implementation Science*, *10*, 113.

Moreno, A., & Klute, M. A. (2011). Infant-toddler teachers can successfully employ authentic assessment: The learning through relating system. *Early Childhood Research Quarterly*, *26*, 484–496. https://doi.org/10.1016/j.ecresq.2011.02.005

National Research Council. (2001). Eager to learn: Educating our preschoolers. Committee on early childhood pedagogy. In Bowman, B. T., et al. (Eds.), *Commission on behavioral and social sciences and education*. (pp. 306–322). National Academy Press.

Neisworth, J. T., & Bagnato, S. J. (2004). The mismeasure of young children: The authentic assessment alternative. *Infants & Young Children*, *17*(3), 198–212.

Office of Head Start. (2012). *Report to congress on the final head start program designation renewal system*. https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/head-start-designation-renewal-system-final-rule.pdf

Priest, J., McConnell, S., Walker, D., Carta, J., Kaminski, R. A., McEvoy, M. A., Good, R. H., III, Greenwood, C. R., & Shinn, M. R. (2001). General growth outcomes for children: Developing a foundation for continuous progress measurement. *Journal of Early Intervention*, *24*, 163–180. https://doi.org/10.1177/10538151010240030101

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE.

Siegler, R. S., & Alibali, M. W. (2005). *Children's thinking* (4th ed.). Prentice Hall.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). SAGE.

Squires, J., & Bricker, D. (2009). *Ages and Stages Questionnaire (ASQ): A parent completed child monitoring system* (3rd ed.). Brooks Publishing Company.

Stichter, J. P., Christ, S. E., Herzog, M. J., O'Donnell, R. M., & O'Connor, K. V. (2016). Exploring the role of executive functioning measures for social competence research. *Assessment for Effective Intervention*, *41*(4), 243–254. https://doi.org/10.1177/1534508416644179

VanDerHeyden, A. M., & Snyder, P. (2006). Integrating frameworks from early childhood intervention and school psychology to accelerate growth for all young children. *School Psychology Review*, *35*(4), 519–534. https://doi.org/10.1080/02796015.2006.12087959

Walker, D., Carta, J. J., Greenwood, C., & Buzhardt, J. (2008). The use of individual growth and development indicators for progress monitoring and intervention decision making in early education. *Exceptionality*, *16*, 33–47. https://doi.org/10.1080/09362830701796784

Walker, D., & Greenwood, C. (2010). The cognitive problem-solving IGDI: Early problem solving indicator. In J. J. Carta, C. R. Greenwood, D. Walker, & J. Buzhardt (Eds.), *Using IGDIs: Monitoring progress and improving intervention results for infants and young children* (pp. 57–73). Brookes.

Willatts, P. (1999). Development of means-end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental Psychology*, *35*, 651–667.

Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research Online*, *8*(3), 1–22. http://www.stat.cmu.edu/~brian/PIER-methods/For%202013-02-28/Readings/building%20tests/mpr122_8.pdf

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum.

Wilson, M. (2013). Using the concept of a measurement system to characterize measurement models used in psychometrics. *Measurement*, *46*(9), 3766–3774.

Zelazo, P. D., Carter, A., & Reznick, J. S. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology*, *1*(2), 198–226.