DOCUMENT RESUME

ED 097 123 PS 007 511

AUTHOR Brainerd, Charles J.; Hooper, Frank H.
TITLE A Methodological Review of Developmental Studies of Identity Conservation and Equivalence Conservation.
PUB DATE [74]
NOTE 29p.

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS *Cognitive Processes; *Conservation (Concept); *Developmental Psychology; *Early Childhood; Elementary School Students; Error Patterns; Kindergarten Children; Measurement Techniques; Preschool Children; *Research Methodology; Sampling
IDENTIFIERS *Piaget (Jean)

ABSTRACT

This report reviews some ostensibly conflicting empirical findings which have been reported in conjunction with Elkind's (1967) conjecture that Piaget's conservation problems tap two distinct concepts. The discrepant findings which report on the order of emergence of identity conservation and equivalence conservation are discussed. An analysis of the procedural details of the conflicting studies reveal that the discrepant findings are probably the results of a measurement error (judgements-plus-explanations response criteria) and a sampling error (older subject samples), routinely committed in studies reporting that identity and equivalence emerge in no fixed order. Some new data from an experiment in which these two errors were controlled provide support for this conclusion. The general consequences of measurement and sampling errors for concept development studies are discussed. (Author/CS)

A METHODOLOGICAL REVIEW OF DEVELOPMENTAL STUDIES OF IDENTITY CONSERVATION AND EQUIVALENCE CONSERVATION

Charles J. Brainerd and Frank H. Hooper

University of Alberta — University of Wisconsin

A METHODOLOGICAL REVIEW OF DEVELOPMENTAL STUDIES OF IDENTITY CONSERVATION AND EQUIVALENCE CONSERVATION

This brief review is concerned with some ostensibly conflicting empirical findings which have been reported in conjunction with Elkind's (1967) conjecture that Piaget's conservation problems tap two distinct concepts. In preparing this review, our principal aim was to determine whether or not it is possible to effect a rapprochement between the conflicting groups of findings via the consideration of certain measurement issues associated with the assessment of cognitive-developmental sequences. The paper begins with an overview of the problem. Next, three procedural variables are discussed which are known to affect the visibility of cognitive-developmental sequences. Finally, some data concerned with these procedural variables are presented.

Background

Theory

According to Elkind, the standard conservation paradigm, which has been employed in so many developmental investigations, should be viewed as assessing two concepts rather than one. Elkind called these two concepts "identity" and "equivalence." Identity was defined as the child's understanding that simple quantitative properties (e.g., length, weight) of single stimulus objects remain constant across irrelevant perceptual transformations. Equivalence was defined as the child's understanding that the equality of pairs of stimulus objects vis-à-vis these same quantitative properties is not affected by irrelevant perceptual transformations of one of the pair members. Thus, the identity concept is concerned with a quantitative relationship between two successive states of the same object, whereas the equivalence concept is concerned with the quantitative relationship between two objects during two successive states.

Elkind went on to argue, primarily on logical grounds, that identity should precede equivalence in children's thinking. More recently, Elkind has suggested that identity normally emerges during the late preschool years and equivalence normally emerges during the early elementary school years Elkind & Schoenfeld, 1972).

Neither Elkind's partitioning of the standard conservation paradigm into identity conservation and equivalence conservation nor his prediction of a developmental lag between the two concepts is acknowledged in Piagetian theory. Concerning the former point, it is clear that identity and equivalence are simply "conservation" from the standpoint of the theory (e.g., Piaget, 1968, chapter 2). Concerning the latter point, a key assumption of Piaget's stage philosophy of mental growth (e.g., Piaget, 1956, 1960, 1971) is that the emergence of the prototypic concepts of each stage is a synchronous and unitary process (cf. also Brainerd, 1973a, 1973c; Flavell, 1971; Flavell & Wohlwill, 1969; Pinard & Laurendeau, 1969). In the explicit case of the conservation concept, it is assumed that the various "components" of the concept emerge in tight synchrony in each quantitative area (e.g., Piaget, 1952).

To avoid subsequent confusion, it is important to issue some terminological caveats at this point. Although Piagetian theory does not acknowledge the distinction just discussed, it does acknowledge a distinction between identity and conservation. However, Elkind's and Piaget's respective versions of "identity" are very different. Elkind's identity quite obviously is a quantitative concept, but Piaget's identity is not. Piaget (e.g., 1968, chapter 2) defines identity as the child's understanding that the transformed stimulus is the "same" object as before the transformation. In contrast, Elkind's identity refers to the child's understanding that the transformed stimulus is the "same amount" as before the transformation. To illustrate, consider the length conservation problem. We begin with two pieces of string of the same length which are placed side-by-side so that their equivalence is evident to casual inspection. One of the strings is bent into a circle. If the subject understands that the transformed stimulus is still "the same piece of string," Piaget's identity concept is present. If the subject understands that the transformed stimulus is still "the same length," then Elkind's identity concept is present. In Piaget's terminology, however, the latter response would simply be "conservation."

There is an important substantive reason for keeping these two versions of identity strictly separated. Although Piagetian theory does not acknowledge a developmental lag between Elkind's identity and equivalence, it does acknowledge a lag between Piaget's identity and conservation (Piaget, 1968, chapter 2). If one does not distinguish between the two identities, therefore, it is possible to argue that Piaget's predictions are the same as Elkind's. In the literature, the two identities are distinguished by appending "quantitative" to Elkind's version and "qualitative" to Piaget's version. In the present paper, however, whenever the term "identity" is employed, we shall understand the referent to be Elkind's version.

Empirical Evidence[1]

Shortly after Elkind's original paper appeared, Hooper (1969a, 1969b) published some data which he viewed as being consistent with the prediction that identity conservation precedes equivalence conservation. Subsequent investigations by Elkind and Schoenfeld (1972), Papalia and Hooper (1971), and Schwartz and Scholnick (1970) provided further support for the prediction in the content areas of discontinuous quantity, liquid quantity, solid continuous quantity, length, and number. Concurrently, however, a series of studies was reported in which the identity → equivalence sequence was not observed. Koshinsky and Hall (1973) failed to find the sequence for discontinuous quantity using a design that was virtually a point-for-point replication of Hooper's (1969a) original study. Moynahan and Glick (1972) failed to find the sequence for weight, number, and liquid quantity. Murray (1970) failed to find the sequence for weight and number. Northman and Gruen (1970) failed to find the sequence for liquid quantity.

Somewhat surprisingly, investigators reporting nonconfirming evidence have not used their findings as a basis for challenging Elkind's orginal argument. Instead, the tendency has been to accept Elkind's argument as logically sound and to advance the ad hominem argument that, for reasons unknown, cognitive development does not conform to the dictates of logic in the special case of conservation concepts (cf. especially Koshinsky & Hall, 1973, p. 423; Moynahan & Glick, 1972, p. 251).

The obvious discrepancy between the supportive and nonsupportive groups of studies, coupled with the significance of Elkind's predicted sequence from the standpoint of Piagetian theory, prompted the present writers to conduct an analysis of the procedural details of the studies. Our general aim was to determine whether or not the two groups of studies tended to differ on one or more of three variables which are known to affect the visibility of asynchronies in the emergence of Piagetian concepts.

Our preliminary hypothesis was that the identity → equivalence sequence actually exists in the referent population and, therefore, some simple design difference(s) between the supportive and nonsupportive studies probably would explain the discrepant findings. Two facts seemed to justify entertaining this hypothesis at the outset. First and most important in our view, the logic underlying Elkind's prediction is ineluctable. Given that we accept his operational definitions of identity and equivalence, as the authors of all the relevant studies do, the prediction follows Q. E. D.: Identity must emerge before equivalence because the operational definition of the former is a component of the operational definition of the latter, but not conversely. Second, the reverse of the predicted sequence (i.e., equivalence → identity) has never been observed to the best of our knowledge. If the identity → equivalence sequence does not exist in the referent population and the concepts actually emerge synchronously, then we would expect on statistical grounds that the sequence and its reverse would be observed with roughly equal frequency. If the predicted sequence does exist, however, it is not at all unreasonable that it would be observed in some studies and that synchrony would be observed in other studies. In fact, if the predicted sequence happens to be a relatively precise one, its visibility will be markedly affected by measurement and sampling errors and, hence, we would expect just such a pattern of findings. Unfortunately, this argument also is ad hominem unless we know what the explicit sources of error are. Three possible sources will now be discussed.

Summary of Analysis

Our analysis was focused on two potential measurement errors and one potential sampling error which have proved troublesome in past research on cognitive-developmental sequences: (a) relative task sensitivity; (b) response criteria; (c) age of subjects. The results of the analysis are summarized briefly by variable.

Relative Task Sensitivity

Flavell (1971) has argued convincingly that failure to equate for relative task sensitivity is a pervasive measureme... error in cognitive-developmental sequence research. According to Flavell, failures to equate for relative task sensitivity can mask sequences which exist in the population and also can manufacture spurious sequences. The argument which supports this claim runs as follows. Suppose we have two concepts A and B such that A invariably emerges before B during cognitive development. That is, there is a real A → B sequence in the population. Suppose we measure A and B in an appropriate sample of subjects using a very insensitive test of A and very sensitive test of B. With these particular tests, the A → B sequence may not be observed because there will be a high rate of Type II error ("false negatives") on the A test. Now, suppose that A and B emerge synchronously during cognitive development and that the same tests are administered to the same sample. This time it is likely that a spurious B → A sequence will be observed in the resulting data. In view of the high false negative rate on the A test, a large proportion of those subjects who actually possess both A and B will pass the B test and fail the A test but not conversely.

A recent review of the role of task sensitivity in developmental studies of transitivity and conservation (Brainerd, 1973a) provides support for Flavell's argument. When the relative sensitivities of transitivity and conservation tests are equated, transitivity is observed to emerge before conservation (Brainerd, 1973a, 1974a, 1974b; Brainerd & Vanden Heuvel, 1974; Toniolo & Hooper, 1974). However, when very insensitive tests of transitivity are used in conjunction with reasonably sensitive tests of conservation, transitivity either is observed to emerge before conservation (e.g., McManis, 1969; Smedslund, 1963) or is observed to emerge synchronously with conservation (e.g., Lovell & Ogilvie, 1961).

Our examination of the identity and equivalence tests employed in the supportive and nonsupportive groups of studies revealed no consistent between-group differences in relative task sensitivity. Although some differences in this variable were noted in virtually all the studies, there was no evidence that (a) blatantly insensitive identity tests had been used in conjunction with reasonably sensitive equivalence tests in the nonsupportive studies or that (b) blatantly insensitive equivalence tests had been used in conjunction with reasonably sensitive identity tests in the supportive studies. Either a or b obviously would tend to explain the discrepancy between the two groups of studies.

Response Criteria

It has become a commonplace in reviews of neoPiagetian research to observe that there is an alarming lack of consensus among investigators concerning the appropriate response criteria for inferring the presence of Piagetian concepts (e.g., Beilin, 1971; Brainerd, 1973a, 1973b, 1974c; Hooper, Goldman, Storck, & Burke, 1971). Two explicit questions about the response criteria employed in the identity/equivalence studies were examined in this portion of the analysis. First, is there any evidence either that (a) the response criteria for identity tend to be more stringent than the response criteria for equivalence in the nonsupportive studies or that (b) the converse tends to be true in the supportive studies? It should be obvious that a would tend to mask a real identity → equivalence sequence and b would tend to manufacture a spurious sequence. Second, is there any evidence either that (c) a response criterion of judgments-plus-explanations was used for both identity and equivalence items in the nonsupportive studies or that (d) a response criterion of judgments-only was used for both identity and equivalence items in the supportive studies?

Concerning the second question, it has been shown elsewhere on theoretical grounds that a judgments-plus-explanations criterion for items on Piagetian concept assessments tends to mask real developmental sequences in the measured concepts (Brainerd, 1973b, 1974c). From a theoretical standpoint, including verbal rationales as an essential component of the response criterion introduces at least two specifiable sources of Type II error. Given two concepts A and B such that A actually precedes B during cognitive development, the introduction of any source of Type II error will tend to reduce the visibility of this sequence. The psychometric rationale for this statement is elementary.

In any given study, inferences about the order of emergence of A and B are based on a comparison of the observed frequencies of subjects who evidence A in the absence of B and subjects who evidence B in the absence of A. Subjects who evidence both concepts or neither concept are not of interest. If the observed frequencies of A/not-B and not-A/B subjects in a given sample differ significantly, then we reject the null hypothesis that the two concepts are not acquired in a fixed order (cf. also Brainerd, 1974d). Assuming a real $A \rightarrow B$ sequence in the population, the effect of any source of Type II error must be to decrease the expected frequency of the A/not-B category relative to the expected frequency of the not-A/B category and thereby reduce the visibility of the sequence:

If A actually precedes B, then the population consists of only three types of subjects--not-A/not-B, A/not-B, and A/B. Assume that the population frequencies of these three categories are P_1, P_2, and P_3, respectively, where $P_1 + P_2 + P_3 = 1$. Also, assume that our method of assessing A and B incorporates a source of Type II error which reduces the probability that either concept will be judged present, given that it is present, by some factor $0 < \underline{x} < 1$. The probability that any not-A/not-B subject will be correctly classified is unity. The probability that any A/not-B subject will be correctly classified is $1 - \underline{x}$ and the probability that he will be incorrectly classified as a not-A/not-B is $\underline{x}$. The probability that any A/B subject will be correctly classified is $1 - 3\underline{x}^2$ and the probability that he will be incorrectly classified as a not-A/not-B or an A/not-B or a not-A/B is $\underline{x}^2$ in each instance. Hence, the expected frequencies of the four possible subject categories following Type II error are

$$E(\text{not-A/not-B}) = P_1 + \underline{x}P_2 + \underline{x}^2P_3 \quad [1]$$

$$E(\text{A/not-B}) = P_2[1 - x] + \underline{x}^2P_3 \quad [2]$$

$$E(\text{not-A/B}) = x^2P_3 \text{ and} \quad [3]$$

$$E(\text{A/B}) = P_3[1 - 3\underline{x}^2]. \quad [4]$$

Note that one effect of Type II error is to decrease the difference between the expected frequencies of the second and third categories, relative to the difference between their population values. Given that P_2, P_3, and $\underline{x}$ all lie between 0 and 1, the difference between the right side of Equation 2 and right side of Equation 3 must be smaller than the difference between P_2 and 0. Therefore, the commission of Type II errors reduces the probability that the null hypothesis mentioned above will be correctly rejected.

By the preceding argument, any response criterion which is known to increase the Type II error rate will reduce the visibility of a real A → B sequence by increasing the observed frequency of the not-A/B category relative to the A/not-B category. In line with this argument, the masking effect of the judgments-plus-explanations criterion has been demonstrated empirically in the case of the well-known number conservation → quantity conservation sequence (Brainerd & Brainerd, 1972; Gruen & Vore, 1972). In these studies, the developmental precedence of number conservation over quantity conservation was much less apparent with judgments-plus-explanations than with judgments-only. In the Brainerd and Brainerd study, the discrepancy was quite large: A binomial test of the null hypothesis that subjects who possess only number conservation and subjects who possess only quantity conservation occur with equal frequency produced a probability value of 7.58×10^{-9} with a judgments-only criterion and a probability value of 3.79×10^{-4} with a judgments-plus-explanations criterion. The difference between these two values is statistically significant.

To return to the two questions posed at the beginning of this section, the answer to both parts of the first question is an unqualified "no." In our review of the response criteria employed in the relevant studies, we observed no discernible tendency to employ identity and equivalence criteria of differential stringency. Although overall criterion stringency differed from one study to another (e.g., compare Koshinsky & Hall, 1973, with Schwartz & Scholnick, 1970), within any single study the typical pattern was to choose some criterion _a priori_ and then apply it consistently to both identity and equivalence responses. On the other hand, the answer to both parts of the second question posed above is a qualified "yes." Concerning the first half of the question, the supportive evidence reported by Elkind and Schoenfeld (1972), Paplia and Hooper (1971), and Schwartz and Scholnick (1970) involved a judgments-only criterion for both identity and equivalence. Concerning the second half of the question, the non-supportive evidence reported by Moynahan and Glick (1972) and Koshinsky and Hall (1973) was based on a judgments-plus-explanations criterion for both identity and equivalence. The previously mentioned consequences of Type II errors suggest that the latter procedure inevitably would reduce the visibility of a real identity → equivalence sequence.

Hooper's (1969a, 1969b) original studies are the only exceptions to the rule that supportive studies tended to employ a judgments-only criterion. However, both of these studies provide much less substantial evidence for an identity → equivalence sequence than the other supportive studies. In the first Hooper study, a total of 108 children between 5 years, 6 months and 8 years, 6 months were assigned to three different testing conditions. Each subject in the first condition was administered either of two tests of identity. Each subject in the second condition was administered either of two tests of equivalence. Each subject in the third condition was administered either of two tests of equivalence, both of which were different from the tests administered in the second condition. After testing, a judgments-plus-explanations criterion was used to partition the subjects in the first condition into conservers and nonconservers of identity. The same criterion was used to partition subjects in the second and third conditions into conservers and nonconservers of equivalence. This procedure permitted a total of 19 chi-square tests of the hypothesis that identity and equivalence tests were of equivalent difficulty. Of these 19 possible comparisons, only one [first condition/both tests (N = 36) vs. both equivalence conditions/both tests (N = 72)] turned out to be significant ($\underline{X}^2 = 5.35$, $\underline{p} < .05$). In other words, the null hypothesis could not be rejected in 18 of 19 cases. With such a large number of interdependent significance tests, it is obvious that the effective alpha level was much higher than the nominal .05. In fact, the probability that one of the 19 tests would be significant at or above the .05 level was almost unity.

The second Hooper study also provides somewhat meager support for an identity → equivalence sequence. A total of 80 subjects were administered tests of identity and equivalence. A judgments-plus-explanations criterion was used to partition the subjects into conservers of both concepts, conservers of only identity, conservers of only equivalence, and conservers of neither concept. Only 11 subjects were classified as members of the second group. Because none of the subjects were classified as members of the third group, however, the null hypothesis that identity and equivalence tests were of equivalent difficulty was rejected ($p < .05$) by the binomial test.

The nonsupportive studies of Northman and Gruen (1970) and Murray (1970) are exceptions to the rule that nonsupportive studies tended to employ a judgments-plus-explanations criterion. In both cases, the subjects were classified as conservers or nonconservers of identity and equivalence via a judgments-only criterion. However, the age of the subjects employed in these studies was sufficiently above the level at which one may reasonably expect to find evidence of an identity → equivalence sequence (cf. the following section) that the choice of criterion may well have been irrelevant.

Age of Subjects

In a recent study, Elkind and Schoenfeld (1972) made the following observations: identity conservation can be arrived at on the basis of preoperational mental structures whereas equivalence conservation requires both the preoperational structures and the concrete operations that develop at about the ages of 6 or 7. To solve the identity conservation problems, the young child can simply call upon accumulated past experience... Young children cannot, however, arrive at equivalence conservation because they lack the deductive apparatus of the concrete operational system. Accordingly, one should find that young children have identity conservation but not equivalence conservation whereas older children should have both ... the differential performance on identity and equivalence tests should hold true for preoperational children but not for concrete operational children [p. 530].

The practical implication of this argument for developmental studies of identity and equivalence is straightforward. If identity and equivalence tests are administered to preschoolers and, perhaps, also to kindergarteners, then the identity → equivalence sequence probably will be observed. On the other hand, if identity and equivalence tests are administered to elementary schoolers (an age level at which most subjects would be considered concrete-operational), then the sequence probably will not be observed.

Note that it is not necessary to accept either the earlier analysis on which Elkind and Schoenfeld's observations are based (Eklkind, 1967) or their invocation of Piaget's cognitive-developmental stages to see that the measurement implication of their observations probably is correct. [Concerning Piaget's stages, for example, it is not at all clear, at least not to the present writers, that vaguely defined typological constructs such as "preoperational child" and "concrete-operational child" have any explanatory power.] Consider two concepts A and B such that A → B during cognitive development. Suppose that A normally emerges during some age range R_1 and that B normally emerges during some later age range R_2. Suppose that tests of A and B are administered to a sample from R_1. Assuming errorless measures, two and only two types of subjects will be observed: first, most subjects will pass A but fail B; second, the remaining subjects will fail both tests. Because the null hypothesis tested in such a study is concerned with observed frequencies of subjects who pass A but fail B and who pass B but fail A, the data of this sample should provide strong support for the A → B sequence. On the other hand, suppose the tests are administered to a sample from R_2. Again assuming errorless measures, two and only two subjects will be observed: first, most subjects will pass both tests; second, the remaining subjects will pass A but fail B. The chances of rejecting the null hypothesis obviously will be much smaller with this second sample than with the first.

By Elkind and Schoenfeld's argument, studies in which partially or primarily preschool samples are employed should tend to find an identity → equivalence sequence more frequently than studies in which only older samples are employed. A review of the age ranges of the subject samples employed in the supportive and nonsupportive studies provides considerable support for this argument. At least some preschoolers were tested in three of the five supportive studies: One-fourth of Schwartz and Scholnick's (1970) subjects were preschoolers; one-third of Papalia and Hooper's (1971) subjects were preschoolers; one-half of Elkind and Schoenfeld's (1972) subjects were preschoolers. Hooper's studies (1969a, 1969b) are the only supportive studies in which no preschoolers were tested. In both studies, the youngest subjects were kindergarteners. However, both studies are subject to the caveats mentioned earlier.

In contrast with the subject samples of the supportive studies, preschoolers were not tested in any of the nonsupportive studies. Murray's

(1970) youngest subjects were 6-year-olds; Moynahan and Glick's (1972) youngest subjects were 6-year-olds; Koshinsky and Hall's (1973) youngest subjects were 5-year-olds. Thus, with the exception of the Koshinsky and Hall study and the Hooper studies, the lower age bounds of the supportive and nonsupportive studies do not overlap.

New Evidence

In this section, some new data are reported from an experiment designed to examine the major conclusions of the preceding analysis. The design was a 3 X 2 X 2 X 2 mixed-model analysis of variance. The first factor was age (preschool/kindergarten/third grade). The last three factors were task (identity/equivalence), criterion (judgments-only/judgments-plus-explanations), and content (length/weight). The four main effects were not of interest from the standpoint of the preceding analysis. Instead, attention was focused on the first-order interactions of age with task and task with criterion. From the preceding section on age of subjects, it follows that there should be an Age X Task interaction such that performance differences between identity and equivalence are more pronounced for younger subjects. From the preceding section on response criteria, it follows that there should be a Task X Criterion interaction such that performance differences between identity and equivalence are more pronounced with judgments-only than with judgments-plus-explanations.

Finally, the first-order interaction of age with criterion was of some interest. An Age X Criterion interaction is anticipated on the ground that the effect of any source of Type II error in concept assessment may be expected to decrease with age (Flavell & Wohlwill, 1969).

Subjects

The present sample consisted of three age levels: preschool (4-year-olds), kindergarten (6-year-olds), and third grade (8-year-olds). A total of 60 preschoolers (29 girls and 31 boys), 60 kindergarteners (30 boys and 30 girls), and 60 third graders (30 girls and 30 boys) were tested. All subjects were pretested for their understanding of the relational terms employed in the concept assessments described below.

Procedure

All subjects were administered identity and equivalence tests in two content areas. The materials used in the four tests were the same as those described by Hooper (1969b). The procedure and questions were the same as those reported by Brainerd (1972, 1973a) and the identity tests were adapted from Hooper (1969a). Each of the four tests involved two basic steps: prediction and transformation. During the prediction phase, a stimulus transformation of some sort was proposed and three questions were posed about the quantitative status of the stimulus after transformation (same? more? less?). During the transformation phase, a transformation was performed on one of two quantitatively equivalent stimuli and three questions were posed about the posttransformation relationship between the two stimuli (same? more? less?). During both phases, subjects were asked to explain (randomly) one of their three answers.

Identity tests. During the prediction phase of the length identity test, a 28.0-cm string was placed in the center of the table at which the subject and experimenter were seated. The experimenter posed the following questions: (a) If I bent this string into a circle, would it still be the same length as it is now? (b) If I bent this string into a circle, would it be longer than it is now? (c) If I bent this string into a circle, would it be shorter than it is now? Each subject was asked to explain (randomly) one of his prediction phase judgments. During the transformation phase, the experimenter bent the string into a circle and posed three more questions: (a) Is the string still the same length as before? (b) Is the string longer now than it was before? (c) Is the string shorter now than it was before? Subjects were again asked to explain one of their three judgments. The weight identity test was the same as the length identity test, except that a clay ball was used as a stimulus and the transformation involved flattening the ball into a "pancake."

Equivalence tests. During the prediction phase of the length equivalence test, 28.0-cm strings were used. After the subject had agreed that the two strings were the same length, three questions were posed: (a) If I bent one of these strings into a circle, would the two strings still be the same length? (b) If I bent one of these strings into a circle, would one of the strings be longer than the other? (c) If I bent one of these strings into a circle, would one of the strings be shorter than the other? Subjects explained one of their three judgments. During the transformation phase, one of the strings was bent into a circle and three questions were posed: (a) Are the two strings still the same length? (b) Is one of the strings longer than the other now? (c) Is one of the strings shorter than the other now? Again, subjects explained one of their three judgments. The weight equivalence tests were the same as the length equivalence tests, except that two clay balls were employed and the transformation involved flattening one of them into a "pancake."

Scoring

In all, each subject made 24 judgments and explained 8 of these judgments. The data were scored by both a judgments-only criterion and a and a judgments-plus-explanations criterion. Concerning the former, subjects were given a point for each item on which they emitted a correct judgment. Concerning the latter, subjects were given a point on each of the 8 explanation items only if they emitted both a correct judgment and a correct explanation. On the 16 nonexplanatory items, subjects were given a point for each correct judgment. As is common in the literature (e.g., Brainerd & Brainerd, 1972; Hooper, 1969a, 1969b; Koshinsky & Hall, 1973), an explanation was considered correct if it fell in one of the usual Genevan categories: addition/subtraction; inversion reversibility; reciprocity reversibility (compensation); tautology.

Results

Insert Table 1 about here

Insert Table 2 about here

A summary of the analysis of variance appears in Table 1. The means and standard deviations for the analysis appear by factor and level in Table 2. Concerning the mean entries in Table 2, the high possible value for each cell is 6.0.

It can be seen in Table 1 that the main effects of age, task, and criterion were large and highly significant. Newman-Keuls tests of the age effect indicated that the tasks were more difficult for preschoolers than they were for kindergarteners ($p < .001$) or third graders ($p < .001$) and that the tasks were more difficult for kindergarteners than they were for third graders ($p < .001$). Concerning the task effect, the equivalence tasks were more difficult than the identity tasks. Concerning the criterion effect, more trials were passed with a judgments-only criterion than with a judgments-plus-explanations criterion.

It also can be seen in Table 1 that the two effects of principal interest from the standpoint of our earlier methodological analysis, the Age X Task interaction and the Task X Criterion interaction, both were observed. Concerning the Age X Task interaction, Newman-Keuls tests indicated that equivalence tasks were more difficult than identity tasks for preschoolers ($p < .001$) and kindergarteners ($p < .001$) but not for third graders. Concerning the Task X Criterion interaction, Newman-Keuls tests indicated that the performance difference between identity and equivalence tasks was significantly greater with a judgments-only criterion than with a judgments-plus-explanations criterion ($p < .001$). Finally, the anticipated Age X Criterion interaction was observed. Newman-Keuls tests indicated that the discrepancy between the two criteria was more pronounced with preschoolers and with kindergarteners than with third graders ($p < .01$ and $p < .025$, respectively). The preschoolers and kindergarteners did not differ.

Conclusions

Both our earlier methodological arguments and the findings just reported tend to substantiate the preliminary hypothesis with which we began. The Age X Task interaction indicates that the identity → equivalence sequence is less visible in older subjects and, hence, the ages of the subject samples employed in the nonsupportive studies militated against finding the sequence. Similarly, the Task X Criterion interaction indicates that the identity → equivalence sequence is less visible with a judgments-plus-explanations criterion and, hence, the response criteria employed in some of the nonsupportive studies militated against finding the sequence. Therefore, it seems reasonable to conclude that identity conservation does indeed precede equivalence conservation. It also seems reasonable to conclude that existing failures to find the sequence are best viewed as consequences of having committed measurement and sampling errors.

In closing, we should like to dwell briefly on the general significance of the measurement and sampling errors discussed in the review portion of this paper for concept development research. These error sources obviously are relevant to all studies concerned with the order of emergence of concepts, not just to studies of identity and equivalence. If these error sources are not carefully controlled in such studies, then discrepant findings of the sort we have just considered will be the inevitable result. This will no doubt seem a trivially obvious point to many readers. However, the point is worth making because cognitive-developmental investigators do not generally acknowledge it. In the concept development literature, the effects of measurement and sampling errors, even ones as gross as those reviewed earlier, are only rarely discussed (e.g., cf. Flavell, 1971; Brainerd, 1973a, 1974d). To illustrate, the elementary question of the general consequences of Type I and Type II measurement errors on the null hypotheses tested in concept development studies has never been systematically examined (Brainerd, 1974d). Many investigators who study the order of emergence of concepts in children's thinking appear to be laboring under the misapprehension that the sequences (or concurrences) which they study are so robust that rigorous control of measurement and sampling error is unnecessary. It is to be hoped that the present demonstration of the pronounced effects of such errors on the identity → equivalence sequence will help dispel this misapprehension.

References

Beilin, H. The training and acquisition of logical operations. In M. F. Rosskopf, L. P. Steffe, & S. Taback (Eds.), Piagetian cognitive developmental research and mathematics education. Washington: National Council of Teachers of Mathematics, 1971.

Brainerd, C. J. The age-stage issue in conservation acquisition. Psychonomic Science, 1972, 29, 115-117.

Brainerd, C. J. Order of acquisition of transitivity, conservation, and class inclusion of length and weight. Developmental Psychology, 1973, 8, 105-116.(a)

Brainerd, C. J. Judgments and explanations as criteria for the presence of cognitive structures. Psychological Bulletin, 1973, 79, 172-179. (b)

Brainerd, C. J. NeoPiagetian training experiments revisited: Is there any support for the cognitive-developmental stage hypothesis? Cognition, 1973, 2, in press. (c)

Brainerd, C. J. Training and transfer of transitivity, conservation, and class inclusion of length. Child Development. 1974, 45, 324-334.(a)

Brainerd, C. J. Structures-of-the-whole and elementary education: Some observations on the Burke-Merkle and Hooper experiment. American Educational Research Journal, 1974, in press. (b)

Brainerd, C. J. Postmortem on judgments, explanations, and Piagetian cognitive structures. Psychological Bulletin, 1974, 81, 70-71. (c)

Brainerd, C. J. On the consequences of Type I and Type II criterion errors in concept development research. Unpublished manuscript, University of Alberta, 1973. (d)

Brainerd, C. J., & Brainerd, S. H. Order of acquisition of number and liquid quantity conservation. Child Development, 1972, 43, 1401-1406.

Brainerd, C. J., & Vanden Heuvel, K. Development of geometric imagery in five- to eight-year-olds. Genetic Psychology Monographs, 1974, 98, 89-143.

Elkind, D., Piaget's conservation problems. Child Development, 1967, 38, 15-27.

Elkind, D., & Schoenfeld, E. Identity and equivalence conservation at two age levels. Developmental Psychology, 1972, 6, 529-533.

Flavell, J. H. Stage-related properties of cognitive development. Cognitive Psychology, 1971, 2, 421-453.

Flavell, J. H. & Wohlwill, J. F. Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell (Eds.), Studies in cognitive development. New York: Oxford, 1969.

Gruen, G. E., & Vore, D. A. Development of conservation in normal and retarded children. Developmental Psychology, 1972, 6, 146-157.

Hooper, F. H. Piaget's conservation tasks: The logical and developmental priority of identity conservation. Journal of Experimental Child Psychology, 1969, 8, 234-249. (a)

Hooper, F. H. The Appalachian child's intellectual capabilities--deprivation or diversity. Journal of Negro Education, 1969, 38, 224-235. (b)

Hooper, F. H., Goldman, J. A., Storck, P. A., & Burke, A. M. Stage sequence and correspondence in Piagetian theory: A review of the middle-childhood period. Research relating to children. Bulletin 28. Washington: U. S. Govenment Printing Office, 1971.

Koshinsky, C., & Hall, A. E. The developmental relationship between identity and equivalence conservation. Journal of Experimental Child Psychology, 1973, 15, 419-424.

Lovell, K., & Ogilvie, E. A study of the conservation of weight in the junior school child. _British Journal of Educational Psychology_, 1961, _31_, 138-144.

McManis, D. L. Conservation and transitivity of weight and length by normals and retardates. _Developmental Psychology_, 1969, _1_, 373-382.

Moynahan, E., & Glick, J. Relation between identity conservation and equivalence conservation within four conceptual domains. _Developmental Psychology_, 1972, _6_, 247-251.

Murray, F. B. Stimulus mode and the conservation of weight and number. _Journal of Educational Psychology_, 1970, _61_, 287-291.

Northman, J. E., & Gruen, G. E. Relationship between identity and equivalence conservation. _Developmental Psychology_, 1970, _2_, 311.

Piaget, J. _The child's conception of number_. New York: Humanities, 1952.

Piaget, J. Les stades du développement intellectual de l'enfant et de l'adolescent. In P. Osterrieth et al. (Eds.), _Le problème des stades en psychclogie de l'enfant_. Paris: Presses Universitaires de France, 1956.

Piaget, J. The general problems of the psychobiological development of the child. In J. M. Tanner & B. Inhelder (Eds.), _Discussions on child development_. Vol. 4. London: Tavistock, 1960.

Piaget, J. _On the development of memory and identity_. Barre, Mass.: Clark University Press, 1968.

Piaget, J. The theory of stages in cognitive development. In D. R. Green, M. P. Ford, & G. B. Flamer (Eds.), _Measurement and Piaget_. New York: McGraw-Hill, 1971.

Papalia, D. E., & Hooper, F. H. A developmental comparison of identity and equivalence conservation. Journal of Experimental Child Psychology, 1971, 12, 347-361.

Pinard, A., & Laurendeau, M. "Stage" in Piaget's cognitive developmental theory: Exegesis of a concept. In D. Elkind & J. H. Flavell (Eds.), Studies in cognitive development. New York: Oxford, 1969.

Schwartz, M. M., & Scholnick, E. K. Scalogram analysis of logical and perceptual components of conservation of discontinuous quantity. Child Development, 1970, 41, 695-705.

Smedslund, J. The development of concrete transitivity of length in children. Child Development, 1963, 34, 389-405.

Toniolo, T., & Hooper, F. H. Micro-analysis of logical reasoning relationships: Conservation and transitivity. Technical Report, Research and Development Center for Cognitive Learning, University of Wisconsin, 1974.

Footnote

[1]We have confined this review to published experiments in which normal children were studied. Pilot studies, unpublished papers, unpublished theses, and studies employing special populations (e.g., retardates) have not been included. Unpublished experiments were excluded because they pose the problem of reader accessability. Studies of special populations were excluded because they pose some obvious interpretational problems. (E.g., if the identity → equivalence sequence fails to show up in retardates, what, if anything, does this tell us about normal cognitive development?) A complete bibliography of pilot studies, major unpublished experiments, and studies employing special populations is available from the authors.

TABLE 1

Summary of Age X Task X Content X Criterion Analysis of Variance.

Source	df	MS	F
Age (A)	2	9572.15	76.68***
Error between	177	124.84	
Task (B)	1	498.34	92.30***
Content (C)	1	3.34	<1
Criterion (D)	1	106.57	118.02***
A x B	2	53.40	9.89***
A x C	2	0.56	<1
A x D	2	4.18	4.63*
B x C	1	0.01	<1
B x D	1	3.68	45.37***
C x D	1	0.43	5.30**
A X B x C	2	0.06	<1
A x B x D	2	0.19	2.37
A x C x D	2	0.32	<1
B x C x D	1	0.00	<1
A x B x C x D	2	0.11	<1
Error within:			
A x B	177	5.40	
A x C	177	11.79	
A x D	177	0.90	
A x B x C	177	1.06	
A x B x D	177	0.08	
A x C x D	177	0.03	
A x B x C x D	177	0.15	

*$p<.05$

**$p<.01$

***$p<.0001$

TABLE 2

Summary Statistics for Age X Task X Content X Criterion Analysis of Variance.

Age	Length		Weight	
	Identity	Equivalence	Identity	Equivalence
		Judgments-only		
Preschool				
M	2.40	1.18	2.37	1.13
SD	1.25	1.28	0.97	1.13
Kindergarden				
M	3.52	2.28	3.63	2.28
SD	1.78	2.23	1.85	2.34
Third grade				
M	5.03	4.58	4.92	4.55
SD	1.59	2.42	1.83	2.47
		Judgments-plus-explanations		
Preschool				
M	1.73	0.40	1.60	0.82
SD	1.22	1.12	1.03	0.97
Kindergarden				
M	3.10	2.05	3.03	1.93
SD	1.77	2.10	1.91	2.20
Third grade				
M	4.68	4.52	4.55	4.35
SD	2.06	2.44	2.01	2.51