

Reliability and validity of the Computerized Comprehension Task (CCT): data from American English and Mexican Spanish infants*

MARGARET FRIEND AND MELANIE KEPLINGER

San Diego State University

(Received 29 January 2006. Revised 7 March 2007)

ABSTRACT

Early language comprehension may be one of the most important predictors of developmental risk. The need for performance-based assessment is predicated on limitations identified in the exclusive use of parent report and on the need for a performance measure with which to assess the convergent validity of parent report of comprehension. Child performance data require the development of procedures to facilitate infant attention and compliance. Forty infants (20 at 1;4 and 20 at 1;8) acquiring English completed a standard picture book task and the same task was administered on a touch-sensitive screen. The computerized task significantly improved task attention, compliance and performance. Reliability was high, indicating that infants were not responding randomly. Convergent validity with parent report and 4-month stability was substantial. Preliminary data extending this approach to Mexican-Spanish are presented. Results are discussed in terms of the promise of this technique for clinical and research settings and the potential influences of cultural factors on performance.

The present paper evaluates the Computerized Comprehension Task (CCT; Friend & Keplinger, 2003) for the direct assessment of infant comprehension. One motivation for the development of this procedure is to facilitate clinical assessment of risk for language delay. Germane to this

[*] We gratefully acknowledge Christina Cisneros, Michelle Foy, Stephanie Velez, Rodrigo Enciso and Marco Ibanez for assistance in data collection, Adrienne Simpson for assistance in data compilation, Pascal Zesiger at the Université de Genève for providing facilities for manuscript revisions, and our reviewers. This research was presented in M. Friend (Chair), 'Picture recognition approaches to comprehension: Neuroscience, cross-linguistic and atypical development perspectives', the X. International Association for the Study of Child Language, Berlin, Germany, July 2005. The research was partially supported by a Blasker grant to the authors from the San Diego Foundation. Address for correspondence: Margaret Friend, PhD, San Diego State University, 6363 Alvarado Ct, Ste. 103, San Diego, California 92103, United States.

purpose are the facilitation of infant compliance, the establishment of convergent validity with parent reports and the extension of the approach across languages.

The measurement of receptive vocabulary is crucial to elucidating the relation between comprehension and production and to identifying children at risk for language delay. Hirsh-Pasek & Golinkoff (1996: 105) draw an analogy between astronomers' fascination with the 'dark' side of the moon and language researchers' interest in studying comprehension, the less visible side of language acquisition. Comprehension provides the earliest window onto children's understanding of word-referent relationships. To discover what a child knows about language, we must study comprehension (Bates, 1993).

The challenges inherent in measuring infants' understanding of words they do not yet say have impeded the study of early comprehension. Infant attention is difficult to maintain and non-compliance has been regarded as a fundamental flaw in infant assessment (Kaler & Kopp, 1990). In contrast, the relative ease of obtaining estimates of child language in a simple, checklist format has made parent report a tempting approach (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994; Rescorla & Alley, 2001). Nonetheless, this approach brings its own set of limitations and concerns have been raised over the exclusive use of parent report in diagnostic contexts (Fenson *et al.*, 1994; Stiles, 1994; Tomasello & Mervis, 1994; Yoder, Warren & Biggar, 1997; Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky & Paradise, 2000; Fenson, Bates, Dale, Goodman, Reznick & Thal, 2000).

In particular, it has been argued that parent report of comprehension is neither sufficiently consistent over time (Yoder, Warren & Biggar, 1997) nor sufficiently predictive of developmental outcomes (Feldman *et al.*, 2000; cf. Heilman, Weismer, Evans & Hollar, 2005, regarding production). Two issues give rise to these limitations: First, it is challenging for parents to discern the specific words that infants know but do not yet say. Over- and under-extensions in early comprehension (McDonough, 2002; Mervis & Canada, 1983; Mervis, 1987; Meints, Plunkett & Harris, 1999) may contribute to parent uncertainty regarding the specific words that infants truly comprehend. Whereas parent report has utility at the summary and group level, it is not consistent at the item level for individual children (Yoder *et al.*, 1997) and may over-predict developmental risk (Klee, Pearce & Carson, 2000). Fenson *et al.* (1993) encourage the use of supplemental measures in diagnostic settings. Second, comprehension estimates from parent report are highly variable across infants. In the absence of converging measures, it is difficult to tease apart variability due to measurement error from variability in comprehension. True variability in comprehension (Fenson *et al.*, 1994, 2000) should be replicable in child performance.

A direct infant assessment would provide a convergent measure of parent report as well as a supplemental behavioral metric in laboratory and clinical settings. Developing a measure for direct infant assessment that is both easy to administer and effective is the focus of this report.

Revisiting the earlier concern that direct assessment is influenced by behavioral compliance, Thal & Friend (2005) presented evidence that compliance itself may prove diagnostic. Specifically, they argue that a fundamental feature of development in the second year of life is obligatory attention to linguistic stimuli. In this sense, failure to be engaged by language does not merely reflect behavioral non-compliance but may indicate that the process of language acquisition and the conceptual development that it represents is at risk. Essential to direct assessments must be the incorporation of design features to minimize distraction and inattention so that behavioral non-compliance can be teased apart from a failure to be captured by language stimuli.

The last decade has witnessed exciting changes in our approach to direct assessment. Pioneering work (Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987; Hirsh-Pasek & Golinkoff, 1996) led to the application of the preferential-looking paradigm to the study of early language. The approach maintains infant attention by presenting carefully timed trials consisting of attractive visual displays and minimizes response requirements by taking visual fixation as evidence of word comprehension. This approach has been a watershed in the study of early comprehension, allowing us to explore infants' early semantic categories (Meints *et al.*, 1999), the role of sentence frames in comprehension (Fernald, Pinto, Swingley, Weinberg & McRoberts, 2001) and the multiple cues that support early word learning (Hollich *et al.*, 2000). A cost of this powerful approach to studying early comprehension is the labor-intensive coding and analysis of the resulting data. Of interest in the present paper is the application of a similarly effective procedure for direct assessment in clinical settings in which resources for data coding and analysis may be limited.

Friend & Keplinger (2003) developed an assessment building on preferential-looking (Hirsh-Pasek & Golinkoff, 1996; Meints *et al.*, 1999; Hollich *et al.*, 2000; Fernald *et al.*, 2001) and picture book approaches (Ring & Fenson, 2000) to address the need for a direct measure of comprehension in the second year of life. As in previous approaches, we presented pairs of high-quality images in a forced-choice format. In the interest of developing a broad measure of comprehension, lexical targets consisting of nouns, verbs and adjectives that vary in frequency of occurrence in infants' receptive lexicons at 1;4 (Dale & Fenson, 1996) were selected from the MacArthur-Bates Child Development Inventories (CDI): Words and Gestures and the CDI: Words and Sentences (Fenson *et al.*, 1993). Together, nouns, verbs and adjectives comprise about 75% (52.3%, 18.8%

and 5.7%, respectively) of infants' receptive vocabularies at 1;4 as assessed on the CDI: WG (Fenson *et al.*, 1994).

The CCT takes into account infants' limited attention capabilities. Images appear on the screen at a standard pace, engaging infant attention. Infants point to or touch images on a 17" kiosk-enclosed screen in response to auditory prompts from an experimenter in which target vocabulary items are embedded (e.g. 'Where is the *shoe*?' 'Touch *shoe*.'). Touching the referent on the touch-sensitive screen produces a reinforcing sound to maintain interest and motivate compliance. A significant contribution of this procedure is that its engaging interface and ease of administration and scoring facilitates assessment before 1;8 (Friend & Keplinger, 2003).

In this paper, we present three studies of the CCT. In the first study, we present data on the relative efficacy of the CCT and the Comprehension Book (CB; Ring & Fenson, 2000) at 1;4 and 1;8, compare child performance with parent report estimates of comprehension and assess short-term test-retest stability. In the second study, we examine a subset of our youngest infants again four months later to evaluate the stability of performance over time. Finally, in the third study, we adapt the CCT to Mexican-Spanish and assess child performance, convergence with parent report and test-retest stability in a preliminary sample of Mexican-Spanish infants.

The adaptation of the CCT to languages other than English is important for several reasons. Parent report inventories have been extended to many languages. The CDI website (www.sci.sdsu.edu/cdi/adaptations_ol.htm) lists 38 adaptations of the CDI inventories. These measures indicate that vocabulary development is a universally significant marker of language acquisition (Dale & Goodman, 2005). It follows that supplementation of parent report with direct child assessment is an important direction for future research across languages.

In the present paper we have extended the CCT to Mexican-Spanish for three reasons. First, Hispanics are a growing demographic in the United States, contributing to a need to provide early assessment in Spanish. In addition, cultural differences in language play and etiquette with unfamiliar adults and objects (Jackson-Maldonado, Thal, Fenson, Marchman, Newton & Conboy, 2003; Marchman & Martinez-Sussman, 2002) may present a special challenge for direct assessment in this population.

Second, this is an ideal population for the study of theoretical issues related to bilingualism. For example, Rescorla & Achenbach (2002) suggest that Hispanic children, on average, may have lower English productive vocabulary than their non-Hispanic counterparts. However, this may be a function of developing vocabulary in two languages simultaneously (Patterson, 2004; Rescorla & Achenbach, 2002). Data on comprehension in English-Spanish bilinguals clarifies these findings.

Kohnert & Bates (2002) found five- to seven-year-olds acquiring Spanish as a first language and English as a second language are relatively balanced in their comprehension across the two languages, supporting Patterson's (2004) and Rescorla & Achenbach's (2002) interpretation. In addition, Umbel, Pearson, Fernandez & Oller (1992) found a statistically significant portion of non-overlapping vocabulary in bilingual five- to eight-year-olds with varying exposure to Spanish and English. Vocabulary across these languages appears to develop somewhat independently, contributing to a composite conceptual lexicon. Little is known about comprehension at younger ages and, in particular, the process of bilingual acquisition early in life.

Finally, the Mexican-Spanish adaptation of the CDI, the *Inventarios del Desarrollo de Habilidades Comunicativas: Palabras y Enunciados* (IDHC: PE; Jackson-Maldonado *et al.*, 2003), shows convergent validity with behavioral measures of PRODUCTION (Marchman & Martinez-Sussman, 2002; Thal, Jackson-Maldonado & Acosta, 2000). The present paper introduces a procedure to facilitate similar comparisons between parent report and child performance in COMPREHENSION. A Mexican-Spanish adaptation of the CCT could be extended to the assessment of early comprehension in monolingual and bilingual infants as early as 1;4 with both practical and theoretical significance.

STUDY 1

METHOD

Participants

Forty infants learning English (20 from 1;4 to 1;5 and 20 from 1;8 to 1;9, samples balanced for sex) recruited through advertisements in local parenting and entertainment magazines participated. A \$10 gift certificate to a local toy store was offered as an incentive.

Procedure

Data were collected in a mixed within/between design in two testing sessions scheduled one week apart. The first testing session was always scheduled within two weeks of the infant's 1;4 birthday for the younger infants and within two weeks of the 1;8 birthday for the older infants. In each session, one vocabulary assessment (CB or CCT) was administered. The order of tasks was counterbalanced across participants.

MacArthur-Bates CDI: Words and Gestures. The CDI: WG is a parent-report checklist of language comprehension developed by Fenson *et al.* (1993). This measure has good test-retest stability and significant convergent

TABLE 1. *Distribution of CCT lexical targets as a function of word class and difficulty level*

	Nouns	Verbs	Adjectives	Total
Easy	10	3	—	13
Moderate	7	5	2	14
Difficult	6	2	6	14
Total	23	10	8	41

validity with an object selection task (Fenson *et al.*, 1994). Of particular interest in the present study was the 396-item vocabulary checklist for comparison with the infants' behavioral data. Parents were mailed the CDI: WG to complete one week prior to their appointment in the laboratory.

Computerized Comprehension Task (CCT). The program presented 41 pairs of images representing nouns, verbs and adjectives. Two images appeared simultaneously at left- and right-center screen during each trial. The pairs were matched on color, size and brightness. Touches to the target image produced a unique, reinforcing, auditory signal but touches to the distractor did not.

There were equal numbers of easy (comprehended by more than 66% of infants at 1;4), moderately difficult (comprehended by 33–66% of infants at 1;4), and difficult word pairs (comprehended by less than 33% of infants at 1;4; Dale & Fenson, 1996). The word pairs were matched on difficulty and word class (nouns, verbs and adjectives). To some extent, difficulty and word class overlap but they do so imperfectly. Difficult words are more likely to include verbs and adjectives (but also unfamiliar nouns, e.g. *giraffe*) and easy words are more likely to include nouns (but also familiar verbs, e.g. *hugging*). See Table 1 for the distribution of lexical targets by word class and difficulty.

There were two forms of the procedure so that both members of each pair of images served as the target referent. In this way we could assess the equivalency of different word–referent pairings. Except for the member of each image pair serving as the target, the two forms were identical. The member of the pair identified as the target was counterbalanced across forms. Within forms, difficulty was matched within pairs and randomized across stimulus presentations. Targets appeared with equal frequency on the right and left sides of the screen. The side on which the target appeared was randomized across presentations with the restriction that targets appear no more than twice in succession on the same side to reduce orientation-bias effects (Hirsh-Pasek & Golinkoff, 1996).

The program began with four practice trials consisting of highly familiar words in English to familiarize infants with the task. The experimenter prompted the child as the images appeared on the screen for each trial:

Where is the ___? Touch ___, for nouns,
Who is ___? Touch ___, for verbs and
Which one is ___? Touch ___, for adjectives.

We collected test-retest reliability data on one-third of the items for children who remained attentive through the last test trial. In the reliability phase, the images appeared in opposite left-right orientation relative to the test trials. The reliability items mirrored the relative proportions of easy, moderate and difficult items and of nouns, verbs and adjectives in the full test.

Comprehension Book (CB). The CB was identical in content to the CCT and both assessments were administered under identical, optimal, testing conditions. During each assessment, infants were seated in the parent's lap. The parent wore dark glasses (the lenses of which were covered in black cardboard) and a pair of headphones over which music played. These precautions prevented parents from cueing their infants in either assessment. The only difference between the tasks was in the method of administration (picture book or touch-screen format).

Coding. Infants were coded correct if they pointed to or touched the target image on either the CCT or the CB and incorrect if they pointed to or touched the distractor. Trials on which infants did not respond but remained compliant and looking at the images were scored as missing. On the CCT only, infants sometimes gave responses that were not unequivocally correct or incorrect; for example, they sometimes touched both images simultaneously or in quick succession. In these cases, we coded the responses as ambiguous.

RESULTS

The results from Study 1 are organized around three central questions. First, we asked whether the CCT results in a significant increment in performance over the conventional picture book assessment. Second, to extend the findings of Friend & Keplinger (2003), we asked whether improvements in performance at 1;4 are maintained at 1;8. Third, we assessed both the test-retest reliability and convergent validity of the CCT.

We conducted three Age (2) \times Task (2) Repeated Measures MANOVAs. Preliminary analyses revealed no main effects of sex. This term was not included in further analyses. The number of attentive trials, during which the infant followed direction by looking at the images regardless of whether they pointed or touched, was the dependent measure in the

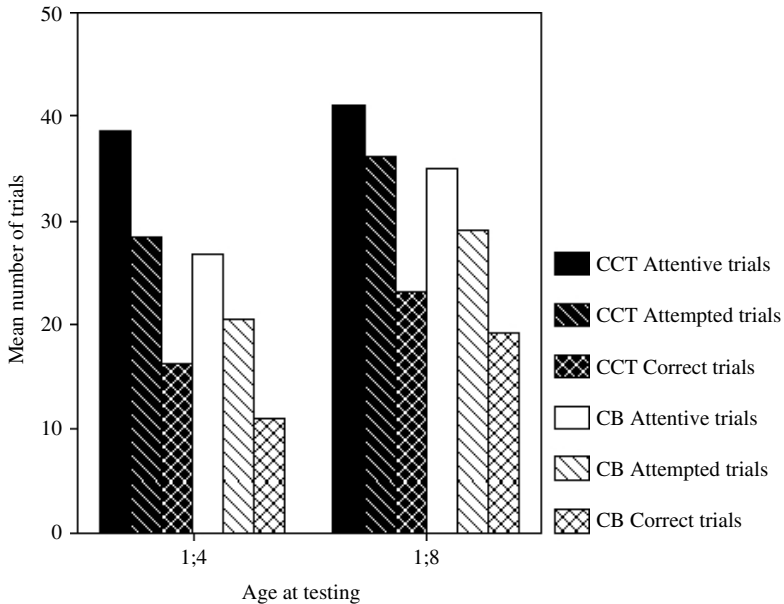


Fig. 1. Infant attention, responsiveness and accuracy in identifying referents as a function of task and age.

NOTE: Within each task, at each age, infants attended to significantly more trials than they attempted and attempted more trials than they completed correctly.

first analysis. There was a main effect of Task ($F(1, 38) = 16.69$, $p < 0.05$) indicating that infants attended to more CCT than CB trials (see Figure 1).

The second dependent measure was the number of trials during which infants actively pointed to or touched an image (inclusive of ambiguous responses) REGARDLESS OF WHETHER THE IMAGE WAS THE TARGET. This provided an estimate of how actively infants participated across tasks. There were main effects of Task and Age ($F(1, 38) = 14.62$ and 7.06 , respectively, $p < 0.05$). Infants pointed/touched on more trials for the CCT than for the standard assessment and older infants responded to more trials, across tasks, than did younger ones (see Figure 1). There was no interaction of Task and Age.

The third dependent measure assessed the number of trials on which infants CORRECTLY touched or pointed to the target. Again there were main effects of both Task and Age but no interaction. Infants touched/pointed to the target on significantly more trials on the CCT relative to the CB ($F(1, 38) = 10.91$, $p < 0.05$) and older infants were correct on more trials, across tasks, than were younger infants ($F(1, 38) = 9.08$, $p < 0.05$;

TABLE 2. Mean proportion of attempted trials correct for American-English infants as a function of word class

	Nouns	Verbs	Adjectives
Trials attempted			
<i>M</i>	0.85	0.68	0.77
<i>SD</i>	0.20	0.31	0.27
Attempted trials correct			
<i>M</i>	0.67	0.54	0.42
<i>SD</i>	0.20	0.29	0.27
Total trials correct			
<i>M</i>	0.58	0.37	0.32
<i>SD</i>	0.24	0.24	0.23

NOTE: The proportion of individual infants performing significantly above chance (65% correct) across attempted trials was 0.63 for nouns, 0.35 for verbs and 0.25 for adjectives. Individual proportions correct ranged from 0.00 to 1.00.

see Figure 1). Bonferroni-corrected comparisons revealed that, on both tasks, infants were attentive on more trials than they were responsive ($t(39)=5.33$ and 5.66 , respectively, $p<0.05$) and were responsive on more trials than they were correct ($t(39)=9.01$ and 11.44 , respectively, $p<0.05$).

Across three measures (attention, number of responses and number of correct responses) infants at 1;4 showed a significant improvement in performance on the CCT relative to the CB assessment. These effects were also significant at 1;8, suggesting that the CCT recruited attention and compliance in older infants. Still, it is possible that by virtue of being engaged in the assessment, infants are responding randomly rather than demonstrating true comprehension. To assess this possibility, we conducted additional analyses of group and individual performance.

First, we calculated the proportion of trials attempted, the proportion of attempted trials correct and the proportion of total trials correct. If the improvement in the number of correct responses from the CB to the CCT was due to increased chance responding, we would expect the proportion of correct responses to hover around 50%. To differ significantly from chance by binomial test an infant would have to respond correctly on at least 65% of the trials attempted. Following this reasoning, we calculated the number of infants who met this criterion. For the CCT, we examined the proportion of correct trials as a function of trials attempted (inclusive of ambiguous responses) as a function of word class. At the group level, the proportion of attempted trials correct is highest for nouns followed by verbs and adjectives. The individual data follows the same pattern. For nouns, performance was significantly above chance for approximately two-thirds of our infants, with this proportion decreasing for verbs and adjectives (see Table 2). This

TABLE 3. *Mean proportion of attempted trials correct for American-English infants as a function of word difficulty*

	Easy	Moderate	Difficult
Trials attempted			
<i>M</i>	0.86	0.73	0.79
<i>SD</i>	0.18	0.29	0.25
Attempted trials correct			
<i>M</i>	0.67	0.64	0.49
<i>SD</i>	0.26	0.23	0.21
Total trials correct			
<i>M</i>	0.58	0.47	0.40
<i>SD</i>	0.25	0.25	0.22

NOTE: The proportion of individual infants performing significantly above chance across attempted trials was 0.60 for easy words, 0.62 for moderately difficult words and 0.28 for difficult words. Individual proportions correct ranged from 0.00 to 1.00.

pattern was also observed when we considered performance as a function of a priori difficulty level: two-thirds of infants performed significantly above chance for easy words, with this proportion decreasing with word difficulty (see Table 3).

To further explore the extent to which infants' responses on the CCT reflect true vocabulary knowledge, we assessed test-retest reliability on the CCT by presenting one-third of the items a second time in opposite left-right orientation. A total of 24 infants (8 from 1;4 to 1;5 and 16 from 1;8 to 1;9) participated in the reliability assessment. The correlation between test and reliability phases was significant ($r=0.70$, $p<0.05$), indicating that infants were not responding at chance during test. Finally, we assessed the convergent validity of the CCT with parent report of vocabulary comprehension on the CDI: WG. Parent report and child performance on the CCT were significantly correlated ($r=0.64$, $p<0.05$).

Child performance on the CCT exceeded that on the CB in attention, responsiveness and vocabulary comprehension, suggesting that this measure facilitates early language assessment. Correct identification of referents on the CCT exceeds chance for nouns and across word classes for easy and moderately difficult words at the individual level. Moreover, there is significant test-retest stability and significant convergence with parent report. The fact that the improved performance from the CB to the CCT at 1;4 is maintained at 1;8 extends the findings of Friend & Keplinger (2003). A remaining concern is the stability of child performance over longer periods of time. This question is addressed in Study 2.

STUDY 2

METHOD

Participants

Fourteen of our infants from Study 1 repeated the CCT four months later (M age at second testing = 1;8, 7 females, 7 males).

Procedure

Data were collected in a single testing session approximately four months after infants participated in Study 1. Eleven infants completed both the test and reliability phases. The administration of the CDI: WG and the CCT was identical to Study 1.

RESULTS

In Study 2, we sought to determine the stability of CCT estimates of comprehension over a four-month interval. In addition, we attempted to replicate our previous findings of short-term test-retest stability.

First, we considered whether infants improved on measures of attention, responsiveness and comprehension from 1;4 to 1;8. Within samples t -tests revealed that, at 1;8, infants knew significantly more words than they had at 1;4 (M at 1;4 = 16.5, SD = 8.23, M at 1;8 = 23.64, SD = 8.76, $t(13) = 3.25$, $p < 0.05$). However, in contrast to Study 1, there were no changes with age in attention or responsiveness. This suggests that the difference in responsiveness observed between 1;4 and 1;8 in Study 1 may have been an artifact of using a cross-sectional design rather than reflecting true age-related change.

To assess the four-month stability of vocabulary comprehension estimates on the CCT, we calculated the correlation between performance at 1;4 and 1;8. The correlation was significant ($r = 0.56$, $p < 0.05$); however, one outlier was noted. This infant correctly identified 11 target referents at 1;4 but identified only 3 targets at 1;8. Removal of this outlier resulted in a modest improvement in the correlation between performance at 1;4 and 1;8 ($r = 0.61$, $p < 0.05$). Short-term test-retest stability in Study 2 mirrored our findings from Study 1 ($r = 0.76$, $p < 0.05$).

As in Study 1, we calculated the number of correct trials as a proportion of trials attempted. The word class analysis revealed a higher proportion of trials correct across classes relative to Study 1. Also, the proportion of infants performing above chance was substantial for nouns and verbs but still low for adjectives (see Table 4). The pattern is almost identical for performance as a function of a priori difficulty level. Infants were most accurate for easy and moderately difficult, relative to difficult, words (see Table 5).

TABLE 4. *Mean proportion of attempted trials correct for American-English infants as a function of word class in Study 2*

	Nouns	Verbs	Adjectives
Trials attempted			
<i>M</i>	0.85	0.68	0.85
<i>SD</i>	0.22	0.32	0.21
Attempted trials correct			
<i>M</i>	0.71	0.75	0.51
<i>SD</i>	0.20	0.27	0.26
Total trials correct			
<i>M</i>	0.64	0.53	0.44
<i>SD</i>	0.25	0.28	0.25

NOTE: The proportion of individual infants performing significantly above chance across attempted trials was 0.71 for nouns, 0.78 for verbs and 0.35 for adjectives. Individual proportions correct ranged from 0.00 to 1.00.

TABLE 5. *Mean proportion of attempted trials correct for American-English infants as a function of word difficulty in Study 2*

	Easy	Moderate	Difficult
Trials attempted			
<i>M</i>	0.82	0.79	0.81
<i>SD</i>	0.21	0.27	0.20
Attempted trials correct			
<i>M</i>	0.77	0.67	0.61
<i>SD</i>	0.22	0.27	0.20
Total trials correct			
<i>M</i>	0.66	0.58	0.49
<i>SD</i>	0.27	0.28	0.20

NOTE: The proportion of individual infants performing significantly above chance across attempted trials was 0.78 for easy words, 0.65 for moderately difficult words and 0.42 for difficult words. Individual proportions correct ranged from 0.00 to 1.00.

Finally, we attempted to replicate our previous finding of convergent validity between parent report and child performance. This relation was not significant with our smaller and older sample in Study 2.

In sum, we replicated our finding that child performance on the CCT is stable across a brief test-retest interval. Further, this stability is maintained at intervals as long as four months. This is promising with regard to our ability to predict developmental outcomes. Although the convergence of child performance and parent report did not replicate, this may be due to a smaller sample size and reduced variability at 1;8 as these children scored higher, relative to younger children in Study 1, on both instruments.

The CCT was designed to overcome the attention and compliance issues that arise with conventional picture book assessments. This approach was effective in a sample of infants acquiring English as a first language. Of primary interest is whether the CCT is a valid measure of vocabulary comprehension in infants acquiring Mexican-Spanish. We have three questions. Do infant attention and responsiveness mirror that observed in our English sample? Are infants' responses non-random and consistent over time? Finally, does performance on the CCT correlate with parent report of comprehension on the IDHC? Study 3 reports preliminary data on the CCT in infants acquiring Mexican-Spanish.

STUDY 3

METHOD

Participants

Seventeen infants acquiring Mexican-Spanish as their primary language between 1;4 and 1;6 ($M=1;6$, 11 males, 6 females) and their parents were recruited through government-sponsored daycare centers and schools in Tijuana, Baja California, Mexico and free weekly newspapers in San Diego County, California, United States. Eight infants participated in Mexico and 9 infants in the US. Parents of all of the infants reported that Spanish was the language spoken in the home. For the infants in Tijuana, parents reported Spanish exposure to be 94% of waking hours on average ($SD=7.5$), whereas in San Diego parents reported Spanish exposure to be 74% ($SD=20.6$). The difference in exposure to Spanish was significant by an independent samples *t*-test for unequal variances ($t(10.3)=2.70$, $p<0.05$). A \$6 gift card to a local store was provided to all participants as an incentive.

Procedure

Vocabulary comprehension was assessed on the Inventarios del Desarrollo de Habilidades Comunicativas: Primeras Palabras y Gestos (IDHC: PG; Jackson-Maldonado *et al.*, 2003) and on the Mexican-Spanish adaptation of Friend & Keplinger's (2003) Computerized Comprehension Task (CCT).

The Mexican-Spanish adaptation of the CCT is comprised of attractive images corresponding to target vocabulary items derived from the IDHC. Target items are 41 pairs of words matched on word class (nouns, verbs and adjectives) and frequency of occurrence in the comprehension vocabularies of monolingual, Mexican-Spanish infants at 1;4 (V. Marchman, personal communication, 2003). As a consequence of cross-cultural variability in the referents that infants encode earliest, there is only partial overlap between the lexical items assessed in the American-English and Mexican-Spanish

adaptations. There is a roughly equal distribution of easy, moderately difficult and difficult word pairs.

Parents completed the IDHC one week prior to testing following the instructions of a trained experimenter. Following completion of the IDHC, parents in Mexico brought their infant to a medical office in Tijuana for testing and parents in the United States brought their infant to a university laboratory in San Diego. Infants were seated on their parents' lap in a quiet, darkened room. Parents wore dark glasses to prevent cueing the infants. Infants who remained attentive at the end of the test trials completed a reliability assessment during which one-third of the items were presented a second time in opposite left-right orientation. The entire procedure was administered in Spanish by a researcher bilingual in Spanish and English.

Pilot-testing revealed that infants acquiring Spanish require additional warming-up, relative to infants acquiring English, in order to comply with the task. Specifically, their parents report that they are reluctant to touch objects that don't belong to them. To compensate for this reluctance, we began each experimental session by demonstrating to the infants that they could finger-paint on the touch-sensitive screen using Microsoft Paint. Once infants engaged in the finger-painting exercise, we initiated the CCT program. The program began with four practice trials consisting of highly familiar words in Mexican-Spanish to familiarize infants with the task.

RESULTS

Preliminary data are presented here for comparison with our larger English sample. First, we were interested in the extent to which the CCT was successful in maintaining infant attention and compliance among Mexican-Spanish infants. Second, we wanted to know the extent to which performance differed significantly from chance, and finally we wanted to assess both the test-retest reliability and convergent validity of the Mexican-Spanish adaptation of the CCT.

We conducted a Sample (2) \times Sex (2) Repeated Measures MANOVA on the number of trials to which infants attended, the number of trials on which they responded and the number of trials on which they were correct. There was no effect of Sample (US or Mexico) or Sex. The absence of an effect of Sample suggests that differences in Spanish exposure did not interfere with Spanish comprehension vocabulary. There was an effect of dependent measure ($F(2, 12) = 129.63, p < 0.05$), indicating that, consistent with our previous studies, infants attended more than they responded ($t(16) = 5.77, p < 0.05$) and responded more than they were correct ($t(16) = 9.45, p < 0.05$; see Figure 2 for means and standard deviations). However, Spanish infants attempted fewer trials and were correct less often

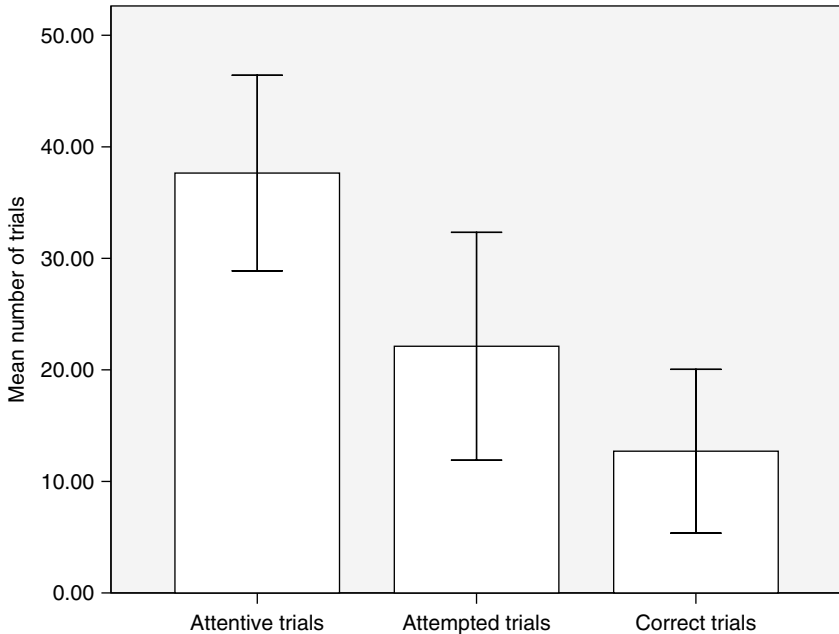


Fig. 2. Mean attentive, attempted and correct trials for Mexican-Spanish infants.
NOTE: Differences between all dependent measures are significant at $p < 0.05$. Error bars represent ± 1 *SD*.

than their English counterparts ($t(55) = 3.72$ and 2.98 , respectively, $p < 0.05$; see Figure 3).

To assess whether performance differed from chance, we repeated the analyses of proportions of correct trials as a function of attempted trials presented in Studies 1 and 2. The pattern of performance across word classes was similar to Study 1. There was a marked attenuation in the proportion of trials attempted relative to infants acquiring English. However, the proportion of attempted trials correct and the proportion of infants performing above chance were similar across languages. Performance was significantly greater than chance for approximately two-thirds of our infants for nouns and this number decreased for verbs and adjectives relative to nouns (see Table 6). However, when we considered performance as a function of a priori difficulty level, the proportion of attempted trials correct and the proportion of infants performing above chance was similar to the English data for easy and difficult words but not for words of moderate difficulty. Two-thirds of infants performed better than chance for easy words and this proportion decreased for moderately difficult and difficult words (see Table 7).

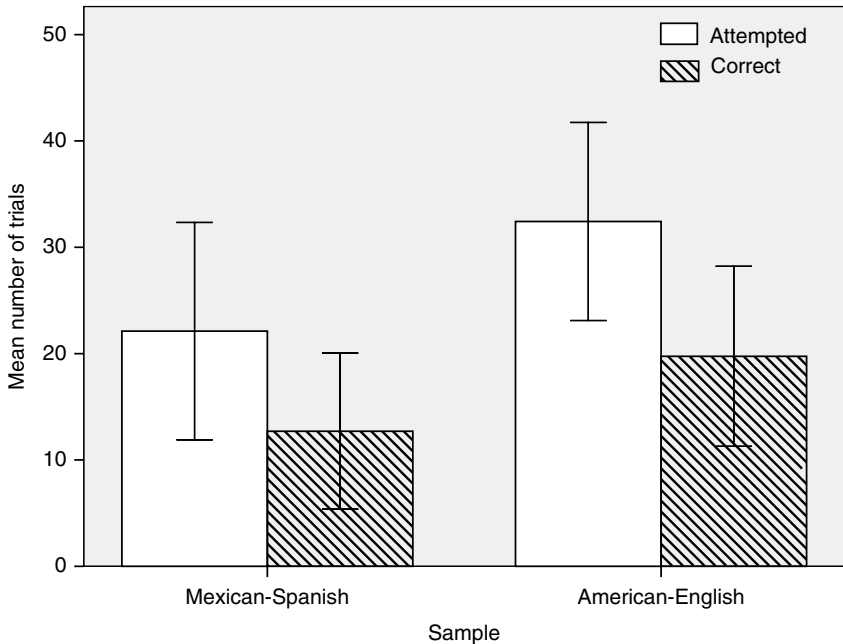


Fig. 3. Comparison of American-English and Mexican-Spanish infant performance on the CCT.

NOTE: Differences between groups are significant at $p < 0.05$. Error bars represent ± 1 SD.

To explore the stability of infants' responses, we assessed test-retest reliability on one-third of the items presented a second time in opposite left-right orientation. A total of six infants participated in the reliability assessment. This sample is too small to yield a reliable correlation coefficient. However, the relation appears strong and positive, indicating that these infants were not responding at chance during the test (see Figure 4). Finally, we considered the relation between child performance on the CCT and parent report of vocabulary comprehension on the IDHC and found that it was not significant.

The Mexican-Spanish adaptation of the CCT was effective in eliciting infant attention but infants in the present sample touched the screen less frequently than the infants in our English sample and, consequently, correctly identified referents less often. Performance at the individual level, when examined as a function of word class and difficulty, differed from chance for approximately two-thirds of the sample for nouns and easy words. This is consistent with what one would expect from the literature and suggests that infant performance reflected true comprehension. Further, the relation between test and retest performance was encouraging.

EARLY VOCABULARY

TABLE 6. *Mean proportion of attempted trials correct for Mexican-Spanish infants as a function of word class*

	Nouns	Verbs	Adjectives
Trials attempted			
<i>M</i>	0.57	0.39	0.51
<i>SD</i>	0.26	0.29	0.30
Attempted trials correct			
<i>M</i>	0.67	0.44	0.44
<i>SD</i>	0.18	0.30	0.30
Total trials correct			
<i>M</i>	0.40	0.16	0.28
<i>SD</i>	0.23	0.15	0.26

NOTE: The proportion of individual infants performing significantly above chance across attempted trials was 0.59 for nouns, 0.24 for verbs and 0.47 for adjectives. Individual proportions correct ranged from 0.00 to 1.00.

TABLE 7. *Mean proportion of attempted trials correct for Mexican-Spanish infants as a function of word difficulty*

	Easy	Moderate	Difficult
Trials attempted			
<i>M</i>	0.56	0.52	0.54
<i>SD</i>	0.30	0.23	0.26
Attempted trials correct			
<i>M</i>	0.69	0.46	0.56
<i>SD</i>	0.24	0.20	0.22
Total trials correct			
<i>M</i>	0.37	0.26	0.30
<i>SD</i>	0.23	0.18	0.20

NOTE: The proportion of individual infants performing significantly above chance across attempted trials was 0.70 for easy words, 0.29 for moderately difficult words and 0.29 for difficult words. Individual proportions correct ranged from 0.00 to 1.00.

In contrast, the finding that child performance did not converge with parent report on the IDHC was surprising.

DISCUSSION

The present research validated a child-performance measure of early vocabulary comprehension with a sample of 40 English-acquiring infants at 1;4 and 1;8. Performance was significantly better on the assessment that employed a touch-sensitive screen, designed to capture and maintain infant attention in the second year of life, relative to the conventional picture book approach characteristic of many extant comprehension measures. Infants

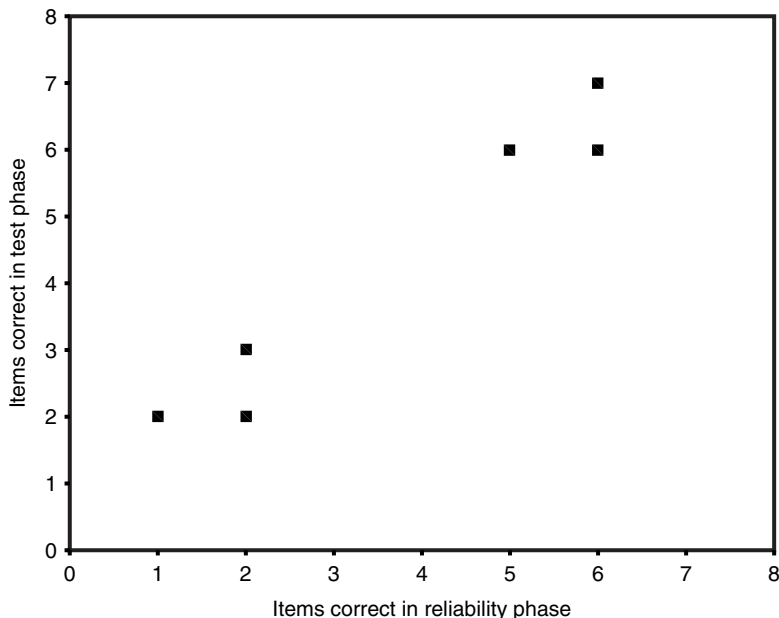


Fig. 4. Test-retest reliability for the Mexican-Spanish CCT.

were more attentive and responsive, and identified more referents on the CCT even though both procedures assessed infant knowledge of the same set of lexical items.

Importantly, increased compliance does not come at the cost of reduced accuracy. The majority of the infants across age and language were performing non-randomly. Infants' best performance was obtained for easy or early appearing words in the lexicon, as would be expected from the literature. For later-appearing words, performance was mixed, suggesting that infants may guess on more difficult trials. In Study 2, the proportion of attempted trials correct across word classes and difficulty levels increased, indicating greater accuracy with age. Whereas the most stable estimates on the CCT may come from early appearing words, comprehension estimates of later appearing words may also have utility. First, infants attempt fewer trials for later appearing relative to early appearing words. Second, proportion correct as a function of difficulty level corresponds reasonably well to parent report data (Dale & Fenson, 1996). Finally, infant performance is stable across immediate and delayed test-retest.

Several measures support the reliability and validity of the CCT with infants acquiring English as a first language: test-retest reliability and four-month stability were substantial and, in Study 1, convergent validity with

parent report was strong. The absence of a relation between parent report and child performance in Study 2 may be attributable to a smaller sample size and to reduced variability in performance at 20 months. In contrast to Study 1, in which 40 infants participated, only 14 infants participated in Study 2 and, as these were older infants, their performance tended toward the high end of both instruments. It is likely that these factors contributed to a reduced correlation coefficient.

Test-retest reliability was strong for infants acquiring Spanish as a first language as well. However, in contrast to the English infants of comparable age in Study 1, there was no convergence of child performance with parent report. As in Study 2, the smaller sample size relative to Study 1 may be a contributing factor. These data are preliminary and it is premature to draw conclusions regarding the relation between parent report and the CCT in Mexican-Spanish. Nevertheless, some parents did appear to overestimate infant vocabulary knowledge and these parents varied with respect to the amount of English and Spanish spoken in the home. Because our demographic data are incomplete, we are unable to characterize the families who evinced this pattern. Thus, it will be important to conduct additional studies of the CCT in Mexican-Spanish with larger samples and with complete demographic information.

Infants acquiring English as a first language performed somewhat better than infants acquiring Spanish. The English sample responded more frequently and correctly identified more referents than the Spanish sample, even though both groups of infants were interested in the task and evinced comparable proportions of attempted responses correct. Infants in our Spanish sample were more reluctant to touch the screen and parents reported that they prohibit them from touching things that do not belong to them. However, this cannot account completely for the observed differences between our samples as infants could also have indicated a referent by pointing. Another possibility is that the cultural experiences of these infants with regard to language games and etiquette with unfamiliar adults may limit their performance on behavioral tasks such as the CCT (Jackson-Maldonado *et al.*, 2003; Marchman & Martinez-Sussman, 2002). Alternatively, they may have less experience interacting with computers. Preliminary data on infants acquiring French in Switzerland show a similar pattern: attenuated responsiveness on an adaptation of the CCT relative to infants acquiring English in the US. Yet the proportions of correct responses are similar to those observed in both English and Spanish (P. Zesiger, personal communication, February 2007).

The fact that performance is consistent across test and retest and differs significantly from chance suggests that infants are providing reliable responses to the task. However, care must be taken to warm infants to the context of the task to obtain optimal performance. Further, even with this

support, typical performance on the CCT may vary across languages. This highlights the need for cross-language assessment with larger samples to construct performance norms for early comprehension.

The significance of the CCT lies in its ability to yield data on early comprehension, which is both elusive and fundamental to language and cognitive development. Whereas significant progress has been made in identifying late talkers on the basis of production measures (Rescorla, Mirak & Singh, 2000; Rescorla & Alley, 2001; Heilman *et al.*, 2005), determining which late talkers are at greatest developmental risk remains problematic. Many late talkers will develop typical language skills. Those late talkers who develop atypically are most likely to also show deficits in comprehension (Thal, 2005). Comprehension assessment is likely to provide an earlier and more definitive prediction of risk for persistent developmental delay.

Thal & Friend (2005) have shown that behavioral assessment of comprehension can mediate interpretations of parent report. Specifically, infants whose parent report scores fall below the 7th percentile but who perform well on a laboratory task are less likely to be at developmental risk than infants who score low on both measures. The problem is to disentangle behavioral non-compliance from a failure to be captured by language stimuli. The CCT significantly improves compliance over conventional picture identification approaches. These data suggest that we can begin to take child performance as an indication of early vocabulary knowledge rather than as a measure of behavioral compliance. This approach may provide valuable information that could supplement parent report in research and clinical settings. Further, preliminary data suggest that the CCT can be productively adapted to languages other than English if cultural influences on child compliance and parent report are taken into account.

Future research will need to establish the long-term predictive stability of this measure as well as its efficacy in predicting both typical and atypical developmental outcomes. In addition, collecting data on sufficiently large samples across languages will be key in establishing appropriate criteria to identify infants who may be at risk for atypical development.

REFERENCES

- Bates, E. (1993). Comprehension and production in early language development. *Monographs of the Society for Research in Child Development* **58** (3–4, Serial No. 233).
- Dale, P. S. & Fenson, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, & Computers* **28**, 125–27.
- Dale, P. S. & Goodman, J. C. (2005). Commonality and individual differences in vocabulary growth. In M. Tomasello and D. I. Slobin (eds), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, 41–78. Mahwah, NJ: Lawrence Erlbaum Associates.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E. & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development* **71**, 310–22.

- Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, S. J. & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development* **71**, 323–28.
- Fenson, L., Dale, P. S., Reznick, 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* **59** (5, Serial No. 242).
- Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethik, S. & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual*. San Diego: Singular.
- Fernald, A., Pinto, J. P., Swingle, D. L., Weinberg, A. and McRoberts, G. W. (2001). Rapid gains in speed of verbal processing by infants in the 2nd year. In M. Tomasello and E. Bates (eds), *Language development: The essential readings. Essential readings in developmental psychology*, 49–56. Malden, MA: Blackwell Publishers.
- Friend, M. & Keplinger, M. (2003). An infant-based assessment of early lexicon acquisition. *Behavior Research Methods, Instruments, and Computers* **35**(2), 302–309.
- Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K. M. & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* **14**, 23–45.
- Heilman, J., Weismer, S. E., Evans, J. & Hollar, C. (2005). Utility of the MacArthur-Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology* **14**, 40–51.
- Hirsh-Pasek, K. & Golinkoff, R. M. (1996). The intermodal preferential looking paradigm: A window onto emerging language comprehension. In D. McDaniel, C. McKee & H. S. Cairns (eds), *Methods for Assessing Children's Syntax*, 105–124. Massachusetts: The Massachusetts Institute of Technology Press.
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., Hennon, E. & Rocroi, C. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development* **65** (3, Serial No. 262).
- Jackson-Maldonado, D., Thal, D. J., Fenson, L., Marchman, V. A., Newton, T. & Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas: User's Guide and Technical Manual*. Baltimore: Brooks.
- Kaler, S. R. & Kopp, C. B. (1990). Compliance and comprehension in very young toddlers. *Child Development* **61**, 1997–2003.
- Klee, T., Pearce, K. & Carson, D. K. (2000). Improving the positive predictive value of screening for developmental language delay. *Journal of Speech, Language, & Hearing Research* **43**, 821–33.
- Kohnert, K. J. & Bates, E. (2002). Balancing bilinguals II: Lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech, Language, & Hearing Research* **45**, 347–59.
- Marchman, V. A. & Martinez-Sussman, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, & Hearing Research* **45**, 983–87.
- McDonough, L. (2002). Basic-level nouns: first learned but misunderstood. *Journal of Child Language* **29**, 357–77.
- Meints, K., Plunkett, K. & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology* **35**, 1072–78.
- Mervis, C. B. (1987). Child-basic object categories and early lexical development. In U. Neisser (ed.), *Concepts and Conceptual Development*, 201–233. New York: Cambridge University Press.
- Mervis, C. B. & Canada, K. (1983). On the existence of competence errors in comprehension: A reply to Fremgen & Fay and Chapman & Thomson. *Journal of Child Language* **10**, 431–40.

- Patterson, J. L. (2004). Comparing bilingual and monolingual toddlers' expressive vocabulary size: Revisiting Rescorla and Achenbach (2002). *Journal of Speech, Language, and Hearing Research* **47**, 1213–15.
- Rescorla, L. & Achenbach, T. M. (2002). Use of the Language Development Survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Research* **45**, 733–43.
- Rescorla, L. & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research* **44**, 434–45.
- Rescorla, L., Mirak, J. & Singh, L. (2000). Vocabulary growth in late talkers: Lexical development from 2;0 to 3;0. *Journal of Child Language* **27**, 293–311.
- Ring, E. D. & Fenson, L. (2000). The correspondence between parent report and child performance for receptive and expressive vocabulary beyond infancy. *First Language* **20**, 141–59.
- Stiles, J. (1994). On the nature of informant judgements in inventory measures: And so what is it you want to know? *Monographs of the Society for Research in Child Language* **59**(5), 180–85.
- Thal, D. (2005). Early detection of risk for language impairment: what are the best strategies? Paper presented at Update on Specific Language Impairment, Urbino, Italy, April.
- Thal, D. & Friend, M. (2005). Prediction of language development at 20-months from parent report and child performance at 16-months of age. In M. Friend (Chair), Picture recognition approaches to comprehension: Neuroscience, cross-linguistic, and atypical development perspectives, the X. International Association for the Study of Child Language, Berlin, Germany, July.
- Thal, D., Jackson-Maldonado, D. & Acosta, D. (2000). Validity of a parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, & Hearing Research* **43**, 1087–1100.
- Tomasello, M. & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. *Monographs of the Society for Research in Child Development* **59**(5), 174–79.
- Umbel, V. M., Pearson, B. Z., Fernandez, M. C. & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development* **64**, 1012–20.
- Yoder, P. J., Warren, S. F. & Biggar, H. A. (1997). Stability of maternal reports of lexical comprehension in very young children with developmental delays. *American Journal of Speech-Language Pathology* **6**, 59–64.