FEATURE ARTICLE

Developing a General Outcome Measure of Growth in Social Skills for Infants and Toddlers

JUDITH J. CARTA AND CHARLES R. GREENWOOD
University of Kansas

GAYLE J. LUZE
Iowa State University

GABRIEL CLINE
Marshfield Missouri School District

SUSAN KUNTZ
University of Kansas

Proficiency in social interaction with adults and peers is an important outcome in early childhood. The development of an experimental measure for assessing growth in social skills in children birth to 3 years is described. Based on the general outcome measurement (GOM) approach (e.g., Deno, 1997), the measure is intended for use by early intervention practitioners to identify children having difficulty acquiring social interaction skills and to monitor children’s progress during the course of interventions. Results from a sample of 57 infants and toddlers demonstrated the feasibility of 1 of 14 possible measures, positive verbal social behavior (PVSB), in terms of its sensitivity to growth over time, psychometric properties, and practicality for use by early interventionists. Implications for future research and practice are discussed.

One of the foremost developmental tasks of early childhood is learning how to establish relationships (see National Research Council and Institute of Medicine, 2000; Rubin, Bukowski, & Parker, 1998). Children's early interaction patterns with adults and peers set the stage for their later social competence or deviance (Kuperschmidt & Coie, 1990) and early social competence affects the way children evaluate their own self-worth and their views of the world as friendly or hostile (Ladd & Price, 1986). Consequently, a major focus of early education and care has been the development of young children's social and emotional competence (National Research Council, 2001).

Contemporary views of the early development of social competency involve the growth of social and communicative capabilities (e.g., Goldstein & Morgan, 2002). Communication skills (verbal and non-verbal) are required to interact socially and are developed in the context of social relationships with others (Odom, McConnell, & McEvoy, 1992). The beginnings of social competence appear in the first few months of life.
Infants respond to the emotional bids of parents, but within a few months move to shared attention to and manipulation of objects, with parents often providing the necessary scaffolding to develop more advanced interactive skills (Eckerman, 1993). Infants anticipate social interactions with their caregivers and patterns of vocal turn-taking emerge (Ladd & Price, 1986; Tronick, Cohn, & Shea, 1986). Parents use play and simple games such as patty-cake or peek-a-boo, to provide a repetitive social routine that helps children learn to take turns and acquire interactive communication skills in a format that can be kept close to the child’s current level of functioning or that can expand to more complex levels as needed (French, Boynton, & Hodges, 1991; Snow, 1984). As children’s verbal skills develop they will engage in verbal interactions with adults, with parents progressively transferring more control of the interaction to the young child (Eckerman, 1993).

By 2 months, babies develop social interest and will stare intently at one another (Eckerman, Whatley, & McGehee, 1979). They first develop skills that coordinate their play with another child’s by imitating, initiating, and responding to simple social overtures by approximately 6 months of age (Eckerman, Davis, & Didow, 1989). By 6 to 9 months of age, they will babble and smile at other babies, and sometimes try to get attention or initiate social bids (Vandell, Wilson, & Buchanan, 1980). Between 1 and 2 years of age, children’s interactions become more sustained and complex. They begin to engage in simple turn-taking during play revealing an emerging ability to understand their playmate’s intent and take on reciprocal roles (Mueller, 1972). Their growing communicative capabilities of turn-taking, sharing meaning with others, understanding others’ intent, and taking on reciprocal roles all provide groundwork necessary for coordinated play (Howes & Matheson, 1992). After 16 months of age children can follow and begin to lead well-practiced interaction rituals; by 24 months children are skilled at imitating the non-verbal play of others and making up simple social games; and by 28 months they add verbal interactions to the nonverbal coordinated play (Eckerman, 1993; Eckerman & Didow, 1989).

While differences exist between the interactions young children have with peers compared to their interactions with adults, important relationships occur in both types of interactions (Beckman & Lieber, 1992). Children tend to follow the same developmental progression when interacting with peers, but verbal peer interaction tends to be somewhat delayed compared to verbal interaction with adults (French et al., 1991). Between 2- and 3 years-of-age, children with experience interacting together usually begin to exhibit episodes of coordinated play and are capable of establishing rudimentary friendships. Children at this age have been shown to be more likely to initiate play and engage in complex interactions with familiar than with unfamiliar agemates (Howes & Matheson, 1992; Rubin et al., 1998).

Several well-known taxonomies have been used to describe the social play of young children in terms of increased levels of competence and increasingly sophisticated levels of interaction with peers (Howes & Matheson, 1992; Parten, 1932). Examples of play levels in these taxonomies include onlooker, parallel play, and simple social play. These taxonomies provide helpful information for understanding the types of interactions children have during play. This information can provide a preliminary basis for understanding when and how to intervene to help children advance in their play and social interaction skills with adults and peers. However, these taxonomies of young children’s social behavior with adults, siblings, and peers often are not capable of providing early interventionists with practical information that can be used to measure growth over a short period of time, which is necessary for evaluating effectiveness of interventions.

While social competence is generally thought to unfold according to a predictable sequence, up to 10% of children have social interaction difficulties severe enough to result in rejection by peers (Asher, 1990) or, in some cases, growth in social competence is uneven. For example, some children are more skilled
in interactions with adults, but have problems interacting with peers (e.g., children with speech and autism spectrum disorders), and vice versa (e.g., children with conduct and social maladjustment problems). Intrinsic factors (e.g., difficulty with attention, temperament, disability, memory) and factors external to the child (e.g., unsupportive caregiving, limited access to peer social networks) affect social competence (Guralnick & Neville, 1997). Determining when the social competence of a child younger than 3 years is deviating from the typical developmental trajectory and in need of intervention support is currently a challenging task. The broad variation of emerging, acceptable social behaviors for children this age makes it difficult to discriminate among individual differences within a typical range versus true maturational delays and specific social-emotional disorders. Norm-referenced, criterion-referenced, and informally developed scales (e.g., Bagnato, Neisworth, Salvia, & Hunt, 1999; Squires, Bricker, & Twombly, 2001) have inherent limitations due to their structure or intended purpose.

One issue is that these measures may not have sufficient sensitivity to help early interventionists detect incremental growth toward a socially valid general outcome over short time periods. Another concern with some measures is that the contexts in which child assessments are made are not authentic, that is, made by strangers in highly unfamiliar settings differing substantially from those in which adult-child, child-peer, and adult-child-peer interactions occur naturally (e.g., care and play routines). Consequently, results assessed in analogue situations may represent different aspects of social development than those of primary interest. Another limitation of existing measures is that many are designed to be re-administered at periods ranging from 6 months to 2 years. In this case, information often cannot be used formatively to reflect recent progress and to help select appropriate interventions or guide intervention decisions.

Collectively, these limitations of existing measures make it difficult to (a) identify accurately when young children are, or are not, following a normative pattern of growth in social competence (National Research Council and Institute of Medicine, 2000), and (b) monitor the effects of early interventions for accelerating growth in social competence, both of which are requirements of Part C of the Individuals with Disabilities Act (IDEA, 1997). IDEA requires evaluations to determine children’s initial and continuing eligibility and their level of functioning across developmental domains. Also required is information relevant to monitoring progress toward goals contained in the Individualized Family Service Plan (IFSP) or Individualized Education Plan (IEP). To ensure that intervention services designed to improve children’s social skills are appropriate and effective, measures that can be used frequently to monitor progress and display growth over time are needed.

An increasingly well-known approach relevant to the goals of measuring the progress of individual children and informing or guiding intervention services is General Outcome Measurement (GOM; Fuchs & Deno, 1991; Greenwood, Luze, & Carta, 2002; McConnell, McEvoy, & Priest, 2002; McConnell, Priest, Davis, & McEvoy, 2001). Deno, Mirkin, and Chaing (1982) described seven criteria that GOMs must meet. First, GOMs should assess key skill elements representative of an important child outcome. Second, GOMs should identify “authentic” child behaviors in natural settings. Authentic behaviors are particularly important for infants and toddlers because they are less able and willing to “perform” specific skills on demand as compared to children who are older. Third, GOM assessments should be standardized and replicable to ensure that data from separate administrations are comparable. Fourth, GOMs must be sensitive to growth over a short period of time so they can be used to evaluate intervention effectiveness. Fifth, GOMs must meet the requirements of technical adequacy, including interobserver agreement and reliability of alternate forms, to provide accurate information that can be interpreted and used for decision making. Sixth, GOMs should be efficient and economical so early interventionists might
gather repeated measurements that are usable for decision-making without adding to their workload. Seventh, GOMs should be sensitive to intervention and intervention effects.

Given these criteria, the purpose of this research was to investigate the feasibility of an experimental social GOM for use by early interventionists with children birth to 3-years-of-age. This work was carried out as part of the larger effort by the Early Childhood Research Institute on Measuring Growth and Development (Priest et al., 2001) to develop GOMs for children birth to 8 years. The overarching goal of the present study was to investigate the technical adequacy of a social GOM. The research questions addressed were: (a) What kinds of social behavior are produced by infants and toddlers? (b) Which social behaviors show promising growth trajectories over time? (c) Can technical adequacy be demonstrated? and (d) Can this measurement be reasonably carried out by early interventionists?

**METHOD**

**Background**

A general outcome statement related to social competence, one of 15 statements for children birth to age 8, was socially validated in a national survey of parents and practitioners (Priest et al., 2001). This outcome statement was “Child interacts with peers and adults, maintaining social interaction and participating socially in home, school, and community” and it was used to guide development of the GOM described below. Preliminary work was completed to (a) identify key social skill elements in the literature for this outcome for children birth to 3 years, (b) determine a set of standard toys and play situations to evoke key skills, and (c) develop a direct observation recording system for recording these skills in real time. Our goal was to develop an instrument usable by a wide range of early interventionists and child care practitioners that would involve familiar adults and peers to engage young children and evoke social behaviors in a standard protocol.

**GOM conceptual framework and taxonomy.**

A review of the social skills literature relevant to infants and toddlers was conducted including a content review of existing instruments (e.g., Bricker, 2002). The review established that (a) the function of social interaction is to achieve a response from another person ranging from proximity, to simple reciprocity, to coordinated action to achieve a common outcome or the friendship-acceptance of peers (Eckerman, 1996; Warren, Yoder, & Leew, 2002); and (b) social interaction in infants develops first with adults, and then with peers (Didow & Eckerman, 2001; Eckerman et al., 1989). Based on this review, a social competence developmental framework for infants and toddlers was mapped leading from the guiding general outcome statement for children birth to 3 years of age. The purpose of this framework was to provide an umbrella for a small but sensitive skill set that would become the social GOM indicator. The aim was not to provide a comprehensive conceptualization of the literature linked to developmental milestones, but rather to represent guiding constructs and key skills. This framework separated positive from negative social behaviors.

With respect to crying in children this age, it was considered acceptable behavior but not coded as either positive or negative. The framework separated refinements and expansions in positive nonverbal behaviors (i.e., smiles, gestures) expected in infants, from positive verbal social behavior (i.e., use of words in greetings, bids to play) expected of children in the 12-month and older age range. These positive nonverbal and verbal social behaviors also were expected to be distributed differently to adults, to peers, and to one or both persons (i.e., nondirected) in a play situation occurring more frequently with adults in younger children and with peers in older children. From this framework, detailed behavioral definitions were developed for recording (see Appendix). In all, seven key skill elements and seven composites based on simple combinations of these elements were generated from this taxonomy.

**GOM Materials**

Play situations that included a familiar adult, a peer, and toys with potential for evoking so-
cial behaviors in young children were evaluated. Because children were observed to have quite variable counts of social behavior in natural classroom situations due to the range of play materials and play partners available at any time (Luze, 2001), a standardized situation was chosen for further study as in prior GOM validation studies for children this age (see Greenwood, Luze, Cline, Kuntz, & Leitschuh, 2002; Luze et al., 2001). Additional pilot testing examined the familiar adult partner's play repertoire for its social behavior evoking potential. In general, we wanted the play partner to comment on and follow the child's lead rather than direct it. When children were not engaged with toys or each other, it was less clear whether the amount of adult social initiation affected total social behavior production. Previous studies have reported that children tended to interact more with each other if the adult played a less active role (e.g., Hops & Greenwood, 1988). Based on an alternating treatments design comparison test of active versus passive adult roles, it was observed that more child social behavior was produced when the adult did initiate during "down" times (Luze, Greenwood, Carta, Cline, & Kuntz, 2002). Therefore, this procedure was included as part of the standard administration procedure.

Following Deno's GOM criteria, a wide variety of common toys were evaluated and a subset selected based on potential for engaging the interest and evoking the social skills of infants and toddlers (Luze, 2001). A review of literature pertaining to children's social interactions, play situations, and toys was conducted (Beckman & Kohl, 1984; Eckerman et al., 1989; Odom & Strain, 1984; Quilitch & Risley, 1973; Sainato & Carta, 1992). Articles that included definitions of social interaction also were examined to identify types of toys and play situations and to develop initial definitional categories of social interaction (Boudrant-Utz & Luciano, 1994; Holmberg, 1980; Strain, 1985; Tremblay, Strain, Hendrickson, & Shores, 1981; Twardosz, Nordquist, Simon, & Botkin, 1983). In addition to empirical evidence that a type of toy promoted social interaction, selected toys (a) were of interest to children across the birth to 3-year age range, (b) were representative of toys typically available to children at home or in childcare settings, (c) allowed for two children to play together along with an adult who facilitated the play, and (d) were sufficiently portable so research assessors could easily take them from one center to another. Additional considerations included safety and developmental appropriateness for children from birth to 36 months.

A number of toy sets with potential for promoting social interaction in children in the target age range resulted from these analyses. After several weeks of piloting toys that could act as equivalent, alternate forms, those found not to encourage social interaction or to keep children's attention over an 8-minute play period were eliminated. Emerging from these pilots was a revised list of toys ranked in demonstrated probability of evoking social behaviors, including the Tub of Toys (TT), Window House (WH), Kitchen with Dishes (KD), and Fun Sounds Garage (FSG). The TT contained an assortment of toys including a bowl and three balls, two Sesame Street® figures in a space ship, and four squeezable animal figures. The WH is a large foam, rubber-filled plastic box made by Environments, Inc.® The WH opens on both ends so that children can crawl through it. It includes two see-through mesh covered windows and one open window. In addition to crawling through, children may stand up in the house, and look out the windows. Balls (two small, one medium-sized) also were added so children could roll and throw them into the house. The KD was Tyco's Sesame Street® Surprise Sounds Kitchen with added dishes (two sets each with cup, glass, plate, bowl, and spoon) and squeaky food. The FSG by Fisher-Price is a multi-level garage with elevator, carwash, food stand, two figures, and two vehicles. Some of the garage elements emit battery driven sounds, including a cash register and gas pump that ring. The FSG is evocative of imaginative play. The TT, WH, KD, and FSG toy sets and the social recording protocol were tested further. Pilot results indicated equivalence in the amount of social behavior evoked by 3 of the 4 toys.

*Carta et al.*
FSG evoked comparatively less social behavior and it was removed from further consideration (Luze, 2001). TT, WH, and KD were used as alternate toy forms in the present study.

**Participants**

Children were recruited at five child care centers serving infants and toddlers located in metropolitan Kansas City. The centers served children of varied racial and socioeconomic backgrounds and three centers also served children with special needs. Centers represented a range of parochial- and private-sponsoring groups and some were affiliated with neighboring high schools serving adolescent mothers. Children in the birth to 36-month age range were eligible to participate in the study. Each eligible child’s parent received a packet of information that included an informed consent form and demographic questionnaire. Any child whose parents returned a signed informed consent participated over the next 9 months.

Sixty-two children initially were enrolled in the study. By the end of the study, 5 children had moved and were dropped from the study. The analysis sample was comprised of 57 children, 32 (56%) males and 25 (44%) females. Based on age at start, these children were assigned to 3 cohorts defined by the first, second, and third year of life. Cohorts 1–3 included 8, 21, and 28 children, respectively. The modal level of mothers’ attained education was 12th grade, with 21% of mothers indicating that this was their highest level of completed schooling. Mothers’ reported levels of attained education beyond high school, based on 43 parents who completed information, included: vocational education, 4 (9%); bachelor’s degree, 18 (42%); master’s degree, 9 (21%); and doctorate, 3 (7%).

With regard to disability status, parents of 9 of the 57 children reported their child had some type of special need. These included five children with Down syndrome (one had a ventricle heart defect; another retinopathy of prematurity, an inflammation of the retina of the premature infant; and another with bronchopulmonary dysplasia experiencing asthma and occasional difficulties breathing); two with pervasive developmental disorder; one with Williams syndrome; and one with a seizure disorder. Each of these nine children had Individualized Family Service Plans (IFSPs) at enrollment. Five additional children were subsequently identified with a social delay by research staff based on social developmental quotient scores at −1.5 SD of the mean or below on the Vineland Social Emotional Early Childhood Scales (Sparrow, Balla, & Cicchetti, 1998b). Three of these five children identified by research staff also had IFSPs.

**Measures**

**Criterion social measures.** To provide a multi-method (direct observation vs. report), multi-informant (professional vs. parent) dimension to the validity analysis (Lewin, Hops, Davis, & Dishion, 1994), two criterion measures appropriate for very young children were used. One was a standardized parent interview widely used in the evaluation of young children, the Vineland Social-Emotional Early Childhood Scales (Sparrow, Balla, & Cicchetti, 1998b); the other was a well-established measure of play, the Howes Peer Play Scale (Howes, 1980).

The Vineland is a measure of social-emotional functioning for children birth through 5 years, 11 months of age administered to a respondent who is familiar with the child’s behavior via a semi-structured interview (Sparrow et al., 1998b). It required 15 to 25 minutes to complete. Subscales used in this investigation included the interpersonal relationships, play/leisure time scales, coping skills, and the composite total. The first two scales were administered to all children; the coping scale was administered only to children 2 years and older. Test-retest reliability on the overall composite score is reported to range from .77 to .79 and subscale score reliabilities ranged from .73 to .79 in a nationally representative sample of 3000 children selected to match the 1980 U.S. Census characteristics of age, gender, community size, geographic region, parent education, and ethnic status (Sparrow et al., 1998b). Concurrent validity score correlations with the Vineland
Adaptive Behavior Scale Survey (Sparrow, Balla, & Cicchetti, 1998a) and the Scales of Independent Behavior (Bruininks, Woodcock, Weatherman, & Hill, 2000) were .65 and .63, respectively (Kush, 2003).

The Howes Peer Play Scale (Howes, 1980, 1987) is an observational measure of children’s interactive play with peers for children aged 12 months to 5 years. An interval recording procedure (20 sec per interval) was used for coding the occurrence of five categories of play with categories ranging from least to most complex (i.e., parallel play, parallel play with mutual regard, simple social play, complementary and reciprocal action, and reciprocal social play, respectively). Children were observed in the classroom during free play or similar activities that allowed children to make choices about play and peer interaction. Each observation lasted 15 min or 45 intervals in all. Scores were calculated as the percentage of peer play at each level of complexity (Howes & Matheson, 1992) by dividing the frequency of intervals in which each play form occurred by the total intervals observed times 100. A total play score was computed as the simple sum across scales.

Interobserver reliability coefficients ranging from .87 to .93 (Md = .89) with Guttman scale coefficients of .90 and .68 for reproducibility and scalability were reported in (a) a cross-sectional sample that included 329 children between 12 and 53 months of age, and (b) two longitudinal samples that included 41 children between 16 and 33 months and another 233 children between 1 and 5 years of age (Howes, 1980, 1987). There were equal numbers of boys and girls in all samples, most from two-parent homes. No indices of validity were reported.

The social GOM: The Early Social Indicator. The experimental social GOM, the Early Social Indicator (ESI), like the Howes, was a direct observational measure. The GOM was administered by a familiar adult using 1 of 3 toy sets to engage the play of the target child and a familiar classroom age-matched peer, and to evoke the social behaviors of interest (e.g., positive nonverbal and verbal social behaviors) during a 6-min session. All observations were videotaped and coded at a later date.

Pilot work supported frequency counts versus duration recording in producing greater variation and more social responding. Frequency counts also appeared more sensitive and easier to use by practitioners and best suited for further investigation (Luze, 2001). Additional pilot testing of an observational coding form confirmed the necessity of recording verbal and nonverbal social behaviors by adult, peer, and nondirected social targets. Due to the infrequency of negative behavior, however, a similar six-element categorization was not needed for negative behavior and only a single total negative tally was recorded. Staff used a paper and pencil recording grid to record the occurrence of these key skill elements (see http://www.lsi.ku.edu/jgprojects/igdi/General_Information.htm for Social Coding form).

The left-most column of the grid indicated each of six, 1-min observation subperiods. Coders used a timer to advance their recording down a row after each minute of observation. Additional columns were labeled by the seven social key skill elements with verbal and non-verbal events nested under each of adult, peer, and nondirected social targets. The right most column was labeled negative. These seven key skills frequency scores were summed to form (a) totals in 6 min and (b) divided by total time observed (i.e., 6 min) to form rate-per-minute scores. From these seven skill element frequencies, seven additional logical composites were computed by combining them across social targets (e.g., positive verbal, positive total social, verbal, non-verbal categories).

**Design and Procedures**

An accelerated longitudinal design was used with three age cohorts (birth–12, 13–24, and 25–36 months) and repeated ESI measures were completed for each child with each observation separated by 4 weeks. The ESI measures occurred between beginning and ending administrations of the criterion measures. After orienting and recruiting prospective child care center staff, informed parental consent was obtained, and the first administration of
Vineland and Howes criterion measures completed. Parents received gift certificates in the amount of $10 after completing the pre- and posttest parent measures including the Vineland. Next a total of 326 ESI observations were collected for 57 children, followed by the second administration of the criterion measures.

With the exception of changes in the alternate toy sets, the administration of the ESI was consistent across occasions. Children were brought to the play room in pairs and played with one of the toy sets, the peer play partner, and 1 of 10 qualified staff members in the role of the adult play partner. All staff had previous experience teaching young children and all had histories of recent work in the child care center in which they were assigned to conduct assessments. Thus, all were familiar to the children whom they were assessing.

The ESI was generally administered in an available room containing the toys and peer located away from the distraction of the classroom. In one child care center, a general meeting room upstairs from the children’s regular classroom setting served this purpose, in the other center, an infant sleeping room was used. Following the protocol, research staff set up the testing session with the toys and a video camera for recording each session. They returned to the classroom to “warm-up” the children in this familiar environment before bringing them into the testing session. In fewer than 10 instances (3% or 10 out of 326 observations), when a child appeared wary of leaving the familiar environment without being accompanied by an adult caregiver, the adult caregiver was allowed to sit on the perimeter of the testing situation where the child could still see the familiar adult. In several additional cases, when children were hesitant to play with research staff during a session, the ESI assessment was conducted with their parent or caregiver as their play partner instead of a staff member.

Two percent of children had only one ESI observation, whereas another 2% completed between two to five observations. Sixty-eight percent of children had between six to eight observations, and 28% of children had all nine observations. The three alternate forms (TT, WW, and KD) were varied systematically across sessions for each child to counter-balance each child’s social performance by these different situations. Each child received a different form on each occasion and the order was repeated beginning with the fourth occasion. This systematic pattern provided a means of assessing alternate forms reliability by blocking on assessment occasions.

The 326 videotaped ESI assessments were coded by 4 of the 10 staff who also served as play partners. Coding of each session required 6 min to complete with an additional 6 min devoted to calculating scores. These records constituted the primary ESI data collection and assessments of interobserver agreement. Scores were entered into a computer database for subsequent use in charting and other statistical software.

Staff measurement training. Research staff received training in the administration procedures for the criterion measures (i.e., Vineland and Howes), and the ESI. Each assessor then met a 90% criterion for administration reliability on each measure conducted over at least 3 days with children of different ages before collecting data for the study.

A protocol was used to guide the actions of the adult play partner and the administration of the ESI and was designed to evoke the most play and interaction possible. Adult play partners were taught to play with the children, but as much as possible not to be directive. Instead, play partners were taught to “follow the child’s lead” in play with the toys by talking about the toy the child was holding rather than trying to interest the child in a new toy. They also learned to encourage peer interaction when appropriate by commenting when children interacted on their own, for example, “You gave your friend a toy, what nice sharing.” Adult partners also were taught to direct the child’s attention to play or interaction when the child was not engaged in either. A second adult was always present to videotape the target child’s interactions and to provide assistance with session timing.

Play partners initially read and discussed
the protocol in a group meeting with the senior researcher. Subsequently, play partners practiced conducting a minimum of three play sessions that were videotaped and viewed by both the senior researcher and the play partner. The researcher provided feedback and suggestions for improving the partners’ repertoire of setting up and ending the session, following the child’s lead, commenting, and, when necessary, directing attention. Based on these sessions and the protocol, the senior researcher qualified each partner to administer the ESI.

Similarly, staff was trained to code the videotaped ESI administrations using the definitions (see Appendix A). After reading the definitions and discussing them with the senior researcher, staff completed practice observations over five sessions in which the senior researcher was the play partner. To qualify as ready for scoring, staff had to code three consecutive sessions at or above 85% agreement for each of the skill elements and the overall total. In cases of disagreement, observers examined tallies that each had made on variables of concern by each minute of observation. This enabled observers to discuss events and understand areas of agreement/disagreement. For training purposes, agreement was calculated using the frequency ratio (Kazdin, 1982). The ratio was computed by dividing the smaller of the two scores by the larger times 100.

**Statistical Analysis**

The general analytic strategy sought to describe empirically the social behaviors produced by infants and toddlers, identify a key skill or composite indicator with sufficient frequency and variation and sensitivity to growth over time, and assess its reliability and criterion validity. Simple descriptive statistics (mean, standard deviation) were used to represent the magnitude of the various measures and individual variation around the mean and over time. Pearson correlation, dependent t-tests, and ANOVA for repeated measures were used for describing relationships and differences and to test indices of reliability and validity. Partial eta-squared was used to estimate effect sizes in ANOVA. Hierarchical Linear Modeling (HLM5: Bryk & Raudenbush, 1992; Bryk, Raudenbush, Cheong, & Congdon, 2000) was used to examine individual and group growth parameters over time. Level 1 HLM analyses were used to compute slope (rate of growth) and intercept parameter values for individuals and the group. Level 2 HLM analyses were used to test the effects of age cohort on slope and intercept. The t-test values in these models were used to estimate effect size correlations.

Examining individual patterns of growth over time is a particularly appropriate application of HLM (Bryk & Raudenbush, 1992; Hatton, Bailey, Burchinal, & Ferrell, 1997; Luze et al., 2001). HLM produces estimates equivalent to multivariate repeated measures (MRM) methods with several advantages: (a) it explicitly represents individual growth; (b) it has generally more flexible data requirements because repeated measurements are considered nested within the person; and (c) compared to MRM, HLM provides a more flexible specification of the covariance structure among repeated observations supporting hypothesis testing (Bryk & Raudenbush, 1992). HLM also is flexible in accommodating missing data.

A unique advantage of HLM analysis is the ability to compute a mean intercept at a single point in time or test for mean differences between groups at a point in time (Bryk & Raudenbusch, 1992). Termed “centering,” calculating the intercept at a specific point in time produces a unique value for the intercept; however, the value of the linear slope is unaffected. Unless otherwise indicated, the intercept means in this study were centered at the 5th measurement occasion, the middle of the total number of measurement occasions, or at 36 months of age, the endpoint of the birth to 36 months chronological age range of interest.

**RESULTS**

**ESI Interobserver Agreement**

Interobserver agreement on the ESI was assessed on a randomly selected 38% of all 326 assessment tapes. Agreement was checked by
### Table 1.
**Interobserver Agreement Using Pearson r and Tests of Differences Between Observers**

<table>
<thead>
<tr>
<th>Social Target</th>
<th>Skill Element/Composite</th>
<th>Pearson r</th>
<th>Observer 1</th>
<th>Observer 2</th>
<th>Paired t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Adult</td>
<td>Positive Verbal</td>
<td>0.92</td>
<td>8.16</td>
<td>8.15</td>
<td>7.84</td>
</tr>
<tr>
<td></td>
<td>Positive NonVerbal</td>
<td>0.91</td>
<td>13.70</td>
<td>7.29</td>
<td>13.72</td>
</tr>
<tr>
<td></td>
<td>Composite</td>
<td>0.93</td>
<td>21.86</td>
<td>12.69</td>
<td>21.57</td>
</tr>
<tr>
<td>Peer</td>
<td>Positive Verbal</td>
<td>0.82</td>
<td>0.53</td>
<td>1.10</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Positive NonVerbal</td>
<td>0.84</td>
<td>1.97</td>
<td>1.93</td>
<td>1.86</td>
</tr>
<tr>
<td></td>
<td>Composite</td>
<td>0.88</td>
<td>2.50</td>
<td>2.53</td>
<td>2.38</td>
</tr>
<tr>
<td>Non Directed</td>
<td>Positive Verbal</td>
<td>0.73</td>
<td>5.27</td>
<td>4.98</td>
<td>5.36</td>
</tr>
<tr>
<td></td>
<td>Positive NonVerbal</td>
<td>0.72</td>
<td>1.50</td>
<td>2.25</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>Composite</td>
<td>0.73</td>
<td>6.78</td>
<td>5.82</td>
<td>6.67</td>
</tr>
<tr>
<td>Summary</td>
<td>Positive Verbal</td>
<td>0.97</td>
<td>13.97</td>
<td>11.32</td>
<td>13.73</td>
</tr>
<tr>
<td></td>
<td>Positive NonVerbal</td>
<td>0.92</td>
<td>17.17</td>
<td>8.01</td>
<td>16.89</td>
</tr>
<tr>
<td>Composites</td>
<td>Positive Total</td>
<td>0.97</td>
<td>31.18</td>
<td>15.56</td>
<td>30.61</td>
</tr>
<tr>
<td></td>
<td>Negative Total</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>All</td>
<td>Positive and Negative</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

A second person who independently recorded the same videotaped session simultaneously with the primary observer. The two observers' frequency counts for key elements and composite scores were examined for interobserver agreement and measurement reliability. Pearson r was used to calculate the correlation between observers' estimates and the paired t-test was used to test for mean differences in estimates as described by Hartmann (1977). Strong to very strong correlation and the lack of a statistically significant difference between two observers' estimates served as evidence of reliability and agreement (see Table 1). Because of the near zero frequencies of negative social behaviors, it was not possible to estimate occurrence reliability. Thus, reliability estimates were only provided for the six key skill elements and six composites without consideration of negative behavior. Overall, Pearson correlations were strong to very strong, and observers' estimates of key elements or composite scores were statistically equal using alpha = .05.

**Social Behaviors Produced by Infants and Toddlers**

*Criterion measures.* At pretesting, the mean chronological age was 22.6 (SD = 8.9, N = 57) months or 13.1, 19.0, and 27.9 months by cohort respectively. Mean developmental age equivalence scores on the Vineland interpersonal and play/leisure scales (N = 57) were 21.3 and 20.2 months, respectively. The composite Vineland standard score was 98.6 (SD = 18.3). Vineland age equivalence scores were only slightly lower than chronological ages at the time of testing. Children in the older compared to the younger cohorts produced higher levels of both interpersonal (F[2,54] = 10.03, p = .0002, partial eta squared = .218) and play/leisure scores (F[2,54] = 7.53, p = .001, eta squared = .271). Post hoc tests indicated that the pattern of significant mean differences was C1 and C2 < C3, but C1 = C2 for both interpersonal and play/leisure scores. A similar test for the Composite was not statistically significant.

Mean performance on the four lowest categories of the Howes was 32.6% for parallel, 13.2% for parallel/mutual, 4.8% for simple social, and 0.1% for complementary/reciprocal (N = 57). None of the children scored in the highest level: reciprocal social. Children's social performance on the Howes was most frequently occurring in the least complex levels (parallel and parallel/mutual). The Howes total composite was 50.6% occurrence (SD = 20.6). None of the Cohort differences tests were statistically significant, yet there was a
pattern of progressively higher scores on parallel/mutual, simple social, and complementary/reciprocal scales by older cohorts (the more advanced categories) and lower scores for older children on parallel play (the least advanced category).

**ESI key skill elements/composites mean levels.** Children in the sample produced a mean rate of 5.04 (SD = 2.7, N = 57) social behaviors (positive and negative) per minute or a total of 30.2 (SD = 16.1) responses in 6 min with considerable variation across children (see Table 2). Negative responding occurred rarely, if at all, at 0.02, SD = 0.28 (negative) in 6 min. The occurrence of positive verbal (M = 2.39 responses/min) versus positive nonverbal (M = 2.65 responses/min) were nearly equal.

Overall, children were more likely to direct their positive social behavior to the adult, to either or both (nondirected), or the peer in rank order (see Table 2). In the case of the adult (2.09 per min vs. 1.37 per min) and peer (0.34 per min vs. 0.11 per min), children were more likely to interact nonverbally than verbally; just the opposite of nondirected interactions, where children were more likely to produce verbal than nonverbal nondirected social behaviors (0.91 per min vs. 0.21 per min). In general, children in the older cohorts produced more of each positive social behavior or composite compared to children in the younger cohorts.

**Sensitivity to Growth Over Time and Appropriateness for Further Investigation**

**Change in total positive social behavior.** The mean intercept for total positive social behavior, M = 28.9 responses per session (SE = 1.6, t[56] = 17.7, p = .0001, ES = .921), was significantly greater than zero, but the slope, M = −0.26 (SE = 0.28, t[56] = −0.93, p = 0.335), was not. These results indicated considerable variation in the amount of group and individual social behavior with a flat, if not slightly declining, trend over 9 months.

**Change in the focus of total positive social behavior.** The change in children’s total positive social responding (nonverbal plus verbal composite) to social targets over time is shown in Figure 1. In general, positive social behavior was directed most often to the adult, next to either adult or peer or both (nondirected), and least often to the peer. More social behaviors occurred in Cohort 3 compared to Cohort 2 compared to Cohort 1. In Cohort 1, positive social responding was increasingly directed to all three social foci over 9 months. With the exception of the responding received by the adult that declined, the same pattern of increasing positive social responding occurred in Cohort 2. This same pattern also was seen in Cohort 3, but, while social behavior with adults occurred most frequently, it declined substantially over time.

**Changes in positive nonverbal and verbal social behavior.** Similar analyses indicated that positive nonverbal social behavior (PNVSB) declined over 9 months and positive verbal social behavior (PVSB) increased (see Figure 2). Children in Cohort 1 displayed growth in PVSB with a higher but declining trend in PNVSB. For children in Cohort 2, PVSB also increased but less rapidly and nonverbal responding declined as in Cohort 1. For Cohort 3, PVSB was increased in frequency above PNVSB for the first time compared to both younger cohorts, but growth in PVSB was flat over measurement occasions. Again, as in the other cohorts, nonverbal responding showed a declining trend over time in Cohort 3.

Given the preceding analyses, the PVSB appeared to be a parsimonious and promising GOM relative to Deno’s criteria because of its relatively high frequency of occurrence and sensitivity to growth compared to other key skills and composites. For example, total responding and nonverbal responding were flat or declining, while verbal social grew. Compared to positive total social composite (verbal + nonverbal) directed to the adult, a peer, or to either, it offered advantages of both growth and simplicity. Thus, the PVSB composite was selected for additional investigation.

**Sensitivity to Age**

**Growth in PVSB composite by age.** HLM analysis of PVSB indicated statistically sig-
Table 2.  
Key Skill Elements Descriptive Statistics (N = 326 Observations of 57 Children)

<table>
<thead>
<tr>
<th>Social Target</th>
<th>Skill Element</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate</td>
<td>Total Responses</td>
<td>Rate</td>
</tr>
<tr>
<td>Adult</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive Verbal</td>
<td>0.25</td>
<td>0.52</td>
<td>1.47</td>
</tr>
<tr>
<td>Positive NonVerbal</td>
<td>1.20</td>
<td>0.83</td>
<td>7.19</td>
</tr>
<tr>
<td>Composite</td>
<td>1.45</td>
<td>1.24</td>
<td>8.67</td>
</tr>
<tr>
<td>Peer</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive Verbal</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Positive NonVerbal</td>
<td>0.10</td>
<td>0.18</td>
<td>0.58</td>
</tr>
<tr>
<td>Composite</td>
<td>0.10</td>
<td>0.18</td>
<td>0.61</td>
</tr>
<tr>
<td>Non Directed</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive Verbal</td>
<td>0.52</td>
<td>0.55</td>
<td>3.11</td>
</tr>
<tr>
<td>Positive NonVerbal</td>
<td>0.17</td>
<td>0.37</td>
<td>1.03</td>
</tr>
<tr>
<td>Composite</td>
<td>0.69</td>
<td>0.71</td>
<td>4.14</td>
</tr>
<tr>
<td>Summary</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive Verbal</td>
<td>0.77</td>
<td>0.96</td>
<td>4.61</td>
</tr>
<tr>
<td>Positive NonVerbal</td>
<td>1.47</td>
<td>0.95</td>
<td>8.81</td>
</tr>
<tr>
<td>Composites</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive Total</td>
<td>2.24</td>
<td>1.70</td>
<td>13.42</td>
</tr>
<tr>
<td>Negative Total</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>All</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Positive and Negative</td>
<td>2.24</td>
<td>1.70</td>
<td>13.42</td>
</tr>
</tbody>
</table>

Significant mean intercept and slope. The mean intercept at 36 months of age was 19.9 verbal social behaviors in 6 min (SE = 1.7), t(56) = 11.6; p = .0001, ES = .840. The mean slope was 0.61 verbal social behaviors per month (SE = 0.09), t(56) = 6.64; p = .0001, ES = .664. As a group, the children increased by an average of .61 verbal social behaviors per month or .15 per week (see Figure 3). Also displayed in Figure 3, are the plus and minus 1 standard deviation linear growth bands around the mean trajectory. Overlaid on these landmarks are the actual trajectories of the five children identified at risk for a social developmental delay on the Vineland (−1.5 SD below the mean). These children consistently fell below the PVSB growth trajectory and at times below the −1.0 SD trajectory. Slopes for these children over time were typically flat or declining.

Reliability of the PVSB

Odd-even reliability. Odd versus even reliability was calculated by consolidating each child's composite PVSB data on odd versus even measurement occasions and correlating these point estimates across children. The odd-even Pearson correlation was r(51) = .85 and the mean point estimates did not differ to a statistically significant degree (13.1 vs. 13.9 PVSBs in 6 min, F[50] = 1.24, p = 0.271).

Alternate forms reliability. The breakdown in the use of toy forms in the 326 observations was 110 (33.7%), 108 (33.1%), and 108 (33.1%) for TT, WH, and KD toys, respectively. Use of the toys by cohorts was well balanced and not significantly different. Alternate forms reliability was calculated by consolidating each child's composite PVSB scores within a common toy form across observations. PVSB Pearson correlations were positive, ranging from strong to moderate in size at .71 (TT vs. WH), .70 (WH vs. KD), and .57 (TT vs. KD), respectively. PVSB mean frequency of occurrence estimates in 6 min were 15.5 versus 15.8 (TT vs. WH), 15.8 versus 10.2 (WH vs. KD), and 15.5 versus 10.2 (TT vs. KD). A one-way ANOVA for repeated measures indicated a statistically significant difference in these estimates, F(1,44) = 13.44, p = .001, partial eta-squared = .233. Test of simple effects between means indicated that KD produced PVSB estimates that were on the order of 5 responses per session lower than that of either TT or WH. These findings suggested that KD was not an equivalent form. Separate HLM analyses of growth in PVSB (slope) indicated a higher slope es-
Table 2.
Extended

<table>
<thead>
<tr>
<th>Cohort 3</th>
<th>Rate</th>
<th>Total Responses</th>
<th>All</th>
<th>Rate</th>
<th>Total Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>2.03</td>
<td>1.48</td>
<td>12.16</td>
<td>8.90</td>
<td></td>
<td>1.37</td>
</tr>
<tr>
<td>2.29</td>
<td>1.18</td>
<td>13.76</td>
<td>7.11</td>
<td></td>
<td>2.09</td>
</tr>
<tr>
<td>4.32</td>
<td>2.15</td>
<td>25.92</td>
<td>12.90</td>
<td></td>
<td>3.46</td>
</tr>
<tr>
<td>0.17</td>
<td>0.27</td>
<td>1.02</td>
<td>1.60</td>
<td></td>
<td>0.11</td>
</tr>
<tr>
<td>0.44</td>
<td>0.39</td>
<td>2.63</td>
<td>2.36</td>
<td></td>
<td>0.34</td>
</tr>
<tr>
<td>0.61</td>
<td>0.58</td>
<td>3.66</td>
<td>3.47</td>
<td></td>
<td>0.45</td>
</tr>
<tr>
<td>1.14</td>
<td>0.95</td>
<td>6.81</td>
<td>5.72</td>
<td></td>
<td>0.91</td>
</tr>
<tr>
<td>0.22</td>
<td>0.36</td>
<td>1.34</td>
<td>2.15</td>
<td></td>
<td>0.21</td>
</tr>
<tr>
<td>1.36</td>
<td>1.07</td>
<td>8.15</td>
<td>6.42</td>
<td></td>
<td>1.12</td>
</tr>
<tr>
<td>3.33</td>
<td>2.01</td>
<td>19.99</td>
<td>12.05</td>
<td></td>
<td>2.39</td>
</tr>
<tr>
<td>2.95</td>
<td>1.40</td>
<td>17.73</td>
<td>8.41</td>
<td></td>
<td>2.65</td>
</tr>
<tr>
<td>6.29</td>
<td>2.63</td>
<td>37.72</td>
<td>15.80</td>
<td></td>
<td>5.04</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>6.29</td>
<td>2.63</td>
<td>37.72</td>
<td>15.80</td>
<td></td>
<td>5.04</td>
</tr>
</tbody>
</table>

timate of .38 responses per month when only using TT and WH observation sessions, compared to .24 responses per month when KD was included. KD proved not to be an equivalent form compared to the amount of PVSB produced by TT and WH and not including it appeared to increase the sensitivity of this social GOM.

Criterion Validity of the PVSB
Vineland and Howes convergence within methods. We first examined convergence within each criterion measure. Children’s interpersonal relations and play/leisure scores on the Vineland, were strongly intercorrelated (r = .91), however, they correlated poorly with the total score that included coping skills (see Table 3). On the Howes, all individual scales were moderately intercorrelated at best (range, .36 to .64) with the exception of the weak relationship between parallel and complementary/reciprocal (r = .18). All Howes scales correlated moderately well with its total composite scale (range, .44 to .87).

Vineland and Howes between method convergence. The convergence between the two criterion measures was moderately large (see Table 3). The best cross-method relationships were seen between the individual scales. For example, interpersonal relations on the Vineland correlated in rank order, .49, .33, .33, .08 to the Howes parallel/mutual, parallel, simple social, and complementary/reciprocal scales. Similar correlations between the Vineland’s play/leisure and the Howes were .44, .38, .27, and .18 for parallel/mutual, simple social, parallel, and complementary/reciprocal. However, the Vineland and Howes composites were not strongly related (r = .11).

Vineland/Howes convergence with the PVSB GOM. Of particular importance was convergence of the PSVB with the Vineland and Howes, and the extent to which PSVB measured social competence relative to these criteria. These cross-measure correlations were positive, ranging from poor to moderate in size. For example, correlations between the Vineland’s interpersonal and play/leisure scales and the PSVB mean intercept value were moderately large at .65 and .62, respectively. Equivalent correlations between these Vineland scales and the PNVSBl intercept value were smaller at .39 and .29, respectively.

On the Howes, only the simple social play scale and overall composite were positive, moderately large correlates of the PVSB mean intercept value (r = .47 and r = .34, respec-
Figure 1.
Growth in positive verbal and nonverbal social behaviors (nonverbal + verbal) to adult, peer, and nondirected over monthly measurement occasions by age cohort.

tively). Similar correlations were observed between the Howes parallel/mutual scale and the overall composite and the PNVSB scale (r = .44 and r = .33, respectively). Social GOM (PVSB and PNVSB) slopes related poorly to all criterion measures.

DISCUSSION
The purpose of this research was to investigate the feasibility and utility of an experimental Early Social Indicator (ESI) designed as a formative indicator of growth and prog-
Figure 2.
Growth in positive verbal and nonverbal social behaviors over monthly measurement occasions by age cohort.

ress for use by early interventionists with children birth to 3 years of age. The intention was not to develop a new social competence conceptual framework or to validate a comprehensive taxonomy of social skills. After describing change in the 7 key skill elements and their 7 composites over 9 months, the effort focused on crafting a social GOM that met Deno’s seven criteria for GOMs (Deno et al., 1982). Results of this study demonstrated
evidence that the PVSB composite met six of these original criteria: (a) key skill elements measured were linked to a socially valid general outcome, (b) authentic social behaviors were assessed, (c) standardized procedures were developed and used repeatedly over time, (d) the technical adequacy of the ESI was confirmed, (e) the sensitivity of the ESI to growth over time was documented, and (f) the efficiency and economy of the measure appeared promising. Future research is needed to demonstrate formative use linked to interventions and to confirm treatment validity.

Initial examination of the occurrence of key social skill elements and composites indicated that negative social behaviors occurred rarely and that considerable variation existed across children in the form of different positive social behaviors and to whom they are directed. The infrequency of negative behaviors was good news generally for children this age and probably reflects developmental levels prior to the emergence of many negative behaviors. It also might be the case that the situations used for this ESI were not those most likely to evoke negative behaviors in this population.

The analogue situation in this research was designed to approximate natural, authentic situations for young children but yet retain integrity as a standardized measure on which other children’s performance could be assessed and compared. Thus, the ESI was administered in a similar but not entirely natural situation. This might be considered a limitation because the loss of naturally occurring stimuli may influence performance in unknown ways. Performance in a common situation or on a standard task of social importance (e.g., manipulating sounds in words, producing sounds or names of letters, orally reading of text), however, is a good examination of the extent to which a child is able to perform a desired skill and a hallmark of the GOM approach (Deno, 1997, 2002). Demonstration that the approach has potential utility for measuring growth in the social skills of infants and toddlers is a unique contribution of this work to the literature.
Initial analysis of total positive social behavior indicated sizeable differences between individual children and by age cohorts; however, trends over time were flat and stable. Differences also were noted in the amount of total positive social behavior that children directed to the adult versus the peer, versus undirected (directed to either or both). The amounts of peer and undirected social behavior increased over time, while the total positive social behavior directed to the adult declined. Analyses separating PVSB from PNVSBS indicated that the former increased over time in the youngest two cohorts and not surprisingly, the latter, PNVSBS, decreased as children improved their spoken language skills.

These findings were generally consistent with expectations based on prior research findings regarding older children and their production of more frequent rates of verbal than nonverbal social behaviors and increasingly greater levels of social involvement with peers. Surprising, however, was finding a level trend in the total amount of interaction observed over 9 months. What the level trend masked, however, was (a) the increasing rates of positive verbal interaction and interaction with peers or undirected, and adults; and (b) the decreasing rates of nonverbal social interactions and interactions with adults.

Results indicated that PVSB scores were reasonably reliable in terms of odd-even and alternate form indices. PVSB scores also attained reasonable criterion validity in relationship to a standardized, parent report measure of social play abilities on the Vineland and to the Howes direct observation in the natural childcare setting. Further convergence between the PVSB and the criterion validity measures was shown for children identified at risk for social delays on the Vineland as indicated by a -2.0 SD discrepancy from expectation. Children with delays this large on the Vineland also showed lower than expected mean levels and rates of growth PVSB. These findings supported the construct validity of PVSB scores as reasonably sensitive, valid, and reliable measures of social competence.

Most important, however, was the PVSB's potential contribution to the formative measurement of social skills proficiency in infants and toddlers compared to existing measures (e.g., McConnell, 2000). A number of issues remain, however, for research and practice.

First, while the three toy forms produced acceptable high levels of intercorrelation, one of the toy forms, kitchen with dishes (KD), evoked significantly lower rates of social behavior compared to the other two forms. Use of this form in combination with the other two forms appeared to add to the "bounce" or variability in each child's growth trajectory as reflected in standard error terms. A desirable feature of GOMs are small ratios between the growth parameters (e.g., mean intercept and slope) and their standard errors of estimate (Fuchs & Fuchs, 1999). These findings suggest not using KD in future work and support additional research designed to replace it with a more equivalent toy form.

Second, this is not the only GOM research to report poor criterion validity correlations with slope (e.g., Greenwood, 2002; Kaminski & Good, 1996). A number of reasons may explain this outcome. One is that the majority of criterion measures of child performance lack a conceptually equivalent measure of growth over time. While mean intercept scores correlate quite well to criterion measure summary status scores, they do not correlate well with GOM slopes. Second, factors that moderate slope are not entirely obvious. For example, individual children's slopes are likely affected by mean intercept level. Children with either a low or a high mean intercept may have flatter slopes than children with intercepts near the mean because they may not be responsive to the intervention being provided.

Third, the logic of the GOM approach is to identify one or several valid key skill elements that can be repeatedly measured over time to reflect growth (Fuchs & Deno, 1991), compared to more traditional approaches of closing the gap between individual and mean level performance (normative, relativistic approach) or to a comprehensive skill attainment endpoint approach (mastery or mastery monitoring approach). Two widely recognized exam-
Table 3.
Criterion Validity Correlations Linking the Vineland, Howes, and Social GOM Positive Verbal and Nonverbal Composites

<table>
<thead>
<tr>
<th>Instruments</th>
<th>Indicators</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at Test</td>
<td>Months</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vineland</td>
<td>Interp Relations*</td>
<td>0.74**</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vineland</td>
<td>Play/Leisure*</td>
<td>0.70**</td>
<td>0.91**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Vineland</td>
<td>Composite*</td>
<td>0.20</td>
<td>0.19</td>
<td>0.19</td>
<td>1.00</td>
</tr>
<tr>
<td>Howes</td>
<td>Parallel*</td>
<td>0.24</td>
<td>0.33*</td>
<td>0.27*</td>
<td>0.08</td>
</tr>
<tr>
<td>Howes</td>
<td>Parallel/Mutual*</td>
<td>0.41**</td>
<td>0.49**</td>
<td>0.44**</td>
<td>0.05</td>
</tr>
<tr>
<td>Howes</td>
<td>Simple Social*</td>
<td>0.34*</td>
<td>0.33*</td>
<td>0.38**</td>
<td>0.18</td>
</tr>
<tr>
<td>Howes</td>
<td>Comp/Recip*</td>
<td>0.13</td>
<td>0.08</td>
<td>0.18</td>
<td>0.04</td>
</tr>
<tr>
<td>Howes</td>
<td>Total*</td>
<td>0.39*</td>
<td>0.46**</td>
<td>0.43**</td>
<td>0.11</td>
</tr>
<tr>
<td>ESI</td>
<td>Verbal Intercept</td>
<td>0.62**</td>
<td>0.65**</td>
<td>0.62**</td>
<td>0.19</td>
</tr>
<tr>
<td>ESI</td>
<td>Verbal Slope</td>
<td>−0.29</td>
<td>0.01</td>
<td>−0.07</td>
<td>0.13</td>
</tr>
<tr>
<td>ESI</td>
<td>Nvnerbal Intctp</td>
<td>0.43**</td>
<td>0.39**</td>
<td>0.29*</td>
<td>−0.13</td>
</tr>
<tr>
<td>ESI</td>
<td>Nvnerbal Slope</td>
<td>−0.36**</td>
<td>−0.19</td>
<td>−0.17</td>
<td>−0.01</td>
</tr>
</tbody>
</table>

Note. * = raw scores, * = p < .05, ** = p < .01, Intern = Interpersonal, Comp = Complimentary, Recip = Reciprocal, Nvnerbal = Non-Verbal, Intct = Intercept.

amples of the GOM approach are the pediatrician’s height and weight charts (Centers for Disease Control and Prevention, 2000) or the Dynamic Indicators of Basic Early Literacy Skills (DIBELS: Good & Kaminski, 1996). Both allow repeated measurement to monitor progress and are used to monitor the effects of interventions and to guide intervention decision-making. Neither, however, is considered to be a comprehensive diagnostic assessment. In fact, concerns over lack of expected growth are typically followed-up with more intensive assessments to define the problem or to prescribe treatment. In both cases, however, success of an applied treatment is monitored using growth charts to monitor acceleration in rate of growth over time resulting from the implementation of an effective intervention.

Fourth, in current work with PVSB, one must question its implications for accurately assessing the social function of children with atypical speech, hearing impairment, language or social developmental delays. Clearly, children with these disabilities will have a high probability of being identified as slow developing by PVSB. The operative questions are (a) Will these measures be sensitive to interventions for children with these problems? and, (b) Are the interventions used with them known to be effective? The use of this measure to evaluate formatively the effects of interventions provided to children with disabilities remains to be completed. Logically, however, the measure should be sensitive to children who are learning or using sign language (i.e., verbal social) to communicate with adults and peers. It also should be sensitive to children with speech and social challenges to detect increased verbal social activity even when verbalizations are not word utterances or where articulation/pronunciation of words is an issue, including those children who might demonstrate negative social behavior. Alternately, another GOM for children this age, the Early Communication Indicator (ECI) might be a better choice because of its measurement of these related key communication skill elements: gestures, utterances, single, and multiple words (Luze et al., 2001).

Fifth, in the present research, children’s data were recorded from videotape. For actual use in child care settings, such data will need to be collected in situ. This will require two persons, one a play partner for the child and the other an observer recording the child’s responding. Practitioners also will need effective and economical forms of training and the necessary administration materials for implementation, analysis, and interpretation. Additionally, a larger sample of children, moni-
Table 3. Extended

<table>
<thead>
<tr>
<th></th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.00</td>
<td>0.54**</td>
<td>0.64**</td>
<td>0.44**</td>
<td>1.00</td>
<td>0.34*</td>
<td>0.07</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>0.59*</td>
<td>0.36**</td>
<td>0.47**</td>
<td>0.24</td>
<td>0.08</td>
<td>0.01</td>
<td>0.21</td>
<td>-0.20</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>0.18</td>
<td>0.36**</td>
<td>0.69**</td>
<td>0.44**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.36**</td>
<td>0.87**</td>
<td>0.28</td>
<td>0.16</td>
<td>0.33*</td>
<td>0.11</td>
<td>-0.13</td>
<td>-0.05</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>0.19</td>
<td>0.24</td>
<td>-0.03</td>
<td>-0.05</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>0.44**</td>
<td>0.13</td>
<td>0.15</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

.tored frequently over time is needed to establish further the representativeness and the practical utility of these findings. In addition to the small sample overall, a limitation in the present study was the small number of children in the youngest cohort.

Results of the present study support the initial feasibility and utility of the ESI. Demonstrations that the ESI is sensitive to specific forms of early intervention designed to accelerate progress towards the general outcome are needed to complete the development of this GOM measure and thus, its potential importance to the field.

REFERENCES


We dedicate this work to our dear friend and colleague, Mary McEvoy. Mary was a great friend of children challenged with behavioral, emotional, and social disabilities and their parents and teachers. Mary was our collaborator on many projects over the years and she pushed us to develop measures and interventions that would make life better for children and families. As a partner on the Early Childhood Research Institute on Measuring Growth and Development, she urged us to focus on the assessment of children’s social and behavioral development, knowing how critical improvements in this area of measurement were, not only for determining when children needed intervention, but also in monitoring children’s progress. In this way, she hoped that we could advance the development of interventions for young children with social and emotional difficulties. Mary continues to inspire us in our work (and our play).

This research was supported by Grant H024560010, funded by the Early Education Programs for Children with Disabilities (EEPCD), Office of Special Education and Rehabilitation Services (OSERS), U.S. Department of Education. The opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement should be inferred. The authors would like to thank the staff of the childcare centers and the children and their families who participated in the project. Thanks are due to the research staff (Paige Bross, Jodi Gager, Kari Manier, Annessa Staab, and Kenneth Parsley) who helped plan and carry out the work. In addition, thanks are due to the following persons who provided invaluable assistance to this project: Deborah Linebarger, Jane Atwater, and Dale Walker.

Address correspondence concerning this article to Judith J. Carta, Juniper Gardens Children’s Project, 650 Minnesota Ave., Second Floor, Kansas City, KS 66101. E-mail: carta@ku.edu.

APPENDIX

Social Key Element Definitions

Social Behaviors

Social behaviors occur when a child attempts to convey a message to a partner requiring one or more behaviors. Social behaviors may last as long as a single behavior (an initiation to play), or as long as an episode/exchange of social behaviors involving several turns (an initiation that is responded to by a peer, that
is responded to by the target child). Social responding is considered ended after 3 s of no responding. For example, an episode begins when a child initiates or responds to an initiation presented by another. The episode ends after a pause of at least 3 s (count to yourself, “one thousand one . . .”). Thus, each single social behavior or each episode separated by a 3-s pause is counted as one event.

Positive versus Negative Social Behaviors
Social behaviors are recorded as either positive or negative. Social behaviors are positive when they are greetings, offers to play, requests, etc. Social behaviors are recorded as negative when they involve aggression, hitting, kicking, threatening, grabbing away another’s toy, or other negative behavior. Crying for children this age is considered to be an acceptable form of social-communication and is not recorded as either a positive or negative social behavior.

Nonverbal Social Behavior (with/to Adult, Peer, or Nondirected)
A. Nonverbal social behaviors are gesture-based attempts to communicate. Examples include smiling at, giving or showing object, rejecting an object by pushing it away, reaching toward, or touching a partner or object the partner is holding, pointing toward an object or person (may or may not be used to establish joint attention), nodding or shaking head to indicate “yes” or “no,” shrugging shoulders.

Non-social examples include play behaviors such as reaching for a toy lying on the ground, physical movements independent from social communication, physical movements showing excitement or pleasure that are not in direct communication with the partner (e.g., waving arms when watching a ball roll away or ignoring directions).

B. Nonverbal social behaviors are recorded according to the person they are directed to: the adult play partner; the peer play partner; or as undirected (in cases in which it is not clear exactly to whom they are directed, or if they are directed to both partners).

Verbal Social Behavior (with/to Adult, Peer, or Nondirected)
Verbal social behaviors are vocal (or sign language) attempts to communicate using non-words, single word, or multiple word utterances. False starts or stutters are counted as one verbalization. For example, “I think this is . . . this looks like a dog” counts as one verbalization. An episode is ended when there is a pause of at least 3 s without vocalizations (count to yourself, “one thousand one . . .”). Again, crying is considered acceptable, but is not recorded.

Like nonverbal, verbal social behaviors are recorded as directed either to the adult or peer, or as nondirected (when it is impossible to tell to whom the social behavior is directed).

Some additional examples of verbal social behaviors are animal sounds (e.g., “moo,” when looking at a cow); transportation/motor sounds (e.g., “vroom,” when pushing a tractor); blows to ask for more bubbles; sequentially naming objects, such as “block, red, phone, girl” (tally for each word); a vocalization in which only one word is understandable; imitation (could be sounds, non-sense words, or sensible words); or standard sign language coded as appropriate for single words.

Non-verbal non-social behavior examples include involuntary noises such as hiccups or coughing or sneezing. These are not social behaviors.

Carta et al. 113
### Key Skill Elements and Composites Summary

<table>
<thead>
<tr>
<th>Number</th>
<th>Social Target</th>
<th>Skill Element/Composite</th>
<th>Number</th>
<th>Social Target</th>
<th>Skill Element/Composite</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Adult</td>
<td>*Positive Verbal</td>
<td>8</td>
<td>*Non Directed</td>
<td>Positive Non-Verbal</td>
</tr>
<tr>
<td>2</td>
<td>Adult</td>
<td>*Positive NonVerbal</td>
<td>9</td>
<td>Non Directed</td>
<td>Positive Composite</td>
</tr>
<tr>
<td>3</td>
<td>Adult</td>
<td>Positive Composite</td>
<td>10</td>
<td>Composite</td>
<td>Positive Verbal</td>
</tr>
<tr>
<td>4</td>
<td>Peer</td>
<td>*Positive Verbal</td>
<td>11</td>
<td>Composite</td>
<td>Positive Non-Verbal</td>
</tr>
<tr>
<td>5</td>
<td>Peer</td>
<td>*Positive NonVerbal</td>
<td>12</td>
<td>Positive</td>
<td>Total</td>
</tr>
<tr>
<td>6</td>
<td>Peer</td>
<td>Positive Composite</td>
<td>13</td>
<td>*Negative</td>
<td>Total</td>
</tr>
<tr>
<td>7</td>
<td>NonDirected</td>
<td>*Positive Verbal</td>
<td>14</td>
<td>All</td>
<td>Positive + Negative</td>
</tr>
</tbody>
</table>

* = Key skill element.