THE QUARTERLY NEWSLETTER OF THE
Institute for Comparative Human Development
WILLIAM S. HALL AND MICHAEL COLE, EDITORS
THE ROCKEFELLER UNIVERSITY July 1978, Volume 2, Number 3

The Effect of Context on Mother-Child Interaction: A Progress Report

ZOË R. GRAVES and JOSEPH GLICK
City University of New York Graduate Center

Much of the literature on mother-child interaction is based on the observation of so-called "naturalistic" behavior. The samples of behavior commonly used for these studies are usually collected by an observer with a tape recorder, videotape camera, or checklist, for later analysis. It is generally assumed by the investigators that what is being captured by these recordings is a sample of typical, "natural," everyday behaviors occurring between mother and child, and that the observer and his equipment are no more than a neutral third eye in the room. Many studies (e.g., Lewis and Rosenblum, 1974; Hess and Shipman, 1965) make general claims about the nature of interactional phenomena across contexts on the basis of such data. The possible effect of the observer's presence on the nature of interactions is rarely investigated or even mentioned, although it seems likely that this may define a particular kind of context for the participants—especially when the observer is known as a psychologist. Subjects' performances under these conditions may differ markedly from those occurring when subjects are alone or unaware of being observed. It may be that a different set of behaviors, linguistic styles, etc. is used by mothers and children relative to the presence of a particular audience and context.

Our interest in the effect of the observer's presence on the nature of interaction arose during a study of maternal gesturing based on an analysis of videotaped interactions of mothers with their young children (Shatz and Graves, 1976). In the course of this work, it became apparent that both mothers and children were orienting to the presence of the camera and experimenter throughout the tapes of these supposedly "natural" interactions. Though subjects may have become "accustomed" to the experimenter's presence, they clearly were not oblivious to it. Both mother and child were obviously aware of and reacting to their audience.

One of the salient issues in recent comparative research is the importance of taking into account the context in which behaviors occur (Cole and Scribner, 1975; Goodnow, 1972; and Glick, 1977). To date, most studies of mother-child interaction assume that the observed behaviors generalize across all situations, and do not attend to contextual factors which might effect the nature of that interaction. This is a particularly serious problem in interpreting comparative research, such as that of Hess and Shipman (1965), in which general claims were made about the linguistic style and interactive competence of Black mothers, based on their performance in an experimental laboratory situation. Here the context of elicitation was not taken into account.

A large body of evidence, however, particularly in the fields of social anthropology and sociolinguistics, shows that different contexts affect various aspects of behavior. Hymes (1962) suggests that social variation in speech use might be due to changes in such factors as the type of speech event being engaged in, the roles of the various parties, the topic of discussion, and style of discussion. Labov (1972) and Hall, et al. (1977) demonstrate that the presence of a particular audience may elicit different linguistic strategies from both adults and children. These data added to our growing suspicion that the presence of an experimenter and/or his equipment might be eliciting a display of particular sorts of behaviors which might not be generalizable across contexts.

We suspected that mothers have ideas about what behaviors are appropriate in a psychological experiment. The particular set of behaviors considered appropriate, however, may vary as a function of such factors as social class and educational level. There seems to be an implicit belief among the White middle class that a child must be actively provided with a social environment in order to develop social skills. Studies by Brazelton, et al. (1974) have shown that, in the experimental situation, mothers assume they are to play the role of socializer, constantly stimulating the child, despite reassurances from the experimenter that such intensive interaction is neither expected nor required. In fact, the children in these studies often reached a point of "oversaturation," when they were forced to turn away from their mothers to avoid the constant stimulation. We had

0160-3361/78/06-41 $1.00 ©The Rockefeller University Press

a further hunch that, for White middle-class mothers, the impression they wish to convey reflects their conception of the role of "good mother" and "good kid." It seemed plausible that the mothers' attempts to have their children perform continuously and effectively were part of a strategic "impression management" (Goffman, 1969), designed, at least partially, for the benefit of the observer. Outside of the experimental context, their behaviors might look quite different.

Therefore, we predicted that White, middle-class mothers would display different sorts of behaviors when they were aware of the presence of an experimenter than those displayed when they assumed they were not being observed. Furthermore, if attempts at producing "normative" behavior (involving the mothers' conception of an ideal socializing agent constantly stimulating the child toward peak performance) are elicited from mothers of this particular social class by the presence of the observer, one would expect maternal behavior to change in the direction of more instruction and greater activity. Differences in maternal behavior should be observable in both speech and action when an observer is present.

Preliminary analyses had indicated that several maternal behavior characteristics might differ across two situations—one in which mothers were aware of the observer and his recording equipment; and another in which they were *not* aware of being observed. The following differences were hypothesized:

1. *Quantity of speech.* Mothers who are actively attempting to maintain their child's attention, and trying to create and maintain interactive contact as much as possible, are expected to produce a greater quantity of speech when they are aware of the observer than when they are not.

2. *Mothers' mean length of utterance* (MLU). Mothers' MLU was expected to change as a function of the experimenter's presence. Newport (1976) proposed that "Motherese" (a speech form characterized by short, simple utterances and spoken with less dysfluency than speech to adults) is produced consistently when mothers talk to their young children.

Although there is no doubt that Motherese exists as a phenomenon and is used in certain contexts, it may not constitute the entire input to the child. Newport's data were tape-recorded by an experimenter in the home, so it seemed possible that the observer's presence was eliciting particular behaviors from the mother's repertoire. It may have been for that reason that her speech was simpler and showed a lower MLU. Because this shorter, simpler speech may, as Newport suggests, be easier for the child to process, it was predicted that mothers aware of being observed would show a lower MLU than when they were unaware of the observation, as it may aid the children in understanding and responding better in a condition where their performance is important to the mothers.

3. *Types of utterances.* We predicted that an analysis of the mother's speech would reveal differences in the types of utterances presented to the child in the two conditions. Some functional speech forms that we thought might show definitive changes in the two conditions were: indirect directives, test questions, evaluative comments, and repetitions. Use of these forms allows the mother to a) command the child without appearing overly demanding; b) elicit and display vocabulary she knows her child has already acquired; c) reinforce behaviors she would like her child to display; and d) give the child a second chance to acquire information from her speech stream, respectively.

4. *Gestures.* Shatz and Graves (1976) noted frequent gesturing on the part of the mothers to their children of language-learning age. It is possible that the presence of the professional observer might elicit "harder work" on the part of the mothers who are attempting to focus their child's attention on the activity at hand, and encouraging him to perform accurately and continuously. It was hypothesized, therefore, that mothers would produce more gestures, and perhaps different types of gestures, when they are knowingly observed, as compared with when observation is unobtrusive.

5. *Naming and action routines.* Naming and action routines are interactive sequences in which observed mother-child dyads seemed constantly engaged (Shatz and Graves, 1976) in preliminary tapes.

These routines may serve as a type of interactional blueprint, or social script, into which the mother could systematically insert any number of vocabulary words or physical tasks for the child to learn. In addition, the routines can act as a functional frame for the rehearsal of previously acquired knowledge, and the display of this knowledge on appropriate occasions.

It is probable that children will "look good" when engaging in naming and action routines, because they are familiar with the format and may also have practiced the material, so it was predicted that mothers would attempt to engage the children in these "games" more frequently when an observer was present than when one was not.

The following study was conducted in order to investigate these hypotheses. Mothers and children engaged in unstructured "free play," and the interactions were videotaped under the following two conditions: first, the mothers and children were aware that they were "on camera"; second, the observer was unobtrusive and mothers and children were unaware that they were being filmed.

Method

Subjects. Subjects consisted of a set of six mother-child dyads, all from White middle-class backgrounds. Children ranged in age from 18 to 25 months with MLU ranging from 1.1 to 1.7.

Materials. Two identical rooms at City University were equipped for the study in the following manner. Each room contained a toy of parallel complexity, a children's picture book, a set of blocks, a children's puzzle, a tumbling mat, a child-sized chair and an adult-sized chair. Both rooms were carpeted and one wall of the room contained a one-way mirror. In both rooms, several pictures were taped to the mirror to disguise its function. The videotape equipment used was a Panasonic port-a-pack camera.

Procedure. Prior to their arrival, mothers were told that the study involved videotaping of the mother-child dyad at play, and that the exact nature of the study and the variables being considered could not be disclosed until completion of the sessions. Each of the mothers agreed to these conditions.

Upon arriving, the mothers and children were shown the two rooms where the taping was to take place. Each subject pair was videotaped for fifteen minutes in each of the two rooms consecutively. In the "observed" condition, subjects were aware that they were being taped—the videocamera, manned by an experimenter, was set up just outside the doorway of the room. In the "unobserved" condition, subjects were not aware that they were being videotaped. Throughout the unobserved session, the experimenter appeared to be in the process of setting up equipment with a colleague outside the door to the experimental room, while subjects actually were being videotaped through the one-way mirror in their room.

The two experimental conditions were counterbalanced to control for sequence effects. Three subject pairs were filmed in the "unobserved" condition, followed by the "observed" condition (Sequence I); the other three subject pairs were videotaped in the reverse order of conditions (Sequence II—observed followed by unobserved).

Mothers were instructed to play as naturally with their children as they would at home, as if there were nobody around, in both the "observed" and "unobserved" conditions.

Results

Initially, speech, action, and gesture for both mother and child were transcribed directly from the videotapes.

Gross quantity analysis. First, the gross quantity of speech (the number of words and utterances produced by the mothers) was compared for each subject in the two conditions.

The results of this initial analysis were striking. The amount of speech produced by the mothers when they were aware of being observed was approximately double that produced in the unobserved condition in the same amount of time. This finding was confirmed for all six subject pairs (see Table I).

Mean length of utterance. There appeared to be a trend for the MLU to decrease in the observed condition as compared with that in the unobserved condition (see Table I). That is, utterances, as predicted, were generally shorter in the observed condition than in the unobserved condition. This finding was confirmed in five of the six subject pairs. However, probably because of the small sample size, this trend was not strong enough to be statistically significant.

Utterance types. Mothers' utterances were analyzed in two ways.

1. Sentence types (e.g., declaratives, imperatives, Wh-questions, etc.): It was found that mothers produced similar proportions of these utterance types in the two conditions. There were no significant differences in the distribution across contexts.

2. Functional categories: Mothers' utterances were also coded into four functional categories—*indirect directives, test questions, evaluative comments* (positive and negative), and *repetitions and repetition variations* of utterances within subject for each conversational turn (see Table I). As anticipated, con-

TABLE I
Data Analysis Averaged Across Mothers (N = 6)

	Word Count	*Utterance Count*	*Mothers' MLU*	*Indirect Directives*	*Test Questions*	*Positive Evaluations*	*Negative Evaluations*	*Repetitions & Variations*	*Gestures*
UNOBSERVED	441	111	4.0	6	7	5	6	22	19
OBSERVED	783	230	3.3	18	19	18	6	47	51

sistent and marked differences were found among these variables across the two conditions.

a. Indirect directives (e.g., "Can you put the elephant on top?" rather than the direct imperative, "Put the elephant on top.") — Differences were found in the number of indirect directives produced in the observed as compared with the unobserved condition. All mothers produced significantly more indirect directives in the observed condition as compared with the unobserved condition, although there were individual differences in the extent of this increase.

b. Test questions (e.g., "What color is the truck?" or "What does the doggie say?") — Similarly, mothers produced more test questions, for which the child is likely to have the answer, in the observed condition than in the unobserved condition.

c. Evaluative comments (e.g., "Very good, Johnny!" or "That's a no-no!") — Evaluative comments, both positive and negative, may serve as feedback for the children, providing them with information about what they are doing or saying right or wrong in a given situation, and whether to continue with a particular activity. For all mothers the total number of evaluative comments, particularly positive ones, increased in the "on-camera" condition.

d. Repetitions or repetition variations — Repetitions refer to a mother's exact repetition of her previous utterance (e.g., "Repetitions refer to mother's exact repetition of her previous utterance.") Repetition variations are slight variations of her previous utterance (e.g., "Variations are slight differences from her previous utterance"). Repetition might be a reflection of the amount of work mothers are doing to insure "pick-up" by the child of some elements of the speech stream and to elicit interaction. All six mothers produced more repetitions and repetition variations in the observed condition.

Mothers' gestures. Mothers' gestures were coded according to the six basic gestural categories noted by Shatz and Graves (1976): pointing, presentation, point-tapping, tapping, illustrations (including iconic actions), and touching. The gross quantity of gesture was noted, as well as the number of individual gestural types. The distribution of gestures was calculated as the percentage of the total number for each type.

The data supported the hypothesis that the amount of gesturing done by the mothers would increase in the observed condition (see Table I). One mother actually produced seven times more gestures in the observed than in the unobserved condition. In addition, mothers seemed to produce different types of gestures in the two conditions.

Naming and action routine—Sequence analysis. Preliminary analysis of interactive "game"-type sequences was performed on the data for two of the mother-child pairs (see Table II). The initiator of the interaction, either mother or child, was noted, as well as the length of the interaction sequence. In addition, instances where there was no joint focus of attention between mother and child were recorded. These were periods during which there was no mutual engagement in action, no mutual gaze, and no common verbal subject matter between mother and child.

Results were as follows: both of these mothers showed a large, significant increase in the percentage of time spent with no joint focus of attention in the unobserved condition, as compared with the observed. One dyad in the unobserved condition showed 53 percent of the recorded time with no joint interactive focus, as compared with only 22 percent in the observed condition. Similarly, the second dyad spent 60 percent of the time in the "off-camera" condition with no joint interactive focus, as compared with only 12 percent of the time in the observed condition. For both mothers, the number of naming and action routines increased radically in the observed, as compared with the unobserved, condition, as was predicted.

Further analysis of the additional videotapes is presently underway to confirm the findings of the "routine" sequence patterns based on these preliminary data.

TABLE II
Naming and Action Routines—Preliminary Analysis:
Data averaged across two mother-child pairs (N=2)

	Number of Naming Routines	*% of Total Interaction Time-Naming R.*	*Number of Action Routines*	*% of Total Interaction Time-Action R.*	*No Joint Focus of Attention*	*% of Total Interaction Time–No Joint Focus*
UNOBSERVED	9	18%	9	23%	13	59%
OBSERVED	20	33%	20	50%	7	17%

Discussion

The data analyzed in this study to date show strong evidence that the presence of an observer who is recording the ongoing mother-child interaction acts as a structuring audience and helps to define a setting in which particular sorts of behaviors are displayed.

When viewed together, the differences in behavior measured in the two conditions form very different global pictures of the nature of mother-child interaction. This is contrary to the consistency of these behaviors across contexts assumed by interactional studies to date.

When the mothers were unaware of the observer, they produced a moderate amount of speech. Some of this speech was directed to the child and some was simply musing to themselves. Their utterances were generally longer than the utterances produced when they were aware that their interaction was being watched. There were periods when many of the mothers appeared apathetic, lost in thought, or involved in a project of their own in the playroom. Repetition of an utterance for their child's benefit was infrequent, as was oral reward for doing a task well or producing a correct response. Mothers maintained a joint focus of attention with their children, and played naming and action games with them, less than half the time. There was a moderate number of test questions and indirect directives produced, as was needed to maintain this lower level of interaction. The mothers would try to elicit some interactions with the children, but with the seeming intent of keeping them occupied and quiet. For most mothers, gesturing in the "unobserved" condition was minimal. Often the child attempted to initiate an interaction with the mother, but would not receive an appropriate response. The child's request may even have been repeated several times, frequently supplemented by the addition of gesture and/or action, before a distracted mother would respond satisfactorily.

When this tableau is contrasted with that formed by the behaviors of these same mother-child dyads when they were aware of observation by the experimenter, the differences are striking. First, the mothers approximately doubled the amount of speech they had produced when unaware of the observation. Their utterances were a bit shorter in general length and almost entirely directed toward the child. Here, the mothers appeared to be working avidly at teaching their children new words and skills, and at getting them to display those they had already acquired. The elicitation of the child's performance was virtually continuous. Sometimes the video camera and the experimenter themselves were the subjects of discussion and investigation by mother and child. Mothers also frequently repeated their prior utterances, giving the child a better chance to process the language and respond appropriately, and the child often was reinforced with a positive evaluative comment when he answered a question correctly or performed a manual task with facility. Mothers in general seemed very involved in the interaction with their children when they were aware that they were being taped. Naming or action "games," where a joint focus of attention was maintained between mother and child, occurred nearly 85 percent of the time. Mothers often initiated interactions with their children intentionally, and were almost always responsive to the child's own attempts at initiating interaction.

These findings pose a problem for some of the work on mother-child interaction that attempts to characterize the nature of the interaction without consideration of the experimental social context which seems to affect behavior so radically. It must be taken into account that a psychological study where measurement is not unobtrusive, whether done in the home or in the laboratory, is a specialized context in which mothers try to put their "best foot forward" according to what they feel is appropriate and advantageous in this particular situation. Generalization from a limited set of observations about the overall developmental milieu for a child, or children in general, seems unwarranted. If one were to reexamine the Hess and Shipman (1965) data in light of this study, it would be interesting to note a similarity in the description of these White, middle-class mothers in the "unobserved" condition, and the lower-class Black mothers in the Hess and Shipman study, whose interactive style was said to be so inadequate. It may be the case that White middle-class mothers utilize some interactional strategies for displaying what they believe the experimenter would like to see—"good mothering." In cases when the experimenter shares a similar socioeconomic background, there may be a match between mothers' and experimenters' expectations of what should constitute this role. If backgrounds differ, so may expectations and strategies for dealing with the world. Whichever the case may be, all analyses of mother-child interaction need to attend to contextual variables that might be affecting the nature of the interaction displayed.

This study has pointed out the great flexibility and variability of behaviors in two situations in which the primary difference was the presence of a particular audience. It may be, in fact, that there are many such changes in context and audience throughout the course of a mother-and-child's day, and that numerous subsets of behaviors are deployed in these various circumstances. Future research will determine the extent to which this is the case. However, if a mother's behavior does change as a result of a change in context, this is a potential way that the child could begin to learn about the types of behavior that are expected and considered appropriate, both for himself and for adults, in the presence of these types of audiences. On a basic level, the shift in quantity and quality of maternal behavior with a change in context may be an initial indicator for the child that human interactional styles do vary, and that certain behavioral patterns are considered appropriate to particular situations, and not to others.

It is clear from this work that the presence of a known observer in a particular scene elicits certain patterned changes in behavior, at least for this small sample of White, middle-class mothers. When they are aware that their behavior is being attended to by a psychologist, although they may have no prior knowledge of the variables he is interested in considering, it seems that these mothers display a specific set of behaviors for his benefit, incorporating the beliefs about what constitutes good mothering and good performance on the part of the child. In the experimental context, the mother expects her child, as well as herself, to be a good performer—attending and responding correctly at the appropriate time. Mothers do much situational management in order to present this picture: they shape the interactional environment by asking their child familiar questions within a familiar format; they gesture profusely in an attempt to hold the child's attention and direct action and speech; they respond quickly and appropriately to the child's attempts at initiating interaction; and they work to maintain contact with the child throughout the interaction period. Therefore, it is important to note that analyses performed on observed mother-child interactions may not characterize the nature of this interaction across contexts.

In addition, these results lead to some interesting speculations about comparative differences. It may be that the direction of change in maternal behavior is related to the values a particular social class places on various child-rearing practices. The conception of the normative "ideal mother" role may, in fact, be quite different for members of different social classes. Perhaps lower-class mothers may not alter their behaviors—or at least not as radically, or perhaps in a different direction from White, middle-class mothers —when confronted with an "authority figure" in experimental and other situations. Different children clearly come into the classroom situation with different preparation and different sets of interactive strategies (Goodnow, 1972), so there may be some relation between their interactional heuristics and those displayed by their mothers in interactions with them in parallel situations. Thus, the way the mothers alter their behavior may affect the child's understanding that certain behavior changes are expected by teachers and/or experimenters with White, middle-class backgrounds in "formal" contexts, such as the classroom. This hypothesis seems worth exploring in future study.

ACKNOWLEDGMENT

Thanks to Virginia Valian and Jan Jewson for help and moral support.

REFERENCES

BRAZELTON, T. B., KOSLOWSKI, B., and MAIN, M. 1974. The Origins of Reciprocity: Early Mother-Child Interaction. In: M. Lewis and L. Rosenblum (Eds.), *The Effect of the Infant on Its Caregiver*. New York: Wiley.

COLE, M., and SCRIBNER, S. 1975. Theorizing about the socialization of cognition. *Ethos*, 3:250–268.

GLICK, J. 1977. Functional and Structural Aspects of Rationality. Presented at Jean Piaget Society Meetings, May 21, 1977.

GOFFMAN, I. 1969. *Strategic Interactions*. Philadelphia: University of Pennsylvania Press.

GOODNOW, J. 1972. Rules and Repertoires, Rituals and Tricks of the Trade: Social and Informational Aspects to Cognitive and Representational Development. In: S. Farnham-Diggory (Ed.), *Information Processing in Children*. New York: Academic.

HALL, W. S., COLE, M., REDER, S., and DOWLEY, G. 1977. Variations in Young Children's Use of Language. In: R. Freedle (Ed.), *Discourse Production and Comprehension*. Vol. 1. New Jersey: Ablex.

HESS, R., and SHIPMAN, V. 1965. Early experience and the socialization of cognitive modes in children. *Child Development*, 36 (4):869–886.

HYMES, D. 1962. The Ethnography of Speaking. In: T. Gladwin and W. Sturdevant (Eds.), *Anthropology and Human Behavior*. Anthropology Society of Washington, pp. 72–85.

LABOV, W. 1972. The Study of Language in its Social Context. In: P. Giglioli (Ed.), *Language and Social Context*. London: Penguin.

LEWIS, M., and ROSENBLUM, L. 1974. *Friendship and Peer Relations*. New York: Wiley.

NEWPORT, E. 1976. Motherese: The Speech of Mothers to Young Children. In: N. Castellan, et al. (Eds.), *Cognitive Theory*. Vol. II. New Jersey: Erlbaum.

SHATZ, M., and GRAVES, Z. 1976. The Role of Maternal Gesturing in Language Acquisition: Do Actions Speak Louder than Words? Paper presented at Boston University Conference on Language Development, October, 1976.

Permissible Inference from the Outcome of Training Studies in Cognitive Development Research

ANN L. BROWN and JOSEPH C. CAMPIONE
University of Illinois

Our aim in this paper is to provide a statement on the reasons for conducting cognitive training studies, and to comment on the interpretations of the outcome of such research. Extensive reviews of the training literature proliferate (Belmont, 1978; Belmont and Butterfield, 1977; Borkowski and Cavanaugh, 1979; Brown, 1974, 1975, 1977, 1978a; Brown and DeLoache, 1978; Campione and Brown, 1974, 1977, 1978; Butterfield, 1977), would suggest that our task should be a simple one. But the task is not simple, mainly because the rationale for conducting such studies, the problems encountered when one does, and the interpretations inspired by the outcomes, all reflect some of the most difficult and controversial issues of comparative research.

The Modal Training Study

One reason that a justification and explanation of training studies is needed is their very prevalence; this genre of research seems to be a basic enterprise of comparative cognitive psychologists, and it is charitable to presume that there is a reason for this uniformity. Consider the modal developmental study (Brown et al., 1978; Brown and DeLoache, 1978). Children of varying age are compared on a single task purported to tap some significant cognitive process. The typical, nonsurprising result (if the paper is published) is that the older child outperforms the younger participant. In the better examples of this genre, there is a somewhat more fine-grained analysis of the components of the task, enabling the authors to diagnose which processes are failing to function appropriately in the less-efficient learner. Other common variants of the situation are those that pit against each other comparative groups that differ not only in age but in terms of putative intelligence level or cultural milieu. The outcome of these endeavors is almost universally a demonstration of poorer performance and/or ineffective processing in the young, the slow, the deprived, the deviant, the non-Western, or, more recently, the unschooled samples.

The majority of comparative studies rest here; having demonstrated a developmental effect, some cognitive infirmity of the poorly performing group is inferred, and the research enterprise terminated. An intrepid minority, however, go farther and question the nature of the group differences. Traditionally, the further question has been posed in the form: Is the deficiency one of production or mediation? To restate the problem, the choice presented is between a "deficit" that is readily remediated or one that is not. Remediation can be effected by such intervention as direct instruction or changing task environments. If either is successful, the original problem is pronounced situationally specific and not due to a stable limitation of the learner. Alternatively, if the deficit appears to be impervious to intervention, it is attributed to a limitation of the learner, one that could be the result of either (or both) inadequate past experience or inherent cognitive potential. Regardless of the etiology, if the deficit is not easily removed, it is taken as an indication of a structural limitation of the organism, perhaps of a permanent nature, or at least diagnostic of current cognitive status (Brown, 1974; Campione and Brown, 1978). A variety of strategies exist that can be used to help distinguish between the oversimplified dichotomy, however, as "cross-situational, cross-cohort, cross-cultural" (Weisz, 1978) comparisons are the exception rather than the rule. The major interest of developmental psychologists concerned with the problem has converged on the training study.

The modal training study can be described as follows: The prerequisite is that the comparative groups of interest differ in terms of efficiency in the use of a particular cognitive process. Training designed to overcome the deficiency is provided for the originally deficient group only, and their performance is then compared to the standard set by the untrained older group. The outcome of such endeavors is encouraging in its stability. If the training is relatively extensive and based on reasonable task and subject analyses, satisfactory improvements in performance occur. In the best examples, retarded children's performance can be improved to the level set by untrained adults (Butterfield, et al., 1973). Again, with appropriate training, the effects of the intervention have been found to persist for at least a year (Brown, Campione, and Murphy, 1974; Brown, Campione, and Barclay, 1978). However, with retarded subjects, transfer to new variants of a training task is rarely found (Brown, 1978a; Campione and Brown, 1978). A similar statement can be made in regard to transfer of training studies conducted with schooled and unschooled populations in other cultures. A typical interpretation of these findings is that generalization of the effects of training across traditional task boundaries is not characteristic of retarded or deprived children in our society and unschooled peoples in non-Western society. It is our major contention in this paper that such a strong interpretation is illegitimate, for the data necessary to support it do not exist.

What would be needed before we could arrive at such a strong conclusion? Several initial steps would be necessary: (a) improvements in the design of training studies; (b) a reconsideration of the types of skills that are the subject of training; and (c) a serious consideration of the relation of the training and transfer tasks to the everyday functioning of the cognitive process under investigation. We will consider these requirements separately.

Basic Requirements of an Adequate Training Study

Earl Butterfield's (1977) detailed analysis of the steps needed to conduct the ideal training study impressed and somewhat daunted investigators in the field. We agree that it is indeed a difficult task to engineer adequate training that could result in practical improvements in performance and in relatively unambiguous theoretical interpretation of the outcome. We would like to add that the situation is even more complex than the Butterfield article would indicate if the aim of training is not only significant enhancement of performance on the initial task, but also durability and generality of the benefits of training. To be judged effective, training should improve the overall level of accuracy and the strategies used to attack the task, thereby permitting reasonable attribution of the component process(es) causing the original deficit. But, in addition, for both practical and theoretical reasons, effective training can be judged as such only if it results in long-term improvement in the subject's performance on both the training task and a class of situations that are similar to the training vehicle. To design an adequate training program that would meet these criteria, the following steps would have to be included.

Diagnosis of the original difference. Before instruction is even contemplated, preliminary steps must be taken to insure that the target of training represents a cognitive skill that proves an intransigent problem for the poor performer. To establish that this is, in fact, the case, it would be necessary to examine the spontaneous operation of the skill in more than one task and in more than one setting (Brown, 1978a). In addition, serious attempts should be made to insure that the poor performers are not simply serving the necessary role of convenient experimental foils (Brown and DeLoache, 1978; Gelman and Gallistel, 1978) by providing a baseline against which a developmental trend can be established. Quite simply, the young should not be faced with proving their ability on a task ill-designed to measure any competencies they may possess. Sufficient task design and analyses should be undertaken to make sure that the particular rules or strategies used by the young child can be detected, so that we do not end up with a total lack of information concerning what the younger subjects are doing when they fail to perform in an "appropriate" manner. Additional refinements that would be nice, if not absolutely necessary, would be that starting, intermediate, and end-state levels of competence could be identified so that instructional routines could be tailored to fit the diagnosis (Brown and DeLoache, 1978).

One last distinction concerning the identification of the initial learning problem concerns the type of knowledge that the child may lack. We refer here to the tenuous separation of task-specific knowledge and the general control of that knowledge. Ginsburg (1977) distinguished between the learning of principles and the learning of strategies. For example, one can acquire either the principle or the practice of addition or both. Knowing that you need to add is the principle, but knowing how to employ a counting routine effectively is the strategy. We have made similar distinctions between the general and specific results of training, a distinction that has been placed in the context of the memory-metamemory literature (Brown, 1978a). For example, in a traditional memory task, one can make the distinction between knowing that one should introduce an active strategy (the principles of being strategic or engaging in self-regulation) and knowing that a particular variant of a rehearsal strategy would be the most helpful. Assessing differences in these types of knowledge would have important implications for what kind of training might be designed. For example, faced with a task in which rehearsal is required, the nonrehearser could be in one of several states of ignorance. He could lack the principle that he must be strategic at all, or he could lack the principle that he should rehearse. Even if he has been known to rehearse in the past, he may lack sufficient mastery over this task-appropriate strategy for a variety of reasons, including: not knowing how to rehearse effectively; not recognizing the new task as one that demands rehearsal; or not knowing how to modify his existing strategy to meet the changed surface structure of the new task. Simply to designate a subject in any one of these dilemmas as a nonrehearser is an inadequate diagnosis of his starting state of knowledge. A more fine-grained initial subject analysis would greatly inform the subsequent task analysis undertaken to design training. A very different form of training would be prescribed for subjects suffering from the varying kinds of ignorance.

Training. The design of a training routine should be based on adequate task, situation, and subject analyses, a point that has been made many times and a practice that has been shown to be successful in a variety of domains (Brown, 1970, 1974; Butterfield et al., 1973; Campione and Brown, 1974; Gelman and Gallistel, 1978). This literature has been reviewed adequately, and we have no space to reiterate it here. One point about training studies that has not received attention, however, is that instruction is rarely provided for the originally adequate. Given that the aim is often a practical one of enhancing the performance of the inefficient, the common practice of training only those in need is obviously economically defensible. But there are interesting theoretical questions that remain unanswered by this procedure, and the simple expedient of also providing training for a subset of the originally competent serves a useful function. Sometimes the tactic established convincingly the exact nature of the original problem. For example, Brown, Campione, and Gilliard (1974) attributed the superiority of older children on a judgment-of-recency task to their ability to capitalize on the presence of relevant background cues that could serve as temporal anchor points. Training an anchor-cue utilization strategy completely removed the developmental effect: trained subjects, young and old alike, performed identically, whereas untrained older children out-performed untrained younger children. In other cases, however, individual or group differences are of the same magnitude before or after intervention (Huttenlocher and Burke, 1976; Lyon, 1977). We believe there are nontrivial theoretical reasons to expect these differences that pertain to the nature of the strategic components of the tasks and the concept of individual differences in intelligence (Campione and Brown, 1978).

Another training tactic that can clarify some problems of interpretation is one that has been employed successfully by two of the major training programs (Brown, et al., 1973; Belmont and Butterfield, 1977). In order to pin down the culprit responsible for an initial performance decrement, one both prevents the originally efficient from engaging in the activity putatively responsible for their effectiveness and induces the poor performers to adopt the appropriate activity. If these manipulations result in comparable patterns of performance, one can be more confident that the original diagnosis was correct.

Maintenance and generalization. Given that one can induce adequate performance on the training task, it is then desirable to show that the benefits of the instruction are durable and generalizable. By durable we mean that the effects of training will persist in the absence of intrusive aids when minor changes in surface structure (e.g., stimulus materials) take place and over significant time intervals. But more than that, successful training must result in appropriate generalization, i.e., transfer to tasks where the new behavior would obviously be helpful, but not to situations where its use would be less than optimal. Generalization involves more than just the use of a trained behavior; it involves discrimination of appropriate and inappropriate settings: the learner must know where and how to transfer and when not to (Brown, 1978a; Campione and Brown, 1974, 1978; Ginsburg, 1977).

Given the proliferation of evidence suggesting that such flexible generalization of training is rarely the outcome of cognitive training studies, are we justified in concluding, as many have, that failure to generalize is a defining feature of the cognitive apparatus of retarded children? We think not, if only because the appropriate tests have yet to be made. Despite the great deal of interest, there has been a dearth of informed experimentation or observation of the conditions that would promote flexible learning. However, we do have enough experience to provide guidelines that we believe would result in appropriate transfer.

Let us begin with the simplest form of transfer failure—the situation in which the child does not even maintain the new behavior unless specifically prompted to do so (Brown, 1978a; Brown, Campione, and Barclay, 1978). One explanation of this finding is that the child simply does not appreciate that the trained strategy improved performance. If it is true that young children are totally unaware of the utility of a strategy, why should they benefit from instruction? If trained to rehearse, they will rehearse only if the situation remains unchanged and they receive continual reminders. But why should they be expected to maintain the behavior, let alone use the new skills insightfully, if the reason for the activity was never made clear? Young children will maintain the effects of training, however, if they can be made aware of the usefulness of the trained strategy (Kennedy and Miller, 1976), and a clear first step in designing training studies should be to include *explicit feedback of the new behavior's effectiveness.*

The aim of training is to achieve not only maintenance but generalization. A major impediment was pointed out by the Soviet psychologist Shif (1969), who identified a consistent problem with training retarded children, i.e., the effect of training seemed to be "welded" to the situation in which it was provided. Given that there are ample examples of such welding in the literature, we suggest that direct attempts to overcome the problem should be initiated. *Training in multiple settings* appears to be the answer. If the learner is instructed in the use of a strategy in more than one appropriate setting, this should reduce the tendency for strict situationally specific learning to occur (Brown, 1974, 1978a; Campione and Brown, 1974). There is some preliminary evidence that the tactic will be successful (Belmont and Borkowski, 1978).

The very young child in a training study is often unaware that the separate phases of the experiment are indeed related (Bullock and Gelman, 1977; Campione, 1973). It is not surprising, therefore, that the utility of a prior solution or strategy to a new variant of the training task does not occur to them spontaneously. But, if the aim of training is flexible generalization, then it seems reasonable to include in the training package *direct instructions concerning generalization.* For example, one could tell the child that the trained behavior could help him on a variety of similar tasks and that the trick is to know which ones. The child could then be exposed to a variety of prototypical tasks (those on which experienced users of the strategy would always attempt to use it), and the utility of the strategy in such situations could be explained and demonstrated. At that point, inappropriate tasks could be considered, and the reason why the trained behavior would not be helpful could be pointed out. Finally, the child could be presented with a generalization test containing new prototypic and inappropriate tasks and his intelligent/unintelligent application of the strategy examined. As far as we know, this procedure has not been attempted, but we are currently investigating the feasibility of such direct generalization training, which incorporates the suggestions made above, together with direct instruction in modifying an overlearned strategy to fit changes in the exact specifications of the new tasks.

Another crucial decision for those who would undertake training studies is the choice of both the training and transfer tasks. It is obviously necessary that the tasks should be related in such a way that efficient users of the cognitive skill in question would adopt it spontaneously on both occasions (Brown, 1978a; Brown, et al., 1977). The tasks must evoke the same underlying cognitive process, but must differ sufficiently in surface structure to provide a reasonable test of generalization. Many of the "successful" training studies have employed transfer tests that vary so little from the training vehicle that one can dispute whether it really is generalization they have achieved or simply maintenance. The investigator must justify the *distinction between maintenance and generalization* on at least practical, and preferably theoretical, grounds.

An additional requirement is that the investigator be able to distinguish between *different potential causes of transfer failures*, for a failure to perform adequately on transfer could be due to the trainees' inability to see the relation of the trained behavior to the new task, to the usual interpretation, or to his inability to execute some other component of the transfer task which neither they nor the investigator fully appreciate. In an early series of studies addressed to identifying transfer mechanisms in a discrimination learning paradigm (Campione and Brown, 1974), we discussed a number of possible reasons for failing to observe transfer. One, termed the interaction-of-components hypothesis, was based on the view that subjects might, in fact, generalize the trained component to a transfer task but be unable to demonstrate the generalization due to a weakness with some other component of the transfer task. As a simple example, an attempt to generalize a trained rehearsal strategy to a new task would be hidden if the subjects could not come up with names or labels for the stimuli. Unless the experimenter is aware of these additional components, it is impossible to interpret transfer failures. In the 1974 paper, we also provide some data consonant with the interaction of components hypothesis. Therefore, our final prescription for designing training is that sufficient task analyses be instituted to enable an unambiguous attribution of failures to the proper cause.

In summary, we have suggested that the basic requirements of an adequate training program must include: 1) careful selection of the cognitive skill to be examined; 2) sensitivity to the actual beginning competence of the learner; 3) stringent analyses of the requirements of the training and transfer tasks so that transfer failures may be interpreted properly; 4) training in multiple settings to alleviate the problem of welding; 5) direct feedback concerning the effectiveness of the trained skill; and 6) direct instruction concerning the generalization of the trained skills. Until a research program can provide answers to these basic questions, it would be premature to conclude that retarded children can, under no circumstances, show flexible effects of training.

Reconsideration of the Skills Trained

Specific versus general skills. Given the impressive evidence of young children's general problems with self regulation and control of their goal-directed activities (Brown, 1975, 1978a; Brown and DeLoache, 1978; Meichenbaum, 1977; Mischel and Patterson, 1976), and the less than optimistic forecast concerning the benefits of training specific cognitive strategies, we believe that it would be judicious to re-think the types of skills that are the subject of intensive training. If it is true that young children in general, and slow-learning children in particular, experience major problems when required to orchestrate and regulate their own attempts at strategic intervention (Brown, 1978a, Brown and DeLoache, 1978; Campione and Brown, 1977, 1978), then an alternative approach to training specific mnemonics would be to train the metacognitive skills that provide the most pronounced difficulties for the immature learner (Brown, 1978a; Brown, et al., 1977). Metacognitive skills, such as checking, planning, asking questions, self-testing, and monitoring ongoing attempts to solve problems are characteristically lacking in retarded children's laboratory learning performance (Brown, 1978a). In addition, these skills are, by their very nature, general ones, in that they apply in a wide variety of problem-solving situations. Furthermore, they are the very control processes implicated in breakdowns of effective transfer of training (Brown, 1974, 1978a; Campione and Brown, 1974, 1977, 1978). As such, the rationale for directing training attempts at metacognitive skills is persuasive.

There is another reason why training attempts directed at general skills might be more likely to result in transfer. One problem with specific skills is that they are just that—specific to a very small class of situations. For learners to generalize the effects of instruction in the use of specific routines, they would have to be able to discriminate the situations in which the routine would be appropriate from those in which it would not. Adequate generalization requires both extended use in novel situations and decisions not to use the trained routines in other situations where it would not be beneficial (Brown, 1978a; Campione and Brown, 1974, 1978). In the case of truly general skills, this discrimination should not be necessary, as the skill or routine could simply be used in a whole battery of problem-solving situations without regard to any subtle analysis of the task being attempted. In this sense, "general metacognitive skills" might be regarded as easy ones to instruct, or at least the most likely to lead to transfer across task boundaries; and it makes sense to begin training with skills which are the simplest.

The types of general problem-solving skills in which we are particularly interested would be the ability to stop and think before attempting a problem; to ask questions of oneself and others to determine if one recognizes the problem; to check solutions against reality by asking not "Is it right?" but "Is it reasonable?"; to monitor attempts to learn to see if they are working or are worth the effort (Brown, 1978a, 1978b). There are serious problems associated with this position, for it is certainly easier to suggest that training should be aimed at showing children "how to organize their knowledge" and "how to solve problems," etc. (Norman, et al., 1976) than it is to incorporate these suggestions in concrete training programs. But, in view of the dismal failure, so far, to induce generalization in slow-learning children, an attempt to address the problem of training general skills seems to be timely.

We have some preliminary evidence that training such general skills will be effective in producing transfer, both from our own laboratory and from a consideration of behavior modification programs aimed at teaching self-instructional routines (Meichenbaum, 1977). The task presented to retarded adolescents was one of estimating their readiness to be tested for rote recall of a list of picture names. We trained a stop-test-and-study routine where the test element was anticipation. Trained students performed more effectively and maintained their superiority over a one-year retention interval. They also transferred the trained routine to a task demanding the same general tactic but differing in surface structure, i.e., a prose recall task where the subject was required to retain the gist of the passage. In comparison to a pair of control groups, the trained students were observed to engage in more overt self-checking behavior, took more time studying, recalled more idea units from the passage, and, further, their recall was more clearly related to the thematic importance of the constituent idea units, a pattern characteristic of developmentally more advanced students (Brown, et al., 1978).

Thus, our major success to date came when we turned to a very simple self-testing routine. We are currently examining the generality of the effects of teaching children to self-interrogate when faced with a certain class of problems (instructions, math problems, a laboratory task, etc.). The type of self-interrogation which we think might work is to provide the child with a routine set of questions to ask himself before proceeding, e.g.: (a) Stop and think! (b) Do I know what to do (i.e., understand the instructions, both explicit and implicit)? (c) Is there anything more I need to know before I can begin? and (d) Is there anything I already know that will help me (i.e., is this problem in any way like one I have done before)? It is with self-instructional routines such as these that those engaged in behavior-modification training and classroom management have achieved their major success in inculcating generalized improvement (Meichenbaum, 1977; Stokes and Baer, 1977).

Transfer appropriate processing. We would like to point out that the preceding discussion applies principally to situations in which the child must acquire and transfer mechanisms of learning that lead to rote recall, or at least gist recall, of material. But this is not the only form of learning, for a fundamental form of learning can be described as "coming to understand something that one had not previously understood" (Bransford and Nitsch, 1976; Brown, 1978b). Bransford's concept of transfer-appropriate processing has much in common with a major tenet of Soviet psychology—that activities are purposive and goal directed. An appropriate learning situation, therefore, is one that is compatible with the desired end-state, and one cannot decide upon appropriate training unless one considers the question: "Appropriate for what end?" (Brown, 1978b). Traditionally, learning studies have relied almost exclusively upon accuracy of memory in assessing the success of learning, but the practice can lead one to neglect some important aspects of learning that are necessary for valuable kinds of transfer. Knowledge in a form that permits optimal memory need not be in an optimal form to be used to understand a novel input. For example, Nitsch (1978) investigated the effectiveness of several study activities for helping subjects learn a set of new concepts. The task consisted of having subjects learn a definition and several examples of a number of concepts. Two of the several study activities were memorization of the definitions or identifying the list of examples. While the group attained equivalent levels of learning as assessed by a memory test of the original examples, they differed reliably in identifying novel examples of the concepts which were introduced in the transfer phase. Thus, in the former case, the study activities appear equivalent, whereas, in the latter case, a clear advantage accrues to the condition where identifying examples was the main activity in the study period. These studies, and others from Bransford's laboratory, illustrate that the optimal learning situation can only be determined relative to the uses to which the acquired knowledge will subsequently be put. Investigation of the compatibility of the learning situation for a variety of subsequent uses would go a long way toward clarifying our knowledge about appropriate training.

Relations of Trained Skills to Real-Life Experience

As a result of both our success and failure in attempting to train retarded children to perform more effectively on common problem-solving tasks, we suggest that the types of cognitive activities which are most suitable for intensive intervention should have certain properties: (a) they should be applicable in a wide range of situations; (b) they should readily be seen by the child to be reasonable activities that work; (c) they should have some counterpart in real-life experience; and (d) their component processes should be well understood, so that effective training techniques can be devised. In this final section, we would like to address (c), the problem of real-life counterparts. This discussion will be brief, because we have run out of space and because detailed discussions of this topic have been featured in recent editions of this *Newsletter* (Cole, Hood, and McDermott, 1978; Ginsburg, 1977; Rogoff, 1978). Here we would just like to confirm the complexity of the problem as illustrated by these prior papers, and to stress that the problem of ecological validity is rarely considered in discussion of the cognitive performance of retarded children (Brown, 1978a).

The issue can be addressed at two levels. First, there is the problem of the cultural context of learning and the design of cognitive tasks. Consider an example. Lave (1977) examined transfer of arithmetic and measurement skills by Liberian tailors with varying amounts of tailoring experience and varying degrees of formal education. Transfer problems were formally equivalent in terms of the underlying computational requirements, but were presented in a surface format that was appropriate either to the familiar context of tailoring or to the context of formal education. Experience was significantly related to transfer efficiency. Tailoring experience led to success across tailoring problems, whereas experience with formal education led to transfer across traditional school problems. Transfer operates within limited contexts and depends on experience and task familiarity. This kind of detailed attention to the conditions promoting generalization has rarely been lavished on the disadvantaged child in our society. But in order to understand fully the retarded student's general failure to be strategic, or transfer learning, on tasks that represent formal educational enterprises, it is necessary for us to consider the processes of learning within realms with which they are experienced (Brown, 1978a).

This problem brings us face to face with the whole issue of ecological validity and the comparative research enterprise (Cole, Hood, and McDermott, 1978) and a level of complexity above the mere problem of providing familiar contexts of learning. As Cole et al. point out, there are serious problems with the assumption that what is important concerning cognitive skills can be identified in the laboratory and then observed in real-life situations, where a familiar variant of the skill operates naturally. It is entirely possible that laboratory tasks are related in only trivial ways to the common exercises of intelligent adaptation. If so, one would need to examine the learning process in cultural manifestation, with which cognitive psychologists themselves are unfamiliar, where the very identification and classification of learning activities would prove troublesome and measurement problems would be severe. Yet, in spite of the complexity of the problem, an awareness of these issues must lead us to consider learning in natural contexts, as well as in laboratory settings, and thereby inform theories of learning in ways that would be precluded by a total reliance on standardized tests and laboratory tasks.

References

Belmont, J. M. 1978. Individual Difference in Memory: The Case of Normal and Retarded Development. In: M. Gruneberg and P. Morris (Eds.), *Aspects of Memory.* London: Methuen.

Belmont, J. M., and Borkowski, J. G. 1968. Instructing Retarded Children on Two Examples of a Memory Method (vs. only One) Improves Maintenance and Generalization of that Method. Paper presented at the Eleventh Annual Gatlinburg Conference on Research in Mental Retardation, Gatlinburg, Tenn.

Belmont, J. M., and Butterfield, E. C. 1977. The Instruction Approach to Developmental Cognitive Research. In: R. V. Kail, Jr. and J. W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition.* Hillsdale, N. J.: Erlbaum.

Borkowski, J. G., and Cavanaugh, J. C. 1979. Maintenance and Generalization of Skills and Strategies by the Retarded. In: N. R. Ellis (Ed.), *Handbook of Mental Deficiency: Psychological Theory and Research,* Second edition. Hillsdale, N. J.: Erlbaum, in press.

Bransford, J. D., and Nitsch, K. E. 1976. Coming to Understand Things that We Could not Previously Understand. Address presented at NICHD conference on "Implications of basic speech and language research to the school and clinic," Bethesda, Md.

Brown, A. L. 1970. Subject and experimental variables in the oddity learning of normal and retarded children. *American Journal of Mental Deficiency,* 75: 142–151.

Brown, A. L. 1974. The Role of Strategic Behavior in Retardate Memory. In: N. R. Ellis (Ed.), *International Review of Research in Mental Retardation*. Vol. 7. New York: Academic.

Brown, A. L. 1975. The Development of Memory: Knowing, Knowing about Knowing, and Knowing how to Know. In: H. W. Reese (Ed.), *Advances in Child Development and Behavior*. Vol. 10. New York: Academic.

Brown, A. L. 1977. Development, Schooling and the Acquisition of Knowledge about Knowledge. In: R. C. Anderson, R. J. Spiro, and W. E. Montague (Eds.), *Schooling and the Acquisition of Knowledge*. Hillsdale, N. J.: Erlbaum.

Brown, A. L. 1978a. Knowing when, where, and how to Remember: A Problem of Metacognition. In: R. Glaser (Ed.), *Advances in Instructional Psychology*. Vol. 1. Hillsdale, N. J.: Erlbaum.

Brown, A. L. 1978b. Theories of Memory and the Problems of Development: Activity, Growth, and Knowledge. In: L. Cermak and F. I. M. Craik (Eds.), *Levels of Processing in Memory*. Hillsdale, N. J.: Erlbaum.

Brown, A. L., Campione, J. C., and Barclay, C. R. 1978. Training Self-Checking Routines for Estimating Test Readiness: Generalization from List Learning to Prose Recall. Technical report, Center for the Study of Reading, University of Illinois.

Brown, A. L., Campione, J. C., Bray, N. W., and Wilcox, B. L. 1973. Keeping track of changing variables: Effects of rehearsal training and rehearsal prevention in normal and retarded adolescents. *Journal of Experimental Psychology*, 101: 123–131.

Brown, A. L., Campione, J. C., and Gilliard, D. M. 1974. Recency judgments in children: A production deficiency in the use of redundant background cues. *Developmental Psychology*, 10: 303.

Brown, A. L., Campione, J. C., and Murphy, M. D. 1974. Keeping track of changing variables: Long-term retention of a trained rehearsal strategy by retarded adolescents. *American Journal of Mental Deficiency*, 78: 446–453.

Brown, A. L., Campione, J. C., and Murphy, M. D. 1977. Maintenance and generalization of trained metamnemonic awareness in educable retarded children. *Journal of Experimental Child Psychology*, 24: 191–211.

Brown, A. L., and DeLoache, J. S. 1978. Skills, Plans and Self-Regulation. In: R. Siegler (Ed.), *Carnegie-Mellon Symposium on Cognition*. Hillsdale, N. J.: Erlbaum.

Bullock, M., and Gelman, R. 1977. Numerical reasoning in young children: The ordering principle. *Child Development*, 48: 427–434.

Butterfield, E. C. 1977. On Studying Cognitive Development. In: G. P. Sackett (Ed.), *Observing Behavior*. Vol. 1: *Theory and Application in Mental Retardation*. Baltimore: University Park Press.

Butterfield, E. C., Wambold, C., and Belmont, J. M. 1973. On the theory and practice of improving short-term memory. *American Journal of Mental Deficiency*, 77: 654–669.

Campione, J. C. 1973. The generality of transfer: Effects of age and similarity of training and transfer tasks. *Journal of Experimental Child Psychology*, 15: 407–418.

Campione, J. C., and Brown, A. L. 1974. The Effects of Contextual Changes and Degree of Component Mastery on Transfer of Training. In: H. W. Reese (Ed.), *Advances in Child Development and Behavior*. Vol. 9. New York: Academic.

Campione, J. C., and Brown, A. L. 1977. Memory and Metamemory Development in Educable Retarded Children. In: R. V. Kail, Jr. and J. W. Hagen (Eds.), *Perspectives on the Development of Memory and Cognition*. Hillsdale, N. J.: Erlbaum.

Campione, J. C., and Brown, A. L. 1978. Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence*, in press.

Cole, M., Hood, L., and McDermott, R. P. 1978. Concepts of Ecological Validity: Their Differing Implications for Comparative Cognitive Research. *Quarterly Newsletter of the Institute for Comparative Human Development*, 2: 34–37.

Gelman, R., and Gallistel, C. R. 1978. *The Young Child's Understanding of Number: A Window on Early Cognitive Development*. Cambridge, Mass.: Harvard University Press.

Ginsburg, H. 1977. Some Problems in the Study of Schooling and Cognition. *Quarterly Newsletter of the Institute for Comparative Human Development*, 1: 7–10.

Huttenlocher, J., and Burke, D. 1976. Why does memory span increase with age? *Cognitive Psychology*, 8: 1–31.

Kennedy, B. A., and Miller, D. J. 1976. Persistent use of verbal rehearsal as a function of information about its value. *Child Development*, 47: 566–569.

Lave, J. 1977. Tailor-made experiments and evaluating the intellectual consequences of apprenticeship training. *Quarterly Newsletter of the Institute for Comparative Human Development*, 1: 1–3.

Lyon, D. R. 1977. Individual differences in immediate serial recall: A matter of mnemonics? *Cognitive Psychology*, 9: 403–411.

Meichenbaum, D. 1977. *Cognitive-Behavior Modification: An Integrative Approach*. New York: Plenum.

Mischel, W., and Patterson, C. J. 1976. Substantive and structural elements of effective plans for self-control. *Journal of Personality and Social Psychology*, 34: 942–950.

Nitsch, K. E. 1978. Structuring Decontextualized Forms of Knowledge. Unpublished Ph.D. thesis, Vanderbilt University.

Norman, D. A., Gentner, D. R., and Stevens, A. L. 1976. Comments on Learning Schemata and Memory Representation. In: D. Klahr (Ed.), *Cognition and Instruction*. Hillsdale, N. J.: Erlbaum.

Rogoff, B. B. 1978. Spot observation: An introduction and examination. *Quarterly Newsletter of the Institute for Comparative Human Development* 2: 21–26.

Shif, Z. I. 1969. Development of Children in Schools for the Mentally Retarded. In: M. Cole and I. Maltzman (Eds.), *A Handbook of Contemporary Soviet Psychology*. New York: Basic Books, pp. 326–353.

Stokes, T. F., and Baer, D. M. 1977. An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10: 349–367.

Weisz, J. R. 1978. Transcontextual validity in developmental research. *Child Development*, 49: 1–12.

Valid Assessment of Complex Behavior: The TORQUE Approach

Judah L. Schwartz
Massachusetts Institute of Technology and *Project TORQUE, Education Development Center*

Edwin F. Taylor
Project TORQUE, Education Development Center

Ordinarily, when we say a test is valid, we mean the test, in fact, tests what it intends to test. Over the years, the notion of validity has been discussed often in the psychological literature. Many different kinds of validity have been defined. Typically, each engages one facet of what is ultimately the only meaningful validation question, i.e.: Does the test elicit the same behavior as would the *same* task embedded in a *real*, noncontrived situation?

Reality is a difficult notion. Both physics and the psychology of perception tell us that human interactions with the real world are usually—some would say exclusively—mediated by models that inform the vision of the beholder. Further, even to speak of the *same* task across contexts requires a model of the structure of the task. In the absence of such a model, one does not know where the equivalence lies.

Now suppose that the task to be tested has been analyzed and a model developed. Is not the validation procedure now obvious? One simply compares performance on the test in question against real activities which have the corresponding tasks embedded in them. The model of the task is employed to analyze and shape the comparison. Consistency of performance between test and real activities means that the test is valid. Lack of consistency means that the test is invalid. An invalid test must either be revised in light of the model of the task or a new model sought or perhaps both. Iteration then improves the test until some reasonable level of validity is attained. End of story.

Or it would be the end of the story except that the validation procedure outlined above rarely works in practice. Everyday experiences that have embedded within them the tasks we are interested in testing are most often too specific and particular to provide a comprehensive assessment of all the subskills of the task. For example, suppose a test-maker has written a test for the skill of measuring length and wishes to validate it against length measurements in real life. One way of doing this is to follow a child around until he or she is observed to measure the length of something spontaneously and see how well he or she does it. Of course, this will not happen very often in the course of a day for most children. Even when it does happen, single measurements of length will rarely make use of more than a few of the many skills involved in the general ability to measure lengths. Did the child attend to a pertinent property? Was the measured length a fraction of the basic unit of length or was it an integer multiple of the unit or was it longer than the ruler? One would like to examine a range of measurements that spans these cases and perhaps others suggested by the model of the task. Relying exclusively on the strategy of validating against real, highly embedded experiences does not provide this coverage.

One way of assuring coverage of the different subtasks of measurement is to devise a set of contrived measurements in advance and ask the child to do them. Children will tolerate some of this, but lengthy sessions of such tasks are boring, and frequently children perform in a perfunctory and superficial manner in order to be rid of pestering adults as quickly as possible.

Our procedure for validating tests is to combine a few "real" tasks with games that the children enjoy playing and that require the skills for which the tests are being validated. The games concentrate the occasions for engaging the subskills of the task as envisioned by the model, while maintaining the enthusiasm of the children so that they perform as well as they are capable. The combination of games and real measurements constitute the validating activities with which performance on the tests is compared.

Lest this all sound too Utopian, we describe below one case in which the procedure has actually been applied and resulted in the development of a valid test.

Project TORQUE (Tests of Reasonable Quantitative Understanding of the Environment) is developing alternative strategies and instruments for the assessment of children's mathematical competence in grades K to 6. The somewhat abstract formulation of the notion of test validity presented above is, in fact, the one employed by Project TORQUE in developing its tests. In order to make the discussion concrete, we believe it would be helpful to highlight the process of developing and validating one particular TORQUE test, i.e., a test of the measurement of length used as an example above.

The task of measuring extensive magnitudes in general and length in particular was analyzed and a general model of the measurement process evolved. According to this model, the major steps in making a measurement are:

1) identifying the attribute to be measured (e.g., *length* of the *longest side* of *which triangle*);
2) choosing an appropriate unit (judging appro-

 · 0160-3361/78/06-54 $1.00

priateness of units amounts to a preliminary estimate of the result);

3) carrying out the mechanics of measuring (e.g., lining up the starting point on the scale with the one end of the object, etc.);

4) judging a level of precision appropriate to the context in which the problem is posed;

5) reporting the results (e.g., correct use of fractions, decimals, units, etc.).

While this model applies in a rather general way to the measurement of all extensive magnitudes, some modifications are necessary when applied specifically to any particular measurable magnitude. In the case of length measurement, the identification of the attribute tends to be a rather trivial issue. In contrast, in the measurement of area the identification of the attribute is far and away the most difficult and subtle part of the task. On the other hand, the measurement of length contains a fair amount of internal structure in the subtask of scale-reading, whereas the measurement of area requires little scale-reading beyond that of length. What we see here is the manifestation of a rather general result: different measurement tasks will apply different pieces of the model with different degrees of sensitivity and different degrees of resolution.

After an initial analysis of the measurement task, a collection of games and other activities involving length measurement were devised collaboratively with children and classroom teachers. The games were deliberately fashioned to play the role of the surrogate activities discussed above, i.e., settings that could concentrate the occasions for the appearance of the fine structure of the skill we wished to probe. The games that seemed most appealing to children and most useful in eliciting children's measuring behavior were given to classroom teachers to try out on their own, well away from the enthusiasm of their developers. At this stage, perhaps 30 to 40 teachers and 1,000 children were involved.

Only after this group of teachers reported back to us about which games were successful with the children did we undertake to design a test of the measurement of length that could be validated against these games and other activities. A one-page written test was fashioned, and the process of validating the test against the games and other activities begun.

The test does not encompass all forms of length measurement. The distance to the sun, the circumference of the earth, and the diameter of an atom are all lengths whose measurement is important. We limited ourselves to the measurement, using rulers, of lengths of objects comparable to the size of parts of the human body. This limitation, we felt, involved instruments easy for children to understand and manipulate, measurements of the kind they could expect to encounter in everyday life, and skills similar to those already being taught in schools.

On the basis of our accumulated experience with children's length measurement, we identified those portions of the measurement model that were appropriate to elementary school children. We took these to be identification of the attribute and scale-reading. Scale-reading, in turn, has a fine structure which we felt it important to observe in some detail. The subskills we examined accounted for the overwhelming majority of difficulties children had with length measurement. These were:

1) identifying the attribute;

2) *a* using appropriate starting point;

b reading the scale for an integer length shorter than the ruler;

c reading the scale for an integer length longer than the ruler;

d reading the scale for a noninteger length.

These five headings shaped our observations and recording of the children's behavior. Our procedure was to ask the children to make four real-length measurements in the classroom; then to take a version of the test; to play a game involving the measurement of length; to go through a small work booklet involving the measurement of length; and finally to take an alternative version of the test.

The validation proceeded in the following manner: through previously established contacts with teachers and administrators in various school systems, TORQUE arranged to visit third-, fourth-, and sixth-grade classrooms. Typically, four staff members would appear at mid-morning with a box of materials. Each staff member (rather than the teacher) would select two children randomly and sit down with them at a table to one side of the on-going classroom activities. The staff member would explain to the children that we were making up tests and were trying to discover with their help whether the tests were any good. We told them that we would be taking notes on observation sheets as they did the test and played some games, that we might not answer all of their questions at the time they asked them, but when the session was over, we would answer all questions and talk with them about what we had learned.

Validation sessions typically lasted about one hour. The total number of children involved was 71, approximately equal numbers of Anglo, Black, and Chicano children in the greater Boston and Los Angeles areas. Three-quarters of the children were 9 and 10 years old and the others 7 to 8 and 11 to 12 years old.

Each observation sheet, the completed written test, and the work booklets carried as full an account as we could manage of the behavior observed while children took the test and performed the validating activities. This account was organized under the five subskills to which we wished to pay particular attention. We distilled these observations into a series of numbers, each one representing the fraction of the opportunities for demonstrating a subskill carried out successfully by the child. Proceeding in this way, we derived, for each child, a series of performance fractions for each of the subskills and for each of the tests and validating procedures. The fundamental criterion of validation was how well each child's performance on the tests compared with that child's performance on the validating activities for each subskill.

For each child and on each subskill, a graph of performance fraction vs. pre-test, validating activities, and post-test was constructed. An example of the summary graphs of child 9004 is shown in Figure 1. Thus, for example, there were seven opportunities to measure a noninteger (number of centimeters) length in the games, seven opportunities to display correct use of the starting point on the ruler (one measures from zero and counts from one) in the test, etc. The Figure shows these summary graphs for all five subskills validated.

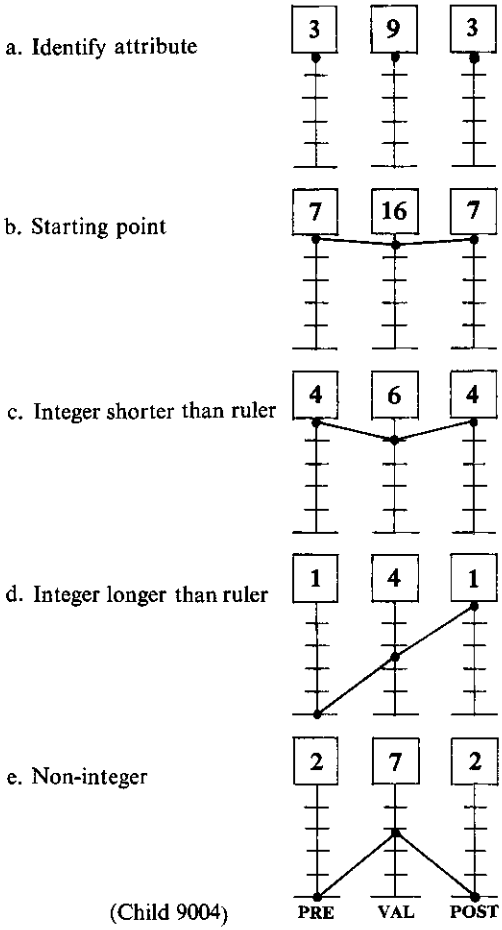

FIGURE 1. *The three vertical axes in each graph* (a *through* e) *refer, from left to right, to the pre-test, the games and other validating activities, and the post-test. The number at the top of each vertical axis indicates the number of opportunities the child had to display that particular subskill.*

Child number 9004 was chosen for Figure 1 because she illustrates several kinds of results we encountered during the validation process. In the top graph (*a*), "Identify attribute," this child gave correct answers on all three opportunities provided on both the pre-test and the post-test, and all nine opportunities recorded during the validation activities.

On the next graph (*b*), "Starting Point," performance was almost equally good, except that on the validating activities she missed one, resulting in a dot on the middle axis at $15/16=0.94$. The inconsistency of performance seems even worse in graph *c*, "Integer [length] shorter than ruler," but this is partly an artifact of the number of opportunities recorded for the validating activities, since the child still missed only one out of six on the validating activities, leading to a dot at $5/6=0.83$.

With graph *d*, "Integer [length] longer than ruler," we encounter an entirely different form of graph, one that rises monotonically from 0/1 on the pre-test through 2/4 on the validating activities, to 1/1 on the post-test. We discuss below the consequences of such a graph for validation of the test for this subskill. Here we note in passing that the test contains only a single opportunity to demonstrate this subskill. This is probably a defect of the test; it may help account for the percentage (19 percent) of items that will be judged to be invalidating in the later analysis of this subskill, greater than for any other subskill.

With graph *e* in Figure 1, "Non-integer [length]," we come to the first case that shows a genuine inconsistency between performance on pre- and post-tests, on the one hand, and performance on the validating activities on the other hand. A teacher predicting this child's ability on this subskill on the basis of our test would have been misled, and a disservice done to the child unless the teacher had investigated further.

The summary graphs for all the children then were categorized as follows:

Horizontal lines connecting the three points tend to validate the test, since they show consistency of performance between tests and validating activities.

Graphs in the form of a "V" or an "inverted V" tend to invalidate the test, since they show an inconsistency between the performance on pre- and post-tests on the one hand and validating activities on the other hand.

Ascending and descending graphs cannot be interpreted easily as either validating or invalidating. In validating the test of length measurement, we found more than four times as many ascending graphs as descending ones, consistent with the assumption that some children learn the various subskills in the course of the validating procedure. We classified ascending and descending graphs as neutral, tending neither to validate nor invalidate the test.

It is important to note that the use of the post-test in this classification has as its only purpose to distinguish the invalidating "V" and "inverted V" graphs from the neutral ascending and descending ones.

Validating cases are those that show consistent performance on the test-activity-test sequence. Consistent performance is validating whether it is at a high level, low level, or somewhere in between. Table I shows the distribution of validating cases for the 71 children who participated in the validation of the length-measurement test. It is clear from the Table that the predominance of validating cases lies in the high and low performance ranges, with few cases in the middle range of performance.

What distribution would one like to see in this kind of table? First of all, it seems to us that the "clean" diagnosis of the subskills should lead to a predominance of validating cases in either the very high or the very low ranges. If the subskills have been adequately distinguished from one another, then for each subskill the children either perform or don't perform. A large number of cases falling in the middle range would imply that a mixture of subskills is being analyzed, some of which were high and some of which were low.

TABLE I
Distribution of Validating Cases,
Length-Measurement Test

	Level of Performance				
Subskill	0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1.0
Identify attribute	0	0	0	0	49
Starting point	5	0	0	1	52
Integer length shorter than ruler	0	1	0	0	63
Integer length longer	14	0	0	0	34
Non-integer length	38	4	0	0	7

The distribution in the Table might be criticized by saying that there is too great a predominance of either high performance (for the first three subskills) or low performance (for the last subskill). According to this view, one wishes to "discriminate" between children, and so an approximately equal number of cases should appear in the very high and the very low performance ranges for each subskill. TORQUE does not agree with this view. The purpose of TORQUE tests is not to discriminate between children, but to identify the important subskills in which they need further instruction. There is a danger that, on the one hand, a test in which all children have high performance may be a trivial one, or, on the other hand, a test in which all children do poorly is too difficult. One has to show, on the basis of other evidence, that the subskills being examined are appropriate, even essential, to the measurement task involved and that, with proper instruction, children can perform them well.

A good diagnostic test of a measurement skill should be one in which the overwhelming majority of children perform at either a high level or a low level for each subskill, with very few children in the middle.

How high is a "high level" of performance and how low is a "low level"? Several re-analyses of the data were performed in order to inspect the sensitivity of the validation results to the definitions of "high" and "low" performance as they affect the number of validating cases. The number of validating cases was found to be insensitive to the level of performance defined as "high" or "low." Somewhat arbitrarily (as permitted by this insensitivity), we divided the performance scale into three regions: "high" is above 75 percent, "low" is below 25 percent, and "middle" is the wide region between 25 percent and 75 percent. Validating cases are those for which subskill performances on pre-test, games, and post-test all lie within one of the three regions. Invalidating cases are judged by using the lines connecting performance values on pre-test, games, and post-test. Invalid cases are those for which these lines cross at least one region boundary twice (creating the "V" or "inverted V" shape described above). Neutral cases (showing generally ascending or descending profiles) comprise all cases that are neither validating nor invalidating.

In summary, then, we consider a test to be validated if two conditions are met:

1) most of the cases are validating for each subskill; and

2) the number of validating cases lying within the broad middle performance range is much less than the number lying within the narrow ranges of high and low performance.

TABLE II

Summary of Validation Results for Test of Linear Measurement

Total: 71	Validating	Neutral	Invalidating	Invalidating % [Inval / (Val + Inval)]
Identify attribute	54	7	10	10/64=16%
Starting point	61	9	1	1/62=2%
Integer shorter than ruler	65	3	3	3/68=4%
Integer longer than ruler	51	8	12	12/63=19%
Non-integer	58	11	2	2/60=3%
Totals	289	38	28	28/317=9%

Table I has shown that condition (2) is satisfied for our length measurement test.

Table II summarizes the results of the validation procedure for the test of linear measurement. The right-hand column of this Table presents an "invalidating percentage" for each subskill, namely, the number of invalidating cases divided by the total number of validating plus invalidating cases, the result expressed as a percentage. These percentages range from a high of 19 percent for "Integer [lengths] longer than ruler" to a low of 2 percent for "Starting Point." In the lower right-hand corner of the Table is an overall figure for the invalidation percentage of the test, calculated by taking the total number of invalidating entries and dividing it by the invalidating plus validating entries. This percentage is 9 percent and might be taken as the best single number estimating the chance that the test will incorrectly predict the performance of a child on any given subskill for linear measurement.

To provide some check on the adequacy of the sample size of 71 children, a subset of 35 was selected randomly and the analysis of the data repeated for the subset. The results of the analysis for the subset are substantially the same as those for the group as a whole.

In analyzing these data, we do not feel we can draw on standard statistical methods which make the assumption that (a) at any given instant there is a *state of knowledge*, and that (b) the myriad factors, that mediate the articulation of that *state of knowledge* as a response, conspire to produce "noisy" responses randomly distributed about some mean which is an indicator of the state of knowledge. Our experience is completely different and does not fit this model. An individual child does not respond randomly to the opportunity to demonstrate a given subskill. On the contrary, it has been seen that if appropriate subskills have been identified, the child most often either performs or does not perform.

Project TORQUE has applied this validation strategy to a number of tests that cover a range of mathematical tasks. Moreover, it has done some preliminary studies of the application of this form of test-validation strategy to the assessment of several linguistic tasks. In each case, the validation procedure described seems to be manageable and implementable with a reasonable range of resource. There are, to be sure, questions that remain. For example, it is not clear how complex the behavior can be before this mode of validation becomes unwieldy. Despite the unanswered questions, we feel that this strategy for validating assessment instruments offers a good deal of promise for tests that cover a wide range of competencies important to society.

The Enhancement of Memory in Infancy

JEROME KAGAN
Harvard University

My colleagues and I have been studying the behavior of infants since 1962 (Kagan et. al., 1978). We have been concerned with such cognitive variables as distribution of attention and play, rather than motor development or emotion, because cognitive processes are central to most psychological phenomena, attention and play are pan-human responses rather than arbitrary reactions which infants are required to generate to accommodate to an examiner's plan, and because duration of attention is easily quantified.

The major questions that guided the initial work were simple. What are the characteristics of those events that recruit and maintain the infant's attention? And does the power of those characteristics change with development? If the latter, what is the growth function that relates class of event to duration of attentiveness, recognizing that the infant's psychological knowledge makes a contribution to this function. One of the first small discoveries was the U-shaped function relating age and duration of attention to masks of human faces across the first three years of life, with a nadir at about seven to nine months of age. Because the incentive stimulus remained constant, duration of attention should have been a function of time to assimilate the mask.

Time to assimilation should become shorter with developmental maturity, so duration of attention should have bottomed out at some low value. As it did not, we were faced with a theoretical problem.

To resolve this and other problems, we adopted two complementary strategies. Experimentally, we tested specific hypotheses about the determinants of attention and we implemented comparative research. Sometimes we compared different ethnic groups within the United States, sometimes groups with different experiences (for example, children in day-care programs versus those who stayed at home), and sometimes we went to other cultures outside the United States to check on the generality of the effects we saw in our Cambridge studies.

In order to test the generality of the original U-shaped function for attention to masks as found in American children, Gordon Finley implemented a replication with children living in Tikal, a rural town in Yucatan. Finley showed two-dimensional, realistic representations of regular and distorted faces to Yucatan children up to three years old, and coded the duration of their attention to these pictures. The same U-shaped function relating attention to age emerged. This replication enhanced the factual status of the original observation and invited a unifying hypothesis.

Our first guess was that, after eight months, the child was activating cognitive strategies in an attempt to relate the discrepant event in his perceptual field to his prior knowledge.

The next advance was again provoked by data. We were trying to determine if the phenomenon of separation distress was more closely related to level of cognitive maturity, as indexed by the quality of the child's play and the age at which he showed the inflection for attention, than to the nature of the caretaker-infant bond. We had gathered data on children's responses to maternal departure in a standard situation in several different cultural settings—the United States, Israel, !Kung San groups in Botswana, and Guatemalan cities and villages. In these studies, the mother and child are brought to an unfamiliar setting and remain until the child is playing happily. At that point, the mother leaves the child alone in the unfamiliar setting, and observers code the occurrence of fretting, crying, and signs of distress, including inhibition of play. The use of varied cultural settings was important, for the closeness of the mother-infant relationship varies across these societies. In all of the settings, separation distress rarely appeared before seven to eight months of age, grew to a peak at fifteen months, and then declined. This finding implied that separation protest emerged at the same time that the trough in attention appeared. This seemed to us to be nonaccidental. We then searched the literature to try to determine what behaviors other investigators had reported as characteristic of this era of development. T. G. R. Bower had written that the growth function for an infant's reaching for an audible object in darkness resembled our attentional function, and Selma Fraiberg had also reported a U-shaped function for reaching for objects among babies who were blind from birth. H. R. Schaffer and his students had found a noticeable increase in inhibition in reaching for a novel stimulus after repeated presentations of a standard. Joseph Campos found no evidence of avoidance of the visual cliff until seven to eight months of age, despite an ability to perceive depth prior to that time. Sandra Scarr and Philip Salapatek found that fearfulness to a variety of events—from strangers to mechanical dogs—also emerged at eight months and peaked during the second year.

These diverse phenomena could not be explained either by our initial notion of activation of schemata or by Piaget's notion of the development of the concept of object permanence. Because Piagetian theory dominates thought about this era of development, and because the object concept was a key construct, Nathan Fox and I initiated a longitudinal study of children across the period six to twelve months, making stage four and stage five of the object concept the focus of study. In one of the procedures, the infant is allowed to find a toy several times at one location, called A. Subsequently, the infant watches the examiner hide the toy at a second location, called B. We found that infants did not make the error of going to location A, after they saw the toy placed at B, if the delay between hiding the toy at location B and the infant's reaching was short, say one to three seconds. However, infants did make the error of going to A if the delay was as long as seven seconds. But the eight-month-old infants who made the "A, not B, error" with a seven-second delay did not make this error when we saw them two months later, indicating that, during even a short course of development, memorial capacity was enhanced. We replicated this finding in a cross-sectional study, using children living in two Indian villages located on Lake Atitlan.

The succeeding investigations were meant to be confirmatory, rather than inductive, in their mode. Mark Szpak demonstrated that during the age period of eight to twelve months, infants become increasingly more proficient at retrieving a toy hidden under one of two cloths, with delays of one, three, or seven seconds, during which an opaque screen blocks visual access to the array. No eight-month-old could retrieve the toy when the screen was lowered for seven seconds; the majority of one-year-olds could. Leslie Brody trained infants eight and twelve months old to touch one of two facial stimuli that had been lit, and varied the delay between the offset of the lighted stimulus and the time when she allowed the child to make an operant touching response in order to receive a visual reinforcement. The delays in her study were three, six, or nine seconds, and opaque screens were used. Some twelve-month-old children remembered which face had been lit after a nine-second delay when the opaque screen was lowered between the child and the faces, whereas those eight months old were unable to remember which face had been lit after the short, three-second delay.

The results of the last two studies imply that the ability to retrieve a representation of a prior event and to hold the information on the stage of active memory while comparing that structure with current experience is enhanced during the last third of the first year. This ability is likely to be a necessary, although not sufficient, condition for display of the phenomena noted earlier. This competence could explain, for example, the inverted-U function for attention. The older infant's attentiveness is due to the fact that he or she can retrieve a representation of the earlier repeated experience, hold it in memory, and try to relate it to the discrepant event in the perceptual field. The appearance of distress to separation could be because the child retrieves a representation of the caretaker when she was in the room, compares it with the present experience of her absence, and tries to relate the two structures. The inability to resolve the inconsistency inherent in the two structures leads to uncertainty and distress.

It is reasonable to suggest that the diverse behavioral changes that appear rather uniformly between seven and twelve months of age are released by structural and/or biochemical events that are essential components of development. Indeed, recent studies of the evoked potential in infants suggest important changes in the characteristic wave forms of the auditory evoked potential during the last half of the first year. Freud seemed to be approaching a similar insight toward the end of his career, for, in a prophetic paragraph, he questioned the formative power assigned much earlier to variation in infant experience with the caretaker. Freud suggested that maturational forces would guarantee that all infants would display some common developmental profile. He wrote:

"The phylogenetic foundation has so much the upper hand over personal accidental experience that it makes no difference whether a child has really sucked at the breast or has been brought up on the bottle and never enjoyed the tenderness of the mother's care. In both cases, the child's development takes the same path" (Freud, 1964, p. 188).

This small corpus of facts from a narrow era of development may have implications for larger themes. If the opening years of life are characterized by the maturation of new cognitive competences which affect the behavioral and emotional attributes of children, then we should see dramatic discontinuities in development as we do in history when an economic depression, a revolution, or the introduction of a technical invention leads to changes in the society, or in biology when the pituitary triggers the gonads to secrete sex hormone at adolescence. The consequences of these material changes are less a derivative of earlier phases than they are novel processes which put a new face on development. Thus, the work on the development of attention with infants may inform the deep theme of continuity versus discontinuity in development which has attracted us for over twenty years.

References

Freud, S. 1964. *Outline of Psychoanalysis.* London: Hogarth.

Kagan, J., Kearsley, R. B., and Zelaso, P. R. 1978. *Infancy: Its Place in Human Development* (with assistance of Cheryl Minton). Cambridge: Harvard University Press.

Subscription price: $5.00 per year.