

Appendix B: Development and Validation

Summary

The High/Scope Infant-Toddler Child Observation Record is reliable and valid when used by caregivers with infants and toddlers in care and education settings. High/Scope developed the instrument with feedback from 38 practicing infant-toddler caregivers. Following this, the instrument's reliability and validity was studied with 20 caregivers, who observed 50 infants and toddlers (4 to 36 months old) in seven child care centers and one family child care home. The caregivers received 2 days of training before beginning their observations.

The alpha coefficients of internal reliability for ratings by caregivers were quite high—.99 for the 28-item scale, .92 or .93 for the items in each of the six categories. To assess inter-observer agreement, 18 of the caregivers worked in 9 pairs. Each pair observed the same children (a total of 42 across this group of centers), with instructions not to share information with each other regarding how they filled out the COR. The correlations between the scores of the two groups of pair members were .93 for the total scale and .83 to .91 for the categories.

The concurrent validity of the Infant-Toddler COR ratings was assessed by examining their correlations with 30 of the same children's scores on the Bayley Scales of Infant Development (Bayley, 1993). Infant-Toddler COR total and category scores were all correlated at .83 or higher with Bayley mental and motor age scores. Removing the effects of age, the Infant-Toddler COR total was still correlated at .36 with Bayley mental age and .26 with Bayley motor age.

The Need for Infant-Toddler Observational Assessment Instruments

The percentage of infants and toddlers receiving care outside the home is steadily increasing, and this growth is accompanied by an urgent need to offer these young children competent and sensitive care that is appropriate to their developmental status. The Infant-Toddler COR assists caregivers in offering optimum care by helping them focus their observations of children and providing a strategy for planning an environment and routine attuned to the individual needs of the children.

Most conventional infant and toddler assessment instruments are not appropriate for use in child care settings. Historically, assessment of infants and toddlers has usually taken place in research or clinical settings. Instruments used in research typically focus on a very specific behavior (for example, the length of time an infant looks at a particular picture) and provide data that are useful when comparing the behavior of groups of infants or toddlers. In contrast, clinicians may use a variety

of assessment tools to evaluate the developmental status of a particular child. The tests used by clinicians may be norm-referenced or criterion-referenced and focus on cognitive, motor, language, social, or sensory skills or some combination of developmental domains. Generally, the tests must be administered one-on-one by highly trained personnel in a clinical setting. The results are then analyzed and interpreted by clinicians and, if necessary, developmental interventions are prescribed. Screening tools are also used to look at the development of infants and toddlers. These instruments are usually easy to administer and may be used by professionals and non-professionals alike, but they provide limited information and are generally employed to identify developmental delays and determine the need for further assessment.

The assessment needs of caregivers in infant-toddler child care settings are very different from those of researchers or clinicians, and the Infant-Toddler COR has been developed with the needs of caregivers in mind. The administration of the COR is based on ongoing observations of infants and toddlers as they go about their daily routine in their natural environment. In this way the COR recognizes the constantly changing nature of development and allows caregivers to document behaviors that reflect children's strengths in a variety of developmental domains.

Infant-Toddler COR Study Procedures

The reliability and validity of the High/Scope Infant-Toddler COR were assessed using the following two-phase strategy. The first phase of the study involved infant-toddler caregivers who were familiar with the High/Scope Child Observation Record for Ages 2½–6, referred to here as the Preschool COR (High/Scope Educational Research Foundation, 1992). These caregivers field-tested a preliminary version of the Infant-Toddler COR and provided feedback to project staff, who reviewed this information and made the necessary changes in the instrument. In the second phase, this revised version of the Infant-Toddler COR was tested for reliability and validity by caregivers who had no previous experience with other versions of the COR.

Phase One: The developmental field test

Project staff recruited 38 infant-toddler caregivers working at 2 private homes and 11 child care and early education centers around the country to pilot the preliminary version of the High/Scope Infant-Toddler COR. The caregivers or their supervisors were previous users of the Preschool COR and thus were familiar with the general format of

and procedures for using the COR. They were given written instructions regarding observation of infants and toddlers and procedures for collecting anecdotes and completing the Infant-Toddler COR. The study participants were then asked to complete the Infant-Toddler COR for at least one child each over a period of 3 months and to record their comments about the content, feasibility, and relevance of the COR for their work with children, families, and staff. Project staff then reviewed their comments and the completed CORs.

The caregivers observed a total of 47 infants and toddlers between 3 and 34 months of age; the age distribution was as follows: 3 children 3–6 months old, 9 children 7–12 months old, 7 children 13–18 months old, 14 children 19–24 months old, 10 children 25–30 months old, and 4 children 31–34 months old.

Project staff examined the collected anecdotes to check them for accurate assignment to COR items and levels and looked at the frequency distributions across levels on each item. Staff also reviewed the users' comments about the content and use of the instrument. On the basis of information gained through these procedures, staff made changes in wording to clarify some items and adjusted the levels of some items to create a more even progression of difficulty. The incorporation of these changes into the Infant-Toddler COR resulted in the version that was tested for reliability and validity.

Phase Two: Design of the reliability and validity study

Sample. To initiate the Infant-Toddler COR reliability and validity study, we sent letters to local infant-toddler caregivers inviting them to participate. Twenty caregivers in seven child care centers and one family day care home agreed to participate. None of these caregivers had previous experience with the High/Scope Preschool COR. Their median age was 28 years; 18 were females and 2 were males.

Table B-1 provides information about the background of the caregivers who participated in the study. As a group, they had varied work histories and educational backgrounds. Their experience working with infants and toddlers ranged from less than 1 year to more than 10 years. Their full-time school experiences ranged from 12 to 18 years, with a mean of 15.5 years. Half the group had bachelor's degrees and two (10%) had graduate degrees. These figures suggest that this group of caregivers was more highly educated than caregivers and preschool teachers in general (Cost, Quality, and Child Outcomes Study Team, 1995; Kisker, Hofferth, Phillips, & Farquhar, 1991). However, the caregivers in this sample had little specialized training in early childhood education or child development. Fifteen of them had no early education

preservice training, three had 3 to 4 years, and two had 8 years; almost half the caregivers had no training in early childhood education or child development.

These 20 caregivers observed a total of 50 infants and toddlers. To assess inter-observer reliability, 18 of the caregivers worked in pairs (9 pairs) to complete the Infant-Toddler COR assessments on 42 children, observing 3 to 6 children per caregiver pair. In one other center, 2 unpaired caregivers completed Infant-Toddler COR assessments on an additional 8 children. Of the infants and toddlers observed, 28 were male and 22 were female. Caregivers received no special instructions on selecting children to observe and thus selected them on the basis of such factors as the frequency and regularity of their program attendance.

The observations took place over a period of 3 months. At the midpoint of the observation period, children ranged in age from 4 to 36 months, with a mean age of 21 months and a median age of 24 months. Not surprisingly, as Figure 1 illustrates, the age distribution for the sample is skewed to the older ages. National surveys

Table B-1
Background Characteristics of Caregivers in the Reliability and Validity Study

Background Characteristic	N	% of Respondents
Years working with infants and toddlers		
0 – 2	2	10
3 – 4	9	45
5 – 6	5	25
10 or more	4	20
Years of full-time schooling		
12 or fewer	2	10
13 – 14	3	15
15–16	8	40
17 or more	7	35
Training in child development		
Yes	11	55
No	9	45
Years of early childhood education training		
0	15	75
1 – 2	0	0
3 – 4	3	15
8	2	10

Note. Twenty caregivers participated in the study.

show that the percentage of children under age 3 years who spend some time in nonparental care each week increases with age (Hofferth, Shauman, Henke, & West, 1998), so it is likely that the caregivers simply had more opportunities to observe children older than 1 year.

Training and follow-up. Project staff conducted a 2-day training for the 20 caregivers. Training began with a focus on anecdotal note taking. Participants practiced note taking as they observed video clips of children, learning how to format notes and keep them objective and concise. Next they practiced sorting the notes into COR categories, items, and item levels, with decreasing assistance from the trainers. This was followed by an inter-rater agreement check on how they sorted the notes. After this, participants completed the Infant-Toddler COR based on observation of a 35-minute video of one child and participated in a discussion of how to use the COR for program planning and during talks with parents. Finally, workshop leaders explained to the caregivers what they would be doing as participants in the project.

Training staff were on call to answer questions throughout the study. At the midpoint of the 3-month observation period, training staff reviewed the progress of each participant by checking the anecdotes collected to date and the assignment of these anecdotes to items and levels on the Infant-Toddler COR.

Reliability and Validity Study Findings

Table B-2 presents means, standard deviations, and ranges for each of the COR items based on the total sample of 50 children and 11 caregivers. Like the age distribution, the item means are skewed toward the top, but the ranges

Table B-2

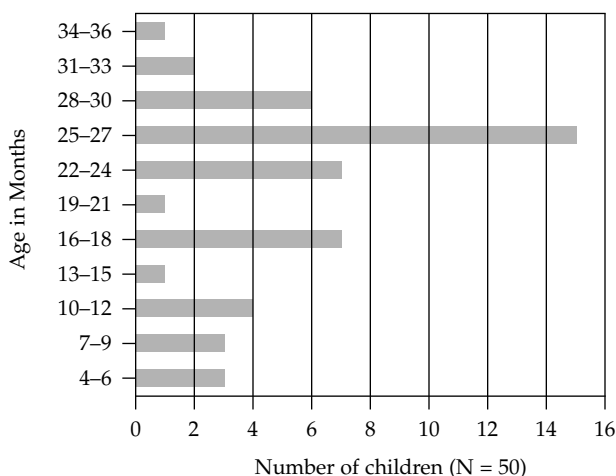
High/Scope Infant-Toddler COR Item Statistics

Category/Item	N	Mean	SD	Range
I. Sense of self	48	4.0	1.0	1.5–5.0
A. Expressing initiative	49	4.2	1.0	2–5
B. Distinguishing self from others	49	3.9	1.2	1–5
C. Solving problems	48	3.8	1.1	2–5
D. Developing self-help skills	48	4.1	1.1	1–5
II. Social relations	39	3.9	1.0	1.8–5.0
E. Forming an attachment to a primary caregiver	49	4.1	1.6	1–5
F. Relating to unfamiliar adults	42	3.9	1.2	2–5
G. Relating to another child	49	3.9	1.2	1–5
H. Expressing emotion	48	3.7	1.1	2–5
I. Responding to the feelings of others	39	3.7	1.1	1–5
J. Playing with others	48	3.9	1.2	1–5
III. Creative representation	46	3.9	1.1	1.7–5.0
K. Pretending	49	3.8	1.1	1–5
L. Exploring building and art materials	49	3.8	1.1	2–5
M. Responding to and identifying pictures and photographs	46	4.2	1.2	1–5
IV. Movement	45	3.9	1.0	1.8–5.0
N. Moving parts of the body	47	3.9	1.2	1–5
O. Moving the whole body	48	4.1	1.5	1–5
P. Moving with objects	45	3.6	0.9	1–5
Q. Moving to music	47	4.1	1.2	1–5
V. Communication and language	47	3.9	0.9	1.8–5.0
R. Listening and responding	49	4.0	0.8	2–5
S. Communicating interest nonverbally	47	3.9	1.2	1–5
T. Participating in give-and-take communication	46	3.9	1.1	2–5
U. Speaking	48	3.8	1.2	2–5
V. Exploring picture books	48	4.0	1.0	2–5
W. Showing interest in stories, rhymes, and songs	47	4.0	1.1	1–5
VI. Exploration and early logic	44	3.8	1.1	1.4–5.0
X. Exploring objects	49	4.1	1.1	1–5
Y. Exploring categories	45	3.8	1.3	1–5
Z. Developing number understanding	48	3.6	1.3	1–5
AA. Exploring space	45	4.1	1.2	1–5
BB. Exploring time	44	3.8	1.0	2–5
Total	39	3.9	1.0	1.8–5.0

Note. Eleven caregivers rated the performance of 50 children. Items have a potential range of 1 to 5, with 5 at the top.

Figure 1

Ages of Children in the Study



indicate that the items fit the sample well, with ranges from 1 to 5 for 18 items and 2 to 5 for 10 items. The table also presents the number of children rated on each item (out of the 50 children in the sample). The caregivers did not rate 1 or 2 children on 16 of the 28 items, 3 to 6 children on 7 of the items, 8 children on one item (*F. Relating to unfamiliar adults*) and 11 children on one item (*I. Responding to the feelings of others*). It is clear enough that infants and toddlers cannot relate to unfamiliar adults unless unfamiliar adults are present during the observation period, and it also appears that it is difficult for caregivers to observe signs of empathy in all infants and toddlers. The numbers of children identified for categories and the total score reflect the smallest number of children identified on any item in the category or total.

Staff examined the mean scores by age for Infant-Toddler COR items and categories. Table B-3 presents descriptive statistics for each of three age groups. The table shows that, as expected, the mean scores increase as children's ages increase. For children under one year, the mean scores for each category ranged from 2.13 to 2.48. Mean category scores for children between the ages of 1 and 2 years ranged from 3.85 to 4.11, and mean scores for children between 2 and 3 years were slightly higher, ranging from 4.51 to 4.70 across categories.

Reliability

The reliability results include intra-observer alpha coefficients (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) that describe the extent to which a single observer rated related behaviors in the same way (Suen & Ary, 1989); and inter-observer Pearson product-moment correlation coefficients that describe the extent to which two observers rated the same behavior the same way.

Alpha coefficients were computed for each of the six categories, as well as for the entire scale. The coefficients were uniformly high: .99 for the entire scale and .92 or .93 for each of the six categories. It is probable that these values are so high in part because of the high correlation between Infant-Toddler COR performance and chronological age. However, when alpha coefficients were computed separately for the three age categories, the values remained quite high: .94 for infants less than 1 year old, .95 for infants 1 to 2 years old, and .78 for toddlers 2 to 3 years old. The high alpha values indicate that the various behaviors of infants and toddlers, as observed with the Infant-Toddler COR, are remarkably consistent, both within and across categories.

To assess reliability across caregivers, 9 pairs of caregivers completed Infant-Toddler CORs on a total of 42 children. The caregivers were instructed not to share anecdotes or any other information regarding COR com-

pletion with each other. In other words, each caregiver worked independently to complete the Infant-Toddler CORs for the children he or she observed.

For the analysis, we arbitrarily divided each pair and called one set of pair members Group 1 and the other Group 2. As Table B-4 illustrates, the Infant-Toddler COR category means of the items completed by each group of caregivers were very similar, with an average category difference of only .05 point.

The Pearson product-moment correlations computed between the two groups of observers were .93 for the overall scale and .83 to .91 for the categories, as follows: **sense of self**, .89; **social relations**, .86; **creative representation**,

Table B-3

High/Scope Infant-Toddler COR Statistics by Age

COR Category/Age	Mean	SD	Range
Sense of self			
Infants less than 1 year	2.48	.55	1.5–3.3
Infants 1 to 2 years	4.11	.68	2.5–5.0
Toddlers 2 to 3 years	4.65	.45	3.5–5.0
Social relations			
Infants less than 1 year	2.37	.58	1.8–3.7
Infants 1 to 2 years	3.86	.59	2.8–4.7
Toddlers 2 to 3 years	4.54	.61	3.4–5.0
Creative representation			
Infants less than 1 year	2.13	.32	1.7–2.7
Infants 1 to 2 years	3.94	.67	3.0–5.0
Toddlers 2 to 3 years	4.70	.48	3.3–5.0
Movement			
Infants less than 1 year	2.19	.52	1.8–3.2
Infants 1 to 2 years	4.11	.54	3.0–3.5
Toddlers 2 to 3 years	4.60	.36	3.8–5.0
Communication and language			
Infants less than 1 year	2.48	.44	1.8–3.2
Infants 1 to 2 years	3.85	.62	2.8–5.0
Toddlers 2 to 3 years	4.61	.41	3.4–5.0
Exploration and early logic			
Infants less than 1 year	2.19	.53	1.4–3.4
Infants 1 to 2 years	3.88	.66	2.4–4.6
Toddlers 2 to 3 years	4.51	.54	3.0–5.0
Total			
Infants less than 1 year	2.33	.45	1.8–3.2
Infants 1 to 2 years	3.94	.55	2.9–4.6
Toddlers 2 to 3 years	4.60	.39	3.5–5.0

Note. Scores have a potential range of 1 to 5, with 5 at the top. The sample of 50 young children had 10 infants less than 1 year old, 16 infants 1 to 2 years old, and 24 toddlers 2 to 3 years old.

.83; **movement**, .88; **communication and language**, .91; and **exploration and early logic**, .86.

These levels of agreement are acceptable for observational instruments. Some have argued that pairs of teachers or caregivers cannot be fully independent when they are working together, despite instructions to the contrary. Such mutual dependence would inflate the levels of agreement to some extent. In any case, these statistics indicate a high level of agreement between caregivers, evidence of the instrument's reliability across observers and the psychometric acceptability this brings.

The percentage agreement between the pairs of caregivers was also computed for each category; the results are presented in Table B-5. Exact agreement within categories ranged from 55 percent to 66 percent; exact or near agreement (ratings within one level of each other) ranged from 89 percent to 94 percent within categories.

Validity

Staff assessed the concurrent validity of COR ratings by examining correlations with a widely respected infant assessment instrument, the Bayley Scales of Infant Development (Bayley, 1993). The Bayley Scales have a long history of use in infant assessment; they are generally used as a developmental assessment tool by specially trained personnel in clinical settings. We computed Pearson product-moment correlations between mental and motor age scores derived from the Bayley on a subsample of 30 children who were part of the reliability study. These children were 8 to 35 months old when tested with the Bayley, with a mean age of 23.6 months. The Bayley assessments were completed immediately after the reliability data were collected, by an experienced examiner who had no knowledge of the results of the Infant-Toddler COR reliability study. The correlations are presented in Table B-6.

Table B-4

High/Scope Infant-Toddler COR Category Means by Two Groups of Caregivers

Category	Group 1	Group 2
Sense of self	4.13 (.96)	4.18 (.99)
Social relations	3.99 (1.0)	3.98 (.99)
Creative representation	4.02 (1.0)	3.88 (1.1)
Movement	3.99 (.98)	4.02 (.93)
Communication and language	4.02 (.92)	3.99 (.99)
Exploration and early logic	3.95 (.98)	4.01 (1.0)
Total	4.02 (.94)	4.03 (.96)

Note. Scores are based on the ratings of a total of 42 children by 9 pairs of caregivers. One member of each pair is in Group 1, the other is in Group 2. Standard deviations are in parentheses.

The Infant-Toddler COR scores are highly correlated with Bayley mental and motor age scores; all of the correlations are .83 or higher. This is partly explained by the fact that both Infant-Toddler COR scores and Bayley age scores are highly correlated with chronological age. To remove the common variance due to chronological age, we computed partial correlations, with the effects of age removed. As expected, this resulted in more modest correlations between the Infant-Toddler COR scores and Bayley

Table B-5

Percentages of High/Scope Infant-Toddler COR Scores on Which Caregiver Pairs Agree

COR Category	% in Exact N	% of Ratings Pairs	
		% in Exact or Near Agreement ^a	Agreement ^b
Sense of self	35	66	94
Social relations	30	64	92
Creative representation	33	63	89
Movement	34	59	94
Communication and language	32	57	94
Exploration and early logic	33	55	92
Total	30	60	93

Note. Statistics are based on ratings of a total of 42 children by 9 caregiver pairs.

^aRatings in agreement divided by all non-missing ratings.

^bRatings in agreement or near agreement (within one level of each other) divided by all non-missing ratings.

Table B-6

Correlations Between High/Scope Infant-Toddler COR Scores and Bayley Scores

Bayley Mental COR Category	Bayley Motor Age	Age-Partialled		
		Bayley Mental Age	Bayley Motor Age	Age
Sense of self	.89	.86	.46*	.34
Social relations	.88	.83	.43*	.27
Creative representation	.89	.87	.22	.26
Movement	.92	.90	.51*	.49*
Communication and language	.88	.83	.22	.10
Exploration and early logic	.89	.85	.44*	.28
Total	.91	.87	.36*	.26

Note. These Pearson product-moment correlations are based on a subsample of 30 children.

**p* < .05.

mental and motor age, ranging from .10 to .51. After removing the effects of age, Bayley mental age was significantly correlated (at $p < .05$) with the Infant-Toddler COR total score and four of its categories—**sense of self, social relations, movement, and exploration and early logic**; Bayley motor age was significantly correlated only with Infant-Toddler COR **movement**, the most similar category.

Study Conclusions

This study indicates that, with two days of training, infant-toddler caregivers can use the Infant-Toddler COR to assess the behavior of infants and toddlers in care and education settings, producing results that are both reliable and valid. The internal reliability and inter-observer agreement of these ratings should be excellent. For the total Infant-Toddler COR score, the validity study found an alpha coefficient of internal reliability of .99 and a Pearson product-moment correlation coefficient between observers of .93. The concurrent validity should be excellent as well. In our study, the total Infant-Toddler COR was correlated .91 with Bayley mental age and .87 with Bayley motor age.

In the study, 20 caregivers observed 50 infants and toddlers. As such, it was large enough to accurately assess the instrument's reliability and validity, but not large enough to establish reliable age norms. As Figure 1 shows, the sample had 10 infants under 1 year old, 16 infants 1 to 2 years old, and 24 toddlers 2 to 3 years old. Many early childhood educators are justifiably wary of age norms because of the variability in the pace of children's development in their youngest years. We recommend that Infant-Toddler COR users assess children's development not against age norms, but in light of their own development over time.

The 20 infant-toddler caregivers who participated in the validity study were typical of all infant-toddler caregivers in age and early childhood training and experience, even though half of them had bachelors'

degrees. Infant-toddler caregivers vary greatly in their early childhood training and understanding. Training and use in the Infant-Toddler COR will advance their understanding of child development, but cannot substitute for preservice and inservice training in early childhood care and education.

The reliability and validity study did not include children with handicapping conditions, although High/Scope staff would like to carry out such a study and would encourage special educators and others who work with children with disabilities to use the Infant-Toddler COR to help identify areas both of strength and challenge in children's development.

References

- Bayley, N. (1993). *Bayley Scales of Infant Development II* (2nd ed.) [Assessment instrument]. San Antonio, TX: The Psychological Corporation.
- Cost, Quality, & Child Outcomes Study Team. (1995). *Cost, quality, and child outcomes in child care centers, public report* (2nd ed.). Denver: Economics Department, University of Colorado at Denver.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- High/Scope Educational Research Foundation. (1992). *High/Scope Child Observation Record (COR) for Ages 2½-6* [Assessment instrument]. Ypsilanti, MI: High/Scope Press.
- Hofferth, S. L., Shauman, K. A., Henke, R. R., & West, J. (1998). *Characteristics of children's early care and education programs: Data from the 1995 National Household Education Survey* (NCES 98-128). Washington, DC: U.S. Government Printing Office.
- Kisker, E., Hofferth, S., Phillips, D., & Farquhar, E. (1991). *Profile of child care settings: Early education and care in 1990*. Washington, DC: U.S. Department of Education. (Eric No. 343702)
- Suen, H. K., & Ary, D. (1989). *Analyzing quantitative behavioral observation data*. Hillsdale, NJ: Erlbaum.