Early Childhood Research Quarterly 28 (2013) 743–758

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Advancing the construct validity of the Early Communication Indicator (ECI) for infants and toddlers: Equivalence of growth trajectories across two early head start samples☆

Charles R. Greenwood*, Jay Buzhardt, Dale Walker, Luke McCune, Waylon Howard

Juniper Gardens Children's Project, University of Kansas, United States

ARTICLE INFO

Article history:
Received 17 December 2012
Received in revised form 3 July 2013
Accepted 11 July 2013

Keywords:
Communication
Infants and toddlers
Progress monitoring assessment
Validity
Latent growth curve analysis
Invariance testing

ABSTRACT

The Early Communication Indicator (ECI) is a measure for universal screening, intervention decision-making, progress monitoring for infants and toddlers needing higher levels of support, and program accountability. In the context of the ECI's long-term wide-scale use for these purposes, we examined the invariance of ECI measurement in two samples of the same Early Head Start (EHS) population differing in the years data were collected. Invariance or equivalence across samples is an important step in measurement validation because making inferences assumes that the measurements are factorially invariant. A number of time-covarying factors (e.g., assessors, children, etc.) can be hypothesized as threats to measurement invariance. Results of latent growth curve analyses indicated similarity in the functional forms (velocity and shape) of the ECIs four key skill trajectories between groups of children and ECI vocalizations, single, and multiple words trajectories met strong factorial and structural invariance. Gestures met only weak factorial invariance. ECI total communications, a weighted composite of the four scales, also met both strong factorial and structural invariance. With one exception, results indicated that the ECI produced comparable growth estimates over different conditions of programs, assessors, and children over time, strengthening the construct validity of the ECI. Implications are discussed.

The purpose of this paper was the psychometric examination of the extent that an increasingly used measure of growth in very young children's early communication skills, the Early Communication Indicator (Carta, Greenwood, Walker, & Buzhardt, 2010), provided comparable measurement in two independent samples of Early Head Start children. Inferences made from a measure need to be grounded in the knowledge of whether or not under such different conditions of assessing different children and observing them across time, that obtained scores are yielding the same attributes (American Educational Research Association, 1999; Cheung & Rensvold, 2002; Little & Slegers, 2005). Developers of measurement systems are obligated to ensure that decisions based on data from the system lead to the intended and not unintended results (Council of Chief State School Officers, 2004).

Converging evidence indicates that oral language is an important early predictor of young children's readiness for school and later success learning to read (Biemiller, 2006; NICHD Early Child Care Research Network, 2005; Shanahan & Lonigan, 2008). Oral language knowledge enables children to label and understand the world, communicate with others, and regulate behavior to achieve goals. Language enables young children to act on what is said to them, and to learn and use new concepts as well as develop a foundation of basic knowledge.

The development of oral language is commonly considered to advance from prelinquistic forms of communication (gestures, vocalizations, babbling) to spoken language (single and multiple word utterances). For example, early gesture use predicts later vocabulary development (Acredolo & Goodwyn, 1988;

☆ The research reported here was supported by grants from the Institute of Education Sciences, National Center for Special Education Research (R324A070085), and the Office of Special Education Programs (H324C040095, H327A060051), U.S. Department of Education to the University of Kansas. Additional support was provided by the Kansas Intellectual and Developmental Disabilities Research Center, National Institutes of Health (HD002528), Schiefelbusch Institute for Life Span Studies, Kansas Social Rehabilitation Services, and the regional Early Head Start Association. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, the National Institutes of Health, or other agencies. We gratefully acknowledge the contributions of colleagues Judith J. Carta, Debra Montagna, Barbara Terry, Christine Muehe, Susan Higgins, Matt Garrett, April Fleming, Constance Beecher, and Chia-Fen Liu. We thank Todd Little, Wei Wu, and Elizabeth Grandfield of the Kansas Center for Research Methods and Data Analysis for assistance with the latent growth curve modeling aspects of the study. We acknowledge our Kansas Head Start program research partners, participating programs, children and families, and Mary Weathers for her support in facilitating partnerships with Early Head Start.

* Corresponding author at: Juniper Gardens Children's Project, University of Kansas, 444 Minnesota Avenue, Suite 300, Kansas City, KS, 66101, United States. Tel.: +1 913 321 3143.
E-mail address: greenwood@ku.edu (C.R. Greenwood).

0885-2006/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ecresq.2013.07.002

Brady, Marquis, Fleming, & McLean, 2004; Iverson, Longobardi, & Caselli, 2003; Rowe & Goldin-Meadow, 2009). Predictive relations between gestures and two-word combinations have been reported (Vandereet, Maes, Lembrechts, & Zink, 2011) and between early word use and later social-emotional development (Vallotton & Ayoub, 2010).

Paul and Roth (2011) reported that early risk factors for future language delays included fewer gestures, vocalizations, babbling, and later onset of speech sounds and word combinations (Dale, Price, Bishop, & Plomin, 2003; Rice, Taylor, & Zubrick, 2008). Delay in the onset of some forms of babbling has been associated with a higher risk for speech and language-related disorders (Oller, Eilers, Neal, & Schwartz, 1999; Yoder, Warren, & Macathren, 1998), as has delay in onset of early words (Fenson et al., 1994) or word combinations (Zubrick, Taylor, Rice, & Slegers, 2007). Children from poverty backgrounds are at particular risk for poor language outcomes as highlighted in the results from the Early Head Start (EHS) National evaluation study (Love, Chazan-Cohen, Raikes, & Brooks-Gunn, 2013).

This critical role played by oral language in development is reflected broadly in the outcome goals of early intervention programs that serve young children. Language and early literacy are mandated school readiness goals for infants and toddlers served by Early Head Start (Early Head Start National Resource Center, 2012). Many early childhood programs emphasize the promotion of young children's ability to communicate (National Early Childhood Accountability Task Force, 2007), and communication outcome goals are required by IDEA Part C programs serving infants and toddlers with developmental delays and disabilities (ECO Center, 2011; Individuals with Disabilities Education Improvement Act, 2004).

In addition to mandated federal and state accountability requirements for child results, the field is increasingly moving toward program-wide, prevention-oriented early intervention systems that discourage one size fits all and encourage individualizing or differentiating children's learning experiences through the use of formative measurement. Measures are needed to screen all children to identify need, modify intervention, and monitor short-term progress over time (Advisory Committee on Head Start Research & Evaluation, 2012).

For example, renewal of Head Start and Early Head Start (EHS) program grants is now contingent on programs use of child results to individualize instructional strategies that support each child's development (45 CFR Sec 1308.3(b)(2)(ii) (Office of Head Start, 2012)). Similarly, the joint position statement of the National Association for the Education of Young Children (NAEYC), Division of Early Childhood (DEC), and the National Head Start Association (NHSA) describes use of universal screening and progress monitoring measurement in the Response to Intervention (RTI) prevention-oriented approach (NAEYC, DEC, & NHSA, 2013). However, few measures exist that meet both high technical standards and are also practical for caregivers and early interventionists to use for informing their intervention and program improvement efforts (Buzhardt et al., 2011; Carta et al., 2002; Moreno & Klute, 2011).

The Individual Growth and Development Indicators (IGDIs) (Carta et al., 2010) address this need across early childhood from birth to age 5. IGDIs measure children's level of performance at a point in time (e.g., 16 months of age) and their rate of growth toward acquiring important general outcomes (i.e., language/communication, cognitive problem solving, movement, and social development (Carta et al., 2002). The Early Communication Indicator (ECI), for example, is a 6-min play-based, formative measure of infants and toddlers (6–42 months of age) growth in expressive communication (Carta et al., 2010). Because most children age-out of EHS and IDEA Part C services at 36 months, we have used 36 months of age as a mean intercept point in past ECI research even though in many cases data have been collected through 42 months of age (Greenwood, Buzhardt, Walker, Howard, & Anderson, 2011; Greenwood, Walker, & Buzhardt, 2010; Greenwood et al., 2013). The ECI, one of the infant-toddler IGDIs, is the focus of this report.

The ECI is usable for universal screening, intervention decision-making, progress monitoring, and program accountability (Carta et al., 2010; Greenwood, Carta, & McConnell, 2011). The ECI taps four foundational early communication skills on a child's developmental path from prelinguistic (i.e., gestures and vocalizations) to spoken language skills (i.e., single and multiple word utterances) (Greenwood et al., 2013; Walker & Carta, 2010). It is intended for use by early childhood personnel, brief to administer, and repeatable enough to provide estimates of short-term growth. The ECI is designed to be administered quarterly for universal screening of all children in a program and as frequently as monthly to monitor progress resulting from a change in intervention (Greenwood et al., 2008). Monthly benchmarks provide normative estimates of expected progress and assist in making intervention decisions.

The original 5-year ECI development and validation effort for infants/toddlers (Greenwood, Carta, & Walker, 2005; Greenwood & Walker, 2010) involved: (a) a national survey of parents of children with special needs and professionals in early childhood (EC) and early childhood special education (ECSE) that socially-validated expressive communication as an important general outcome of early intervention for young children (Priest et al., 2001), (b) studies documenting the psychometric properties and feasibility of the ECI, including sensitivity to growth over time (Greenwood, Buzhardt, & et al., 2011; Greenwood, Carta, Walker, Hughes, & Weathers, 2006; Greenwood et al., 2010; Luze et al., 2001), and (c) studies of the ECI's sensitivity to the effects of language promoting interventions (Buzhardt et al., 2011; Harjusola-Webb, 2006; Kirk, 2006).

The ECI's reported criterion validity with the Preschool Language Scale-3 (Zimmerman, Steiner, & Pond, 1992) was $r = 0.62$ and $r = 0.51$ and a parent rating of the child's language skills (Luze et al., 2001). Split-half reliability for the ECI's total communication score was $r = 0.89$ for mean level and $r = 0.62$ for slope. Alternate forms reliability was reported to be $r = 0.72$. Interrater agreement on the scoring of ECI assessments was 90% overall for total communication, ranging from a mean of 70% for single words to 81% for gestures (Greenwood & Walker, 2010).

In a sample of 5883 children ages 6–36 months served in EHS programs in two states, we reported that the shape of the mean ECI total communication trajectory was positively accelerating (Greenwood et al., 2010). The ECI total communication mean score, that is, a weighted composite of gestures, vocalizations, single and multiple words, at 6 months of age was 2.6 communications per minute. Total communication grew to 21.6 communications per minute (or 130 total communications in 6 min) by 36 months of age.

The shapes of trajectories of gestures, vocalizations, and single words we reported to be positively accelerating to a peak or inflection point followed thereafter by attenuation to a lower level over time (Greenwood et al., 2010). Growth in multiple word utterances was also positively accelerating and had not reached a peak by 36 months of age. Consistent with theory and prior reports, we observed a pattern of sequential ordering of adjacent skills from gestures to vocalizations, vocalizations to single words, and single words to multiple words suggestive of a continuum of skill development over time (Greenwood et al., 2013). Regarding the ECI total communication sensitivity to home-based language interventions, we reported effect sizes of $d = 0.47$ after 6 months of home-based intervention and $d = 0.79$ after 9 months between randomized groups (Buzhardt et al., 2011, 2010).

Website technology supports wide-scale access to the ECI and usability at reasonable costs and time investment (e.g., training, certification, protocols, data services including tools for director's

supervision and management, and individual- and program-level reporting (Buzhardt et al., 2010). Web-based reports are available to program staff at the individual child level and when aggregated across children in a program at the program level. Reports are also available at the project level when multiple programs within a state or region, for example, are using the ECI. Training and certification of new assessors are supported by materials and instructions for learning the administration and scoring protocols. Video calibration tools on the website are used to certify that an assessor's administration of the ECI has met a fidelity standard and that their coding of ECI skills is reliable compared to a calibration standard.

As described, prior work demonstrated that the ECI can be used to describe growth of early communication with demonstrated good construct and criterion validity, and accordingly, is being used in an increasingly large number of early childhood program, EHS in particular. However, the long-term, large scale use of an instrument like the ECI by program practitioners can be seen fraught with threats to quality of inferences made from the data. A few examples include (a) history (e.g., change in EHS policies and practices and services), (b) improvements in assessor training and certification, (c) drift in assessor's scoring of protocols, (d) improvements in the measurement instrument (e.g., definitional changes, revised norms, benchmarks for decision-making), and (e) improvements in internet and website infrastructure/technology, etc. We hypothesized such threats, perhaps expressed differently in the two groups participating longitudinally over this period that might influence invariance.

In the context of the ECI's wide-scale use for these purposes over the past 10 years, the opportunity arose to extend existing evidence supporting the psychometric properties of the ECI by examining the invariance of ECI measurement in two samples of the same Early Head Start (EHS) population. The samples differed in the time periods that the data were collected. The original sample consisted of individuals with a history of early involvement in use of the ECI as it first became available and supported by the first version of the website. The replication sample consisted of individuals with a much later history of involvement using an improving website. Of interest was whether or not the ECI measurement was factorially and structurally invariant given hypothesized threats, and whether or not current ECI training and implementation procedures appeared to be sufficiently robust with regard to maintaining measurement invariance. Thus, we investigated the invariance of each ECI skill (i.e., gestures, vocalizations, single and multiple words) and weighted total communication when estimated unambiguously at each measurement occasion, including ruling out floor and ceiling effects in children at younger and/or older ages on the measure (Embretson, 2006; Widaman, Ferrer, & Conger, 2010).

1. Method

1.1. Participants

Early Head Start (EHS) programs ($N = 15$), staff members ($N = 394$), and enrolled children with ECI data ($N = 5478$) in one Midwestern state participated during the 10-year period between February 12, 2002 and February 14, 2012 (see Table 1). The children assessed with the ECI during this period were divided into two samples based on a cut point of March 7, 2007 a date near the middle of the full 10-year period. The original sample consisted of those children whose ECI data collection began prior to March 7, 2007 versus the replication sample of those children whose ECI data collection began after this date. There was no particular significance to selection of this date, such as association with such factors as EHS programs, policy, updates to the online data system, assessor training or ECI protocol changes. It was associated however; with funding of research requiring collection of the replication sample data as part of a large ECI investigation.

EHS programs enrolled families as early as pregnancy, and any time prior to 36 months of age, consistent with federal and state EHS policies. Children were eligible for ECI assessments beginning at 6 months through 42 months of age (see Table 1). As can be seen, programs contributed different proportions of children in each sample, ranging from 0.0 to 14.7% in the original, versus 0.2–24.0% in the replication samples. Local program directors originally had adopted the ECI for accountability reporting purposes relative to the language development domain, as well as a means of addressing the emerging "individualization" mandate in EHS curriculum (Siegel, 2000). Over the study period, this original mandate loosened as interpreted by programs in Kansas to the point where individual programs were able to choose whether or not to replace the ECI with other measures of the domain for accountability purposes. This accounted for the imbalance in program-level proportions, namely the drop in proportion of children in the replication compared to the original sample (i.e., Programs 1, 7, 8, and 11) and the increase in Program 15 that started ECI assessments for the first time after March 7, 2007. These children and their programs were diverse regionally (i.e., urban, suburban, and rural).

The children's overall mean age at first ECI was 16.9 months ($SD = 9.9$), and the vast majority were 6–36 months of age (see Table 1). The distribution by gender was 52.9% boys versus 47.6% girls, while 9% had Individual Family Service Plans (IFSPs) indicating their eligibility for Part C early intervention services. Twenty percent of children were dual language learners of which the vast majority heard Spanish spoken at home according to parent report. Overall, the samples did not differ appreciably on these attributes.

1.2. Settings

EHS is a national child development program serving low-income families with infants and toddlers. Family income at or below the federal poverty guidelines was a requirement for EHS participation. Additionally, EHS policies make 10% of openings in programs available to children receiving Part C early intervention services under the Individuals for Disabilities Education Act (IDEA). The objectives of EHS are to enhance children's growth and development, strengthen families as primary caregivers, provide education, health, and nutritional supports, link child and family to services, and ensure well managed programs include parents in decision-making. The majority of participating programs (87%) used a home-visiting model in which program staff worked with parents on intervention goals at home and where the ECI was administered.

1.3. Measurement procedures

1.3.1. The Early Communication Indicator (ECI)

As previously described, the ECI progress monitoring measure was used exclusively for measuring language outcomes. Administration of the ECI involved an adult (i.e., the assessor or parent trained by the assessor) familiar to the child who was taught to interact as a play partner with a child centered around either the Fisher-Price Barn® (Form A) or Fisher-Price House® (Form B). Familiarity between the child and adult was important in the administration of the ECI in avoiding reactivity due to the stranger effect common in very young children (Greenwood et al., 2008). In cases in which adults were not familiar with the child, for example new home visitors, they engaged in a number of play sessions prior to ECI assessments to become familiar to the child. Familiarity was reached when the child would willingly engage in play with the adult.

Table 1
Socio-demographic characteristics.

Variable		Statistic	Samples		
			Original (n = 2931)	Replication (n = 2547)	All (N = 5478)
Age at start	0–12	%	44.1	47.1	45.5
	13–24	%	29.1	28.3	28.7
	25–36	%	23.3	22.9	23.1
	37–42	%	3.6	1.7	2.7
		M	17.2	16.5	16.9
		SD	10.1	9.7	9.9
Gender	Male	%	52.8	52.0	52.5
	Female	%	47.2	48.0	47.6
IFSP status	No	%	90.0	92.1	91.0
	Yes	%	10.0	7.9	9.0
Home Language	English	%	83.7	75.5	79.9
	Spanish	%	15.3	22.5	18.6
	Other	%	1.0	2.1	1.5
Programs ID	1	%	7.7	0.4	4.3
	2	%	7.5	7.0	7.3
	3	%	6.0	7.9	6.9
	4	%	3.3	3.4	3.3
	5	%	3.4	3.8	3.6
	6	%	5.3	7.9	6.5
	7	%	10.0	0.2	5.5
	8	%	8.5	2.4	5.7
	9	%	7.2	6.0	6.6
	10	%	14.7	24.0	19.0
	11	%	10.4	13.9	12.0
	12	%	8.6	0.4	4.8
	13	%	5.4	6.6	6.0
	14	%	2.1	11.2	6.3
	15	%	0.0	5.0	2.3
	All	%	100.0	100.0	100.0

These two toy sets were shown in prior research to be equivalent alternate forms for evoking early communication during play (Luze et al., 2001). The play sessions took place in a convenient setting with few distractions present in the home or early education program. The assessor timed the session duration for 6 min using a digital timer. The play partners' role during an ECI session was to encourage the child's communication by following the child's lead and commenting on the child's actions and words. Because the goal is to capture the child's typical communication performance, assessors did not direct or take the lead but instead supported the child's communicative behavior through encouragement and interest. ECIs were coded either live in real time or via videotape later at the office. The frequency of occurrence of each skill element was recorded on a paper and pencil data sheet. These raw data were entered into the IGDI online data system for data management and reporting (Greenwood, Carta, Walker, Buzhardt, & Baggett, 2006).

Accommodations recommended for children with sensory and/or physical impairments included: moving toys closer to the child, supported positioning for the child in a manner that orients the child toward the toys and enables best access to them, and where needed, and using toys that are larger and more identifiable. In these situations, the adult play partner may carry out the following: introduce the toys to the child allowing him/her to touch and manipulate them, provide more movement of toys, and tell the child where the toys have been placed if needed.

Procedures for children whose primary home language was not English required a play partner who could engage with the child in his/her primary language, and a coder who was fluent in both English and the child's home language capable of discriminating utterances from single words and multiple words as well as utterances that were not words in both languages (Walker & Buzhardt, 2010). The percentage of ECIs that were reported as administered in the home language of the child was 90% for the original and 92% for the replication samples. Note that the scoring of the ECI was not intended to provide results in separate languages, rather simply the sum of all words said regardless of language used. If a child used a word in Spanish and subsequently in English it was counted as two words.

1.3.2. Program data collection and staff management

ECI assessors were EHS local program staff assigned by their program's director to conduct ECI measures. Using the program management tools in the IGDI website, local program directors enrolled their program staff (e.g., home visitors and other service providers) giving them log in credentials so that they could enter their security protected area as needed. The credentials enabled program staff to enroll children, and add, and manage their ECI data. Local directors often assigned a staff member to enter data for all their staff that administered and scored the ECI with children on their caseload. Otherwise, staff entered their own data.

Local EHS staff administered and scored ECI assessments quarterly for all children in these years linked to each child's entry date into the program and in some cases more frequently for the lowest performing children. The frequency of progress monitoring for low-performing children was a local program decision. Using the website, it was possible for program staff to view individual children's ECI growth charts. Local program directors/coordinators could view and print children's growth charts, as well as program-wide summary reports. State directors had similar data-viewing privileges including a consolidated state-wide report and program-by-program reports (Greenwood, Carta, Walker, Hughes, & et al., 2006). They all used these tools, and in some cases appointed coordinators to manage the data collection process.

The expectation was that data would be entered into the system shortly after collection so that home visitors, program directors, and state EHS directors could view current information online.

Thus, ECI scores were nested under children, who were nested under EHS programs within states. Program staff could only access data for their own program. All EHS programs in this report reached program-wide ECI implementation and maintained data collection during at least half of this entire period (see Table 1).

1.3.3. ECI assessor training and certification

EHS staff members learned to use the ECI in a trainer-of-trainers model supported by the ECI developers and website resources (Walker & Carta, 2010). At least one local staff member from each program attended a workshop by the ECI developers where they were certified to administer the ECI, code children's communicative events, and interpret the results. These staff returned to the local program and certified additional staff using materials given to them at the workshop or from the website.

Certification required accomplishing two tasks meeting standards of calibration (Walker & Buzhardt, 2010). The first was coding or scoring of children's communication skills. The second was fidelity of administration. To certify as an ECI coder, trainees coded two certification videos each with an 85% reliability standard based on exact percentage agreement. Trainees accessed the two videos at the website (Walker & Buzhardt, 2009), coded them, and entered results into the online data system. The system provided them an immediate calculation of their reliability (overall and for each key skill element) in comparison to master codings. In our experience, most trainees achieved this standard within two to four attempts per video.

Finally, trainees were required to demonstrate administration fidelity at an 80% level. ECI trainers observed each trainee's administration of the ECI either live or videotaped and evaluated their use of at least 13 of the 16 ECI administration steps. Most assessors achieved this standard on their first administration given instructions and coding of two administrations previously discussed. Overall, certification was completed in 2.5–5 h: 1–2 h for learning the coding and administration guidelines, 1–2 h for coding certification, and 30–60 min for administration certification.

1.3.4. ECI score reliability

Assessors in local programs were encouraged to conduct annual paired inter-assessor coding agreement checks. These checks were a dual coding of a videotaped administration of the same child on the same date such that agreement between a reliability assessor and a primary assessor could be checked. Not all programs conducted reliability checks, however.

In all, 390 paired interrater assessments, 223 from the original sample versus 167 from the replication sample, were available for analysis (see Table 2). Given that assessor training had required meeting both (a) administration fidelity and (b) videotaped coding calibration to a standard, the analysis of interrater reliability for this report focused on ECI score reliability rather than percentage agreement (Hartmann, 1977; Walker & Buzhardt, 2010). Pearson r was used to estimate the primary and reliability raters mean scores on each of the ECI scales. We also used equivalence confidence intervals (CI = 95%) of the mean difference to assess similarity of observer groups' mean estimates (Seaman & Serlin, 1998). Overall, there was a strong pattern of reliability between rater groups' scores in both samples. Interrater score reliability was above r = 0.90 for all ECI scales in both samples. According to Seaman and Serlin (1998) the equivalence of groups' mean estimates is indicated when a 0 is included within the confidence interval. This was the case for 7 of the 10 comparisons. Single words and total communication in the original sample and gestures in the replication sample did not meet this test. However, the size of these mean differences in practical terms was small. All the key skills were less than 0.13 of a communication per minute in a metric ranging from 0 to 5 per minute while for total communication it was 0.55 of a communication per minute in a metric ranging from 0 to 30 per minute.

1.4. Statistical approach

The basic raw data across all children were comprised of quarterly ECI assessments across the ages of 6–42 months in cases were programs continued providing services beyond 36 months. Because most children aged out of services at 36 months, we have used 36 as an intercept point in past research (Greenwood, Buzhardt, & et al., 2011; Greenwood et al., 2010, 2013). Thus, complete quarterly data for any one child in a program consisted of 11 quarterly occasions (separated by 3 months starting at 6 months of age). A small percentage of children identified as below benchmark on ECI total communication in these screens were assessed more frequently by program staff as part of progress monitoring.

The total number of ECI assessments collected was 20,740 including 12,150 in the original sample and 8590 in the replication sample. The mean number of ECI assessments available per child was 4.1 (SD = 3.1, range = 1–20) and 3.4 (SD = 2.7, range = 1–16) in the original and replication samples, respectively. Across ages and sample, the amount of data collected was well balanced. However, variations in terms of the numbers of children measured at each occasion did occur because children entered programs at various ages after 6 months and exited before 37 months of age, thus it was not possible to collect these data. As a result, the treatment

Table 2
Correlations and equivalence of raters' scores by samples.

	Original (N = 223)						Replication (N = 167)						Total (N = 390)
		Primary	Reliability		95% Confidence Interval			Primary	Reliability		95% Confidence Interval		
ECI score	r	M (SD)	M (SD)	$Mdiff$	Low	High	r	M (SD)	M (SD)	$Mdiff$	Low	High	r
Gestures	0.90	1.41 (1.29)	1.38 (1.27)	0.03	−0.04	0.11	0.91	1.71 (1.28)	1.59 (1.20)	0.12	0.04	0.20	0.91
Vocalizations	0.95	2.84 (2.22)	2.82 (2.27)	0.02	−0.07	0.11	0.94	2.54 (2.01)	2.58 (2.03)	−0.04	−0.15	0.07	0.95
Single words	0.92	1.59 (1.92)	1.46 (1.82)	0.13	0.03	0.23	0.93	1.65 (1.98)	1.65 (1.95)	0.00	−0.11	0.12	0.92
Multiple words	0.96	1.25 (2.06)	1.19 (2.04)	0.06	−0.02	0.13	0.98	1.38 (2.41)	1.43 (2.45)	−0.05	−0.12	0.03	0.97
Total communication	0.94	11.34 (8.63)	10.79 (8.56)	0.55	0.15	0.94	0.96	11.91 (9.36)	11.93 (9.41)	−0.02	−0.40	0.36	0.95

of missing data was a consideration in planned analyses described below.

When aggregated by monthly occasions, a cross-sectional, longitudinal dataset was available for analysis. In this report, we included data from 6 to 37 months of age. The mean number of ECI assessments available per month of age between the ages of 6 and 37 months was 357 (*SD* = 64), ranging from 185 to 444 in the original sample and 260 (*SD* = 54), ranging from 170 to 417 in the replication sample (see Table 3). Prior to analysis, all variables of interest in the original data set were screened for accuracy using the SAS 9.3 program. Results indicated that data did not contain extreme collinearity.

1.4.1. Outliers

Given our knowledge of the appropriate scaling of ECI metrics in EHS programs (Greenwood, Buzhardt, & et al., 2011), we have routinely screened for outliers for both statistical purposes, as well as an indicator of "reasonableness and accuracy" in using the ECI. Outliers greater than three standard deviations above the age mean were removed to prevent biasing of results because of extremes due to data entry errors or subpopulations (Greenwood, Buzhardt, & et al., 2011). Doing so ensures that analyses and findings reported herein are consistent with prior work and metrics in earlier reports. For example, we previously reported identifying a program where the majority of ECI scores exceeded expected upper score limits suggesting issues of drift, weak training and lack of supervision of assessors. We also identified a program who against our advice and training, hired outside evaluators (rather than EHS home visiting staff assessors) unfamiliar to the child to administer the ECI. They complained that their children were scoring systematically lower than children at comparable ages in other programs. This was because their assessors were "unfamiliar reactive strangers" who were actually biasing the assessment. When the program changed this practice and used their familiar home visiting staff as assessors, the problem went away.

In the present study, such screening identified less than 1% of all values (0.89%), a trivial number that were removed from the dataset. Doing so resulted in removal of only 20 children from a total of 5498 children (0.36%) because of outliers. And, no program had a concentration of outliers, suggestive of poor ECI implementation. After removing univariate outliers, there were no multivariate outliers (according to Mahalanobis distance-squared).

1.4.2. Missing data

The current design of ECI measurement in these EHS programs is inherently an incomplete data design because children may enter and leave EHS at different ages and thus, are not able to be assessed on all occasions. Evaluation indicated that 36% was missing due to all reasons not related to program enrollment. Because the most likely missing data patterns accounted for only 5% or fewer children within each group, it suggested random fluctuations and that the data were missing at random (Enders, 2006, 2010).

Full Information Maximum Likelihood (FIML) is an appropriate imputation approach given the observed missing data rate (Collins, Schafer, & Kam, 2001; Graham, Olchowski, & Gilreath, 2007; Littvay, 2009; Newman, 2003; Schlomer, Bauman, & Card, 2010; Zhang & Walker, 2008). Review of the missing data literature suggests that FIML performs as well (Schlomer et al., 2010) or better (Graham et al., 2007; Yuan, Yang-Wallentin, & Bentler, 2012) than multiple imputation (MI) techniques. Both are favored over traditional methods for addressing missing data (Collins et al., 2001; Enders, 2010) and they arrive at similar solutions when the models are identical in analyses and auxiliary variables (Buhi, Goodson, & Neilands, 2008). When checked, FIML produced similar estimates and standard errors in comparison to MI. We used the FIML procedure in the M*plus* version 6.1 software package (Muthén & Muthén, 1998–2010) to address missing data.

1.4.3. Invariance testing

To address the question of the invariance of ECI trajectories, Latent Growth Curve Modeling (LGCM) invariance testing was used (McArdle, 1988). In planned analyses, each ECI scale was considered a construct (e.g., vocalizations) and each construct was comprised of one indicator at each age or occasion of measurement. Given the positive accelerating shape of multiple words and total communication, a one-piece LGCM analysis was used. However, given the accelerative and decelerative trajectories of gestures, vocalizations, and single words, a two-piece ("spline") LGCM growth model was used.

LGCM was an appropriate choice for computing individual ECI growth patterns because it allows one to investigate mean growth and the variability around said growth, as well as the associations between growth parameters (i.e., slopes and intercepts) (Preacher, Wichman, MacCallum, & Briggs, 2008). Separate LGCMs were fit for each ECI key skill and also total communication to reduce complexity and to isolate each key skill trajectory individually in terms of its invariance across samples. The intercepts of these spline LGCMs were determined by standards used in previous studies that located an inflection point (Greenwood et al., 2013). For gestures, the intercept (inflection or peak value) was placed at 12 months of age, vocalizations = 18 months, single words = 24 months so that slope indicated growth and decline from peak values at this point in time. The intercepts for multiple words and total communication were placed at 36 months of age so slope indicated growth to this point in time.

In these models, the mean level or intercept was interpreted as the peak status reached prior to an inflection in the two-piece models (gestures, vocalizations, and single words) and at 36 months of age in the one-piece models (multiple words and total communication). The slope factor represented the change in the growth trajectory over time both before (slope 1) and after (slope 2) the inflection point in the two-piece spline models. The loadings on the slope factors (other than those corresponding to anchor time points at the beginning and end of the slope) were allowed to be estimated via a level-and-shape model (McArdle, 1988) to further account for the nonlinear nature of the growth trajectories.

To assess the similarity of growth trajectories between the original and replication samples, we used factorial and structural invariance testing of the LGCMs. Invariance testing, recommended for use in construct validation, involves fitting a sequence of models that add increasingly more stringent constraints between multiple groups in order to test for differences in the models and parameters between the groups (Dimitrov, 2010).

Factorial invariance testing involves estimating (1) configural (or baseline) invariance, (2) weak (i.e., metric) invariance, and (3) strong (i.e., scalar) invariance. The configural invariance model tests whether or not both groups have a similar structure of the model under investigation. The weak invariance model constrains the factor loadings to be equal between groups, testing the expectation that a unit increase in the latent factor is associated with the same unit change between groups in the measured variables. The strong invariance model adds an additional restriction by constraining the measured variable intercepts to be equal across groups. It tests whether or not, at a particular value of the latent variable, the measured variable means are the same between groups. A tenable strong invariant model indicates that the measured items are related to the latent factors in the same way across groups. Alternatively, lack of strong invariance indicates differential functioning of the measure (Dimitrov, 2010).

Structural invariance testing is concerned with how characteristics of the latent slope and intercept factors (i.e., mean, variance,

Table 3
ECI total communication descriptive statistics per month of age.

Months of age	Original cohort			Replication cohort			All		
	M	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*
6	2.8	1.8	443	3.2	1.9	417	3.0	1.9	860
7	2.9	2.1	215	3.2	2.2	189	3.1	2.1	404
8	3.3	1.9	281	4.1	2.3	296	3.7	2.1	577
9	3.7	2.1	387	4.5	2.5	346	4.1	2.3	733
10	4.1	2.2	258	4.6	2.5	237	4.3	2.4	495
11	5.0	2.7	305	5.4	2.9	278	5.2	2.8	583
12	5.3	2.5	437	5.7	2.9	355	5.5	2.7	792
13	5.4	2.9	297	5.9	2.9	239	5.6	2.9	536
14	5.9	3.4	353	6.5	3.3	257	6.1	3.3	610
15	6.4	3.3	411	6.3	3.3	334	6.4	3.3	745
16	6.6	3.4	333	6.7	3.5	248	6.7	3.4	581
17	7.4	4.0	353	7.9	4.2	278	7.6	4.1	631
18	7.9	4.2	427	8.1	4.2	308	8.0	4.2	735
19	8.6	4.8	337	8.7	4.9	238	8.6	4.9	575
20	9.7	5.3	375	10.2	5.7	284	9.9	5.5	659
21	10.8	5.9	422	10.4	5.6	269	10.6	5.8	691
22	11.4	6.1	353	11.7	6.6	202	11.5	6.3	555
23	12.5	6.9	344	13.1	7.2	315	12.8	7.1	659
24	14.0	7.6	441	14.0	7.3	268	14.0	7.5	709
25	14.6	8.5	355	14.7	7.9	246	14.6	8.3	601
26	14.8	8.1	364	16.5	8.8	283	15.6	8.4	647
27	16.7	7.9	418	17.0	7.9	233	16.8	7.9	651
28	17.9	8.9	370	18.5	9.4	242	18.1	9.1	612
29	18.3	8.6	374	20.1	9.8	275	19.1	9.2	649
30	19.5	8.9	444	19.8	9.4	255	19.6	9.1	699
31	19.8	9.1	300	19.9	9.7	225	19.8	9.3	525
32	19.4	9.4	386	21.7	9.1	257	20.3	9.3	643
33	21.1	9.7	428	20.7	9.7	222	21.0	9.7	650
34	19.9	9.4	340	20.8	9.1	199	20.2	9.3	539
35	20.9	9.6	350	23.4	9.9	207	21.9	9.8	557
36	22.0	9.3	341	22.5	9.5	170	22.2	9.4	511
37	21.5	8.7	185	22.9	9.9	174	22.2	9.3	359
Grand M	11.9	5.9	357	12.5	6.1	261	12.1	6.0	618
SD	6.6	2.9	64	6.9	3.0	54	6.7	2.9	104

and covariance) were the same or different between groups. When a model has achieved both factorial and structural invariance, growth trajectories are the same between groups. While factorial invariance evaluates the model in terms of its indicator-to-factor relationships, structural invariance precisely investigates the qualities of the factors themselves. For instance, once weak invariance modeling has established a similar scale for the variance, the latent variances and covariance can be tested for equality between samples. In the case of LGCM, invariance of the latent variances and covariances indicates that growth and peak values vary with the same magnitude across groups, and that growth and peak values are similarly related between groups. Where latent variances were not equivalent (i.e., not on the same metric), phantom variables were used to transform variances onto the same standardized metric to test equivalence of the latent correlations.

The other piece of structural invariance, latent mean invariance, can be tested after strong factorial invariance has been established. When latent mean values of the LGCMs are invariant, the mean growth and peak of a key skill trajectory is shown to be equivalent between samples. Both latent variance/covariance invariance and latent mean invariance are important in establishing construct stability and validity across samples. We did not attempt to test for strict (i.e., uniqueness) invariance, because it postulates the unrealistic expectation that item-specific error and random error do not change between groups (Brown, 2006; Little, Preacher, Selig, & Card, 2007). Together, these factorial invariance procedures were used to progressively test whether or not the measured ECI variables related to the latent variables of intercept and slopes in the same way across samples.

In these analyses, we used the following goodness of fit indicators: the root mean square error of approximation (*RMSEA*), the non-normed fit index (*NNFI*), and the comparative fit index (*CFI*). For *RMSEA*, values less than or equal to 0.08 are preferred. Values greater than 0.90 are generally considered acceptable for the *NNFI* and *CFI*. The chi-square fit statistic was also reported. For the tests of factorial invariance, we used the change in *CFI* of >0.01 rule as criterion for the failure of invariance (Cheung & Rensvold, 2002). When factorial invariance failed, individual parameters were tested for invariance using nested chi-square difference tests with a significance cutoff of $p < 0.001$ to account for the large sample size and multiple testing (French & Finch, 2006; Little, 1997). Structural invariance was tested with nested chi-square (Table 4).

2. Results

2.1. Were ECI trajectories invariant?

2.1.1. Gestures

The configural invariant model for growth in gestures fit the data closely [χ^2 (98) = 247.73, $p < 0.001$, *RMSEA* = 0.024 (90% *CI*: 0.020, 0.027), *CFI* = 0.986, *TLI* = 0.980] (see Table 5). This indicated that the two samples had similar factor structures (i.e., the same general spline structure). In addition, given the lack of significant decrease in model fit for the weak invariant model [χ^2 (106) = 266.58, $p < 0.001$, *RMSEA* = 0.024 (90% *CI*: 0.020, 0.027), *CFI* = 0.985, *TLI* = 0.981], the pattern of factor loadings was shown to be the same between samples (see Fig. 1). Fitting the strong invariant model led to a significant drop in model fit from its baseline model [χ^2 (101) = 371.18, $p < 0.001$, *RMSEA* = 0.031 (90% *CI*: 0.028, 0.035), *CFI* = 0.975, *TLI* = 0.966], with all intercepts showing significant differences between groups except for the intercept corresponding to gesturing at 35–37 months of age [$\Delta\chi^2(1) = 9.81$;

Table 4
Gesture rate original-replication invariance testing.

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	*RMSEA*	*RMSEA* 90% *CI* Lower	*RMSEA* 90% *CI* Upper	*CFI*	*NNFI*	Tenable
Configural Invariance	247.73	98	<0.001	–	–	–	0.024	0.020	0.027	0.986	0.980	–
Loading Invariance[a]	266.58	106	<0.001	–	–	–	0.024	0.020	0.027	0.985	0.981	Yes
Intercept Invariance[a,b]	371.18	101	<0.001	–	–	–	0.031	0.028	0.035	0.975	0.966	No
6 mos. to 7 mos.[c]	260.88	91	<0.001	53.61	1	<0.001	0.026	0.022	0.030	0.985	0.976	No
8 mos. to 10 mos.[c]	279.68	91	<0.001	72.41	1	<0.001	0.028	0.024	0.031	0.983	0.973	No
11 mos. to 13 mos.[c]	259.07	91	<0.001	51.80	1	<0.001	0.026	0.022	0.030	0.985	0.976	No
14 mos. to 16 mos.[c]	242.79	91	<0.001	35.52	1	<0.001	0.025	0.021	0.028	0.986	0.979	No
17 mos. to 19 mos.[c]	227.37	91	<0.001	20.10	1	<0.001	0.023	0.020	0.027	0.988	0.981	No
20 mos. to 22 mos.[c]	236.69	91	<0.001	29.42	1	<0.001	0.024	0.020	0.028	0.987	0.980	No
23 mos. to 25 mos.[c]	238.97	91	<0.001	31.70	1	<0.001	0.024	0.021	0.028	0.987	0.979	No
26 mos. to 28 mos.[c]	237.01	91	<0.001	29.74	1	<0.001	0.024	0.020	0.028	0.987	0.979	No
29 mos. to 31 mos.[c]	224.07	91	<0.001	16.80	1	<0.001	0.023	0.019	0.027	0.988	0.981	No
32 mos. to 34 mos.[c]	225.06	91	<0.001	17.79	1	<0.001	0.023	0.019	0.027	0.988	0.981	No
35 mos. to 37 mos.[c]	217.08	91	<0.001	9.81	1	0.002	0.022	0.019	0.026	0.989	0.982	Yes
Homogeneity of Variances and Covariances[c]	421.13	112	<0.001	154.55	6	<0.001	0.032	0.029	0.035	0.972	0.965	No
Homogeneity of Variances[c]	336.91	109	<0.001	70.33	3	<0.001	0.028	0.024	0.031	0.979	0.973	No
Slope 1 Variance[c]	266.87	107	<0.001	0.29	1	0.590	0.023	0.020	0.027	0.985	0.981	Yes
Intercept Variance[c]	330.94	107	<0.001	64.36	1	<0.001	0.028	0.024	0.031	0.980	0.973	No
Slope 2 Variance[c]	269.30	107	<0.001	2.72	1	0.099	0.024	0.020	0.027	0.985	0.981	Yes
Homogeneity of Latent Correlations[c,d]	273.55	109	<0.001	6.97	3	0.073	0.023	0.020	0.027	0.985	0.981	Yes

Note. Each nested model contains its constraints, plus the constraints of all previous, tenable models.
[a] Evaluated with the CFI Model Test (Cheung & Rensvold, 2002).
[b] Evaluated against a separate baseline model.
[c] Evaluated with the χ^2 Difference Test.
[d] Uses phantom variables to standardize variance (see Little, 1997; Rindskopf, 1984).

$p = 0.002$]. Further investigation showed that the replication sample had higher intercepts (between 0.303 and 0.495 gestures per minute greater), indicating that most of the means of the replication sample measured variables (e.g., gesturing at 11–13 months of age) were greater than were the means of the original sample. Therefore, we were unable to compare the two samples' latent means because the different pattern of mean intercepts (i.e., different latent mean scaling) would confound any interpretation of mean differences.

Building from the weak invariant model, constraining the variances and covariances of the latent factors [χ^2 (112) = 371.13, $p < 0.001$, *RMSEA* = 0.032 (90% *CI*: 0.029, 0.035), *CFI* = 0.972, *TLI* = 0.965] resulted in a significant drop in model fit [$\Delta\chi^2(6) = 154.55$; $p < 0.001$]. Testing the variances individually revealed that the peak value (i.e., intercept) at 12 months for gesturing varied significantly more for the original sample (variance = 1.412) than for the replication sample (variance = 0.644) [$\Delta\chi^2(1) = 64.36$; $p < 0.001$]. Given the difference in variance scaling, phantom variables were used in order to standardize the variances to test for equivalence of the latent correlations (Little, 1997). With standardized variances, the correlations between slope and intercept factors were invariant [$\Delta\chi^2(3) = 6.97$; $p = 0.023$], indicating that the associations between growth and peak values for gesturing were the same across samples. This final model including invariant loadings, variances of the slopes, and latent correlation is seen in Fig. 1. Overall, while the mean scaling of growth and intercept values was different between samples, changes in the latent growth and peak values were equivalently assessed by the measured variables between the two groups. The exception was the variance of the intercept, the latent constructs varied and covaried at the same magnitude between samples.

2.1.2. Vocalizations

Model fit statistics indicated close fit for the vocalizations configural invariant model [χ^2 (98) = 384.81, $p < 0.001$, *RMSEA* = .033 (90% *CI*: 0.029, 0.036), *CFI* = 0.974, *TLI* = 0.962], supporting a similar factor structure between both the original and replication samples (see Table 5). The weak invariant restrictions did not result in a significant decrease in model fit [χ^2 (106) = 395.79, $p < 0.001$, *RMSEA* = 0.032 (90% *CI*: 0.028, 0.035), *CFI* = 0.974, *TLI* = 0.965], nor did the strong invariant restrictions result in a decrease in model fit from its baseline model [χ^2 (101) = 308.37, $p < 0.001$, *RMSEA* = 0.027 (90% *CI*: 0.027, 0.034), *CFI* = 0.974, *TLI* = 0.967]. These results demonstrated the indicator-to-factor invariance between samples of vocalizations, meaning that the values of the measured variables

Table 5
Vocalization rate original-replication invariance testing.

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	*RMSEA*	*RMSEA* 90% *CI* Lower	*RMSEA* 90% *CI* Upper	*CFI*	*NNFI*	Tenable
Configural Invariance	384.81	98	<0.001	–	–	–	0.033	0.029	0.036	0.974	0.962	–
Loading Invariance[a]	395.79	106	<0.001	–	–	–	0.032	0.028	0.035	0.974	0.965	Yes
Intercept Invariance[a,b]	308.42	101	<0.001	–	–	–	0.027	0.024	0.031	0.981	0.974	Yes
Homogeneity of Variances and Covariances[c]	399.33	112	<0.001	3.54	6	0.739	0.031	0.027	0.034	0.974	0.967	Yes
Latent Mean Invariance[c]	415.53	115	<0.001	16.20	3	0.001	0.031	0.028	0.034	0.973	0.966	Yes
Slope 1 Mean[c]	401.84	113	<0.001	2.51	1	0.113	0.031	0.027	0.034	0.974	0.967	Yes
Intercept Mean[c]	399.36	113	<0.001	0.03	1	0.862	0.030	0.027	0.034	0.974	0.967	Yes
Slope 2 Mean[c]	403.93	113	<0.001	4.60	1	0.032	0.032	0.028	0.035	0.973	0.967	Yes

Note. Each nested model contains its constraints, plus the constraints of all previous, tenable models.
[a] Evaluated with the CFI Model Test (Cheung & Rensvold, 2002).
[b] Evaluated against a separate baseline model.
[c] Evaluated with the χ^2 Difference Test.

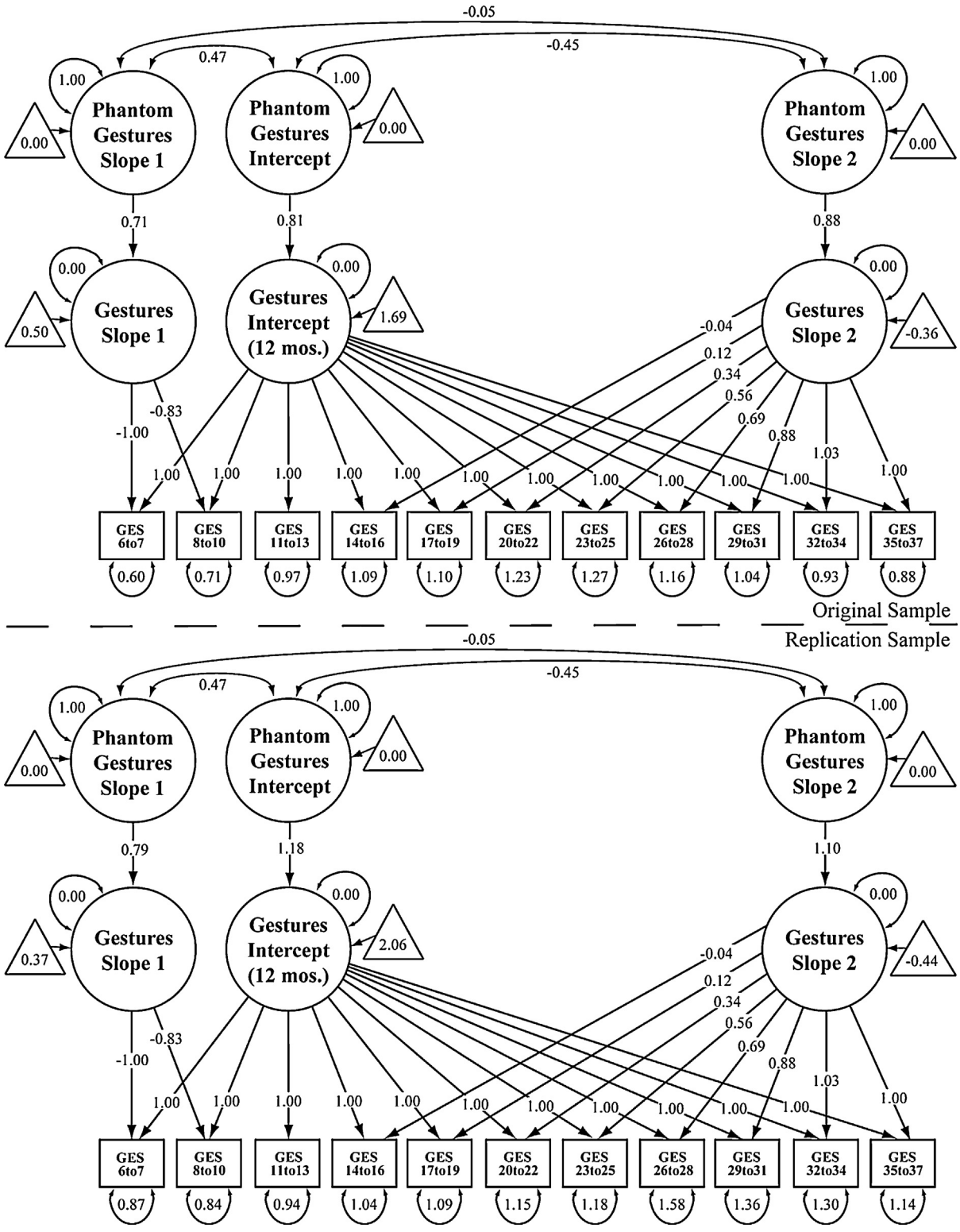

Fig. 1. Latent correlation invariant model for gestures. Model Fit: χ^2 (109, n = 5478) = 273.55, p < 0.001, *RMSEA* = 0.023 (0.020 −0.027), *CFI* = 0.985, *NNFI* = 0.981.

were affected in the same way by changes in the latent growth and intercept values, and the mean values of the measured variables are equal at given values of the latent variables.

Testing the structural components of the vocalization latent growth curve model, constraining the variances and covariances did not result in poorer model fit [$\Delta\chi^2(6) = 3.54$; $p = 0.739$], suggesting that growth and peak value of vocalizations varied at the same magnitude between samples. Further, in testing the latent means, it was shown that the mean levels of growth, peak, and decline in vocalizations did not differ between the original and replication samples [$\Delta\chi^2(3) = 15.66$; $p = 0.001$]. Achieving strong factorial invariance as well as the invariance of latent variances, covariances, and means (see Fig. 2), vocalizations evidenced strong construct validity across samples.

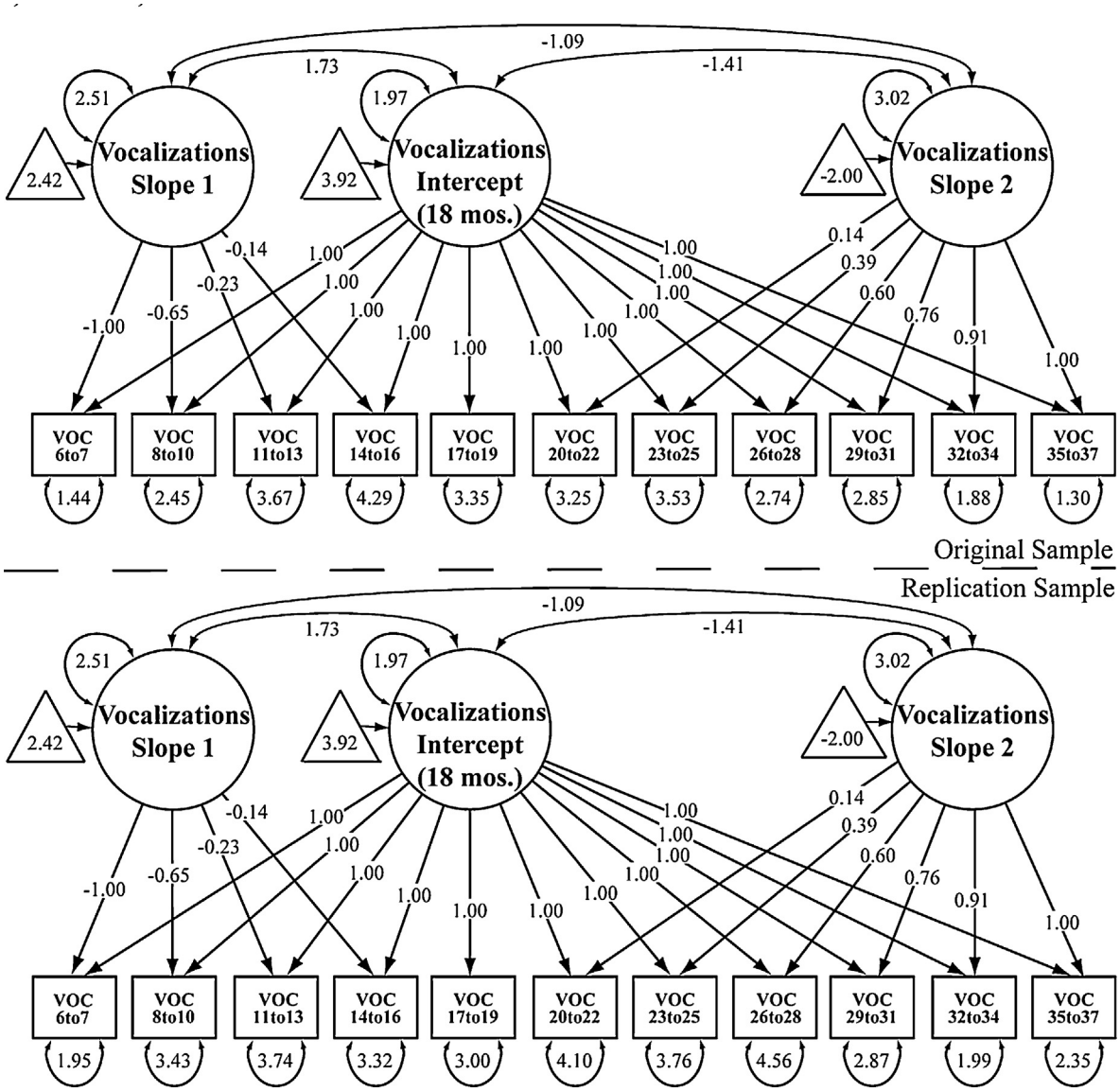

Fig. 2. Latent mean, variance, and covariance invariant model for vocalizations.
Model Fit: χ^2 (115, n = 5478) = 415.53, p < 0.001, *RMSEA* = 0.031 (0.028 −0.034), *CFI* = 0.973, *NNFI* = 0.966.

2.1.3. Single words

The close fit of the single words configural model [χ^2 (60) = 291.30, p < 0.001, *RMSEA* = 0.038 (90% *CI*: 0.033, 0.037), *CFI* = 0.989, *TLI* = 0.974], like those of the prior skills also indicated a similar factor structure between the original and replication samples (see Table 6). The fit statistics of the weak invariant model [χ^2 (66) = 295.37, p < 0.001, *RMSEA* = 0.036 (90% *CI*: 0.032, 0.040), *CFI* = 0.989, *TLI* = 0.976] did not significantly differ from the foundational configural model, indicating an equivalence of the indicator-to-factor variance relationships between groups. And, fit for the strong invariant model was not significantly worse than the strong invariant model's baseline model [χ^2 (63) = 294.31, p < 0.001, *RMSEA* = 0.037 (90% *CI*: 0.032, 0.041), *CFI* = 0.989, *TLI* = 0.975]. Collectively, these single words results illustrated the factorial invariance of the two samples, and the similarity of their indicator-to-factor relationships.

While constraints on all latent variances and covariances led to a significant drop in single words model fit [$\Delta\chi^2$(6) = 25.24; p < 0.001], tests of the individual variances and latent correlations revealed no significantly different parameters. This leaves one to assume that the variances and covariances of the latent growth and intercept variables were indeed the same between samples. When testing the latent means, the same divergence between omnibus and specific contrast tests held true. Specifically, while constraints on all three latent means resulted in a significant decrease in model fit [$\Delta\chi^2$(3) = 34.27; p < 0.001], tests of each individual latent mean showed no significant differences. Therefore, it was safe to assume that the latent growth, peak, and decline of single word usage were the same between samples (see Fig. 3), showing complete factorial and structural invariance between groups.

2.1.4. Multiple words

The configural model showed moderately close fit to the data [χ^2 (114) = 875.52, p < 0.001, *RMSEA* = 0.067 (90% *CI*: 0.063, 0.071), *CFI* = 0.961, *TLI* = 0.943], and the weak invariant model did not produce significantly divergent fit statistics from the configural model [χ^2 (73) = 885.99, p < 0.001, *RMSEA* = 0.064 (90% *CI*: 0.060, 0.068), *CFI* = 0.961, *TLI* = 0.948] (see Table 7). Further, the fit for the strong invariant model was not significantly different from its baseline model [χ^2 (68) = 540.28, p < 0.001, *RMSEA* = 0.050 (90% *CI*: 0.046, 0.054), *CFI* = 0.977, *TLI* = 0.968]. Overall, these results of factorial invariance for multiple words demonstrate the equivalence of the

Table 6
Single word usage rate original-replication invariance testing.

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	*RMSEA*	*RMSEA* 90% *CI* Lower	*RMSEA* 90% *CI* Upper	*CFI*	*NNFI*	Tenable
Configural Invariance	291.30	60	<0.001	–	–	–	0.038	0.033	0.042	0.989	0.974	—
Loading Invariance[a]	295.37	66	<0.001	–	–	–	0.036	0.032	0.040	0.989	0.976	Yes
Intercept Invariance[a,b]	294.31	63	<0.001	–	–	–	0.037	0.032	0.041	0.989	0.975	Yes
Homogeneity of Variances and Covariances[c]	320.61	72	<0.001	25.24	6	<0.001	0.036	0.032	0.040	0.977	0.956	No
Homogeneity of Variances[c]	300.80	69	<0.001	5.43	3	0.143	0.035	0.031	0.039	0.989	0.977	Yes
Slope 1 Variance[c]	297.99	67	<0.001	2.61	1	0.106	0.035	0.031	0.040	0.989	0.977	Yes
Intercept Variance[c]	299.66	67	<0.001	4.29	1	0.038	0.036	0.032	0.040	0.989	0.976	Yes
Slope 2 Variance[c]	297.19	67	<0.001	1.82	1	0.178	0.035	0.031	0.040	0.989	0.977	Yes
Homogeneity of Latent Correlations[c]	298.80	69	<0.001	3.43	3	0.330	0.035	0.031	0.039	0.989	0.977	Yes
Intercept/Slope 1[c]	298.03	67	<0.001	2.66	1	0.103	0.035	0.031	0.040	0.989	0.977	Yes
Intercept/Slope 2[c]	296.26	67	<0.001	0.89	1	0.347	0.035	0.031	0.040	0.989	0.977	Yes
Slope 1/Slope 2[c]	296.39	67	<0.001	1.02	1	0.313	0.035	0.031	0.040	0.989	0.977	Yes
Latent Mean Invariance[c]	357.70	75	<0.001	34.27	3	<.001	0.037	0.033	0.041	0.974	0.951	No
Slope 1 Mean[c]	324.37	73	<0.001	3.76	1	0.052	0.035	0.032	0.039	0.977	0.956	Yes
Intercept Mean[c]	326.27	73	<0.001	5.66	1	0.017	0.036	0.032	0.040	0.977	0.955	Yes
Slope 2 Mean[c]	323.57	73	<0.001	2.96	1	0.085	0.035	0.032	0.039	0.977	0.956	Yes

Note. Each nested model contains its constraints, plus the constraints of all previous, tenable models.
[a] Evaluated with the CFI Model Test (Cheung & Rensvold, 2002).
[b] Evaluated against a separate baseline model.
[c] Evaluated with the χ^2 Difference Test.

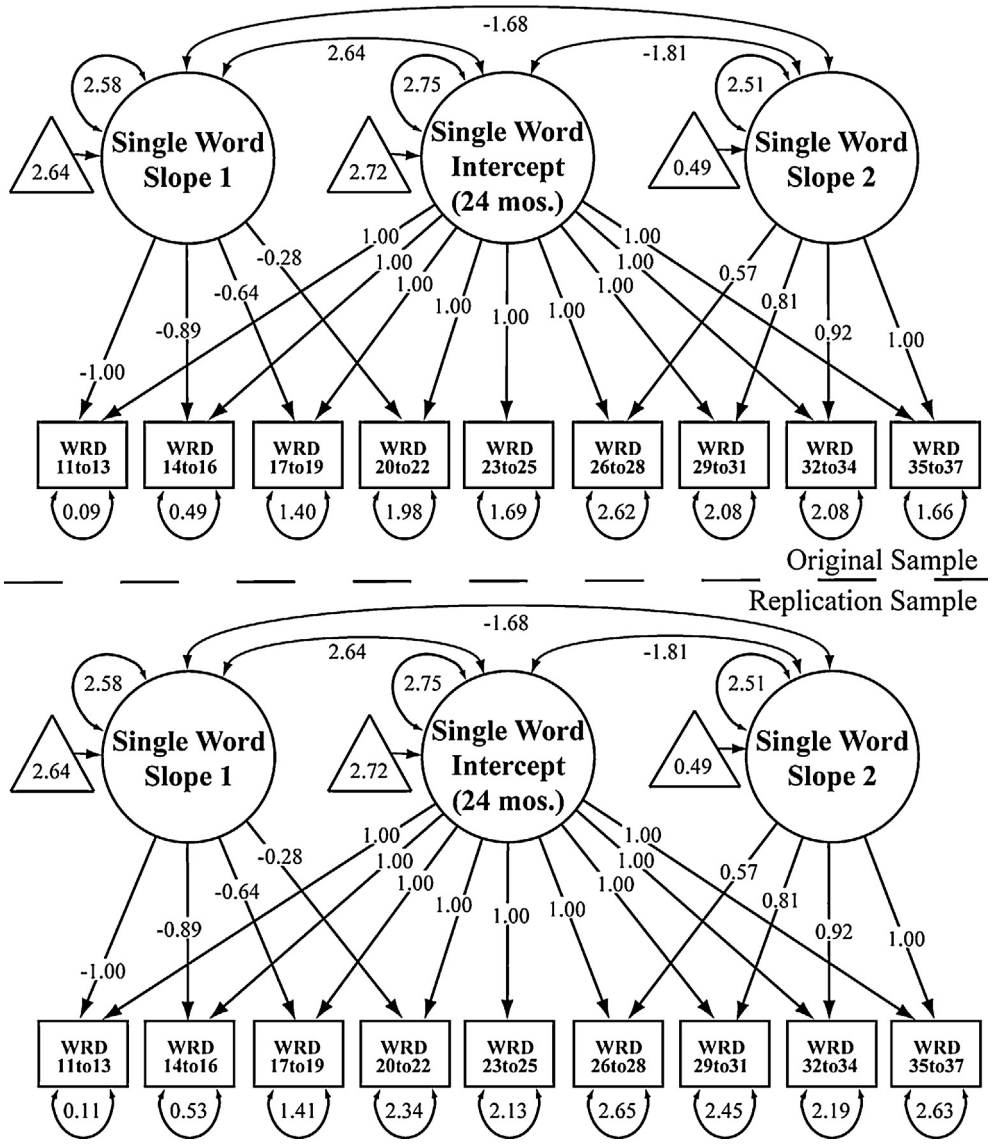

Fig. 3. Latent mean, variance, and covariance invariant model for single words.
Model Fit: χ^2 (75, n = 5478) = 357.70, p < 0.001, *RMSEA* = 0.037 (0.033 −0.041), *CFI* = 0.974, *NNFI* = 0.951.

Table 7
Multiple word usage rate original-replication invariance testing.

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	*RMSEA*	*RMSEA* 90% *CI* Lower	*RMSEA* 90% *CI* Upper	*CFI*	*NNFI*	Tenable
Configural Invariance	875.52	66	<0.001	–	–	–	0.067	0.063	0.071	0.961	0.943	–
Loading Invariance[a]	885.99	73	<0.001	–	–	–	0.064	0.060	0.068	0.961	0.948	Yes
Intercept Invariance[a,b]	540.28	68	<0.001	–	–	–	0.050	0.046	0.054	0.977	0.968	Yes
Homogeneity of Variances and Covariances[c]	977.22	76	<0.001	91.23	3	<0.001	0.066	0.062	0.070	0.956	0.945	No
Homogeneity of Variances[c]	887.69	75	<0.001	1.70	2	0.427	0.063	0.059	0.067	0.961	0.950	Yes
Homogeneity of Covariance[c]	886.92	74	<0.001	0.93	1	0.335	0.063	0.060	0.067	0.961	0.949	Yes
Latent Mean Invariance[c]	981.65	78	<0.001	4.43	2	0.109	0.065	0.061	0.069	0.957	0.946	Yes

Note. Each nested model contains its constraints, plus the constraints of all previous, tenable models.
[a] Evaluated with the CFI Model Test (Cheung & Rensvold, 2002).
[b] Evaluated against a separate baseline model.
[c] Evaluated with the χ^2 Difference Test.

mean and variance relationships between the measured variables and the latent variables.

For the omnibus test of the structural latent variances and covariances, we see a significant decrease in model fit [$\Delta\chi^2(3) = 91.23$; $p < 0.001$]. However, in testing the variances and covariance individually, neither test was significant; both the variance of the intercept and slope were the same across groups [$\Delta\chi^2(2) = 1.70$; $p = 0.427$], and the covariance between slope and intercept also was equivalent between groups [$\Delta\chi^2(1) = 0.93$; $p = 0.335$].

The latent means are shown to be invariant across samples [$\Delta\chi^2(2) = 3.72$; $p = 0.155$], meaning that the mean level of growth and peak for multiple word usage is the same between the original and replication samples. The construct validity of multiple word usage was supported by the complete factorial and structural invariance (see Fig. 4). Both the make-up and the characteristics of the growth and intercept factors were equivalent between samples.

2.1.5. Total communication

The configural invariant model for total communication rate indicated close fit to the data [χ^2 (104) = 440.18, $p < 0.001$, *RMSEA* = 0.034 (90% *CI*: 0.031, 0.038), *CFI* = 0.917, *TLI* = 0.913], denoting strong fundamental similarities between samples (see Table 8). Fitting of the weak invariant model did not significantly worsen it [χ^2 (113) = 455.60, $p < 0.001$, *RMSEA* = 0.033 (90% *CI*: 0.030, 0.036), *CFI* = 0.916, *TLI* = 0.918]. Therefore, in terms of total communication rate, a change in latent growth or intercept factors was met with the same change in the measured variables across the two groups.

Strong invariant restrictions similarly did not result in a decrease in model fit from its own baseline model [χ^2 (106) = 405.61, $p < 0.001$, *RMSEA* = 0.032 (90% *CI*: 0.029, 0.035), *CFI* = 0.926, *TLI* = 0.935], indicating equivalence of the indicator-to-factor mean relationships between samples. At given values of the latent growth and intercept factors, the mean values of the measured variables for total communication will be the same between samples. With strong invariance, total communication reached factorial invariance across samples.

The structural components of the total communication growth model also showed invariance across groups. Constraining the variances and covariances to equality between the original and replication samples did not result in a significant decrease in model fit [$\Delta\chi^2(3) = 6.45$; $p = 0.092$]. Restricting the intercept and slope means to be equal between samples did result in a significant drop in fit [$\Delta\chi^2(2) = 74.23$; $p < 0.001$]. However, testing each of the means individually did not reveal a significant difference in either the intercept mean [$\Delta\chi^2(1) = 5.00$; $p = 0.025$] nor in the slope mean [$\Delta\chi^2(1) = 0.08$; $p = 0.777$]. Therefore, the latent means were shown to be invariant through the individual contrasts and invariance held for all structural elements. With both factorial and structural invariance denoting equivalent intra- and inter-factor relationships (see Fig. 5) tests supported the validity of the ECI total communication.

3. Discussion

The aim of this work was to add new knowledge to the validity of the ECI used to assess the communication growth and development of infants and toddlers. Original development of the ECI focused on designing a formative measure practitioners could use and interpret that was easy to collect, repeatable, and sensitive to individual short-term growth over time. In early reports, we provided evidence that ECI measurement was reliability and measured communication skills (criterion validity) as was intended (Luze et al., 2001). Development of the website and open Internet access enabled replication and scaling up to EHS programs (Buzhardt et al., 2010), and within increasing participation and sample size provided improved estimates of local program normative growth (Greenwood, Carta, Walker, Hughes, & et al., 2006).

Subsequent validity work in even larger samples provided better normative child-level estimates for children in EHS programs and benchmarks for determining levels of risk and identification of likely candidates for language intervention (Greenwood et al., 2010). Using these ECI benchmarks in a randomized trial to identify children at risk, we reported positive effect sizes in growth in ECI total communication as a function of parent's implementation of language promoting strategies at home guided by a home visitor using the ECI for intervention decisions making (Buzhardt et al., 2011).

Also with larger samples, we were able to examine program-level variations in ECI scores as function of differences in programs' sociodemographic make-up and their fidelity of ECI measurement implementation (Greenwood, Buzhardt, & et al., 2011). We also expanded the construct validity in support of the ECI's key skill trajectories over time, reporting complex patterns of development with and across skills in support of a continuum of change with some skills growing early to a peak or intercept point (gestures and vocalizations) and then declining as single words and multiple word utterance become more frequent, and growing functional communication skills (Greenwood et al., 2013).

In the current investigation, we increased confidence in the inferences made from ECI data collected by practitioners in the context of large scale and sustained use by examining the equivalence (invariance) of ECI measurement. The original and replication groups were separated in time and history and were thus, exposed to variations in potential threats to measurement invariance. Results indicated that the ECI was factorially and structurally invariant for four of the five ECI scales (i.e., vocalizations, single words, multiple words, and total communication). ECI measurement for these ECI scales was the same between samples in terms

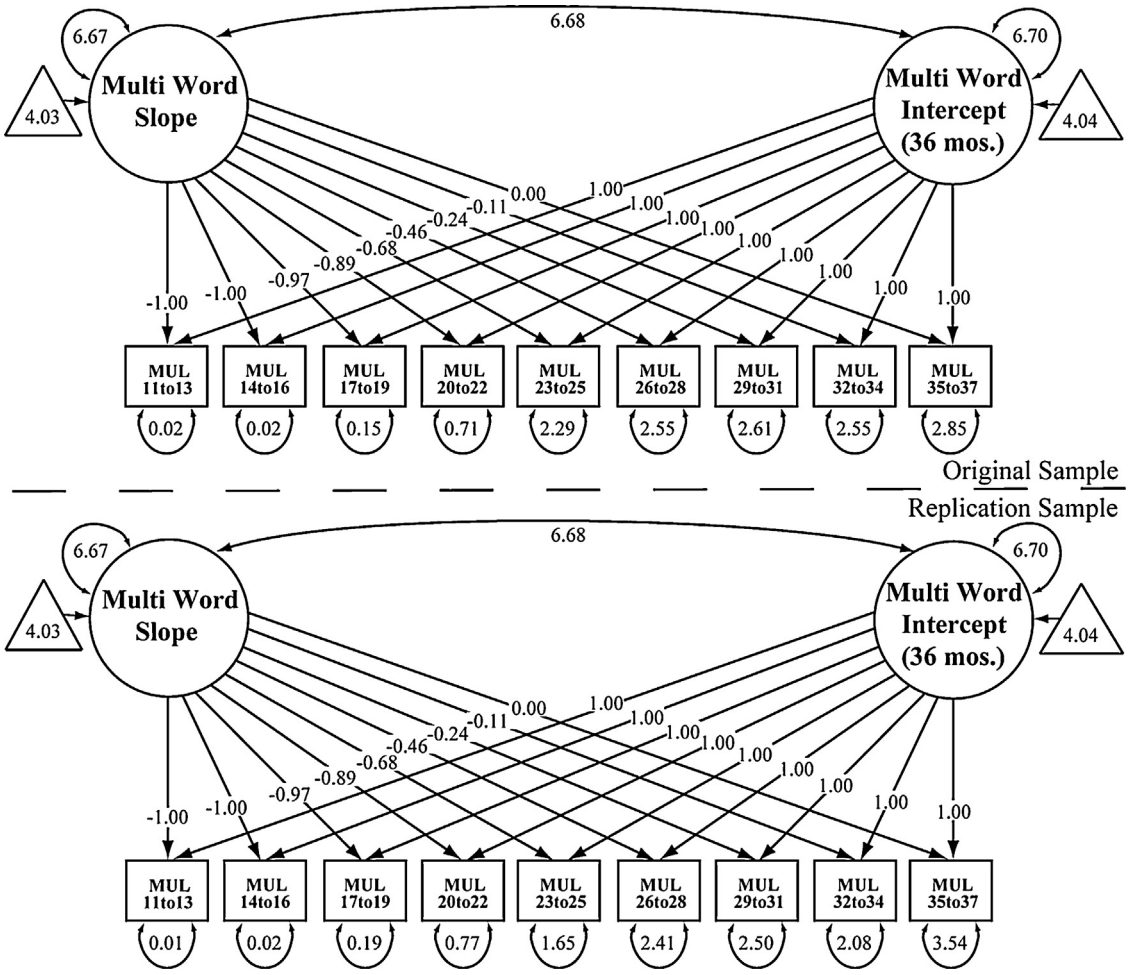

Fig. 4. Latent mean, variance, and covariance invariant model – multiple words.
Model Fit: χ^2 (78, n = 5478) = 981.65, p < 0.001, *RMSEA* = 0.065 (0.061 −0.069), *CFI* = 0.957, *NNFI* = 0.946.

of the mean observed score and the corresponding intra- and the inter-factor characteristics also were equivalent.

Gestures however, lacked strong invariance (i.e., invariance of the manifest intercepts), and so cannot be regarded as factorially invariant in these models. This means that its latent growth constructs were not defined the same across the replication and original samples suggesting significant differences in the measurement properties of gestures across the time frame investigated. There were several historical events in play for the replication sample that could explain this outcome, (a) changes made in coding definitions for gestures and (b) improvements to the website in particular.

Conversations with assessors had pointed to a lack of clarity surrounding the coding of signing, and whether or not signs should be coded as a gesture or a single word? Signing, and teaching signing to young children as a means of facilitating early communication for even those without hearing loss had become a much more common practice in the replication sample period. A similar issue emerged regarding the coding of toy movements directed at the play partner and whether or not they should be coded as gestures. We clarified

Table 8
Weighted total communication rate original-replication invariance testing.

Model	χ^2	*df*	*p*	$\Delta\chi^2$	Δdf	*p*	*RMSEA*	*RMSEA* 90% *CI* Lower	*RMSEA* 90% *CI* Upper	*CFI*	*NNFI*	Tenable
Configural Invariance	440.18	104	<0.001	–	–	–	0.034	0.031	0.038	0.917	0.913	–
Loading Invariance[a]	455.60	113	<0.001	–	–	–	0.033	0.030	0.036	0.916	0.918	Yes
Intercept Invariance[a,b]	405.61	106	<0.001	–	–	–	0.032	0.029	0.035	0.926	0.935	Yes
Homogeneity of Variances and Covariances[c]	462.05	116	<0.001	6.45	3	0.092	0.033	0.030	0.036	0.915	0.919	Yes
Latent Mean Invariance[c]	536.28	118	<0.001	74.23	2	<.001	0.036	0.033	0.039	0.897	0.904	No
Intercept Mean[c]	467.05	117	<0.001	5.00	1	0.025	0.033	0.030	0.036	0.914	0.919	Yes
Slope Mean[c]	462.13	117	<0.001	0.08	1	0.777	0.033	0.030	0.036	0.915	0.920	Yes

Note. Each nested model contains its constraints, plus the constraints of all previous, tenable models.
[a] Evaluated with the CFI Model Test (Cheung & Rensvold, 2002).
[b] Evaluated against a separate baseline model.
[c] Evaluated with the χ^2 Difference Test.

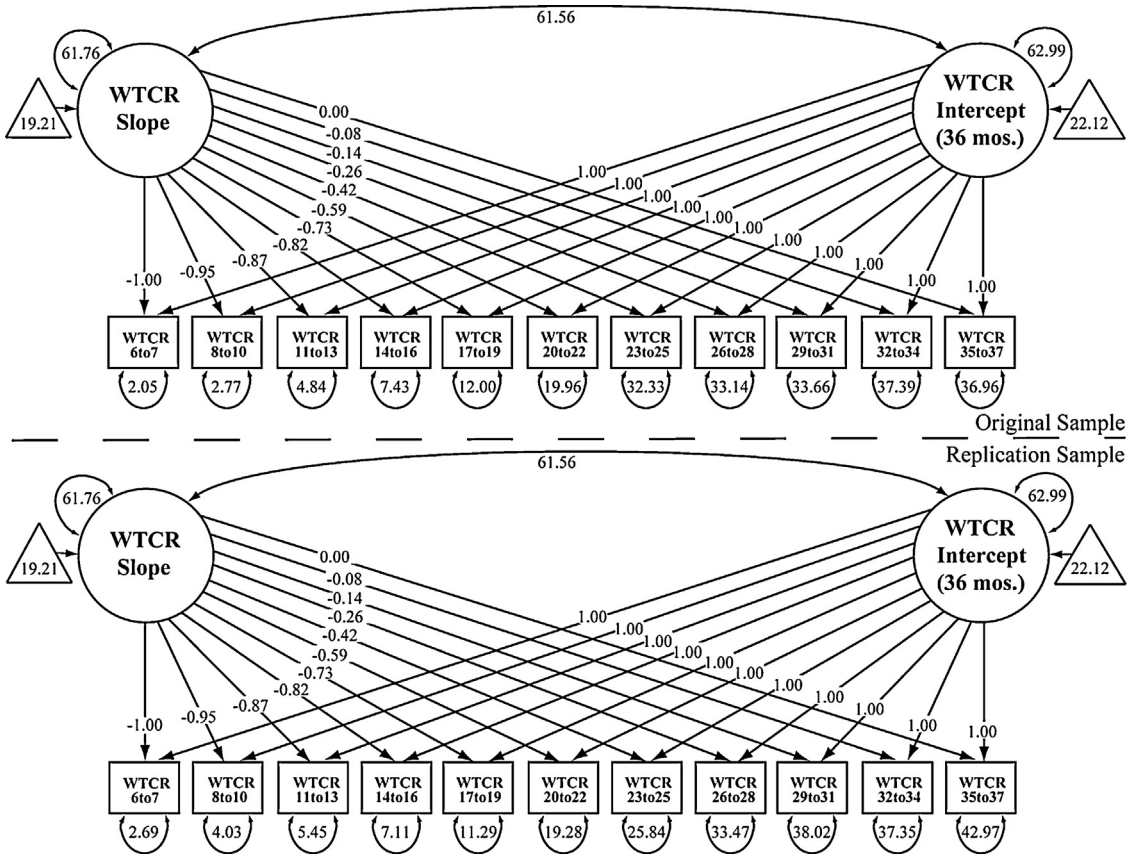

Fig. 5. Latent mean, variance, and covariance invariant model for weighted composite total communication rate.
Model Fit: χ^2 (118, n = 5478) = 536.28, p < 0.001, *RMSEA* = 0.036 (0.033 −0.039), *CFI* = 0.897, *NNFI* = 0.904.

these two aspects in our ECI definitions, training materials, and calibration standards for certification.

Specifically, signs were to be coded as words and not gestures, and gestures in the form of toy movements directed at the play partner were scored either with or without associated eye contact. Changes included: (a) revised coding definitions using additional examples and non-examples, (b) additional practice videos for teaching gesture coding, and (c) an updated set of calibration master scores against which trainees' scores from video were compared for agreement and certification. These changes were implemented after the March 7, 2007 cut point, affecting measurement in the replication sample. At this time, we also modernized the online data system adding data entry validation checks and warnings to check data entry when extreme values (e.g. three *SDs* above the mean) were entered. This also may have influenced gesture coding at data entry. Other than these changes, we could not identify other reasonable time-covarying threats to gestures invariance.

It was interesting to note that even though gestures did not meet strong factorial and structural invariance, this did not adversely impact the invariance of ECI total communication of which gestures was an aggregate part. We note that gestures overall was the least frequently occurring communication skill overall, and its relatively flat trajectory compared to the other three skills that had more acceleration and in some cases (vocalization, single words) also deceleration (see Figs. 1–5). Also gestures were scored with a weight of 1 as compared to single and multiple words that were scored with weights of 2 and 3 respectively. Vocalization was also weighted only 1 but it occurred relatively more frequently as well. Thus, gestures may have simply washed out its contribution to the total communication score in favor of the greater influence of the other more frequent and heavier-weighted skills. In summary, we believe that the replication sample findings based on data collected with the improved set of tools now represent the correct scaling of gestures. However, future demonstration is needed.

3.1. Limitations

The data analyzed in this report were both cross-sectional and longitudinal in nature. Interpreting developmental trajectories from data that are not entirely longitudinal involves a degree of caution in that cross-sectional data are known not to be fully representative of longitudinally collected data (Kraemer, Yesavage, Taylor, & Kupfer, 2000; Maxwell & Cole, 2007). However, this concern was attenuated somewhat in our combined design by the fact that most children had more than one ECI measurement occasion.

Analyses in this investigation were based on child-level data and did not account for program-level cluster effects. With the small number of 15 program clusters coupled with the complexity of the current multi-group latent growth models and their limitations (Hoxx & Maas, 2001), teasing out cluster effects await future research. For similar reasons, we also did not conduct subpopulation analyses.

3.2. Implications for future research and practice

This study is important to the field because it strengthens the construct validity of the ECI as a measure of growth and change in children's early communication skills across the 6–37 months age span. In the context of two large samples of young children learning to communicate, strong evidence was provided that ECI measurement consisted of similarly shaped growth trajectories over time, similar variations around the mean, and similar peak

estimates across the groups. These findings replicated prior reports regarding the functional form of the ECI key skills (Greenwood et al., 2010, 2013). The present findings also cross-validated the sequential ordering of development in cross-skill trajectories in terms of their invariance properties in both samples (Greenwood et al., 2013).

Important in these analyses was that they held true even though the two samples of children compared were historically distinct with measurement carried out uninterrupted over a total span of 10 years by changing groups of similarly trained local staff assessors during this period. In terms of the ECI total communication score, the appropriateness of using it for determining risk and monitoring progress decisions for individual and groups of children was further validated. The key skills invariance findings for vocalizations, single, and multiple words supported future work using them for risk determination and progress monitoring. Overall, the data in these two samples were trustworthy, and comparable, and the ECI protocol and its procedures for measurement by program staff over long periods of time produced high-quality data without direct oversight of a research staff. This insured that the decisions based on data from the system, the total communication score in particular, led to their intended and not unintended results (American Educational Research Association, 1999; Council of Chief State School Officers, 2004; Widaman et al., 2010). The results also illustrate how changes and improvements in a measure can be evaluated over time and scale.

The present findings also demonstrated that the ECI's scales performed relatively well accommodating measurement without ceiling effects and capturing the behavioral changes of both the youngest and oldest children in the range. The inclusion of the four key skills in the standard ECI protocol ensured that at least one if not several measures were available to capture growth and change in communication at each age occasion.

These findings also contributed meaningfully to the theory and empirical basis of what we know about young children's of language development. Because the ECI taps indicators of both prelinguistic and spoken language, present findings confirmed and replicated much of what is reported in the language literature about children's growth in early communication skills. Specifically, growth in earlier prelinguistic skills preceded growth in spoken language in a sequential pattern of acquisition to a peak level followed by decline in earlier skills in favor of new, emerging skills (i.e., gestures followed by vocalizations followed by single words followed by multiple word utterances).

Overall, this is very good news because the ECI data were collected by practitioners in EHS programs longitudinally at large scale and not by professional assessors. Of course, the design and intention of progress monitoring measurement is for practitioners to administer and use the information to make intervention decisions for individual children. Therefore, the concern that practitioner data collection introduces bias in results intended for use in intervention decision-making and accountability reporting appeared in this case to be attenuated.

Present findings in combination with what is currently known, further strengthens the construct validity of the ECI used for its intended purposes of universal screening, intervention decision-making, progress monitoring of infants and toddlers, and program accountability.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, *59*, 450–466.

Advisory Committee on Head Start Research and Evaluation. (2012). *Final report*. Retrieved from. http://www.acf.hhs.gov/sites/default/files/opre/eval_final.pdf

American Educational Research Association. (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. In D. K. Dickinson, & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2) (pp. 41–51). New York, NY: Guilford.

Brady, N., Marquis, J., Fleming, L., & McLean, L. (2004). Prelinguistic predictor of language growth in children with developmental disabilities. *Journal of Speech, Language and Hearing Research*, *47*, 663–677.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior*, *32*(1), 83–92.

Buzhardt, J., Greenwood, C. R., Walker, D., Anderson, R., Howard, W. J., & 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. R., Walker, D., Carta, J. J., Terry, B., & Garrett, M. (2010). Web-based tools to support the use of data-based early intervention decision making. *Topics in Early Childhood Special Education*, *29*(4), 201–214.

Carta, J. J., Greenwood, C. R., Walker, D., & Buzhardt, J. (2010). *Using IGDIs: Monitoring progress and improving intervention results for infants and young children*. Baltimore, MD: Brookes.

Carta, J. J., Greenwood, C. R., Walker, D., Kaminski, R., Good, R., McConnell, S. R., et al. (2002). Individual growth and development indicators (IGDIs): Assessment that guides intervention for young children. *Young Exceptional Children Monograph Series*, *4*, 15–28.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*(4), 330–351.

Council of Chief State School Officers. (2004). *A framework for examining validity in state accountability systems*. Retrieved from. http://www.ccsso.org/Documents/2004/Framework_For_Examining_Validity_2004.pdf

Dale, P., Price, T., Bishop, D., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research*, *46*, 544–560.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validiation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149.

Early Head Start National Resource Center. (2012). *School readiness goals for infants and toddlers in Head Start and Early Head Start programs: Examples from the Early Head Start National Resource Center*. Retrieved from. http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/ehsnrc/Early%20Head%20Start/early-learning/curriculum/school-readiness-goals-infants-toddlers.pdf

ECO Center. (2011). *SPR reporting requirements. Washington, DC*. Retrieved from. http://www.ed.gov/about/reports/annual/2007plan/program.html

Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist*, *61*, 50–55.

Enders, C. K. (2006). Analyzing structural equation models with missing data. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 313–342). Greenwich, CT: Information Age Publishing.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

Fenson, L., Dale, P. S., Resnick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, Serial No. 242*, *59*(5), 1–185.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analyic procedures for the determination of measurement invariance. *Structural Equation Modeling*, *13*(3), 378–402.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, *8*(3), 206–213.

Greenwood, C. R., Buzhardt, J., Walker, D., Howard, W. J., & Anderson, R. (2011). Program-level influences on the measurement of early communication for infants and toddlers in Early Head Start. *Journal of Early Intervention*, *33*(2), 110–134.

Greenwood, C. R., Carta, J. J., Baggett, K., Buzhardt, J., Walker, D., & Terry, B. (2008). Best practices in integrating progress monitoring and response-to-intervention concepts into early childhood systems. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology* (Vol. 5) (pp. 535–548). Washington, DC: National Association of School Psychology.

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.

Greenwood, C. R., Carta, J. J., & Walker, D. (2005). Individual growth and development indicators (IGDIs): Tools for assessing intervention results for infants and toddlers. In W. J. Heward, H. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, R. Gardner III, L. D. Peterson, & S. B. Hersh (Eds.), *Focus on behavior analysis in education: Achievements, challenges, and opportunities* (pp. 103–124). Columbus, OH: Pearson/Prentice-Hall.

Greenwood, C. R., Carta, J. J., Walker, D., Buzhardt, J., & Baggett, K. (2006). *Individual indicators of growth and development (IGDI): The infant and toddler website*. Retrieved from. http://www.igdi.ku.edu

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.

Greenwood, C. R., & Walker, D. (2010). Development and validation of IGDIs. In J. J. Carta, C. R. Greenwood, D. Walker, & J. Buzhardt (Eds.), *Using IGDIs: Monitoring progress and improving intervention for infants and young children* (pp. 159–177). Baltimore, MD: Brooks.

Greenwood, C. R., Walker, D., & Buzhardt, J. (2010). The Early Communication Indicator (ECI) 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. J., McCune, L., & Anderson, R. (2013). Evidence of a continuum in foundational expressive communication skills. *Early Childhood Research Quarterly*, *28*, 540–554.

Harjusola-Webb, S. M. (2006). *The use of naturalistic communication intervention with young children who have developmental disabilities*. Lawrence, KS: University of Kansas. Unpublished Doctoral Dissertation.

Hartmann, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis*, *10*, 103–116.

Hoxx, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, *8*(2), 157–174.

Individuals with Disabilities Education Improvement Act. (2004). 20 U.S.C., § 1400 et seq.

Iverson, J., Longobardi, E., & Caselli, M. C. (2003). Relationship between gestures and words in children with Down's syndrome and typically developing children in the early stages of communicative development. *International Journal of Language Communication Disorders*, *38*, 179–197.

Kirk, S. (2006). *Using outcome measures and progress monitoring to guide language-promoting interventions in Early Head Start Programs*. Lawrence, KS: University of Kansas. Unpublished Doctoral Dissertation.

Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How can we learn about developmental processes from cross-sectional studies, or can we? *American Journal of Psychiatry*, *157*(2), 163–171.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, *32*, 53–76.

Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, *31*, 357–365.

Little, T. D., & Slegers, D. W. (2005). Factor analysis: Multiple groups. In B. Everitt, D. Howell, & D. Rindskopf (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 2) (pp. 617–623). Chichester, UK: Wiley.

Littvay, L. (2009). Questionnaire design considerations with planned missing data. *Review of Psychology*, *16*(2), 103–113.

Love, J., Chazan-Cohen, R., Raikes, H., & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development*, *78*(1), 1–173.

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

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23–44.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., Vol. 2, pp. 561–614). New York, NY: Plenum.

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.

Muthén, L. K., & Muthén, B. O. (1998–2010). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

NAEYC, DEC, & NHSA. (2013). Frameworks for response to intervention in early childhood education: Description and implications. Jointly Developed by the Division for Early Childhood of the Council for Exceptional Children, the National Association for the Education of Young Children, and the National Head Start Association. Retrieved from http://www.dec-sped.org/uploads/docs/about_dec/position_concept_papers/DEC_NAEYC_NHSA%20Joint%20Paper%20on%20RTI%20in%20Early%20Childhood_final.pdf

National Early Childhood Accountability Task Force. (2007). *Taking stock: Assessing and improving early childhood learning and program quality. The report of the National Early Childhood Accountability Task Force*. Retrieved from. http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/Pre-k_education/task_force_report1.pdf

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods*, *6*(3), 328.

NICHD Early Child Care Research Network. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology*, *41*, 428–442.

Office of Head Start. (2012). *Report to congress on the final Head Start program designation renewal system*. Retrieved from. http://eclkc.ohs.acf.hhs.gov/hslc/mr/rc/Head_Start_Designation_Renewal_System_Final_Rule.pdf

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders*, *32*, 223–245.

Paul, R., & Roth, F. P. (2011). Characterizing and predicting outcomes of communication delays in infants and toddlers: Implications for clinical practice. *Language, Speech, and Hearing Services in Schools*, *42*, 331–340.

Preacher, K. J., Wichman, A. L., MacCallum, R. C., & Briggs, N. E. (2008). *Latent growth curve modeling*. Los Angeles, CA: Sage.

Priest, J. S., McConnell, S. R., Walker, D., Carta, J. J., Kaminski, R., McEvoy, M. A., et al. (2001). General growth outcomes for children: Developing a foundation for continuous progress measurement. *Journal of Early Intervention*, *24*(3), 163–180.

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year old children with or without a history of late lanaguage emergence at 24 months. *American Speech and Hearing Association*, *51*, 394–407.

Rindskopf, F. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*, 37–47.

Rowe, M. L., & Goldin-Meadow, S. (2009). Early gesture selectively predicts later language learning. *Developmental Science*, *12*, 182–187.

Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Jounrnal of Counseling Psychology*, *57*(1), 1–10.

Seaman, M. A., & Serlin, R. C. (1998). Equivalence confidence intervals for two group comparisons. *Psychological Methods*, *3*, 403–411.

Shanahan, T., & Lonigan, C. J. (2008). *Developing early literacy: A report of the National Early Literacy Panel*. Retrieved from. http://www.nifl.gov/publications/pdf/NELPReport09.pdf

Siegel, W. C. (2000). Individualization: An essential element of the curriculum. *Policy, Head Start Bulletin*, (67), 1–36. Retrieved from. http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/eecd/Individualization/Learning%20Styles/edudev_art_00039_071005.html

Vallotton, C. D., & Ayoub, C. C. (2010). Symbols build communication and thought: The role of gestures and words in the development of engagement skills and social-emotional concepts during toddlerhood. *Social Development*, *19*, 601–626.

Vandereet, J., Maes, B., Lembrechts, D., & Zink, I. (2011). The role of gestures in the transition from one-two-word speech in a variety of children with intellectual disabilities. *International Journal of Language Communication Disorders*, *46*, 714–727.

Walker, D., & Buzhardt, J. (2009). *ECI training videos*. Retrieved from. http://www.igdi.ku.edu/training/ECI_training/ECI_videos/

Walker, D., & Buzhardt, J. (2010). IGDI administration: Coding, scoring, and graphing. 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. 23–35). Baltimore, MD: Paul H. Brookes.

Walker, D., & Carta, J. J. (2010). The communication IGDI: Early communication 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. 39–56). Baltimore, MD: Paul H. Brookes.

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factoral invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10–18.

Yoder, P. J., Warren, S. F., & Macathren, S. B. (1998). Determining spoken language prognosis in children with developmental disabilitie. *American Journal of Speech-Language Pathology*, *7*, 77–87.

Yuan, K., Yang-Wallentin, F., & Bentler, P. M. (2012). ML versus MI for missing data with violation of distribution conditions. *Sociological Methods and Research*, *41*(4), 598–629.

Zhang, B., & Walker, C. M. (2008). Impact of missing data on person-model fit and person trait estimation. *Applied Psychological Measurement*, *32*(6), 466–479.

Zimmerman, I. L., Steiner, V. G., & Pond, R. V. (1992). *Preschool language scale - 3*. San Antonio, TX: The Psychological Corporation.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research*, *50*, 1562–1592.