

# The Development of Reversible Operations in Judgments of Ability, Effort, and Performance

Colleen F. Surber

University of Wisconsin—Madison

SURBER, COLLEEN F. *The Development of Reversible Operations in Judgments of Ability, Effort, and Performance*. CHILD DEVELOPMENT, 1980, 51, 1018–1029. The degree to which judgments of ability, effort, and performance conform to the predictions of Harold Kelley's theory of causal schemata was examined at 4 age levels (college, fifth grade, third grade, and kindergarten). Subjects made 3 kinds of judgments: (1) judgment of performance, given ability and effort; (2) judgment of effort, given performance and ability; and (3) judgment of ability, given performance and effort information. Kelley's causal schema concept requires that effort and ability should be judged to be inversely related (when performance is held constant), and that the pattern of judgments (i.e., whether additivity or interactions are found) should be consistent across judged dimensions. The results showed inconsistent use of inverse compensation in judging ability and effort between kindergarten and fifth grade. At the adult level inverse compensation was consistent, but there was no evidence that ability, effort, and performance are integrated into a fully reversible causal schema. 2 possible explanations of the children's inconsistent use of inverse compensation are suggested: (1) it may reflect a normal step in the development of attribution processes for both social and nonsocial events, or (2) it may reflect different interpretations of the meaning of ability and its relationship to effort. Also discussed is the issue of whether the failure of the adults to show a completely reversible causal schema should be attributed to lack of the reversible operations or to the nature of the social ecology with respect to achievement.

Piaget proposed that one of the important developments of the concrete operational period is mental reversibility, that is, the ability to think about relationships among events in a bidirectional manner (Inhelder & Piaget 1958). The present research examines the development of reversible thinking for social concepts by examining the causal schema for ability, effort, and performance. Kelley (1973) has characterized a causal schema as "an assumed pattern of data in an analysis of variance framework," or a set of assumptions that allows one to make inferences on the basis of partial information. The schema for ability, effort, and performance can be examined by collecting three different types of judgments: (1) prediction of a person's performance, given ability and effort information; (2) inference of effort

expended, based on information about ability and the level of performance attained; and (3) inference of ability, given performance and effort.

If ability, effort, and task performance are judged using a single causal schema, it is predicted that if judgments of task performance are an increasing function of both effort and ability, then in judgments of effort and ability these cues will be inversely related. This requires a reversible operation that can be termed "inverse compensation." A second prediction is that the *pattern* of results for the three types of judgments should be consistent. If an interaction is obtained in judgments of performance, then qualitatively consistent interactions should be obtained for judgments of ability and ef-

This work is based in part on a thesis submitted to the University of Illinois at Urbana-Champaign in partial fulfillment of the requirements of the Ph.D. The author would like to thank the members of the dissertation committee, Joseph Campione, Carol Dweck, Glenn Kleiman, and Robert Wyer. The chair of the committee, Michael Birnbaum, provided helpful advice during the course of the research and on drafts of the manuscript. The author would also like to thank Sam L. Witryol, University of Connecticut, for his encouragement, and Lola Lopes, University of Wisconsin, for comments on the manuscript. Computing support was received from the University of Connecticut Research Computing Center. The research was supported in part by grant HD00244 from the U.S. Public Health Service, grant HD 06864 to Ann L. Brown, and grant HD 05951 to Joseph C. Campione and Ann L. Brown. Requests for reprints should be addressed to Colleen F. Surber, Psychology Department, University of Wisconsin, Madison, Wiscon 53706.

fort. This prediction can also be viewed as requiring a reversible cognitive operation.

### *Developmental Change in Attributions of Ability and Effort*

Developmental changes in judgments of ability and effort might be expected because inverse compensation for judging transformations of physical quantities develops between the ages of approximately 4 and 7 years (Larsen & Flavell 1970). Second, the concepts of ability and effort may develop at different rates because the effects of variation in one's own effort can be directly experienced while the effects of ability must be apprehended indirectly (Kun 1977). Previous research has found mixed evidence for age differences in judgments of ability and effort (Karabenick & Heller 1976; Kun 1977). Kun found that third- and fifth-grade children showed inverse compensation in judging effort, while first graders showed a *positive* relationship between ability and the judged level of effort. However, effort did not have a significant effect on judgments of ability at any age level examined.

Karabenick and Heller (1976) asked subjects to compare two hypothetical persons who were described as succeeding at a task and differing in ability (or effort). Over 90% of subjects as young as first grade indicated that the less able person worked harder to succeed at the task, a finding inconsistent with Kun's results for first-grade children's effort judgments. Karabenick and Heller also found that approximately 70% of the children and 90% of the adults chose the person who worked the least as having the most ability. Procedural differences between the two studies may account for the different findings. Karabenick and Heller held performance constant, perhaps making lesser demands on the children's memory. Age differences in recency effects have been found in achievement and moral judgments, giving some support to the memory interpretation (Feldman, Klosson, Parsons, Rholes, & Ruble 1976; Kun, Parsons, & Ruble 1974).

The present research was designed to resolve the inconsistency in past findings by employing procedures to help the youngest children remember the stimuli. In addition, subjects judged single story characters as well as differences between pairs of characters, allowing evaluation of use of inverse compensation in making the two types of judgments.

### *Causal Schema Hypothesis*

Past research concluded that adults and 10-year-olds combined ability and effort multi-

plicatively while 6-year-olds' predictions of performance showed no effort  $\times$  ability interaction (Kun et al. 1974). The developmental change in judgments of performance predicts that the pattern of judgments of effort and ability should also change, assuming the judgments are based on a causal schema. Research with adults, however, suggests that it may not be possible to represent judgments of ability, effort, and performance by a single causal schema. Anderson and Butzin (1974) found that judgments of ability and motivation did not show interactions that would agree with the multiplicative rule for judgment of performance. An alternative interpretation is that response-scale nonlinearity produced the inconsistent pattern (this possibility has been discussed in detail by Surber, Note 1). Similarly, developmental changes in judgments of performance could be due to changes in use of the response scale rather than in the way ability and effort are subjectively combined.

Birnbaum (1974, 1978; Birnbaum & Viet 1974) has proposed methods for separating the process of combining stimulus information (the integration function) from the process of translating the result into a response on the experimenter's scale (the judgment function). In the present work, judgments of differences between pairs allow application of Birnbaum's "scale-free" method to the question of whether the results of previous research can be attributed to nonlinear judgment functions. (Details of the scale-free method are in the Appendix.)

### *Information Integration Processes*

Both Anderson and Butzin (1974) and Kun et al. (1974) concluded that the integration function for judgments of performance was a multiplicative one. Neither of the previous experiments was designed to test the possibility that an averaging model might adequately represent the data, although Anderson and Butzin suggested it merited further research.

Since an averaging model in which the weight of information varies with the scale value can account for bilinear interactions, such an interaction alone does not distinguish between the multiplying and averaging models. An averaging model predicts that the effect of information presented alone will be greater than its effect when presented in combination with other information. (The details of the predictions of the averaging model may be found in Butzin and Anderson [1973] or Surber [1977].) In contrast to the averaging model, the multiplying model does not predict that the

effect of a cue in isolation should be greater than when it is combined with another stimulus.<sup>1</sup> By obtaining judgments of performance based on only ability information or only effort information, the multiplicative and averaging model can be competitively tested. Similar arguments can be made for judgments of ability and effort.

In summary, there were three purposes to the present research: (1) to clarify developmental trends in the use of inverse compensation in attribution of ability and effort, (2) to apply Birnbaum's scale-free methods to the question of whether there are developmental changes in the combination of ability and effort in predicting performance and to the question of whether it is possible to represent adults' judgments as based on a fully reversible causal schema, and (3) to examine formal algebraic representations of the information integration process competitively.

## Method

### Subjects

Participating in the experiment were 113 adults, 72 children from mixed fifth-sixth grade homerooms (referred to below as fifth graders), 85 children from mixed third-fourth grade homerooms (referred to below as third graders), and 72 kindergarten children. The adults were volunteers from a developmental psychology class, and the children were public school students who received parent permission. The mean ages of the children were 6.04, 9.19, and 11.69 years. Each person made judgments of only one of the three dimensions. The numbers of individuals of each age group who judged ability were 35, 35, 38, and 21; numbers judging effort were 42, 14, 27, and 26; and the numbers judging performance were 36, 23, 20, and 25, for adult through kindergarten, respectively.

<sup>1</sup> A multiplying model really makes no predictions about how a subject should respond to a single piece of information. If it is assumed that the missing piece of information is replaced by the identity operator, however, then the multiplying model predicts that the curve for a single source of information should not cross the curves for the ability-effort combinations. The single-source curve should appear as another curve in the bilinear fan. This adaptation of the multiplying model is equivalent to dropping terms from the equation that pertain to information that is not presented. It is possible for the multiplying model to predict a greater slope for the curve for the judgments of a single type of information (e.g., if the scale values of the other source of information are less than 1.0), but the curve should not cross the other curves. The predictions of a multiplying model are ordinarily the same as the predictions of an additive model. The averaging model predicts that the single-cue curve should both be steeper and cross the other curves. These predictions for both models require that the judgment function be constant over judgments of both single cues and combinations, and that the values of the parameters (scale values and weights) not vary between judgments based on single cues versus combinations. It is not required that the judgment function be linear, however.

### Design

*Judgments of individuals.*—Subjects judged the ability, effort, or performance of hypothetical 8-year-old boys in a physical education class weight-lifting contest. The design for each rated dimension was a  $4 \times 4$  factorial. For example, for performance the design was a 4 (ability)  $\times$  4 (effort) factorial. Each piece of information was also presented separately (randomly interspersed with the other trials), bringing the number of trials to 24. The four stimulus levels of ability were: very very weak, kind of weak, kind of strong, and very very strong. The levels of effort were: didn't try at all, tried a little bit, tried pretty hard, and tried very very hard. The levels of performance described the weight that was lifted: very very light, kind of light, kind of heavy, and very very heavy.

*Judgments of differences between pairs of individuals.*—The design for judgments of differences was the upper triangle of a stimulus boy 1 ( $2 \times 2$ )  $\times$  stimulus boy 2 ( $2 \times 2$ ) factorial design. Each stimulus boy was described by two pieces of information (e.g., ability and effort), either the highest or lowest value of the four levels used in judgments of individuals. There were six trials in the design. For performance, for example, the trials were: low ability-high effort versus low ability-low effort, high ability-low effort versus low ability-low effort, high ability-high effort versus low ability-low effort, high ability-low effort versus low ability-high effort, high ability-high effort versus high ability-low effort, and high ability-high effort versus low ability-high effort.

### Procedures

*Third grade, fifth grade, and adult.*—The stimuli were printed in booklets for the adults and fifth and third graders, who participated in classroom groups. The instructions and stimuli were read aloud to the fifth and third grad-

ers, who followed in their booklets. The adults read the instructions and stimuli and worked independently. For judgments of individuals, each stimulus was printed on a separate page of the booklet with the labeled rating scale. The pages for the 24 trials were arranged in one of two different random orders, and were preceded by four practice trials. For judgments of the differences, the descriptions of each pair of boys were printed on opposite edges of a page with the rating scale between the two. These were preceded by two practice trials, and the order of judgments of individuals and judgments of the differences between pairs was counterbalanced over classroom groups.

*Rating scales.*—The rating scales for judgments of individuals consisted of the numerals 1–7, with the verbal labels for 1, 3, 5, and 7 corresponding to the lowest through highest stimulus levels for the dimension to be rated (e.g., 1 = “didn’t try at all,” for ratings of effort). The rating scale for differences consisted of the numerals 1–9. For ability differences, for example, 5 was labeled “about the same strength,” 1 was printed under the left-hand stimulus and was labeled “very very much stronger,” and 9 was under the right-hand stimulus and was labeled “very very much stronger.” The instructions specified that one of the numerals 1–4 should be circled if the boy described on the left were stronger than the boy on the right, etc. The rating scales for differences in effort and performance were labeled similarly.

*Kindergarten.*—The stimuli were presented pictorially and verbally to the kindergarten children, who were tested individually. The levels of strength were represented by four line drawings of males, who were identical except for the size of their muscles, and were labeled verbally exactly as for the older children and adults. When ability was the variable to be rated, three additional drawings were added to make a series of seven. Performance was represented by line drawings of different-sized barbells, and effort was represented by squares varying in size. The largest square was verbally labeled “tried very very hard,” etc. The kindergarten children made their ratings by pointing to one of the seven pictures for whatever dimension was to be rated.

To maximize attention to the stimuli, the child was required to select from an array of the four stimuli the pictures that were appropriate for each stimulus combination presented by the experimenter. For example, if the stim-

ulus was described as “very very strong” and “didn’t try at all,” the child would choose the drawing of the strongest boy and the smallest square, and would then make a judgment.

For the kindergartners’ difference judgments, the stimuli for each member of the pair were placed on opposite edges of the table. The child was first asked whether one boy was stronger (or tried harder or lifted more), followed by rating of “how much greater” by pointing to one of a series of four lines differing in length. The order of judgments of individuals and of differences was counterbalanced over subjects.

## Results

### *Development of Inverse Compensation*

*Effort judgments.*—The mean judgments of weight-lifting effort in figure 1 show that judged effort decreases as the level of ability increases at each age level. The main effect of ability in a 4 (age)  $\times$  4 (ability)  $\times$  4 (performance) analysis of variance was significant,  $F(3,315) = 52.15$ ,  $p < .01$ , while the age  $\times$  ability interaction was not,  $F(9,315) = 1.65$ ,  $p > .10$ . In separate analyses of variance, the main effect of ability was significant at each age level with the exception of the fifth grade, where it approached significance,  $F(3,123) = 55.77$ ,  $F(3,39) = 3.50$ ,  $F(3,78) = 9.54$ ,  $F(3,75) = 6.39$ , for adult through kindergarten,  $p$ 's  $< .01$ , except fifth grade,  $p < .05$ . There were also interactions of performance  $\times$  age,  $F(9,315) = 3.74$ ,  $p < .01$ , and performance  $\times$  ability,  $F(9,945) = 3.05$ ,  $p < .01$ . The performance  $\times$  age interaction appeared to be due to the fact that the performance effect for the kindergartners was greater than for the other age groups.

*Analysis of individual data patterns.*—The direction of the ability effect was examined for individuals by taking the difference between the ratings of the highest and lowest levels of ability at each level of performance. If the majority of the four differences were negative, the individual was classified as having a negative slope, and if the majority were positive, the individual was classified as having a positive slope. Individuals with two positive differences and two negative differences, or with all zero differences, were excluded. Almost all of the adults judged effort to be an inverse function of ability (37 out of 42), as did the fifth graders (11 out of 14), third graders (21 out of 27), and kindergarten children (19 out of 26). A  $\chi^2$  test for age differences in positive and negative effects of ability was nonsignificant,

$\chi^2(3) = 3.511, p > .10$ . Thus, the individual data agree with the conclusions based on the mean judgments.

*Effort difference judgments.*—The judgments of differences between pairs that differ in ability but not performance should reflect the inverse effect of ability found for judgments of individual characters. For all age levels these comparisons were significantly in the predicted direction from the midpoint of the response scale, all  $p$ 's  $< .01$ ,  $\bar{X}$ 's = 6.755, 6.465, 7.050, 6.520, for adult through kindergarten, respectively). Regardless of whether the children are asked to judge the effort of an individual or the difference in effort be-

tween pairs, children of kindergarten age and older use the principle of inverse compensation when memory for the stimuli is assured. Thus the results replicate Karabenick and Heller's finding that young children use inverse compensation in judging effort, and extend the finding of inverse compensation to judgments of single individuals as well.

*Ability Judgments*

The mean judgments of ability are presented in figure 2. In a 4 (age)  $\times$  4 (effort)  $\times$  4 (performance) analysis of variance, the significant age  $\times$  effort interaction showed that there are age differences in the way effort is used to infer ability,  $F(9,375) = 4.42, p < .01$ .

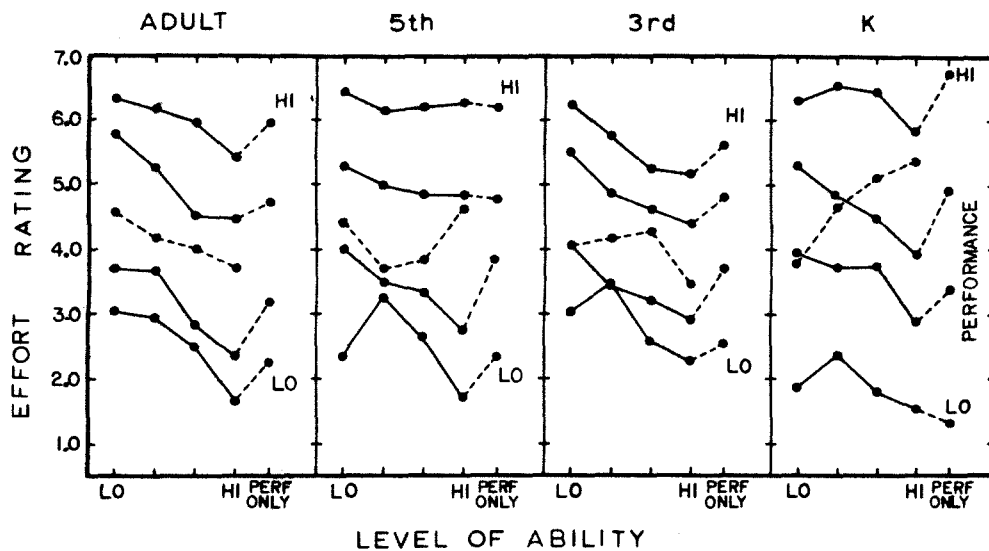


FIG. 1.—Mean judgments of weight-lifting effort. The data of each age group are presented in a separate panel, with ability on the abscissa and a separate curve for each level of performance. Points connected by dashed lines represent the mean judgments based on single cues. Points connected by solid lines represent the mean judgments based on ability and performance combined.

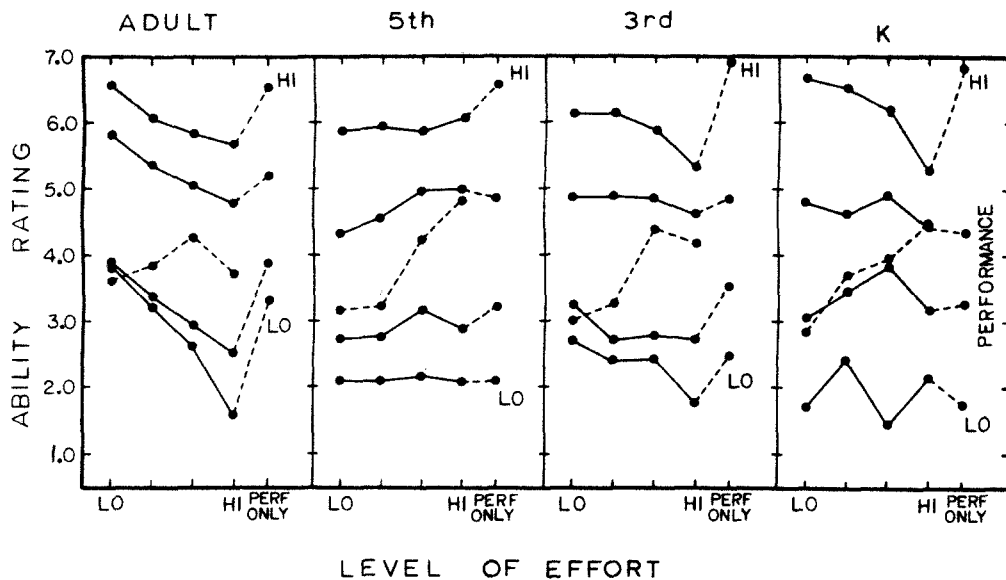


FIG. 2.—Mean judgments of weight-lifting ability. The data of each age group are presented in a separate panel, with effort on the abscissa and a separate curve for each level of performance. Points connected by dashed lines represent the mean judgments based on single cues. Points connected by solid lines represent the mean judgments based on effort and performance combined.

Separate analyses showed a significant main effect of effort at only the adult level,  $F(3,102) = 28.16$ ,  $F(3,102) = 1.25$ ,  $F(3,108) = 3.23$ ,  $F(3,60) = 0.69$ , for adult through kindergarten, respectively, although the third-grade effort effect approached significance,  $p < .05$ . Only the adults show strong evidence of use of inverse compensation in judging ability.

The age  $\times$  effort  $\times$  performance interaction was also significant,  $F(27,1125) = 2.57$ ,  $p < .01$ . Separate analyses revealed significant effort  $\times$  performance interactions for the adults and kindergartners,  $F(9,306) = 5.13$ ,  $F(9,180) = 3.03$ ,  $p$ 's  $< .01$ . The interaction of effort and performance in the adult data appears to be due to the fact that the curves diverge as the level of effort increases.

*Individual data patterns.*—Individuals were classified as showing a positive or negative effect of effort on their ability judgments in a manner analogous to that used in the effort ratings. A  $\chi^2$  test for age differences in positive and negative slopes was significant,  $\chi^2(3) = 16.78$ . The mean ability judgments of the positive- and negative-slope groups at each age level are presented in figure 3. (The

adults were excluded, since only a few did not show inverse compensation.) One group judges ability to increase as effort increases, while the other group judges ability to decrease as effort increases. Analysis of variance at each age level showed significant group  $\times$  effort interactions,  $F(3,90) = 62.47$ ,  $F(3,96) = 31.75$ ,  $F(3,51) = 11.81$ ,  $p$ 's  $< .01$ , for grades 5 through K, respectively.<sup>2</sup>

*Ability difference judgments.*—The difference judgments of the slope groups reflect the effort effects seen in figure 3. For fixed levels of performance, the difference judgments of the negative-slope groups are larger than for the positive-slope groups,  $\bar{X}$ 's = 6.50, 6.75, 6.86 for negative and 4.10, 4.77, 5.06 for positive groups, grades 5 through K, respectively. This holds at all three age levels, and the group main effects of the fifth and third graders were significant, while the kindergarten effect approached significance,  $F(1,29) = 6.34$ ,  $F(1,27) = 5.80$ ,  $F(1,19) = 4.03$ ,  $p$ 's  $< .05$ , .05, .10, for grades 5 through K, respectively.<sup>3</sup>

#### Judgments of Performance

The mean judgments of performance are presented in figure 4. The curves appear to

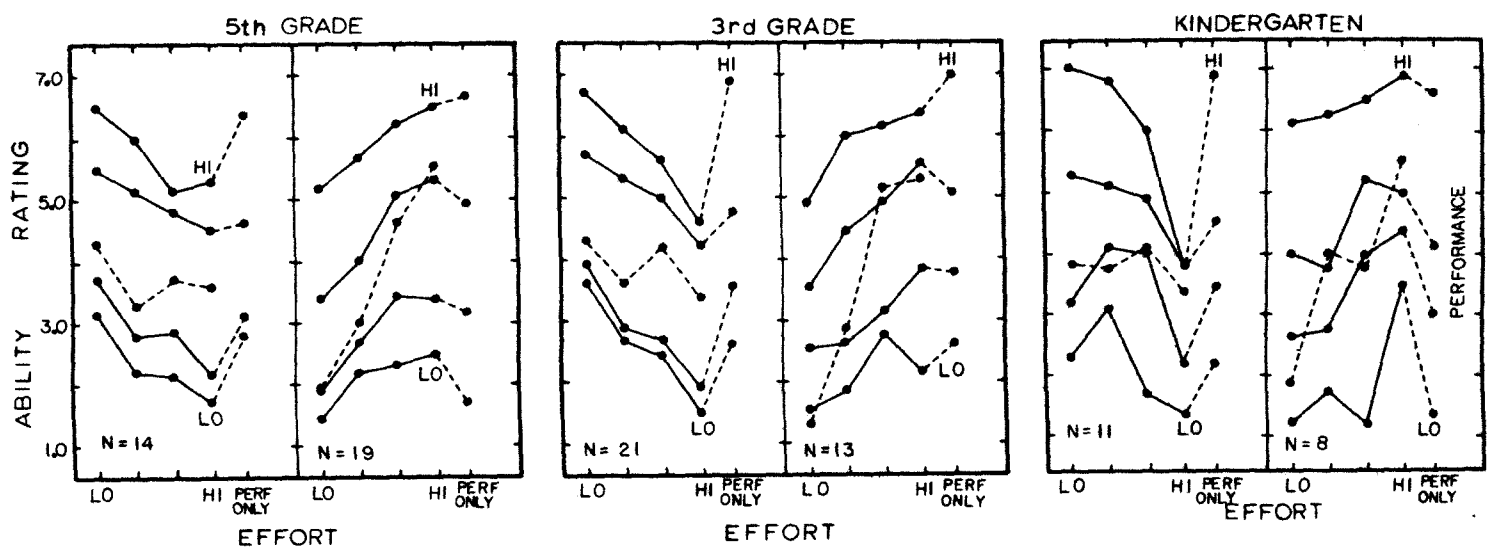


FIG. 3.—Mean judgments of weight-lifting ability for the kindergartners and third and fifth graders. The left-hand panel for each age group presents the mean judgments of the negative-slope group, and the right-hand panel presents the mean judgments of the positive-slope group. Points connected by dashed lines represent the mean judgments based on single cues. Points connected by solid lines represent the mean judgments based on effort and performance combined.

<sup>2</sup> It could be argued that the individual differences in use of effort to infer ability are due to chance. If the classification capitalizes on chance, an analysis of variance excluding the highest and lowest levels of effort (on which the classification was based) should show no group  $\times$  effort interaction. These interactions were significant at each age level, however,  $F(1,31) = 39.00$ ,  $F(1,32) = 9.52$ ,  $F(1,17) = 4.66$ ,  $p$ 's  $< .01$ , .01, .05, for grades 5 through K, respectively. Two other sources of evidence described below converge in supporting the slope classifications: (a) judgments of differences in ability and (b) judgments of ability based on effort information alone.

<sup>3</sup> For analyses of the individual difference groups,  $\alpha = .05$  is adopted, since power is reduced in these comparisons.

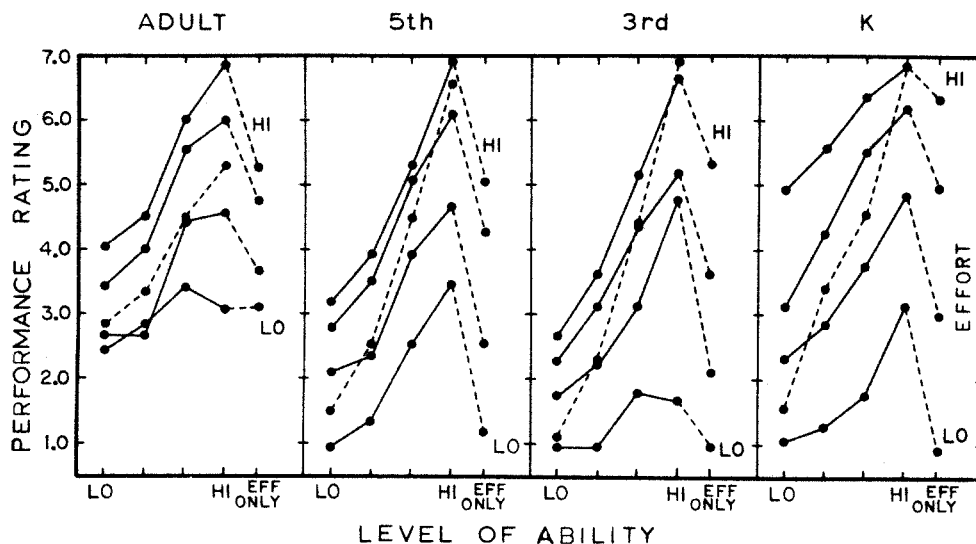


FIG. 4.—Mean judgments of weight-lifting performance. The data of each age group are presented in a separate panel, with ability on the abscissa and a separate curve for each level of effort. Points connected by dashed lines represent the mean judgments based on single cues. Points connected by solid lines represent the mean judgments based on ability and effort combined.

diverge as the level of ability increases, except in the kindergarten data, where the curves diverge at the middle levels of ability and converge again at the highest level of ability. The age  $\times$  ability  $\times$  effort interaction was significant,  $F(27,873) = 2.75$ ,  $p < .01$ . In separate analyses the ability  $\times$  effort interaction was significant at the adult and third-grade levels,  $F(9,315) = 8.63$ ,  $F(9,153) = 10.18$ ,  $p$ 's  $< .01$ ; approached significance at the kindergarten level,  $F(9,216) = 2.10$ ,  $p < .05$ ; and failed to attain significance in the fifth-grade data,  $F(9,189) = 1.85$ .

According to a multiplying model of judgments of performance, the ability  $\times$  effort interaction should be concentrated in the bilinear component. The test of the bilinear component was significant for the adults,  $F(1,35) = 18.61$ ; fifth graders,  $F(1,21) = 8.80$ ; and third graders,  $F(1,17) = 48.82$ ; but was not significant for the kindergarten children,  $F < 1$ . The bilinear component accounted for the majority of the interaction for all except the kindergarteners (percentage of interaction in bilinear

components = 68, 68, 82, and  $< 1$ , for adult through kindergarten, respectively). These percentages are close to those reported by Kun et al. (1974). The test of the residual from the bilinear component was nonsignificant for the fifth graders,  $F(8,168) < 1$ , but reached significance for the adults,  $F(8,280) = 4.52$ , and kindergarteners,  $F(8,192) = 2.64$ , and approached significance for the third graders,  $F(8,136) = 2.15$ .<sup>4</sup> The residuals for the adult and third-grade groups could be caused by the small nonmonotonicities at the lowest level of effort. These effects might disappear if either more practice trials or more replications were used. Nevertheless, the residual  $F$ 's are not large compared to the bilinear  $F$ 's for the adults and third graders. For the kindergarteners, the significant residual is expected, since their data showed no bilinear component.

It is important to note that a significant bilinear interaction can be produced in several ways. One possibility is a linear integration rule (e.g., additive or constant weight averaging) combined with an exponential judgment

<sup>4</sup> The coefficients used for the bilinearity test were the differences between the grand mean and the marginal means for each age group (Shanteau's [1977] POLYLIN program was used). These coefficients allow for the subjective spacing of the stimuli at the group level. These same coefficients were used in computing the  $SS_{S \times L \times L}$  to be used in the error term for the  $F$  test. The test of the residual from bilinearity was

$$(SS_{Res}/8) / \left[ \left( \sum_{\text{subjects}} SS_{Res} - SS_{Res} \right) / (8)(N - 1) \right],$$

where  $SS_{Res}$  in the numerator is equal to  $SS_{A \times B} - SS_{L \times L}$ . The denominator of this test is  $MS_{S \times Res}$ . Each individual's  $SS_{Res}$  is equal to that subject's  $SS_{A \times B}$  minus  $SS_{L \times L}$  for that subject. This analysis does not allow for individual variability within age group in subjective spacing of the stimuli. However, an exact solution that does allow subjective spacing for individuals is not available.

function (Birnbau 1978; Birnbau & Mellers 1978). Another possibility is some sort of non-linear integration rule (e.g., a multiplicative model or certain types of differentially weighted averaging models) in combination with a judgment function that is either a linear or a power function (Birnbau 1978). Thus, the further constraints provided by the scale-free test are needed in order to decide between the general class of linear versus nonlinear integration rules.

*Performance difference judgments.*—If the developmental changes in judgments of performance reflect the manner in which the information is combined rather than change in use of the response scale, then the results of the scale-free test should be similar. The difference judgments of the adults were significantly in the direction predicted by the data in figure 4,  $F(1,105) = 8.21$ ,  $p < .01$ ; those of the fifth graders approached significance,  $F(1,63) = 5.42$ ,  $p < .05$ ; but those of the third-grade and kindergarten groups failed to reach significance,  $F(1,42) = 1.05$ ,  $F(1,72) = 3.51$ ,  $p$ 's  $> .05$ . At the adult and fifth-grade levels the scale-free test can be interpreted as evidence that the interaction in the judgments of the performance of individuals (fig. 4) is not due to nonlinearity in the rating scale.

For the third-grade and kindergarten samples, the difference judgments are in the direction predicted by a divergent interaction. Since there was no grade  $\times$  stimulus interaction in an analysis of variance of the difference judgments,  $F(3,285) = 1.40$ ,  $p > .10$ , the best conclusion appears to be that the interactions in the third-grade and kindergarten samples are *not* due to a nonlinear response scale. In order to be more definitive, however, a study is needed in which the assumptions behind the scale-free method can be tested developmentally.<sup>5</sup>

#### *Representation of Integration Processes*

*Judgments of performance.*—The predictions of an averaging integration process can be evaluated by comparing the points connected by dashed lines with the points connected by solid lines in figure 4. With the exception of the adults, the effects of the ability and effort cues presented in isolation (dashed lines) are greater than when they are combined (solid lines) and cross over the solid curves. This comparison provides an ordinal

qualitative test that is inconsistent with both additive and multiplying models, but is predicted by an averaging representation.

At the adult level, the effects of the ability and effort cues in isolation are approximately the same as when presented in combination. These data are consistent with both the multiplicative and additive models. For the adult data, the multiplicative representation seems preferable, since it can readily account for the interaction of ability and effort as well as the judgments based on single cues.

*Effort.*—The adult data are inconsistent with an averaging integration process, since the effects of both ability and performance in isolation are approximately the same as their effects when they are combined (see dashed curves in fig. 1). For the children, the effect of performance alone is approximately the same as the effect of performance in combination with ability, and the effort judgments based on ability information presented in isolation show no consistent trend (except at the kindergarten level, where there is a positive trend). Thus, the children's judgments of effort are not readily described by either an averaging or an adding model.

*Ability.*—The adult judgments of ability in figure 2 also present difficulties for either an averaging or an adding integration function. The adults' judgments of ability based on effort alone (points connected by dashed lines in fig. 2) show neither a distinct positive nor a negative trend, a finding inconsistent with both additive and averaging integration processes. The judgments of ability based on performance in isolation (points connected by dashed lines at the far right of the adult panel) also do not agree with an averaging model. The data of those children showing a negative effect of effort in their ability judgments resemble the adult results and create the same difficulties for adding and averaging integration processes (see left-hand panel for each age group in fig. 3).

For those children showing *positive* effects of effort on their ability judgments, a different pattern emerges. The effects of effort information alone (dashed curves) and performance alone are greater than when presented in combination, as predicted by an averaging inte-

<sup>5</sup> The design for the difference judgments allows a minimal ordinal test of the subtractive model. No ordinal violations were found in the performance difference judgments. Across all three judged dimensions and four age groups, there was only one ordinal violation of the subtractive model. This compares well with research conducted with adults.



gration process. These results show that the transition to inverse compensation brings a change in the nature of the integration process. When a positive effect of effort is observed, the data agree with an averaging model. Once inverse compensation is present, however, the data do not agree with an averaging process.<sup>6</sup>

### *Evidence Relevant to the Causal Schema Hypothesis*

At the adult level the results are inconsistent with the hypothesis that judgments of ability, effort, and performance can be represented by a single causal schema. The interaction in judgments of performance (see fig. 4) is supported by the scale-free test using judgments of differences and is unlikely to be due to a nonlinear response scale. There was also an interaction in the adult judgments of ability, but it did *not* agree with the interaction in the performance judgments. The performance interaction predicts that the ability judgments should converge as the level of effort increases, while the data diverge. The scale-free test using the judgments of differences in ability was consistent with the interaction in figure 2,  $F(1,102) = 6.64$ ,  $p < .025$ . At the adult level, the data of both judgments of individuals and differences between pairs are inconsistent with the hypothesis that judgments of ability, effort, and performance are based on a single causal schema.

## Discussion

### *Reversible Operations*

The findings question the existence of "fully reversible" thinking even at the adult level, since the interactions in the adult ability and performance judgments are inconsistent with each other. Although Anderson and Butzin (1974) concluded that adults' ability, effort, and performance judgments were inconsistent across dimensions, the present work extends their conclusion by decreasing the plausibility of response-scale nonlinearity as a source of the inconsistency. Recent findings for adult judgments of velocity, distance, and time are also similar to the present findings (Wilkening 1979). Together, the results suggest that even college-age subjects may lack the cognitive

mechanisms to form a completely reversible cognitive structure.

It is possible that in the present research the adults failed to show a single schema because the information they have acquired from their social experience is inconsistent in some ways. This hypothesis could be indirectly tested with a nonsocial causal inference task in which experiences with the variables and their interrelationships are controlled. If under such conditions adults still fail to show a single causal schema in their judgments, then it would begin to be reasonable to attribute the lack of a single schema for achievement judgments to lack of necessary cognitive operations.

The results also suggest that children's judgments are not based on a single causal schema. In the children's data, ability is inversely related to judgment of effort, while effort information is not necessarily inversely related to judged ability. Inconsistent application of inverse compensation shows that, although neither the children's nor the adults' data are completely consistent with a single causal schema, there is developmental progression in cognitive organization. It is possible that at an age younger than kindergarten, children are consistent in *failing* to use the principle of inverse compensation. Such a result would not necessarily imply use of a single schema in making the judgments. Instead, it might suggest a lack of the concept of "causation" and a reliance on the notion of covariation.

### *Inverse Compensation and Concepts of Ability and Effort*

The present research clarifies the developmental course of use of inverse compensation in judging ability and effort. By assuring memory for the stimuli, and by having children judge both individual story characters and the differences between pairs, the inconsistency between Kun's and Karabenick and Heller's results has been resolved. By kindergarten age, children are aware of the inverse relationship between ability and effort, but they do not necessarily apply the skill in judging both variables. A larger proportion of children showed inverse compensation when judging effort than

<sup>6</sup> The failure of the judgments of effort based on ability only (and vice versa) to conform to the predictions of simple algebraic models could indicate any of a number of possible violations of the assumptions of possible models. In some cases the assumption of parameter invariance is obviously violated, raising questions about the cognitive processes underlying the judgments. For example, it is possible that different models apply to judgments of combined stimuli and judgments of single stimuli.

when judging ability.<sup>7</sup> One possible explanation for the inconsistency across dimensions is that inverse compensation for ability is acquired later because ability is experienced only indirectly by children (Kun 1977). Unfortunately, this would not explain why most of the children infer effort from ability by using inverse compensation, but only approximately half of them infer ability from effort by using inverse compensation.

Another possibility is that children are conceptualizing ability differently when judging it versus when they are using it to infer effort. Perhaps children who infer ability to be a direct function of effort are expressing a belief that higher effort results in higher ability (this was also suggested by Kun [1977]). An athlete who trains hard will, on the average, develop or *acquire* more ability (strength, speed, etc.) than an athlete who does not train as hard. Thus, the concept of ability as a person's current potential in a given area is not being judged. When asked to infer effort from ability and performance, however, the influence of current effort on future ability is irrelevant, since the child is given information about the story character's *current* ability.

Another approach to explaining the asymmetry in judgments of ability and effort is to propose that it may be based on a general cognitive development in the understanding of causal relations. If inconsistent use of inverse compensation is a normal step in the development of causal inferences for both social and nonsocial events, the development of mature concepts of effort and ability may depend on the cognitive developments involved in causal thinking. Undoubtedly, both cognitive development and social experiences influence the development of social attributions, and disentangling these influences is an important goal for future research.

#### *Developmental Change in Integration Processes*

The judgments of performance for all except the adults are consistent with an averaging integration process. This questions the conclusions of earlier research that by approximately 10 years of age the integration process is a multiplicative one (Kun et al. 1974). Although interactions were obtained and were supported

by the scale-free test, the results of the judgments of performance based on effort alone or ability alone are difficult to reconcile with the multiplicative model, except at the adult level. Other recent research suggests that the averaging model may provide the best account for the performance judgments of adults as well (Surber, Note 2). Developmental change in predictions of performance might then be parsimoniously represented by a change in the way the weights vary with the scale values.

A novel aspect of the present work is the attempt to extend algebraic models of information integration to attributions of ability and effort. For subjects who showed inverse compensation, the judgments of ability and effort are difficult to reconcile with simple algebraic models of information integration. For example, the predictions of the additive, ratio, and averaging models are contradicted by the adult judgments of ability based on effort alone. It seems clear that mature subjects bring different cognitive processes to bear on attributions based on only a cause (e.g., ability only or effort only) than on attributions based on both a cause and the effect. These different cognitive processes may be producing variation in the parameters between judgments based on only a cause and those based on both a cause and an effect. Mature causal inference processes may involve an information integration process that is essentially *configural*—one in which the interpretation of a piece of information depends on the other information presented with it. This provides an interesting puzzle for future research. In contrast, for the children who judged ability to be directly related to the given level of effort, the data are predicted by an averaging representation.

The shift in the integration process with the developmental shift to inverse compensation could be interpreted as a change from correlational thinking to more genuine causal thinking. Piaget describes concrete operational children as dealing with causal inference tasks by formulating "correspondences" or "simple correlations" between events, while early formal operational children consider "necessary relations" (Inhelder & Piaget 1958, p. 18). The present data, as well as those of Kun and of Karabenick and Heller, seem to suggest that

<sup>7</sup> It could be argued that individual children are consistent and that the apparent inconsistency across dimensions could be due to a random sampling difference in the two conditions. Both Kun's and Karabenick and Heller's data showed intraindividual variation in use of inverse compensation. Thus, it seems unlikely that the inconsistency across dimensions is due to sampling differences in the present study.

consistent use of inverse compensation in the achievement realm is not firmly established until approximately the early formal operational level.

#### Note on Judgment Methodology

It is occasionally argued that, in order to discover structural changes, methods in which the child is asked for verbal explanations are preferred (Larsen 1977; Livesley & Bromley 1973). The prototypical example is the issue of whether to use "judgments only" or "judgments plus explanations" in determining whether a child conserves (e.g., Brainerd 1974; Reese & Schack 1974). The present research, as well as that of Butzin (1979) and Cuneo (Note 3), shows that appropriate use of judgment methodology can reveal qualitative stagelike shifts in children's use of information. The shift to using inverse compensation in judging ability and effort appears to be a stagelike transition. It is also worth noting that algebraic representations of the information integration process potentially allow description of both gradual quantitative changes (as change in parameter values) and stagelike or structural changes (as a change in the form of the equation). Thus, formal models of human judgment provide one possible method for describing the two major hypothesized types of developmental change.

Important issues in the "judgments only" or "judgments plus explanations" controversy are whether the methods lead to different conclusions, and if they do, why. Nicholls (1978) has suggested that the two methods *do* lead to different conclusions about children's concepts of ability and effort, and that asking children for explanations is more likely to reveal use of inverse compensation than is judgments alone. (See Brainerd [1977] for a discussion of the issue of measurement error in use of judgments versus judgment plus explanations.) The present research, as well as that of Karabenick and Heller (1976), converges with Nicholls's conclusion that 6-year-old children do understand something about inverse compensation between effort and ability. Nevertheless, future research might profitably examine the empirical relationship between verbal explanations of concepts (such as those collected by Nicholls) and the same concepts as revealed by judgments.

#### Appendix

An additive model for task performance judgments based on ability and effort can be written:

$$R = J(w_A s_{Ai} + w_E s_{Ej}),$$

where  $w_A$  and  $w_E$  are weights that are constant over all levels of  $s_{Ai}$  and  $s_{Ej}$ . The  $s_{Ai}$  and  $s_{Ej}$  are the levels of ability and effort, and  $J$  is any monotonic function. The multiplicative model can be written

$$R_{ij} = J(s_{Ai} \times s_{Ej}).$$

If subjects are asked to judge the differences between pairs of stimuli, where each stimulus is an ability-effort combination, one need only assume that the judged differences are monotonically related to the psychological differences; that is,

$$R_d = J(\Psi_{ij} - \Psi_{kl}),$$

where  $\Psi_{ij}$  and  $\Psi_{kl}$  are the psychological impressions of the stimulus compounds (where each stimulus compound is an ability-effort combination). Under this assumption, the additive model predicts that, for pairs of stimuli involving equal ability information, the judgment should depend only on the difference in effort:

$$\begin{aligned} R_d &= [(w_A s_{Ai} + w_E s_{Ej}) - (w_A s_{Ai} + w_E s_{Ek})] \\ &= J[w_E (s_{Ej} - s_{Ek})]. \end{aligned}$$

An analogous conclusion holds for pairs of stimuli differing only in ability. This implies that if ability and effort are combined in an additive fashion, then the judgment of the difference between high ability-low effort and high ability-high effort should be equal to the judged difference between low ability-low effort and low ability-high effort, for example.

In contrast, the multiplicative model predicts that the judged difference will depend on *both* the difference in effort and the level of ability:

$$\begin{aligned} R_d &= J[(s_{Ai} \times s_{Ej}) - (s_{Ai} \times s_{Ek})] \\ &= J[s_{Ai} (s_{Ej} - s_{Ek})]. \end{aligned}$$

Thus, the judged difference should vary directly with the value of ability as well as with the difference in effort.

The above predictions of the difference judgment task are ordinal predictions that do not depend on the form of the  $J$  function, as long as it is monotonic. Thus, the multiplicative (or any nonadditive) integration function can be differentiated from an additive integration function at each age level, regardless of possible developmental changes in the form of the  $J$  function. The difference judgment task does make two other assumptions, however. First, it assumes that the process of comparing stim-

uli to judge the difference can be represented by a subtractive model (i.e.,  $R_d = J[\Psi_{ij} - \Psi_{kl}]$ ). Second, it assumes that the process of combining ability and effort to predict performance (for example) is not altered by asking for difference judgments. Extensive experimental work with adults has provided support for the subtractive model for difference judgments (see Birnbaum 1978).

## Reference Notes

1. Surber, C. F. Organization in social inference: is there a schema for judgments of ability, effort, and task performance? Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, 1978.
2. Surber, C. F. Effects of information reliability in predicting task performance using ability and effort. Paper presented at the annual convention of the Midwestern Psychological Association, St. Louis, May 1980.
3. Cuneo, D. H. Judgments of numerosity: developmental changes in the integration of length and density. Paper presented at the meeting of the Psychonomic Society, Washington, D.C., November 1977.

## References

- Anderson, N. H., & Butzin, C. A. Performance = motivation  $\times$  ability: an integration-theoretical analysis. *Journal of Personality and Social Psychology*, 1974, **30**, 598-604.
- Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology*, 1974, **102**, 543-561.
- Birnbaum, M. H. Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory*. Vol. 3. Hillsdale, N.J.: Erlbaum, 1978.
- Birnbaum, M. H., & Mellers, B. A. Measurement and the mental map. *Perception and Psychophysics*, 1978, **23**, 403-408.
- Birnbaum, M. H., & Viet, C. Scale-free tests of an additive model for the size-weight illusion. *Perception and Psychophysics*, 1974, **16**, 276-282.
- Brainerd, C. J. Postmortem on judgments, explanations and Piagetian structures. *Psychological Bulletin*, 1974, **81**, 70-71.
- Brainerd, C. J. Response criteria in concept development research. *Child Development*, 1977, **48**, 360-366.
- Butzin, C. A. Children's moral judgments of ulterior motives. *Abstracts of Individual Papers, Society for Research in Child Development, Biennial Convention*, 1979, **2**, 208.
- Butzin, C. A., & Anderson, N. H. Functional measurement of children's judgments. *Child Development*, 1973, **44**, 529-537.
- Feldman, N.; Klosson, E.; Parson, J.; Rholes, W.; & Ruble, D. Order of information presentation and children's moral judgments. *Child Development*, 1976, **47**, 556-559.
- Inhelder, B., & Piaget, J. *The growth of logical thinking from childhood to adolescence*. New York: Basic, 1958.
- Karabenick, J. D., & Heller, K. A. A developmental study of effort and ability attribution. *Developmental Psychology*, 1976, **12**, 559-560.
- Kelley, H. H. The processes of causal attribution. *American Psychologist*, 1973, **28**, 107-128.
- Kun, A. Development of the magnitude-covariation and compensation schemata in ability and effort attributions of performance. *Child Development*, 1977, **48**, 862-873.
- Kun, A.; Parsons, J.; & Ruble, D. Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology*, 1974, **10**, 721-732.
- Larsen, G. Y. Methodology in developmental psychology: an examination of research on Piagetian theory. *Child Development*, 1977, **48**, 1160-1166.
- Larsen, G. Y., & Flavell, J. H. Verbal factors in compensation performance and the relation between conservation and compensation. *Child Development*, 1970, **41**, 965-977.
- Livesley, W. J., & Bromley, D. B. *Person perception in childhood and adolescence*. New York: Wiley, 1973.
- Nicholls, J. G. The development of the concepts of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development*, 1978, **49**, 800-814.
- Reese, H. W., & Schack, M. L. Comment on Brainerd's criteria for cognitive structures. *Psychological Bulletin*, 1974, **81**, 67-69.
- Shanteau, J. POLYLIN: a FORTRAN IV program for the analysis of multiplicative (multilinear) trend components of interactions. *Behavior Research Methods and Instrumentation*, 1977, **9**, 381-382.
- Surber, C. F. Developmental processes in social inference: averaging of intentions and consequences in moral judgment. *Developmental Psychology*, 1977, **13**, 654-665.
- Wilkening, F. Integrating time, distance and velocity information. *Abstracts of Individual Papers, Society for Research in Child Development, Biennial Convention*, 1979, **2**, 208.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.