Applied Psycholinguistics **22** (2001), 45–60
Printed in the United States of America

Narratives validate Communicative Development Inventories

MÅRTEN ERIKSSON
University of Gävle

ADDRESS FOR CORRESPONDENCE
Mårten Eriksson, University of Gävle, 801 76 Gävle, Sweden. E-mail: marten.eriksson@hig.se

ABSTRACT
The study explores the criterion-related validity of the Swedish version of the Communicative Development Inventories–Word & Sentences (SECDI-W&S). In two follow-up procedures, SECDI-W&S was used to assess vocabulary and grammar skills in 32 children at ages 1;10 or 2;4. At the first follow-up, 14 months later (at ages 3;0 or 3;6), two grammar tasks from the SECDI-W&S and one narrative task (frog story) were given to 27 children. Predictive validity ranged from .50 to .60 for most assessed skills over the 14-month period, whereas concurrent validity scores at ages 3;0 or 3;6 ranged from .48 to .70. The second follow-up at 26 months (at ages 4;0 or 4;6) was conducted with the narrative task only. Predictive validity from ages 1;10 or 2;4 for the second follow-up remained high for some of the measures; however, for other measures it diminished, which could be partly accounted for by ceiling effects. The overall results confirm that the criterion-related validity of the SECDI is sound.

The MacArthur Communicative Development Inventories (CDI) (Fenson et al., 1993, 1994) is an instrument that uses parent reports to evaluate communicative skills in children aged 0;8 to 2;6. The major components of the instrument assess the use of gestures, vocabulary comprehension, vocabulary production, grammar skills, and the mean utterance length of the child's three longest utterances (MaxLU). The CDI has become a powerful tool in the study of early language development, enabling researchers to describe the general development and variation of a diversity of linguistic features in large samples. CDI-based studies have also addressed questions concerning development in slow and fast talkers (Thal & Tobias, 1994; Weismer & Murray-Branch, 1994); bilingualism (Zurer Pearson, & Fernández, 1994); relations between early gestures, word comprehension, and word production (Eriksson & Berglund, 1999; Fenson et al., 1994; Thal & Tobias, 1994); relations between word production and early syntax (Berglund & Eriksson, 2000; Fenson et al., 1994; Thal, Bates, Zappia, & Oroz, 1996); gender differences (Berglund & Eriksson, 2000; Eriksson & Berglund, 1999; Fenson et al., 1994); and language development in children with developmental disorders (Bates, Dale, & Thal, 1995; Berglund, 1998; Berglund, Eriksson, & Johansson, 1998; Miller, Sedey, & Miolo, 1995). The inventories are widely used throughout the world and have been adapted to numerous lan-

© 2001 Cambridge University Press 0142-7164/01 $9.50

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

guages and cultures in addition to the original American version, including Chinese (Wu, 1997), Croatian (Kovacevic, Babic, & Brozovic, 1996), Finnish (Lyytinen, Poikkeus, & Laakso, 1997; Lyytinen, Poikkeus, Leiwo, Ahonen, & Lyytinen, 1996), French (Poulin-Dubois, Graham, & Sippola, 1995), German (Grim, Doil, Müller, & Wilde, 1996), Hebrew (Maital, Dromi, Sagi, & Bornstein, 1998), Icelandic (Thordardottir & Weismer, 1996), Italian (Camaioni, Caselli, Longobardi, & Volterra, 1991), Japanese (Ogura, Yamashita, Murase, & Dale, 1993), Spanish (Jackson-Maldonado, Thal, Bates, Marchman, & Gutierrez-Clellen, 1993), and Swedish (Berglund & Eriksson, 2000; Eriksson & Berglund, 1999).

Because of the prevalence of these instruments, studies exploring the validity of different versions are essential. The present study focuses on the criterion-related validity of the Swedish Early Communicative Development Inventories–Word & Sentences (SECDI-W&S).

Reliability of different versions of the CDI

The reliability of an instrument is relevant because the consistency or stability of a measure sets the upper limit for the validity of the instrument. Two types of reliability have been reported for the CDI – internal reliability of different dimensions or scales, as measured by coefficient alpha, and test–retest reliability. Fenson et al. (1993) reported alpha values of .95 or above for all scales except the total gesture scale (infant form), which yielded an alpha of .39. Recent studies on the Swedish version of the CDI, however, reported alpha values of .83 or above for all scales (including the gesture scale) (Berglund & Eriksson, 2000; Eriksson & Berglund, 1999).

Test–retest reliabilities have been conducted after 4 weeks (Jackson-Maldonado et al., 1993), after 6 weeks (Fenson et al., 1993), and after 8 weeks (Berglund & Eriksson, in press). Reliabilities for the major scales ranged between .70 and .97 across different versions and time intervals. These findings suggest that the original CDI and its Spanish and Swedish versions are reliable.

Validity of different versions of the CDI

Content validity essentially incorporates two elements: face validity and logical (or sampling) validity (Allen & Yen, 1979). Both have been considered during the development of the CDI and are discussed at length by Fenson et al. (1993). When the CDI is adapted to a new language or culture, logical validity can be examined using several procedures, including comparing the vocabulary checklist with the words in culturally relevant speech corpora (Berglund & Eriksson, in press; Jackson-Maldonado et al., 1993; Thordardottir & Weismer, 1996), with items on existing language or intelligence tests (Jackson-Maldonado et al., 1993), or with the words children were reported to use in their three longest utterances the first time the instrument was employed (Berglund & Eriksson, in press).

Studies of construct validity (more specifically, convergent validity) have also

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

been performed for versions of the CDI specifically designed to accommodate other cultures or languages. Comparisons of developmental patterns and correlations between different sections in instruments adjusted to different languages and cultures with those of the original CDI provide a sensible means to estimate convergent validity. Developmental patterns of diverse language skills and interrelations between different skills have proven to be robust across many versions of the CDI (e.g., Berglund & Eriksson, 2000; Camaioni et al., 1991; Caselli, Bates, Casadio, Fenson, Sanderl, & Weir, 1995; Eriksson & Berglund, 1999; Jackson-Maldonado et al., 1993; Thordardottir & Weismer, 1996). Moreover, evidence of discriminant validity with a near-zero correlation between use of crying/vocalizations at 1;0 and vocabulary size at 1;8 has been reported (Camaioni et al., 1991). To summarize, studies on content and construct validity have produced satisfactory results for the original version of the CDI as well as for the Icelandic, Italian, Spanish, and Swedish versions.

Several studies on the validity of the CDI and its variants have concerned criterion-related validity (i.e., concurrent and predictive validity). Concurrent validity has been assessed in relation to measures derived from naturalistic speech (typically a vocabulary count or MLU score), in relation to vocabulary counts from diaries, and in relation to scores from structured tests. Most correlations are in the range of .60 to .86 (see upper portion of Table 1). Considering the nonperfect reliability of the CDI score and of the different criteria, many values are close to their theoretical upper limit.

A number of studies have investigated the predictive validity of the CDI. These studies have concerned follow-up test periods between 2 and 10 months. CDI scores or standardized test scores have served as criteria on the second measuring occasion. Correlations have ranged between .38 and .83 for different measures and periods. Predictions from children younger than one year are uncertain (see lower portion of Table 1). To our knowledge, no studies on the predictive validity of CDI have concerned predictions of complex linguistic tasks. Predictions over periods longer than 10 months are scarce.

Objective of the present study

The objective of this study was to investigate the criterion-related validity of the SECDI-W&S in relation to different language measures derived from a narrative task given to young children. Narratives have previously been used with considerable success as a measure of communicative competence (e.g., Hemphill, Picardi, & Tager-Flusberg, 1991). Compared with language samples of free play, narratives require a more structured and linguistically advanced competence. It is commonly recognized that testing language in preschoolers by traditional testing procedures involves many problems associated with extralinguistic factors (e.g., Scarborough, 1990). To minimize these undesirable factors in the narrative task, we used a procedure in which the parents tape recorded their child's speech at home according to our instructions.

First, we investigated the concurrent validity of the grammar score and the MaxLU score of the SECDI-W&S. Second, we evaluated the predictive validity of the grammar and vocabulary measures of the SECDI-W&S at ages 1;10 or

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Table 1. *Studies on criterion-related validity of different versions of the CDI*

Study	Criterion	Age		Result
		From	To	
Concurrent validity				
Bates, Bretherton, & Snyder (1988)	Naturalistic speech	1;6		.83
Jackson-Maldonado et al. (1993)	Naturalistic speech	0;8–1;4		.84
Pine, Lieven, & Rowland (1996)	Naturalistic speech	1;1		.45–.64
Dale (1991)	Naturalistic speech	2;0		.53–.79
Dale (1991)	EOWPVT[a]	2;0		.73
Dale (1991)	Sentence memory test[b]	2;0		.75
Thordardottir & Weismer (1996)	Naturalistic speech	1;3–3;0		.90–.94
Thordardottir & Weismer (1996)	EOWPVT	1;3–3;0		.73
Reznick & Goldfield (1994)	Diaries	1;2–1;8		.64–.86
Predictive validity				
Bates, Bretherton, & Snyder (1988)	PPVT[c]	1;6	2;4	.83–.51
Dale, Bates, Reznick, & Morisset (1989)	Bailey's (BSID)[d]	1;6	1;8–2;2	.39–.55
Fenson et al. (1993)	CDI	1;4	2;0	.62–.71
Fenson et al. (1993)	CDI	0;10–1;4	1;4–1;10	.62
Fenson et al. (1993)	CDI	0;8–0;10	1;2–1;5	.38–.44
Lyytinen et al. (1996)	Finnish CDI	1;2	1;6	.76
Lyytinen et al. (1996)	Reynell (RDLS)[e]	1;2	1;6	.53–.60

[a]Expressive One-Word Picture Vocabulary Test.
[b]Memory for Sentence Subtest of the Stanford–Binet Intelligence Scale.
[c]Peabody Picture Vocabulary Test.
[d]Bailey Scales of Infant Development.
[e]Reynell Developmental Language Scale.

2;4 relative to different language skills after 14 and 26 months. Selection of outcome measures in the children's narratives was made in order to display clear developmental trends between 3- and 4-year-olds and to reflect a variety of linguistic skills. The total number of word types (WT), mean length of turn (MLT) in words, mean length of utterances (MLU) in morphemes, use of the indefinite form (INDEF) when introducing new characters (cf. Bamberg, 1985, 1994), and attribution of cognitions (COGNI) to the characters of the story (cf. Berman & Slobin, 1994) fulfilled these criteria.

To summarize, the present study examined the criterion-related validity of the SECDI-W&S, where different linguistic measures derived from children's narratives served as criteria.

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Table 2. *Task administration, age, and gender of participating children at T1, T2, and T3*

	T1	T2	T3
Task administration	SECDI-W&S: vocabulary SECDI-W&S: grammar SECDI-W&S: MaxLU	Narrative SECDI-W&S: grammar SECDI-W&S: MaxLU	Narrative
First cohort			
Age	1;10	3;0	4;0
Girls	7	6	3 (2)
Boys	12	12	8 (8)
Second cohort			
Age	2;4	3;6	4;6
Girls	6	5	4 (3)
Boys	7	4	5 (2)

Note: Age is within two weeks. MaxLU = mean utterance length (in morphemes) of the child's three longest utterances. All children participated at T1. Parentheses indicate the number of children who participated at T2 and T3.

METHOD

Design

Berglund and Eriksson (2000) collected 900 parent reports on the communicative skills of 336 children in two randomly selected nationwide cohorts. The SECDI-W&S was given either at ages 1;4, 1;7, and 1;10 (first cohort) or at 1;10, 2;1, and 2;4 (second cohort). Following completion of that study, the families were asked to indicate their interest in participating in a follow-up that would take place one year later. Data from 32 children of the last administration (at ages 1;10 or 2;4), comprising a vocabulary score, a grammar score, and a MaxLU score, constitute the T1 measures of the present study. After 14 months (T2), the grammar scale and the MaxLU task of the SECDI-W&S were administered again. At T2 and at T3 (after 26 months), the children also produced a narrative (see Table 2).

Procedure

A mail-in procedure was used on all three occasions. This procedure has previously been used successfully with the SECDI forms. It was now extended to include a speech sample in the form of a narrative from the child at T2 and T3. Copies of a picture book, a cassette tape for a tape recorder, and detailed instructions were mailed to the parents, who were asked to record their child telling the story from the book. The child and the parent were instructed to look

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

through the book once before recording. The children were told to look at one page at a time and not to turn back. The parents were told to minimize their interaction with the child and only to do what was needed to elicit a story. However, to make the narrative "story-like," many of the parents provided the child with prompts and filled in vital information. This form of prompting is reflected in the low MLT.

Subjects

A total of 67 families conveyed an interest to participate in the study. Families that had expressed interest but failed to submit material at the first follow-up reported the following explanations: no access to a tape recorder or technical problems with the recording ($n = 19$), noncooperative children ($n = 10$), and intervening external incident ($n = 11$).

Thirty-two children were included at T1. Of these, 27 completed the one-year follow-up (after 14 months at ages 3;0 or 3;6) at T2, and 20 children completed the two-year follow-up (after 26 months at ages 4;0 or 4;6) at T3. Fifteen children contributed data on all three occasions. Table 2 shows the distribution over age and gender. The children at the one-year follow-up had a median vocabulary corresponding to the 60th percentile (189 words) of the national norms at age 1;10. Median vocabulary at age 1;10 for the children in the two-year follow-up was 182.5 words. Five children were within the lower quartile of the national norms in both follow-ups. Nine children in the first follow-up and five children in the second follow-up were within the upper quartile.

Materials

The narratives were based on the wordless 24-page picture book, *Frog, where are you?* (Mayer, 1969). The plot revolves around three characters: a little boy, a dog, and a frog. The boy keeps the frog in a glass jar. One night, the frog escapes. Following a series of adventures in search of the frog, involving bees, a gopher, and an owl, a deer throws the boy into a pond, whereupon he finds a frog family. The story is temporally structured in a set of goal-based events and tests a number of linguistic skills. The book was initially used by Bamberg (1985) and has been utilized in many studies within a variety of languages (see Berman & Slobin, 1994).

Derivation of scores from the SECDI-W&S. The vocabulary section of the SECDI-W&S consists of a checklist of 710 words organized into 21 semantic categories. The vocabulary score corresponds to the sum of all words on the list that the parent reported the child produced. The grammar scale of the SECDI-W&S comprises six questions. Five questions concern morphological markers for the possessive form, definite singular, definite plural, plural marking, and past tense/supine. Parents were also asked to indicate if their child combined words. For each question, "never" indicated a score of 0, "sometimes" indicated a score of 1, and "often" indicated a score of 2. Hence, the summed scale spanned from 0 to 12. The MaxLU score is the mean length of utterances com-

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

puted on the child's three longest utterances that the parent could recall. A MaxLU score of 13 was the limit to the 90th percentile, and this was set as the maximum score. A score of 13 was also assigned if the parent had reported that his or her child could make utterances of "any length." If the parent provided two of three requested examples, one was subtracted from the total score before dividing by two. Similarly, if the parent only provided one example, the score was decreased by two. None of these tasks had reached ceiling at age 2;4 (see Berglund & Eriksson, 2000).

Derivations from the narratives. Audiotapes were transcribed and coded using the Codes for the Human Analysis of Transcripts (CHAT) system (MacWhinney & Snow, 1990). An experienced transcriber, independent of the project group, checked all transcriptions. MLU, MaxLU, MLT, and the total number of different word types were first computed with the Computerized Language Analysis (CLAN) programs (MacWhinney & Snow, 1990) and then corrected manually for nonwords, variations of the same stem, and so forth. Outliers (the top 10%: three children at T2 and two children at T3) were substituted with values corresponding to the 90th percentiles. The number of times each child spontaneously introduced a new character into the narratives in the indefinite form was counted, with a maximum score set at six (which is the number of different animal species in the story). Children's spontaneous attributions of the cognition of the story characters were noted (e.g., "knows," "believes," "thinks," and "is curious"). Only different types of attribution were calculated regarding the same subject in the narrative; a maximum of three cognitive attributions was set for each narrative.

Reliability of coding

The only measures that required hand scoring were INDEF and COGNI. Two independent coders scored a random sample of the material constituting 20% of the total material. The reliability scores were $r = .87$ for INDEF and $r = .85$ for COGNI.

RESULTS

First, the general level of performance in terms of means, standard deviations, and developmental trends is presented in order to assert the developmental properties of the employed measures. Next, concurrent validity is analyzed by correlating the grammar score and the MaxLU score of the SECDI-W&S at T2 with corresponding assessments from the narratives. In addition, each child's spontaneous use of the features assessed by the grammar scale of SECDI-W&S in the narrative is compared with the child's results on the corresponding items from the grammar scale of the SECDI-W&S. Finally, the predictive validity of the grammar score, the MaxLU score, and the vocabulary score of the SECDI-W&S at T1 is correlated with measures from the narratives at T2 and T3.

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Table 3. *Performance on language measures at 1;10 or 2;4 (T1), at 3;0 or 3;6 (T2), and at 4;0 or 4;6 (T3)*

	T1		T2		T3	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Vocabulary score (SECDI)	349	209	–	–	–	–
Grammar score (SECDI)	5.97	3.79	10.33	2.43	–	–
MaxLU (SECDI)	4.53	2.31	8.78	3.00	–	–
WT (narrative)	–	–	91.8	34.1	95.1	29.7
MLU (narrative)	–	–	3.74	1.01	5.50	1.70
MLT (narrative)	–	–	6.92	8.88	15.7	20.3
INDEF (narrative)	–	–	1.37	1.21	2.48	1.89
COGNI (narrative)	–	–	–	–	.500	.95

Note: $n(\text{T1}) = 32$, $n(\text{T2}) = 27$, $n(\text{T3}) = 20$. MaxLU = mean utterance length (in morphemes) of the child's three longest utterances; WT = word types; MLU = mean length of utterances (in morphemes); MLT = mean length of turns; INDEF = use of the indefinite form when introducing new characters; COGNI = the attribution of cognitions to the characters of the story.

General level of performance

Table 3 lists the means and standard deviations (*SD*) for the language measures used in this study. Independent *t* tests confirmed significant developmental trends between T1 and T2 for grammar score, $t(60) = 5.38$, $p < .05$, and for MaxLU, $t(59) = 6.14$, $p < .05$. Significant developmental trends were also confirmed between T2 and T3 for MLU, $t(45) = 4.43$, $p < .05$; for MLT, $t(46) = 2.03$, $p < .05$; for INDEF, $t(46) = 2.46$, $p < .05$; and for COGNI, $t(45) = 2.76$, $p < .05$; but not for WT, $t(45) = .34$, $p > .05$. No children at T2 attributed cognitions to the characters of the narratives – an ability that was just starting to emerge at T3. The use of the indefinite for introducing a new character was also rare, although it did increase from T2 to T3. The grammar score from the SECDI-W&S approached an upper limit (ceiling effect) at T2 with a mean of 10.33 (maximum = 12). The decreasing *SD* in the grammar score between T1 and T2 signaled that the scores approached ceiling. *SD* also decreased in WT between T2 and T3, indicating that the scores were close to the ceiling.

Concurrent validity

MaxLU from the SECDI-W&S correlated significantly with the corresponding MaxLU, MLU, and WT derived from the narrative. The SECDI-W&S grammar scale also correlated with the MaxLU, MLU, and WT derived from the narrative (Table 4). The correlation between the grammar scale and the narrative MaxLU was higher (.53) than the correlation obtained between the two grammar scores from the SECDI-W&S, the grammar scale and the MaxLU measure (.35). The correlation between MaxLU and MLU reached .80 in the narrative.

A qualitative analysis was undertaken to compare each child's spontaneous

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Table 4. *Concurrent correlations for the SECDI measures and measures derived from the narrative at T2*

	Narrative measures		
SECDI measures	MaxLU	MLU	WT
MaxLU	.557**	.544**	.695**
Grammar scale	.530**	.543**	.480**

Note: MaxLU = mean utterance length (in morphemes) of the child's three longest utterances; MLU = mean length of utterances (in morphemes); WT = word types.
**$p < .01$, one-tailed.

Table 5. *Presence of the features from the SECDI grammar scale in the narratives at T2*

	In SECDI	In narrative	In SECDI and narrative	In narrative but not in SECDI	In SECDI but not in narrative
Possessive forms	27	11	11	0	16[a]
Definite form singular	24	26	24	2	0
Definite form plural	22	18	16	2	6[b]
Plural forms	25	22	21	1	4[c]
Past tense	25	21	20	1	5[d]
Combine words	27	27	27	0	0

Note: All tabulations are based on 27 children.
[a]This occurred on both occasions the parent indicated "sometimes" rather than "often."
[b]This occurred on all four occasions the parent indicated "sometimes" rather than "often."
[c]One of these children produced the correct plural in the definite form in the narrative.
[d]This occurred on three of the seven occasions the parent indicated "sometimes" rather than "often."

use of the six morphological features in the narrative (measured by the grammar scale of SECDI-W&S) with the child's results on the corresponding items on the grammar scale. As revealed in Table 5, the overall agreement was high, with slightly more indications of the child's mastery of some forms found in the grammar scale of the SECDI-W&S than what was spontaneously used in the narratives (potential false positives). However, it was evidently possible to engage in the narratives without using the first five features of the grammar scale. For instance, all children were reported as having used possessive forms, although only 11 children actually used possessive forms in their narrative. Conse-

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

quently, it would be inequitable to consider all the remaining 16 cases as (true) false positives. The other form of mismatch (false negatives), in which the child produced grammatical constructions not reported by the parents, occurred a total of six times and never more than twice in response to a single feature. In contrast to false positives, false negatives are evidently instances of imperfect validity. However, the infrequent examples of false negatives support concurrent validity of the grammar scale.

Predictive validity

Predictions from T1 to T2. The grammar score, the MaxLU score, and the vocabulary score of the SECDI-W&S correlated highly with one another at T1, ranging from .66 to .71. They all significantly predicted the grammar score at T2 and predicted the various criteria of the narrative at T2 about equally well. The MaxLU score at T1 also correlated with itself at T2. However, the grammar score and especially the vocabulary score at T1 did not correlate with the MaxLU score from SECDI-W&S at T2, although they did correlate highly with the corresponding MLU measure from the narrative at T2. We interpret this to mean that parents could not reliably recall their child's three longest utterances at ages 3;0 or 3;6. The MaxLU score (a task extracted from the SECDI-W&S) was unreliable at T2 (but not at T1).

The grammar score, the MaxLU score, and the vocabulary score of the SECDI-W&S at T1 were each able to predict most measures from the narratives at T2 except for the COGNI scores (which no child scored on at T2). Most of the correlations were .50 or above (Table 6).

Predictions from T1 to T3. The MaxLU score and the vocabulary score at T1 correlated with MLT over the 26-month period from T1 to T3. The grammar score at T1 was also found to correlate with the emerging COGNI score at T3. All measures at T1 exhibited low correlations with the WT, MLU, and INDEF scores at T3 (Table 6).

The low correlations between predictors at T1, WT, and MLU at T3 are likely attributable to ceiling effects. WT was the only criterion that failed to meet our initial aspiration of a significant increase between T2 and T3 for the selected criteria. (WT was included because it is a measure from the narratives that reflects the child's vocabulary, just as the vocabulary score from SECDI-W&S does at T1. Moreover, WT was predicted at T2 by all SECDI-W&S measures at T1.) Many of the MLU scores were most likely suppressed at T3, although the increase in MLU from a mean of 3.74 at T2 to 5.50 at T3 was statistically reliable. Scarborough, Rescorla, Tager-Flusberg, Fowler, and Sudhalter (1991) showed that the MLU is associated curvilinearly to other measures of grammatical complexity, such as the IPSyn score. The MLU curve levels off around 4.0 to 4.5 when plotted against IPSyn. Therefore, chances to predict MLU at T3 with a mean of 5.5 are small, even if the predictors are good.

To introduce new characters in a story by using the indefinite form requires mastery of the grammatical distinction between definite and indefinite forms, as

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Table 6. *Predictive validity of SECDI scores at T1 on SECDI scores at T2 and on narrative measures at T2 and T3*

	From SECDI		From narrative				
Measure	Grammar	MaxLU	WT	MLU	MLT	INDEF	COGNI
T2							
Grammar score	.488**	.28	.599**	.488**	.318	.620**	–
MaxLU	.475**	.559**	.578**	.630**	.539**	.548**	–
Vocabulary score	.406*	.060	.377*	.466**	.330*	.588**	–
T3							
Grammar score			.121	.168	.290	.081	.472*
MaxLU			.104	.235	.520**	.186	.201
Vocabulary score			–.202	–.118	.438*	.065	.318

Note: n(T1–T2; SECDI) = 28–32, n(T1–T2; narrative) = 25–27, n(T1–T3) = 19–20, n(T2–T3) = 18–19. MaxLU = mean utterance length (in morphemes) of the child's three longest utterances; WT = word types; MLU = mean length of utterances (in morphemes); MLT = mean length of turns; INDEF = use of the indefinite form when introducing new characters; COGNI = the attribution of cognitions to the characters of the story.
*$p < .05$, one-tailed; **$p < .01$, one-tailed.

well as a pragmatic sensitivity to the listener's needs. The 3- and 4-year-olds in our data used both forms but were not consistent in their applications. Although we found that the children at T3 introduced new characters in the indefinite form with significantly greater frequency than at T2, they also used the definite form for this purpose slightly, though not significantly, more often at T3 than at T2. Therefore, it is likely that the INDEF score in this study also reflects the child's more basic ability to name the characters on his or her own initiative. This account might explain why INDEF was predicted at T2. However, the mixture of two abilities of varying difficulty might particularly suppress the associations from T1 to T3.

COGNI – but particularly MLT – showed moderate to high correlations at T3 with the present language measures at T1. The increase in MLT was one of the most apparent changes between the narratives at T2 and T3, often reflecting a shift from short answers in response to parents' prompts to attempts at telling the story themselves, an accomplishment that must be regarded as a major communicative achievement. The COGNI score, on the other hand, might appear to reflect more of a detail in the children's narratives. However, attributions of the cognition of characters demand a level of inference that must be abstracted from what is explicitly shown in the pictures. This capacity is rarely found among younger children (Berman & Slobin, 1994). Nonetheless, such attributions are vital for the understanding of the plot in terms of plans and goals of the protagonists.

In conclusion, all calculated measures at T2 except the COGNI score (which no child scored on at T2) were successfully predicted from the SECDI-W&S

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

scores at T1. All measures at T1 were highly correlated with MLT at T3 and moderately to highly correlated with the emerging COGNI score at T3. However, the correlations between measures at T1 and the WT, MLU, and INDEF scores at T3 were low. The overall findings support the view that valid predictions concerning rather complex linguistic skills can be made over periods of one to two years. Nevertheless, many simple measures of linguistic development display ceiling effects at around 4 years of age. Moreover, the grammar score, the MaxLU score, and the vocabulary score from the SECDI-W&S at T1 all seemed to serve about equally well as predictors.

DISCUSSION

The present study investigated the criterion-related validity of the Swedish version of the CDI. Various measures derived from a narrative task served as criteria. Concurrent validity for MaxLU and the grammar scale of the SECDI-W&S at ages 3;0 or 3;6 (T2) ranged between .48 and .70. Predictive validity of scores from the SECDI-W&S at ages 1;10 or 2;4 (T1) and the criteria 14 months later (T2) varied between .50 and .60 for most measures. Many of the measures were found to approach ceiling at ages 4;0 or 4;6 (T3). Nonetheless, the MaxLU and the vocabulary score of SECDI-W&S at T1 predicted MLT in the narratives at T3 and the grammar score of SECDI-W&S at T1 predicted the COGNI score in the narratives at T3.

Many studies on the concurrent validity of different CDI versions have reported slightly higher correlations in comparison with those observed in the present study (Table 1). One reason for the more moderate, although solid, correlations of concurrent validity obtained in our study may be attributed to a curvilinear relation with language competence for both MLU and the grammar scale that is noticeable at ages 3;0 or 3;6. This phenomenon functions to suppress the correlations. However, the more qualitative analysis reported in Table 5, showing that children rarely produced grammatical constructions not reported by the parents, supports additional confidence in the concurrent validity of the grammar scale of the SECDI-W&S.

Studies on the predictive validity of different versions of the CDI have involved periods of 2 to 10 months, and the predictive validity figures have ranged between .38 and .83 (Table 1). Generally, the most favorable condition for achieving high predictive validity is to reuse the same task after a short interval. Under such a condition, the predictive validity converges with the test–retest reliability. Similarly, the concurrent validity usually sets the upper limit for the criterion-related validity. Hence, weaker correlations are expected for studies that cover longer periods of time or employ new criteria. One exception to this general argument may be noted. When a criterion different from the first assessment is used, which is often the case in developmental studies, an earlier ability can be strongly associated with abilities that develop later. The pattern of associations between the grammar score at T1 and the COGNI scores at T2 and T3 are excellent illustrations of this phenomenon. However, the general reasoning described here might nevertheless serve as a guideline in evaluating the magnitudes of the correlations reflecting predictive validity, displayed in Table 6. In

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

this perspective, the magnitude of many correlations in this study after 14 and 26 months, at around .50, is impressive.

Several studies (Bates et al., 1995; Berglund & Eriksson, 2000; Fenson et al., 1994) have shown that vocabulary production and grammar are strongly related skills for children ages 1;4 to 2;6, in contrast, say, to word comprehension and word production for children ages 0;8 to 1;4. The present study supports the notion of a close relationship between vocabulary and grammar of children at ages 3;0 or 3;6 by showing that measures of vocabulary and grammar at T1 predict the same skills at T2, as shown from the narratives.

Methodological issues

Finding solid criteria for studies of predictive validity is always a problem – one that tends to augment with the length of the follow-up. Many standard measures of linguistic achievements, such as WT and MLU, worked out well at T2 but not at T3. MLT, on the other hand, largely measures a pragmatic skill and so proved successful at both T2 and T3. The ability to attribute cognitions to the characters in the story, as measured by the COGNI score, is certainly an important narrative skill. The ability to draw inferences from others' mental states usually emerges between the ages of 3 and 5 years and is commonly referred to as "theory of mind" (see Flavell, 1999, for a review). Associations between language skills and theory of mind are commonly acknowledged, though different versions of causation have been proposed (de Villiers & Pyers, 1999). Therefore, it is quite possible that the COGNI score primarily measures a general cognitive ability rather than a specific linguistic one.

Problems with criteria that display a short span between floor and ceiling are associated with the extraction of single features from children's narratives. Good standardized tests that are made up of a set of different skills varying in difficulty span longer periods. Unfortunately, Swedish versions of such tests do not exist at present. However, many linguistic skills (e.g., skills associated with pragmatics) are rarely included in standardized tests of any language. Moreover, the item reflecting pragmatics in the present study, the MLT measure, proved to be the criterion that was best predicted at both T2 and T3. Therefore, different validation studies with some variation in the applied criteria are desirable. The use of narratives for this purpose has proved fruitful, and analyses can focus on different features as relevant for divergent age groups.

The mail-in procedure of the present study had some advantages compared with a professionally elicited narrative. One such advantage is that children were not disturbed by such factors as inattentiveness and shyness. However, the mail-in method, in comparison with the narrative situation, was also associated with less control. In particular, we found that the amount of prodding and support from the parents varied, although we did try to consider this in our analysis. Moreover, we had no control over whether the parents complied with our instructions and looked through the booklet once as requested. It may be that some parents ignored our instructions and looked through the book several times or not at all. Some noise in our data can certainly be attributed to this source. The mail-in procedure also involved a selection effect in that not all of the

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

families interested in participation had a suitable tape recorder. Nevertheless, the comparison of the productive vocabulary of the study group with that of the normative group at age 1;4, placing the study group in the 60th percentile, did not indicate that this selection was related to the language competence of the children.

CONCLUSIONS

The present study demonstrated concurrent validity of the SECDI-W&S, in which reasonably high correlation coefficients were obtained. Good predictive validity was achieved at ages 1;10 or 2;4 to WT, MLU, and MLT and to INDEF at ages 3;0 or 3;6. Strong correlations were also observed between scores on the SECDI-W&S at ages 1;10 or 2;4 and scores on the MLT and COGNI at 4;0 or 4;6. The use of different linguistic measures from children's narratives made it possible to compute more advanced linguistic skills, including the MLT, INDEF, and COGNI scores, which would have caused significant problems from naturalistic recordings of communication during free play.

ACKNOWLEDGMENTS

This research received financial support from the Research and Development Department of the County Council of Gävleborg and from the Center for Caring Sciences, Uppsala University. The author is indebted to Åsa Nordquist and students at the Department of Linguistics at Gothenburg University for transcriptions of children's narratives. Eva Berglund kindly assisted in reliability coding. Finally, Igor Knez of the Laboratory of Applied Psychology, Royal Institute of Technology, and two anonymous reviewers provided valuable comments on the manuscript.

REFERENCES

Allen, M. J., & Yen, M. W. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole.

Bamberg, M. (1985). *Form and function in the construction of narratives: Developmental perspectives*. Unpublished doctoral dissertation, University of California, Berkeley.

(1994). Development of linguistic forms: German. In R. A. Berman & D. I. Slobin (Eds.), *Relating events in narrative. A cross-linguistic developmental study* (pp. 189–238). Hillsdale, NJ: Erlbaum.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Bates, E., Dale, P., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 96–151). Oxford: Blackwell.

Berglund, E. (1998). *The relative risk of low vocabulary scores among children with Down syndrome and additional deficiencies* (No. 847). Stockholm University, Department of Psychology.

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16–28 months olds. The Swedish early communicative development inventory – Words and sentences. *Scandinavian Journal of Psychology, 41*, 133–144.

(in press). Validity and reliability of the Swedish early communicative development inventories. *Logopedics Phoniatrics Vocology*.

Berglund, E., Eriksson, M., & Johansson, I. (1998). *Spoken language in children with Down syn-*

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

drome 1 to 5 years old: Developmental trends, individual variations, and sex differences (No. 846). Stockholm University, Department of Psychology.

Berman, R. A., & Slobin, D. I. (Eds.) (1994). *Relating events in narrative: A cross-linguistic developmental study*. Hillsdale, NJ: Erlbaum.

Camaioni, L., Caselli, M. C., Longobardi, E., & Volterra, V. (1991). A parent report instrument for early language assessment. *First Language, 11,* 345–359.

Caselli, M. C., Bates, E., Casadio, P., Fenson, L., Sanderl, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10,* 159–199.

de Villiers, P., & Pyers, J. (1999, July). *Complementation and false belief representation*. Paper presented at the Eighth International Congress for the Study of Child Language, San Sebastian, Spain.

Dale, P. (1991). The validity of a parent report measure of vocabulary and syntax at 24 months. *Journal of Speech and Hearing Sciences, 34,* 565–571.

Dale, P., Bates, E., Reznick, S., & Morisset, C. (1989). The validity of a parent report instrument of child language at 20 months. *Journal of Child Language, 16,* 239–249.

Eriksson, M., & Berglund, E. (1999). Swedish Early Communicative Development Inventories: Words and Gestures. *First Language, 19,* 55–90.

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*.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D. J., Bates, E., Hartung, J. P., Pethick, S. J., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing Group.

Flavell, J. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology, 50,* 21–45.

Grim, H., Doil, H., Müller, C., & Wilde, S. (1996). Elternfragebogen für die differentielle Erfassung früher sprachlicher Fähigkeiten. *Sprache and Kognition, 15,* 32–45.

Hemphill, L., Picardi, N., & Tager-Flusberg, H. (1991). Narrative as an index of communicative competence in mildly mentally retarded children. *Applied Psycholinguistics, 12,* 263–279.

Jackson-Maldonado, D., Thal, D., Bates, E., Marchman, V., & Gutierrez-Clellen, V. (1993). Early lexical development of Spanish-speaking infants and toddlers. *Journal of Child Language, 20,* 523–549.

Kovacevic, M., Babic, Z., & Brozovic, B. (1996). *A Croatian language parent report study: Lexical and grammatical development*. Paper presented at the Seventh International Congress for the Study of Child Language, Istanbul, Turkey.

Lyytinen, P., Poikkeus, A. M., & Laakso, M. L. (1997). Language and symbolic play in toddlers. *International Journal of Behavioral Development, 21,* 289–302.

Lyytinen, P., Poikkeus, A.-M., Leiwo, M., Ahonen, T., & Lyytinen, H. (1996). Parents as informants of their child's vocal and early language development. *Early Child Development and Care, 126,* 15–25.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17,* 457–472.

Maital, S., Dromi, E., Sagi, A., & Bornstein, M. (1998). *The Hebrew CDI: Language specific properties and cross-linguistic generalizations*. Manuscript submitted for publication.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Press.

Miller, J., Sedey, A., & Miolo, G. (1995). Validity of parent report measures of vocabulary development for children with Down syndrome. *Journal of Speech and Hearing Research, 38,* 1037–1044.

Ogura, T., Yamashita, Y., Murase, T., & Dale, P. S. (1993, July). *Some preliminary findings from the Japanese early communicative development inventories*. Paper presented at the Sixth International Congress for the Study of Child Language, Trieste, Italy.

Pine, J., Lieven, E., & Rowland, C. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language, 23,* 573–589.

Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The contribution of parental labeling and infants' categorization abilities. *Journal of Child Language, 22,* 325–343.

Reznick, J. S., & Goldfield, B. A. (1994). Diary vs. representative checklist assessment of productive vocabulary? *Journal of Child Language, 21,* 465–472.

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press

Scarborough, H. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11,* 1–22.
Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied Psycholinguistics, 12,* 23–45.
Thal, D., Bates, E., Zappia, M., & Oroz, M. (1996). Ties between lexical and grammatical development: Evidence from early talkers. *Journal of Child Language, 23,* 349–368.
Thal, D., & Tobias, S. (1994). Relationships between language and gesture in normally developing and late-talking toddlers. *Journal of Speech and Hearing Research, 37,* 157–170.
Thordardottir, E., & Weismer, S. (1996). Language assessment via parent report: Development of a screening instrument for Icelandic children. *First Language, 16,* 265–285.
Weismer, S. E., & Murray-Branch, J. (1994). A prospective longitudinal study of language development in late talkers. *Journal of Speech and Hearing Research, 37,* 852–871.
Wu, J. (1997). *Language, play and general development for Chinese infant-toddlers: Using adapted assessments*. Unpublished doctoral dissertation, University of Colorado, Boulder.
Zurer Pearson, B., & Fernández, S. C. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning, 44,* 617–653.

https://doi.org/10.1017/S0142716401001035 Published online by Cambridge University Press