Order Number 9521275

Development of peer evaluations: Controllability, effort and ability

Goldhammer, Eva A., Ph.D.

City University of New York, 1995

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

DEVELOPMENT OF PEER EVALUATIONS: CONTROLLABILITY, EFFORT AND ABILITY

by

EVA A. GOLDHAMMER

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1995

This manuscript has been read and accepted by the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Jan. 4, 1995 — Herbert D. Saltzstein

Date — Professor Herbert D. Saltzstein
Chair of Examining Committee

January 4, 1995 — Kay Deaux

Date — Professor Kay Deaux
Executive Officer

THE CITY UNIVERSITY OF NEW YORK

Abstract

DEVELOPMENT OF PEER EVALUATIONS:

CONTROLLABILITY, EFFORT AND ABILITY

by

Eva A. Goldhammer

Adviser: Professor Herbert D. Saltzstein

Children in grades 1, 3 and 6 observed and evaluated a peer's performance on a concrete game-like task under two conditions: a) where outcome was under the partial control of the player (skill) and b) where it was not (luck). There were two difficulty levels in each condition, and at each level, outcome was matched between the skill and luck conditions by means of a yoking procedure.

Younger children rewarded more if the outcome was successful, regardless of whether the task involved luck or skill. Older children rewarded more for performance in the skill than in the luck condition, and only within the skill condition according to level of difficulty. The results are discussed in terms of children's understanding of causality and personal control.

ACKNOWLEDGMENTS

I would like to thank my dissertation advisor, Professor Herbert D. Saltzstein, for the many years of patient and supportive training and encouragement, which culminated in this project. He is a uniquely intelligent and decent man. I could not have done it without him.

I am also grateful to my husband, parents and parents-in-law for all their help and moral support. Much love to my sons, Bennett, Andrew and Michael Schwarz who were all born and raised during my journey as a graduate student. They are always there to inspire me by bringing my textbooks about child development to brilliant, glowing life.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Figures-Continued

Chapter I

INTRODUCTION

Children regularly evaluate each others' performance. Such evaluation occurs across a range of tasks, including games of skill and of chance. These evaluations include tangible and social rewards, based on assessment of performance. Performance evaluations are rendered by all members of society, for example, by employers, supervisors, teachers, consumers, voters and parents. The appropriateness of adult assessments and rewards contributes to the more efficient functioning of every aspect of society. Indeed, it can be assumed that many people would not improve their performance without appropriate feedback from the environment and would not maximize their efforts without the expectation of contingent rewards. This is how development interacts with the environment, injecting much needed informativeness. This is one example of acting upon the environment, exacting changes in both the environment and developmental changes in oneself, as Piaget described it. Therefore, how the assessment of achievements develops becomes consequential.

There is evidence of a systematic developmental change in how children evaluate their peers in the moral judgment domain, where it has been shown that young (pre-operational)

children frequently consider the outcome to be the most relevant feature of an event. Many studies, including those reviewed by Keasey (1977), have demonstrated a shift from an earlier reliance sometimes on the outcome of an act and sometimes on the actor's intentions or motives, to a more exclusive reliance on the actor's intentions. This trend was brought to general notice by Piaget (1932). He had children of different ages comparing the child who broke a single teacup as he was performing an act which was harmless but self-gratifying (climbing to an upper shelf to reach for jam), as opposed to the child who broke many teacups as he was obeying his mother. Younger children were found to use the amount of property damage (one cup vs. many cups - clearly an outcome criterion), to evaluate the level of "naughtiness" of the actor child. Older children more often (though not exclusively) evaluated the child in the story according to his motivations or intentions (to gratify himself vs. to help his mother).

There are several reasons for shifting the focus of the investigation to achievement evaluations. First, there is a common underlying cognitive structure which these two domains share. This is causal reasoning. It has been argued that changes in evaluative judgments are driven by changes in children's conceptions of causality. For example, Saltzstein (1987) reported a relationship between children's moral judgments of acts with unintended outcomes and their causal

attributions of these outcomes. However, these findings were based on children's evaluations of hypothetical situations where the actual level of causality (controllability) was not under experimental control, thus making it unclear whether causal reasoning affected evaluation, or vice versa.

Furthermore, moral judgments and achievement evaluations have certain features in common as well. Both demonstrate young children's initial reliance on the outcome, which is concrete. This is represented as good or bad in moral judgment; success or failure in achievement evaluations. Later, as children develop, they shift to more complex as well as more inferential criteria, such as intentions (in moral judgment) and levels of skill and effort (in achievement evaluations).

In order to determine whether a child was at fault for breaking teacups, or any other undesirable outcome the immediate underlying cause has to be determined. Either it is internally attributed to the child, in which case he or she is to blame, or it is externally attributed to some other cause, in which case he or she is not to blame. By the same token, when evaluating another child's achievement, it must be determined whether that outcome should be internally attributed (i.e., reliant on the skill or motive of the performer), or externally attributed (i.e., reliant on task difficulty or luck).

It should also be acknowledged that in moral judgments,

the determiners of the child's judgments are relatively simple. Outcomes are good or bad. They are intended or not, caused or random. However, in the domain of achievement evaluations, the determiners of the child's judgment are far richer and more complex. Outcomes can be good, or bad, to various degrees, depending on the level of difficulty built into the task. Obviously, the more difficult a task is, the "better" the accomplishment is likely to be perceived. They can be intended or not, controllable or not, or partially so.

There are several difficulties inherent in the previously described hypothetical moral judgment situations, and the legions of similar ones that have been reported. First, moral judgment situations can remind children of situations they have been involved in, and perhaps even been punished for, generating anxiety. Second, any hypothetical presentation of material relies on a child's linguistic processing skills, which puts younger children at a decided disadvantage. Third, hypothetical situations may not feel as immediate to children, tend to rely on their memory, and therefore, their responses are less ecologically valid and because of this, less generalizable.

Therefore, the strategy behind the current study was to focus on evaluations of achievements, and not on moral judgments. Additionally, it was distanced from the hypothetical format, as described above. This study

investigated children's evaluations of other children and their performance on a concrete task, which the observer child had witnessed. Of course, it can not be assumed that the achievement and moral domains are equivalent, only that they both share common elements, that causal reasoning and reasoning about issues of controllability are inherent in each.

In response to the parallels mentioned above, between moral judgment and achievement evaluations, and the further benefits that the latter offers above the former, a study was conducted using a concrete game situation. The game was non-threatening and involved a clear successful/unsuccessful outcome on a concrete achievement task. It was designed to easily introduce variations in the controllability of the outcome from absolutely no control to partial to almost complete controllability. This, in turn, allowed assessment of the role of the children's developing understanding of controllability as a basis for developmental changes in their criteria for evaluating others. This concrete achievement situation represents improvements on all of the difficulties and limitations of classical hypothetical moral judgment situations that were enumerated above, in addition to enabling the experimenter to fine-tune the appropriate level of control by the actor over the outcome. Again, it is not assumed that the achievement and moral domains are equivalent, but only that both domains offer the possibility of studying the

relationship between evaluative judgment and causal reasoning.

It was assumed that children (termed, Observers) at three age levels (grades 1, 3 and 6), spanning pre-operational thought through concrete operational thought, would reward and evaluate a peer (termed, Actor) performing a task under conditions where the outcome was or was not under the Actor's control (i.e., under conditions of skill or luck). This game was expected to shed light on several developmental and other psychological questions, which are discussed below.

Analysis of the Issues

The following are the experimental questions addressed in this study, and the theoretical issues relevant to each of them.

Question 1- What criteria, e.g., outcome, effort, ability, task difficulty, controllability, etc, do children of different ages use in overtly evaluating (demonstrated by rewarding) the performance of a peer? Are they similar to criteria used in making moral judgments? Do they change with development in the same way as in the moral domain?

In the realm of moral judgments, there is much documented evidence that children develop from an emphasis on outcome to an emphasis on intentions. In order to shift the focus to achievement evaluations, it is necessary to translate the

components of the moral situation into an achievement situation. In an achievement situation such as the one being described here, intention is indicated by the amount of effort being expended, since effort is what the actor has within full control. The components of ability and personal controllability are very similar in the two domains.

In the realm of achievement evaluations, Nicholls & Miller (1984) examined the development of children's reasoning about effort and ability when their own achievements were involved and when another person's achievements were. Most relevant to this discussion is their finding of a developmental progression in children's rules for evaluating other children's performance. At Level 1, effort or outcomes were treated as equivalent to ability. For example, people who had tried harder were seen as more capable, even if the outcome was worse. At Level 2, effort was seen as the cause of outcomes. There was still no acknowledgment of ability as an independent feature of action. At Level 3, there was a partial, though frequently inconsistent differentiation of effort and ability. At Level 4, ability was seen as a potential component of action. The effect of effort was understood to be constrained by ability. In this case, there was a clear demarcation between effort and ability, and a multiplicative relationship was hypothesized to obtain between the two. Generally, the skill-related concepts of task difficulty, effort and ability are imperfectly differentiated

in young children and become clearly differentiated from one another only at about age 12.

To this experimenter's knowledge, only one study, that of Weiner & Peter (1973) has simultaneously studied the developmental shift in evaluative criteria in the moral and achievement domains. The situations differed according to the intent (effort) and ability of the actor, as well as in the objective consequences of the behavior. Intent, ability and outcome were all criteria used systematically in both domains. However, in both domains, outcome was the prevalent criterion, to be replaced with intent at a later age. This shift followed a parallel pattern in both domains. However, changes were studied using hypothetical materials. As noted, the current study focused on this shift using more concrete and real-life methods.

Hypothesis regarding development of children's evaluation criteria as they are used to evaluate a peer's performance:

The various criteria available to children are outcome, effort, task difficulty and skill level.

The first hypothesis is that younger children emphasize outcome as a criterion for evaluating achievement, while older children increasingly emphasize the amount of effort expended, a practical synonym for intention, or the level of task difficulty, but only on the skill-relevant task. The main study compared children's responses in skill and luck, easy

and hard conditions to supply information regarding this issue.

Hypothesis regarding integration rules among the various criteria:

It is also hypothesized that older children demonstrate development of the integration rules among these criteria in contrast to younger children. In other words, while the youngest children focus on only one dimension (most likely, outcome), somewhat older children add one criterion to another, and the oldest children exhibit the multiplicative rule, and ultimately, the compensatory relationship as described above. The auxiliary study gave children an opportunity to reason about effort and ability, and how they contribute to skill, supplying information to support or contradict this hypothesis.

Question 2- How do children at different ages understand whether an event is under the control of the Actor or not? At what ages and by what mechanism do they make a distinction between outcomes which are generated by an Actor's skill, as opposed to outcomes which are not, but occur randomly? Most critical, does the development of the understanding of controllability mediate development in the criteria that children use to evaluate other children's achievements?

It is reasonable that the determination is made about whether something was intended or not according to Shultz and

Wells' (1985) matching rule. According to these researchers, information about an actor's intention statement is matched against information regarding outcome of the action. Developmentally, this is the first rule, used even by their youngest Subjects (age groups of 3 years and 7 years). This matching rule is utilized to ascertain whether an event was intended or not. Such a matching model predicts more frequent intentionality assessments in situations where outcome had matched the stated intention than where it had not, regardless of the actual control over the outcome.

An alternative possibility to the matching rule is that an inference of intentionality depends on a determination about controllability of the event. This can be done by concluding whether it had been a skill-based or luck-based event.

Controllable vs. random events

Heider's (1958) theory of naive psychology is applicable in this discussion. He made reference to social perception as a process between the center of one person and the center of the other. To him, the underlying causes of events, including the motives of an actor, are the relevant invariances in the environment. He postulated that we first make inferences about whether an actor could have, and did in fact, control the event, and, if so, to what extent? If the event was under the actor's control, he is held accountable for its positive or negative outcome. Following Heider and others, including

Saltzstein & Goldhammer (1986), it has been documented that the most fundamental distinction which older children make is between controllable events and those which occur randomly, i.e., are not under anyone's control.

There is some evidence in the literature which indicates that children gradually distinguish between controllable events and those which occur randomly, i.e., are not under anyone's control. Saltzstein & Goldhammer (1990) reported a developmental finding where younger children evaluated a peer's performance based on the very concrete outcome, but older children evaluated a peer's performance based on the intention of the acting peer, and the amount of control possible over the outcome.

According to Weiner et al., (1971), people use four causal elements (ability, effort, task difficulty and luck) to interpret the outcomes of achievement-related events.

King (1971) found that the ability to distinguish accidents from intended outcomes increases with age. Whether an event had happened by design or accident is yet another way of determining the controllability of the event. It is the current contention that the importance assigned to outcome is diminished as the understanding that it may not have been attributable to the Actor (that it was accidental) develops.

It is hypothesized that, paralleling similar findings in the moral domain, in the achievement domain, young children

reward heavily for a positive outcome, that is, success on an achievement task, irrespective, or perhaps unaware, of the level of difficulty of the task, or the level of potential internal control that the Actor has over the outcome. It is hypothesized that older children reward more heavily for a positive outcome that is more difficult to control than for one which is easy to produce. Furthermore, older children are expected to systematically develop reward patterns reflecting a distinction between the positive outcome which is more difficult to achieve, either because it is randomly generated or more difficult, and that which is skill-generated or easier, and therefore, at least somewhat controllable by the Actor. In a later study conducted by Weiner (1985), children were asked by him to match patterns on flash cards in one of two conditions: with the patterns showing, or when cards were overturned, giving no clue to the patterns. They found that the amount of effort children put into the task which depended on luck, decreased with age. That is, older children did not try hard to match patterns on the overturned cards. Conversely, the amount of effort put into the task which relied on skill (where the patterns on the cards were visible), increased with advancing age (during childhood). In other words, in the condition where the patterns were visible, and skill was useful in matching elements of the pattern, there was increasing effort found as children developed. The general principle to be extracted from this study is that for

events which are understood to be random in nature and not under the control of the actor, skill is irrelevant. In contrast, events which are skill-based and therefore can generally be attributed to the actor, skill and effort are seen as relevant, and deserving of invested effort. In the current study, the distinction between luck and skill-based tasks is expected to be absent or at least less distinct for younger children because of their immature understanding of causality. The necessity for causal reasoning to aid in the formation of this conclusion is self-evident.

Evidence will be added to that from previous research which has indicated a developmental progression in the understanding of controllability as opposed to randomness, and in the criteria used to evaluate performance by others. In the current study, children were also given the opportunity to reason about whether the events (outcomes) they had witnessed were randomly produced, or under the control of the Actor. Their reasoning took place in the concrete, yet non-threatening context that was described above. A developmental progression in the use of this criterion is predicted. It is also expected that the developmental progression will correlate with changes in evaluation of the Actor.

Hypothesis regarding development of the cognitive distinction between controllable and random events

The youngest children are not able to distinguish between

those events which are under the Actor's control, and those which are not. Older children are able to make this distinction. This should be evident both in their justifications and in their actual rewarding behavior in the luck-based as compared to the skill-based conditions.

In the current main study, contrasts between young Subjects' reward patterns in the luck and skill conditions were expected to demonstrate little distinction in reward patterns. By comparison, the same contrast between older Subjects' reward patterns in the luck and skill conditions are expected to show a distinction, which would support the preceding hypothesis. Additionally, there was a comparison between children's rewarding behavior in the easy and difficult version of the skill task. It is hypothesized that older children recognize the fact that a more difficult task represents less control, and an easier one represents more control. Younger children are not expected to make this distinction. Younger and older Subjects' causal attribution statements in the mentioned conditions are also expected to demonstrate a development in the understanding of differing levels of control in the luck and skill tasks, and in the easy and difficult tasks.

In the Auxiliary procedure, the child's age, and his or her ability level implied therein, was the feature which was under lesser or greater control. It was expected that younger Subjects would not make a significant distinction between

comparable performances of older and younger performers. By contrast, older Subjects can make a significant distinction both in reward patterns and justifications when rewarding older and younger (naturally more or less capable) performers.

Question 3 - Does the previously documented development of an understanding of the compensatory relationship between ability and the amount of effort expended, as posited by Heider (1958), Weiner et al (1985), and demonstrated on hypothetical tasks, also occur in a concrete game situation?

The issue here is whether children understand the compensatory relationship between an actor's ability and the amount of effort he or she must expend, to produce a certain effect, as posited by Heider (1958), Weiner (1985), and others.

As adults, we recognize that people with less ability must compensate by trying harder, and those with more ability can achieve the same outcome with less effort. Effort and ability are seen as important personal or internal causes of success and failure in Heider's "Naive Theory of Action". Ability is seen as the limiting factor with regard to how much improvement in performance can be gained from extra effort. Nicholls (1978) found that young children tend to see all good things such as high ability, great effort and positive outcome as belonging together. This phenomenon has also been called a Halo Scheme by Kun (1977). In her study, it was found that

young children scored the harder working of equal scoring actors to be more able. In other words, according to her formulation, in the less mature stages, effort equals ability. In later stages, ability is recognized as an independent factor. At the next level, effort is seen as the exclusive cause of outcome. Ultimately, ability is seen as capacity, completely differentiated from effort.

Among the numerous variables (effort, ability, luck, task difficulty, skill, etc.,) that pertain to achievement judgments, even the youngest children (age 6) have been shown to consider more than one of these variables at a time (Kun, Parsons & Ruble, 1974). However, in their study, all Subjects were told the level of ability and effort of a hypothetical Actor. Even in this hypothetical format, a developmental progression was found in how the various features were integrated. An additive rule was said to characterize young children's responses, while a multiplicative rule characterized older children's and adults' styles of integrating the various components of an achievement judgment. As previously mentioned, one of the dimensions in Weiner's theory is that of stability/instability. This dimension categorizes variables such as the amount of effort expended, which can change from task to task, making it an unstable variable, in contrast to amount of ability, which is a stable, unchanging internal trait. It should be added for purposes of this analysis that ability too, can be seen as unstable, since

each individual's ability is different in different spheres of endeavor. Weiner completes the picture by describing the development of a compensatory relationship between effort and ability. This parallels the compensatory relationship necessary to solve certain traditional nonsocial tasks, such as Piaget's balance beam task or his conservation tasks. All of these tasks rely on the child's developing capability to decenter, consider more than one dimension of a problem simultaneously, and combine them in an inverse or compensatory relationship with one another.

Hypothesis regarding the development of an understanding of the compensatory relationship between an Actor's ability and the amount of effort he or she expends:

The older children demonstrate an understanding that an Actor with more ability can achieve the same performance level with less effort than an Actor with less ability, i.e., the compensatory relationship between effort and ability, in contrast to younger children. The auxiliary procedure, as described in Appendix C, provides children with a forum to demonstrate their ability to use this compensatory principle.

In an effort to enable a finer-grained analysis of the development of the understanding of variables which go into skill, a study by Wimmer, Wachter & Perner (1982) will be replicated, and each Subject child will be exposed to this during the same session as the main study. This study was

chosen because it supplemented the issues being investigated in the main study, that is the less advanced focus on outcome, with more mature emphasis on intention, expressed as effort. Additionally, the capability to generate a finer grained analysis of the variables which are incorporated into skill was seen as a valuable addition to the main study. The relevant features of this, termed the Auxiliary study, were outcome (better or worse), effort (less or more) and ability (less or more).

In addition to the measures mentioned above, children in the original study also were given a Piagetian conservation of mass task. Children in the replication study were asked to balance a beam on a fulcrum using various weights, a version of Inhelder & Piaget's (1958) balance beam task, as described by Siegler (1978). There were two reasons for adding the latter task. The primary aim was to determine the children's general level of cognitive reasoning. Only children who had performed in an age-appropriate manner continued on to be Subjects in this study. Additionally, this task permitted assessment of the same kind of compensatory relationship (between the weights being used and the distance from the fulcrum), as was being investigated in the auxiliary procedure between effort and ability.

In young children, there is generally a failure to understand and coordinate these two components which go into skill. Later, they understand that a surfeit of one can allow

a paucity in the other. In other words, skill level is understood to be a sum of these two variables. In order to accomplish comparable tasks, as the level of one falls, the level of the other presumably rises. This is the compensatory nature of the thinking which develops as skill is evaluated, perhaps parallel to the compensation necessary to respond at the highest stage level in Inhelder & Piaget's balance beam task, as described by Siegler (1978).

Summary

The repeating theme in the studies which have been cited above is that generally, there seems to be a developmental progression from simplicity and preference for one or two dimensions to more complex integration styles among numerous dimensions. Additionally, the dimension which is most preferred may also be different from age to age. Some of the steps implied in this developmental attainment, as documented by previous research may be outlined as follows: 1) Focus on only one dimension at an early stage. That dimension in both studies reported here, as well as in Goldhammer & Saltzstein's (1992) study, is outcome. 2) At a more advanced stage of reasoning, the distinction is made between controllable and uncontrollable events. This is evidenced in both the distinction between the luck-dependent and skill-dependent conditions and the easier and harder versions of the skill-dependent condition. 3) The most advanced level of reasoning

calls for an understanding that even within the skill condition, the factors of effort and ability interact in some way. First, an additive model holds, and ultimately, there is a recognition of compensation between these two factors.

A child's developing understanding of the distinction between events which had occurred randomly and those which had occurred through the Actor's effort and ability (skill) seems to be a crucial issue. It is the first step in the group of determinations which needs to be made by the evaluator. Even before determining whether skill <u>is</u> involved, it is necessary to determine whether skill <u>could</u> be helpful in that situation. In other words, is the outcome reliant on skill (under the Actor's control) or purely random (not under the Actor's control). If skill is found to be necessary, one needs to utilize the criteria of level of ability of the actor, amount of effort put into the achievement, and level of difficulty of the task. It has been found that children do indeed come to utilize these criteria in an increasingly sophisticated fashion. (First, singly, then additively, and finally, acknowledging the compensatory relationship between these factors).

Evidence will also be produced supporting the hypothesis that there is a development of the understanding of the concepts of effort and ability as they contribute to skill. In this study, the distinction between luck and skill (involving the issue of controllability) was directly manipulated by

varying the details of the task. Therefore, if outcomes are held constant, and children nevertheless reward differently in the skill-based than in the luck-based tasks, this constitutes evidence of the important mediating role of controllability in children's evaluations of other's performance. The role of task difficulty will also be investigated in this study by creating two levels of difficulty in each condition. Differential levels of reward, signifying differences in the Observers' assessments, would indicate that level of difficulty is a factor in children's assessments of each other.

The child's understanding of the relationship between effort and ability, which has also previously been cited as a factor in how children judge an outcome, was studied in the Auxiliary Procedure (fence study).

Two features that make this group of manipulations distinctive are: First, the fact that the development of the concepts, as discussed above were tested in a concrete, realistic situation. This has the added advantage of not relying exclusively on children's verbal skills to either comprehend the situation, or to use that comprehension to justify their rewarding behavior.

Second, while each of the concepts has been investigated with various procedures by other researchers singly, in this case, the endeavor was to draw a more complete map of the most relevant of the various concepts (controllability, i.e.,

random and caused, effort, ability, luck and skill, as well as task difficulty), which contribute to and mediate the developmental changes in how children assess each other. The method of using converging procedures to gain insight into comparable cognitive capabilities has the advantage of increasing the level of confidence in the results.

The hypotheses which will be tested can be summarized as follows:

1) Hypothesis regarding development of children's evaluation criteria as they are used to evaluate a peer's performance:

The various criteria available to children are outcome, effort, task difficulty and skill level.

2) Hypothesis regarding integration rules among the various criteria:

It is also hypothesized that older children demonstrate development of the integration rules among these criteria in contrast to younger children. In other words, while the youngest children focus on only one dimension (most likely, outcome), somewhat older children add one criterion to another, and the oldest children exhibit the multiplicative rule, and ultimately, the compensatory relationship as described above.

3) It is hypothesized that, paralleling similar findings in the moral domain, in the achievement domain, young children reward heavily for a positive outcome, that is, success on an

achievement task, irrespective, or perhaps unaware, of the level of difficulty of the task, or the level of potential internal control that the Actor has over the outcome. Older children reward more heavily for a positive outcome that is more difficult to control than for one which is easy to produce. Furthermore, older children are expected to systematically develop reward patterns reflecting a distinction between the positive outcome which is more difficult to achieve, either because it is randomly generated or more difficult, and that which is skill-generated or easier, and therefore, at least somewhat controllable by the Actor.

4) The youngest children are not able to distinguish between those events which are under the Actor's control, and those which are not. Older children are able to make this distinction. This should be evident both in their justifications and in their actual rewarding behavior in the luck-based as compared to the skill-based conditions.

Chapter II

METHOD

Design

The following variables were measured in the main procedure:

1. The non-verbal measure: The number of stars each Observer child awarded an Actor child, a) in conditions where the observed child had succeeded or failed on a trial; b) in conditions where the outcome had been easier or more difficult to achieve; c) in conditions where the outcome was or was not under the Actor's control.

2. Verbal justification of his or her own star allocation by the child, indicating level of reasoning, a) where the Actor, or observed child had succeeded or failed on a trial; b) where the outcome had been easier or more difficult to achieve; c) where the outcome was or was not under the Actor's control.

3. Observer children's (verbal) assessment of the Actor's intentionality, a) where the observed child had succeeded or failed on a trial; b) where the outcome had been easier or more difficult to achieve; c) where the outcome was or was not under the Actor's control.

4. Verbal assessment of the controllability of the outcome a) where the observed child had succeeded or failed on a

trial; b) where the outcome had been easier or more difficult to achieve; c) where the outcome was or was not under the Actor's control.

Luck or Skill conditions, Level of Difficulty (easy-hard), Age of Subject and order of presentation of trial types constituted the four varying factors. The first, third and fourth variables were between-Subject variables, and the second was a within-Subject variable. Two levels of difficulty in the skill-based condition were created by changing the difficulty of the tactile discrimination task. Once proportions of successes (matches) between intent and outcome and failures (mismatches between intent and outcome) had been generated by Subjects in the skill-based condition, the comparable luck-based conditions were created by a yoking procedure.

Children participated in one of two conditions: luck or skill-based outcomes. Within each of these conditions, there was a harder and easier version of the task. All children who were assigned to the skill-based task were interviewed first. The skill-based task required each child to first state an intention about his or her performance. This was regarding the actual task, which was to retrieve a yellow or blue poker chip from an opaque bag. In order to aid in the discrimination, there were disks of coarse or fine textured sand paper affixed to both sides of each poker chip. Yellow chips all had fine textures, and blue ones all had coarse textures. Therefore,

the actual skill needed to translate an intention into an outcome was one of tactile discrimination.

The luck-based task, on the other hand, required a child to pick a yellow or blue chip, as he decided, blindly from a bag. There were no textures present to guide his/her choice. Therefore, despite the child's statement of intent, the outcome was purely random, i.e., not controllable. In the luck-based procedure, the number of yellow and blue chips which were placed in the bag corresponded to the number of coarse-textured and fine-textured poker chips which had been correctly retrieved by the matched Subject in the skill-based condition. In other words, the approximate level of success possible for each individual Subject in the luck-based condition was directly yoked to the level of success of a matched individual Subject in the skill-based condition. The difference was that since there were no textures present on the yellow and blue chips, so the Subjects' success or failure in selecting the chip color of choice was actually random, or not under the Actor's control. In this way, proportion of successful outcomes was held quite constant between groups of Subjects while the amount of control exerted by each group differed. Both Actor and Observer children were familiarized with all textures to be used on a task.

Three grade levels (1, 3, and 6) assumed to be roughly representative of three ages and levels of development, constituted the third factor. General levels of cognitive

development were assessed by having each child in the original group tested on a standard Piagetian task, conservation of mass, to help confirm that age is an adequate measure of children's cognitive levels. A replication group of Subjects was tested on another standard Piagetian task, the balance beam task, in order to test whether grade level and age corresponded well with children's cognitive levels. The Piagetian tasks were performed first in all cases, because they were utilized as a screening device, ensuring that each child who would actually participate in the study was making cognitive responses which corresponded with his or her chronological age.

Subjects

One hundred and sixty eight pairs of Subjects participated in the study. Each pair consisted of an Actor and an Observer. There were 28 pairs of Subjects, half male, from each of grades 1, 3 and 6 in the skill condition, and another 28 pairs from each grade in the luck condition. The original pairs were randomly selected from a pool of children who were attending an Orthodox Jewish day camp. The only restriction on the pairing of player and observer was that although they were from the same grade and of the same sex, they were not to be from the same class in school or day camp group. This precaution was instituted to avoid prior acquaintanceship within pairs.

Materials and Procedure

Materials

Materials were varying numbers of yellow-colored and blue-colored poker chips, the exact numbers to be determined by conditions. In the skill condition, these colored chips were also distinguishable by fine, medium and coarse textures. This variation in textures was effected by gluing circles of different textured sanding paper on the poker chips. There was also a bag to contain the poker chips, from which the player was asked to select a chip without benefit of visual cues.

Procedure

In the skill-dependent condition, Observers in each age group observed, evaluated and explained their evaluations of peer Actors playing the following game: The player (Actor) stated an intention to obtain the yellow or blue chip, indicated by fine or coarse textures. He or she then reached into the bag and (blindly) selected a chip from therein. The Actor was relying exclusively on tactile senses to find one which matched his or her stated intention. The Actor was not asked any further questions, and the Observer was then asked to reward the player that he of she had been watching with a number of stars (0 to 3), and to explain to the Experimenter why that many stars were awarded. Observers' responses were coded into categories, to be described below. The Observer was also asked attribution questions. These questions included

"Did the Actor mean for that chip to appear?" If the response was negative, it was followed with "What made that chip appear?"

The game continued until at least three individual instances of match and mismatch between intention and outcome had occurred. An example of a match in the skill-dependent condition was when the player stated an intention of obtaining a yellow chip, and a yellow chip with a smooth texture was, in fact, chosen. On the other hand, if he or she intended to draw a blue, coarse chip, but drew a yellow, smooth one instead, it was coded as a mismatch. When the patterns of match and mismatch had occurred enough times to provide all the necessary data (e.g., three instances each of matches and mismatches), and that pattern recurred, the Observer rewarded the player, but he or she was not asked to justify or explain the evaluation. In other words, only the <u>first</u> three instances of each pattern were used as data. All responses were recorded by hand at the time of the interview. Also, randomly chosen sessions were tape-recorded, so as to make an assessment of inter-rater reliability possible.

There were two phases to this study: one requiring moderate skill, distinguishing between coarse and fine textured chips, and one of greater difficulty, distinguishing between coarse and medium textured chips. All Subjects observed and judged both of these conditions, half of them in the order of difficult-easy, and half of them in the order of

easy-difficult. Children were assigned to one of these groups randomly. However, all Subjects were informed at the outset of the session that there will be an easy and a difficult level of the task, and they were introduced to the various textures which comprise each condition in the skill-based condition. (See Appendix A for verbatim instructions.) This was done to minimize the order effects which were anticipated, specifically, floor and ceiling effects.

In the luck-dependent condition, a different group of Subject pairs in grades 1, 3 and 6 were enlisted. Here, too, Actor children played a similar game as was previously described. That is, they stated an intention to retrieve a blue or yellow chip from the bag, and proceeded to retrieve it. The crucial difference in this procedure was that in the luck-based condition, there were no textures affixed to the chips, and therefore, no possibility for a tactile discrimination between the two kinds of chips. This ensured that the chip which actually was retrieved was done so randomly. As previously mentioned, each Actor child in this group was yoked to a same-aged child who had participated in the skill-based condition. The proportion of yellow to blue chips which were placed in each child's bag in this condition was equivalent to the proportion of successes to failures which the child he was yoked to in the skill-dependent condition had achieved. This ensured comparable proportions of successes to failures in the two groups, while maintaining

controllability in the skill-based condition and eliminating it in the luck condition.

This study represents an effort to underscore the developing understanding of varying skill levels, and developing valuation of situations that are skill-controlled, as opposed to luck-controlled.

Auxiliary Procedure

This procedure exposed each of the Observer Subjects (half of them before the main procedure, and half of them after), to a group

of hypothetical situations that included the elements of effort, ability, and outcome. The reason for varying order of presentation of the main procedure and auxiliary procedures, was to ensure that reasoning about these issues did not affect and facilitate more mature reasoning or behavior in the main procedure.

The procedure was as follows:

The Experimenter read a basic story (see Appendix C), with accompanying visual aids in the form of large cards with drawings (see Appendix D), along with explanations by the Experimenter of the significance of each visual component. Ability was represented by a card showing a young boy (who could be assumed to have less ability to complete the assigned task), or of an older boy (who could be assumed to have more ability to complete the assigned task).

Effort was represented by a card which either showed a

can of paint and a brush (indicating a high level of effort expended toward accomplishing the task), or a flash card showing a baseball and bat (indicating less effort expended on completing the task).

Outcome was represented either by a flash card showing a fully painted fence (a good outcome), or a fence which was painted only halfway (a poor outcome).

All Subjects were told that the child in the hypothetical story had been instructed by a parent to paint the fence on a particular afternoon. They were then shown various combinations of the three stimuli, each scenario representing a separate trial. The eight kinds of trials which emerged are listed in Table 1A. They were presented in one random order of trials for half of the Subjects, and in the reverse order for the other half of the Subjects.

Children were instructed to reward with cookies at the end of each trial, and to justify their choice of how many cookies, as well as to rate the hypothetical actor's ability and effort and accomplishment, as well as to make a causal attribution of the outcome, which was coded to reflect their attitudes about effort, ability, and skill which contributed to outcome.

The question arose about whether the children in grades 1, 3 and 6 respectively, were actually consistent in terms of their intellectual capabilities. With this issue in mind, the original group of Subjects (16 pairs of children in each grade

and condition, as previously described) were also asked a standard Piagetian conservation of mass question. They were presented with two comparable balls of clay. The Experimenter rolled one of them into a snake-shape. Then, Observer children were asked which one was bigger. Responses such as "they are the same" or "they're equal" were recorded as a correct response. Answers indicating that either one of them is larger than the other were recorded as an incorrect response. This was done to provide an independent assessment, apart from grade and age level, of each child's reasoning level. This measure provided only a bi-polar (conserver/non-conserver) measure. Therefore, the third grade, who as a group performed inconsistently on this measure, were not easily categorizable. (Interestingly, there did not seem to be a direct correlation between their ability to conserve and the level of social reasoning which third graders demonstrated on the above-described task).

The group of children who were enlisted to replicate the study (12 pairs of children in each grade and condition as previously described) were distinguished from the original group by virtue of their participation in another reasoning task of Piaget's - the balance beam task, as it was described by Siegler (1978). The balance beam task provides a fuller, more fine-grained analysis of children's reasoning levels. According to Siegler's interpretation, it provides for four levels of reasoning, roughly consistent with four age

categories. The children in this replication study were representative of the first three of the age levels, and fell into the corresponding three categories. In this task, children were asked to balance a beam on a fulcrum. They did this with the aid of several sized pairs of weights. Children were coded as having conformed to Levels 1, 2 or 3. Only those first grade children who were coded as Level 1 responders on this task were placed into the first grade group to participate in the main and auxiliary procedures. Only those third graders who were coded as Level 2 responders on this task were placed into the third grade group to participate in the main and Auxiliary procedure. Only those sixth graders who were coded as Level 3 responders on this task were placed into the sixth grade group to participate in the main and Auxiliary procedures.

Altogether, five children were eliminated from the study. Two of them were first graders who were not age-appropriate for their grade. The third child was a third grader who did not respond at Level 2 (but at Level 3) in the balance beam task. Two of them were sixth graders, one who did not focus well on the interview, and one who was not age-appropriate for her grade.

Gathering evidence that any or all of the previously mentioned factors is (are) the mediator(s) to the developing assessment of others was undertaken as follows: the various groups, representing different ages, and in different

conditions were statistically compared against each other, using One-Way Analyses of Variance where applicable. These were followed up with local contrasts, as warranted. A Multiple Analysis of Variance was also conducted where it fit the data, to clarify the evidence of the effects of the various manipulations (eg., luck/skill, effects of age, task difficulty). Furthermore, there was evidence which was retrieved from the semi-structured interviews conducted in a Piagetian clinical style with the children after each trial, which was coded so as to provide information regarding the development of these concepts, and whether they develop at the same ages as the assessment criteria are changing, or perhaps, slightly before. The aim here was to determine whether the development of understanding of luck vs. skill and the compensatory relationship between effort and ability are co-incidental to the development of a more mature style of evaluating peers, or even whether the two former variables are the underlying mechanism for the development of the third.

There is also possibility here for documentation with convergent operations. Together, the two sources of evidence allowed the conclusion to be drawn that each of the factors discussed above contributes to, or is a pre-requisite for the changing criteria used to assess peers as children mature intellectually.

Experimenter is attempting to add evidence to previous research which has indicated a developmental progression in

the understanding of controllability as opposed to randomness, and in the criteria used to evaluate performance by others.

Chapter III

Results

The data were collected in two phases: The original and replication phases. The factor of group affiliation (original group or replication group) was inserted as a dummy variable, in order to ascertain whether there were any systematic differences in responding between the original and control groups. There were no consistent differences in responding between the first and second group of Subjects. Therefore, all data reported were pooled for purposes of analysis. Also, Subjects' gender was also included as a dummy variable. No systematic response differences were found between males and females. Henceforth, data will be discussed describing the pooled population of males and females.

All data were collected during game-observation and interview sessions either at the day camp, or in individual Subjects' homes. Experimenter took notes regarding the number of stars or cookies awarded in each trial, and verbal justifications which were garnered were recorded in writing in coded form. Approximately 10% of the interviews were also tape-recorded, and one of two alternative raters listened to these tapes and coded childrens' responses in order determine an inter-rater reliability with the Experimenter. The rate of inter-rater reliability over-all was found to be 86%. The data in the main procedure were analyzed by a MANOVA option

included in the SPSS-X program. Some of the data collected with the Auxiliary procedure were analyzed by Univariate Analyses, also included in the SPSS-X program. The rest of the data in the Auxiliary procedure were analyzed with the help of Chi-Square analyses, also furnished by the SPSS-X program. The following includes the most significant findings:

NON-VERBAL MEASURE: NUMBER OF STARS AWARDED

Main effects: Grade affected the number of stars given as a reward, $F(2,164)=53.26$, $p<.001$. Collapsing across all conditions, Subjects in grade 1 awarded more stars ($M=5.45$ of a possible 9) per trial type, than third graders ($M=5.13$), and significantly more than sixth graders ($M=3.79$). (Refer to Figure 1.0).

There was also a strong effect of level of task difficulty, $F(1,165)=24.51$, $p<.001$. Collapsing on other conditions, Subjects in all grades awarded a mean of 5.08 stars in the hard condition, but only 4.49 stars in the easy condition, on the average.

The effect of outcome was also highly significant, $F(1,165)=87.3$, $p<.001$. Collapsing on all other conditions, the mean number of stars awarded by all Subjects in match conditions (conditions where outcome matched stated intention) was 7.23 but only 2.34 in the miss conditions (conditions where outcome did not match stated intentions). (Note Figures 2.0 and 3.0).

As shown in Table 1, there was also a first-order

interaction effect between grade and outcome, $F(2,163)=11.75$, $p<.001$ in the rate of stars awarded. It is interesting to note that even though younger children awarded more stars overall, both for success (matches) and failures (misses), the ratio of stars awarded was approximately 1:3 between misses and matches in all grades.

A first-order interaction was also found between the skill and luck conditions (or task versions) and outcome. Collapsing on the hard and easy conditions, children in general awarded fewer stars for a skill-dependent miss ($M=1.37$) than for a luck-dependent miss ($M=1.75$). They also awarded more stars for a skill-dependent match ($M=5.0$) than for a luck-dependent match ($M=4.63$), acknowledging the presence of controllability as a relevant factor. The ratio of awards for misses vs. matches in the skill condition was 1.37:5.0, which was greater than in the luck condition - 1.9:5.0. This supports the second hypothesis, which contends that (at least, older) children determine whether the outcome was controllable, and include this information in their assessment of a peer's performance. (See Figure 4.0).

Second-order interactions- The second order interaction between grade, skill and outcome, $F(2,)=11.78$, $p<.001$) superseded the above. First graders awarded a similar number of stars for a bad outcome, regardless of whether it was luck or skill dependent. (Ms = 4.9 and 5.9 in skill and luck conditions, respectively). They also awarded virtually the

same number of stars in luck and skill dependent good outcome (match) conditions. (Ms=15.0 in both conditions). Sixth graders, however, awarded significantly more stars in the skill-dependent, good outcome condition than in the luck-dependent good outcome condition (Ms=14.5 and 9.2, respectively. Conversely, they awarded more stars in the luck-dependent bad outcome condition than in the skill-dependent bad outcome condition, (Ms=4.7 and 2.4, respectively). This further supports the hypothesis that only older children take the controllability of the outcome into consideration when evaluating a peer's performance. (Refer to Figure 5.0).

JUSTIFYING REASONS

The term "justifying reason" refers to how children justified the magnitude of rewards, that is, the number of stars, they gave. The use pattern of each individual justifying reason is discussed below:

REASON 1 - OUTCOME

Justification responses were coded as outcome-oriented when they focused exclusively on whether the event was a positive or negative one (the Actor either succeeded or did not succeed at retrieving the chip of his color choice).

Collapsing across grades, there was a main effect $F(1,166)=48.37$, $p< .001$ in the use of outcome to justify magnitude of rewards, in both match and miss (good and bad

outcome) conditions. A significant effect of grades was also found $F(2,164)=99.15$, $p<.001$. Children in the youngest group (grade 1) used this criterion most often ($M=2.48$), third graders used it at an intermediate rate ($M=1.94$), and the oldest children (grade 6) used it least often ($M=1.09$). This provides partial support for the first hypothesis, which is that younger children focus more on outcome than older ones do, and older ones focus increasingly on controllability .

A significant second-order interaction was also found among skill (hard-easy distinction), outcome (match or mismatch) and grade, $F(2,163)=43$, $p<.01$. Among first graders, there was a significant main effect of outcome $F(1,53)=3.83$, $p<.05$. That is, first graders used outcome as justification for how many stars they awarded an average of 2.61 times after a match, but only 2.06 times after a miss. Among third graders, there were no significant main effects of any experimental variables on their use of outcome-oriented justifications, although they utilized this justification a great deal overall. Among sixth graders, there was an effect of level of task difficulty on use of outcome-oriented justifications, $F(1,53)=4.34$, $p<.04$. That is, the outcome was mentioned more often as justification for reward in difficult conditions than in easy task conditions. Specifically, in the hard task condition, the good outcome (match) was mentioned more often ($M=1.23$) as justification for reward than it was in the easy task condition ($M=.95$). (Note Figure 6.0).

REASON TWO - LEVEL OF SKILL REQUIRED

This kind of justification described the kind and amount of action necessary to bring about the outcome, or lack of it.

Collapsing across grades, there was a main effect of the skill-luck distinction, $F(1,165)=35.5$, $p<.001$. In skill-dependent conditions, this kind of justification was used a mean of 0.44 times, but in the luck-dependent condition, only 0.09 times, on average. There was also a more moderate effect of grade $F(2,164)=5.607$, $p<.004$. Sixth graders utilized this type of justification 2.5 times as often as first graders. There was also an interaction between the skill/luck distinction and grade level $F(2,163)=10.2$, $p<.001$. In other words, as was hypothesized, older children paid more attention to the necessary act that was required, than did younger children.

First graders did not seem to recognize the relevance of the skill-luck distinction on the issue of controllability by the act. Among third graders, there was a modest but significant effect of the skill or luck condition in use of references to skill or luck as a focus for justification, $F(1,53)=4.53$, $p<.035$. Third graders used this type of justification a mean of .34 times in the skill condition, but only a mean of .11 times in the luck condition. Among sixth graders, there was a far stronger effect of the skill or luck condition, $F(1,53)=39.9$, $p<.001$. This type of justification

was used a mean of .75 times in the skill condition and only .05 times in the luck condition. That is, the oldest children appeared to comprehend the fact that the necessary act (the skill involved) only made a difference in the skill condition, not where outcomes occurred randomly. In the two latter grades, we saw a developing understanding, demonstrated by increasing use of the skill justification in the skill-dependent condition and not in the random, luck-dependent condition, as expected.

REASON 3 - DIFFICULTY LEVEL OF TASK

This kind of justification referred to different difficulty levels of the task.

Collapsing across grades, there were first-order effects of the actual difficulty level, $F(1,165)=4.79$, $p<.05$, and a very strong grade effect ($F(2,164)=69.87$, $p<001$). In other words, older children justified magnitude of their reward by mentioning the difficulty or ease of task, whereas younger children (grades 1 and 3), did not.

In grade 1 there were no significant differences found between the two task levels (easy and hard) in their use of difficulty level justifications. In grade 3 there was a significant effect of the actual level of task difficulty $F(1,53)=5.23$, $p<.02$. This kind of justification was used a mean of .25 times in the hard condition, but only .05 times, on the average, in the easy condition. In grade 6, there was

an even more significant effect of the actual level of difficulty, $F(1,53)=4.56$, $p<.04$. (Ms=.38 and .25, for hard and easy conditions, respectively). Additionally, there was an even stronger effect of the luck and skill conditions, $F(1,53)=121.7$, $p<.001$, (Ms=.69 and .83 in the skill and luck conditions, respectively). Overall, the expected developmental trends were found.

There was also found a first-order interaction between the skill and luck condition and level of difficulty. In other words, first graders made no differentiation between the two levels of task difficulty in their use of this reason (Ms=.00,.05), whereas third graders exhibited a significant understanding of the two degrees of difficulty (Ms =.25 in the hard condition .05 in the easy condition); and sixth graders not only understood the issue of different difficulty levels, but also recognized the different levels of control in the luck and skill conditions. They used the difficulty level as justification .69 times in the skill condition and .83 times in the luck condition. Although these findings were not in the hypothesized direction, both of these means were significantly higher than use of this reason in younger grades, (M=.31).

REASON 4 - POTENTIAL CONTROLLABILITY OF OUTCOME

This type of justification referred to potential control over the outcome in the skill condition, or the lack of control, in the luck condition. Collapsing across grades,

there were significant effects of the skill-luck condition distinction $F(1,165)=85.3$, $p<.001$, and of grade, $F(2,164)=36.6$, $p<.001$, in the use of this kind of justification, as expected. There was a significant simple interaction effect between the skill-luck distinction and grade level, $F(2,163)=39.64$, $p<.001$, as well as a two-way interaction among the skill-luck condition distinction, level of task difficulty, and grade level, $F(2,162)=3.07$, $p<.05$.

In the luck-dependent condition, lack of controllability was offered as justification more often by children in grade 6 than by children in either grade 3 or in grade 1. (Ms=.08, .23, and .97 in grades 1, 3 and 6, respectively). This supports the hypothesis that older children distinguish between luck and skill and recognize their implications for controllability more than younger children do. (See Table 4). Interestingly, when using this criterion, only third graders made the distinction between the positive outcome (match between intention and outcome) and the negative outcome (miss between the intention and outcome), referring to controllability in the miss condition more than four times as often as in the match condition. That is, they justified not awarding fewer stars in a miss condition, citing lack of controllability by the Actor more often than they justified awarding more stars in a match trial.

In grade 1, there were no significant condition effects found with regard to the controllability-centered justifying

reason. In grade 3, there was a somewhat significant effect of the skill-luck distinction $F(1,53)=4.24$, $p<.04$ in the use of this type of justifying reason, as well as a significant effect of the positive or negative outcome as a variable in use of this kind of justifying reason, $F(1,53)=4.24$, $p<.04$. In grade 6, there was a very strong main effect of the luck/skill distinction, $F(1,53)=121.65$, $p<.001$, and a smaller effect of task difficulty, $F(1,53)=4.56$, $p<.04$. There was also a first-order interaction effect between the skill/luck distinction and difficulty level, $F(1,52)=4.56$, $p<.04$. To clarify, sixth graders used the criterion of controllability more in luck than in skill conditions (referring to lack of controllability), more in difficult than in easy conditions (again referring to less control over outcome), and most often in the difficult/luck condition. (Refer to Figures 8.0, 9.0, 10.0, and 11.0).

REASON 5 - UNCODEABLE JUSTIFICATIONS

On those occasions when a child's response did not conform to the coding scheme, or was purposefully evasive, it was counted as an uncodeable justification.

Collapsing across grades, there were significant effects of the skill/luck condition distinction, $F(1,165)=8.67$, $p<.003$, of the outcome type, $F(1,165)=5.64$, $p<.02$, and of grade, $F(2,164)= 5.22$, $p<.006$. There was a first-order interaction between the skill and luck condition distinction

and grade levels, $F(2,163)=7.91$, $p< .001$, and a second-order interaction between the skill/luck distinction, good or bad outcome, and grade level, $F(2,162)=3.58$, $p<.03$. Overall, first and sixth graders justified their awarding behavior in an uncodeable fashion about the same amount (Ms=.27 and .28, respectively). Third graders made uncodeable justifications far less often (M=.09).

In grade 1, there was a significant effect of outcome type, $F(1,53)=10.24$, $p<.002$ on amount of uncodeable justifications. In other words, first graders made fewer uncodeable justifications in good outcome conditions (M=.13) than in bad outcome conditions (M=.44). There was also a one-way interaction between the skill and luck conditions and outcome type $F(1,52)=8.29$, $p<.005$) in the use of uncodeable criteria. In grade 3, there were no significant effects found of any of the variables. In grade 6, the effect of the luck-skill distinction on the use of uncodeable justifications was significant, $F(1,53)=19.07$, $p<.001$. More uncodeable justifications were made by sixth graders in luck-dependent conditions (M=.40) than in skill-dependent conditions (M=.28).

NUMBER OF CORRECT INTENTION STATEMENTS

Children's responses in the affirmative to the question "did he mean to pull out the yellow or blue chip?" were coded either as correct if the chip retrieved matched in color to the stated intention of the Actor, or incorrect if the chip

retrieved did not match in color to the stated intention of the Actor about which colored chip he will try to retrieve. There were significant main effects of grade $F(2,164)=13.17$, $p<.001$, skill condition, $F(1,165)=11.26$, $p< .001$ and of outcome, $F(1,165)=34.26$, $p<.001$, in the number of correct intention statements made. There were also first-order interactions between grade and the skill and luck condition distinction, $F(2,163)=8.41$, $p<.001$, and grade and outcome, $F(2,163)=24.35$, $p<.001$.

In grade 1, there was a main effect of outcome, $F(1,53)=47.89$, $p<.001$, as well as an interaction between skill and outcome, $F(1,52)=7.1\ell$, $p<.008$. On the average, they made 2.81 (out of a possible 3.0) correct intention statements per match condition, but only 1.84 correct intention statements per miss condition. The typical error was that if he did it, he must have meant it. In grade 3, there was also an effect of outcome, $F(1,53)=9.60$, $p<.002$. Third graders were also more likely to make a correct judgment about the intention behind the outcome in a match (good outcome) condition ($M=2.88$) than in a miss (bad outcome) condition ($M=2.50$). Sixth graders, on the other hand, made more correct statements about intention in the skill than in the luck condition ($Ms=3.00$, 2.39, respectively). This was a significant finding, $F(1,53)=63.55$, $p<.001$. The number of correct intention statements was affected in the younger grades only by the outcome. In the highest grade, it was affected mainly by the skill condition,

representing controllability of outcome. This pattern is parallel to the hypothesized one. Younger children utilized the matching rule, as Shultz (1984) described it, to infer intendedness. That is, only outcomes which matched previously stated intentions were judged as being intended. Older children relied instead on their assessment of controllability. (See Table 5).

NUMBER OF INTERNAL ATTRIBUTIONS

When the Observer attributed the outcome to the Actor, it was coded as an internal attribution. There were significant main effects of grade, $F(2,165)=16.62$, $p<.001$, the skill and luck condition distinction, $F(1,166)=35.3$, $p<.001$ and outcome, $F(1,166)=13.37$, $p<.001$ for number of internal attributions. There were also first-order interactions between grade and skill, $F(2,164)=58.64$, $p<.001$ and between skill and outcome, $F(1,165)=6.9$, $p<.01$.

Interestingly, children in grade 1 actually made significantly more internal attributions in the luck condition ($M=2.61$) than in the skill condition ($M=2.06$), $F(1,53)=10.2$, $p<.003$. In grade 3, this trend reversed to the more expected form: These children made an average of 2.44 internal attributions in the skill conditions and only 2.11 internal attributions in the luck condition. While this is only marginally significant, $F(1,53)=4.01$, $p<.05$, the trend is in the intuitively correct direction. Children in grade 3

also exhibited a significant effect of outcome, $F(1,53)=9.05$, $p<.003$. That is, in the match conditions (successful outcomes), they made more internal attributions ($M=2.56$) than in the miss conditions (unsuccessful outcomes), ($M=1.95$), which were more likely to be attributed externally. In grade 6, children exhibited a large main effect of the skill and luck condition distinction, $F(1,53)=236.27$, $p<.001$. They made 2.67 internal attributions in the skill condition, but only .83 internal attributions in the luck condition, on average. This supports the hypothesis that only older children recognize that skill-dependent outcomes are internally attributable to the Actor, while luck-dependent outcomes are externally attributable.

All grades attributed internally more often in the match condition than in the miss condition, but only third graders made a large enough distinction to be statistically significant. There was also a U-shaped function in children's distinctions between the skill and luck condition. That is, grades 1 and 6 made the distinction between luck and skill, and third graders did not. However, first-graders made more internal attributions in the luck condition. In contrast, sixth graders made more internal attributions in the skill condition, obviously the more mature, correct response. (See Table 6.0).

NUMBER OF EXTERNAL ATTRIBUTIONS

Children were asked what made the outcome happen. If they responded that it was some circumstance other than the Actor, it was coded as an external attribution. Collapsing across grades, there was a strong main effect of the skill and luck condition distinction, $F(1,165)=34.62$, $p<.001$, of outcome, $F(1,165)=12.95$, $p<.001$, and of grade, $F(2,164)=16.9$, $p<.001$. There was also a very strong skill condition by grade interaction $F(2,163)=38.8$, $p<.001$. It should be noted that the third graders made the largest distinction between the match and miss conditions, utilizing the external attribution .44 times, on average, in the match condition and 1.02 times in the miss condition. Older children attributed externally more often in the luck-dependent condition than in the skill-dependent condition, as was expected.

AUXILIARY STUDY

Number of cookies awarded by condition

There were 8 story types used, all revolving around a basic theme. The variations in the theme included outcome, effort and ability, high and low ratings for each. The range of combinations, yielding a total of eight conditions, will henceforth be referred to as H or L, representing high and low; and A, E, and O, representing ability, effort and outcome, respectively. (refer to Table 1A).

Effect of Outcome

In almost all cases, (except for grade 6 in the LA, LE ,HO condition), the positive outcome generated greater rates of rewarding, $F(1,167)=419.7$, $p<.001$. The aforementioned exception conforms to the hypothesis that as criteria such as ability and effort become salient, the outcome criterion loses importance. Across all grades, the average number of cookies awarded per child in all conditions with a positive outcome was 7 out of 9. The average number of cookies awarded per child in all conditions with a negative outcome was 6. The magnitude of this difference was relatively small because there were strong attenuating influences of effort and ability in selected conditions and grades.

Developmental Trend

Across all conditions, there was a powerful interaction between grade and outcome, $F(2,166)=113$, $p<.001$. The above effect of outcome was most powerful and very resistant to features of ability or effort in grade 1. The effects of other variables became increasingly important with age as evidenced by the decreasing difference between positive and negative outcome conditions in the higher grade.

In grade 1, the average number of cookies awarded per child in all positive outcome conditions was 7.49, compared to 6.46 in grade 6. The average number of cookies awarded per child in all negative outcome conditions was 2.34 in grade 1, compared to 4.46 in grade 6, indicating the expected

developmental de-emphasis of outcome as criterion.

Effect of Ability

There was a significant effect of ability on children's rewarding behavior. $F(1,165) = 68.4$, $p < .001$. All groups actually rewarded more heavily for trials which included high ability. (M=4.6 in high ability conditions, and 3.9 in low ability conditions) Various interaction effects provide additional details. For example, there was an ability X grade X effort interaction in the LA, LE, HO condition, $F(2,164) = 5.05$, $p < .009$. First graders awarded heavily in this condition, (M=6.0), ostensibly in response to the positive outcome, third graders awarded much less, (M=1.4), and sixth graders least of all (M=.75), basing the reward rate on the low effort and/or ability. The interaction effects between ability and effort will be discussed below.

Effect of Effort

There was a strongly significant effect of effort across all grades $F(1,165) = 497$, $p < .001$. There was also a developmental trend found in the effect of effort on the reward patterns of children. $F(2,164) = 142$, $p < .001$. An effort X grade interaction made itself evident in various conditions, most prominently in the HA, LE, HO condition $F(2,164) = 142$, $p < .001$. In this condition, the average number of cookies awarded in grade 1 was 7.4, obviously in response to the

positive outcome, not the negative effect of effort. In grade 6, on the other hand, the average number of cookies awarded was 2.0, in response to the low effort expended, despite the positive effect of outcome. Third graders awarded an average of 5.2 cookies, with their unexpected focus on the 'disobedience' of the story Subject, who did not paint, despite parental instructions to do so. Although the condition of high or low effort did not make much difference in the pattern of children's justifications, except for the outcome, effort was used as justification the most, by all groups, across condition.

Compensatory Relationship

There were several examples where the effect of ability interacted with level of effort to form a compensatory relationship. This was evident in children's choices of justifying reasons. The HA LE HO, LA HE LO, HA LE LO and LA HE HO conditions all have the common feature that there is a mismatch between effort and ability. In these conditions, all children mentioned both effort and ability, but only older children (grade 6) also added that a child with less ability has to try harder, or that a child with more ability can get away with less effort to accomplish the same task. There was virtually no mention of this compensatory relationship among first and third graders, in any condition. This supports the hypothesis that the compensatory relationship develops as a

child matures intellectually. Even though this phenomenon was evident in children's different justification choices, it did not evidence itself to a significant degree in their awarding behavior.

JUSTIFYING REASONS IN AUXILIARY PROCEDURE

HA HE HO CONDITION

In the condition where there was a good outcome, high ability and high effort expended, the following was found: This kind of trial type had no effect on children's choice of justifications in either grade 1 or 3. First graders justified their awarding behavior 94.6% of the time, with the outcome. They also referred to the effort expended 5.4 % of the time. Third graders justified their awarding behavior 84.6 % of the time with the outcome, 11.6% of the time with reference to effort, and made only minimal reference to the Actor's ability (3.6% of the justifications). Sixth graders, on the other hand, used a greater variety of justifications. Their use of justifications was much more flexible, and therefore influenced to a significant degree, $F(5,51)=90.25$, $p<.001$, by trial type. Overall, they used outcome justifications only 38.4% of the time, effort justifications 24.1 % of the time, ability justifications 20.3% of the time, and compensatory relationship type justifications 17.2% of the time. This confirms the expectation that one of the things which occurS with development is that children acquire a wider

array of potential justifications to choose among.

HA LE HO CONDITION

In the condition where there was a high ability, low effort, and good outcome, first graders used outcome to justify the magnitude of their rewards .77 times on average and the (lack of) effort .33 times on average. Third graders used the outcome justification slightly less (M=1.40) and the effort justification slightly more (M=.60). Sixth graders, on the other hand, used each of these justifications less (M=.90 for outcome; M=.30 for effort). They made occasional mention of the Actor's strong ability (M=.20). The most interesting finding is the predicted one: Sixth graders made .60 references, on average, to the compensatory relation between effort and ability in this kind of trial type, and no mention of it was made in younger grades.

LA LE HO CONDITION

In this kind of condition, there was low ability, low effort, but nonetheless a positive outcome. The outcome justification was used most by first graders, less by third graders and least by sixth graders (M=1.70, 1.21 and .80, respectively). The use of effort as justification was least common in first graders, and increased in popularity with increasing grades M=.21, .45 and .72 in grades 1, 3 and 6, respectively). The ability justification was used least by first graders and more with increasing grades (M=.09, .25 and .36 in grades 1,3 and 6, respectively).

LA HE HO CONDITION

In the condition where there was a low ability but a high effort and good outcome, the following justifications were found: Youngest children (grade 1) used the outcome justification most, as expected. Third graders used it somewhat less, and sixth graders least of all (M's=1.76, 1.22 and .70 for grades 1, 3 and 6, respectively). Consequently, children's reference to effort increased with age (M=.24, .45 and.51, respectively). There was little reference to ability, and that little, only by third and sixth graders (M=.21 and .29 in grades 3 and 6, respectively). Again, there was a striking developmental trend found in reference to the compensatory relationship between effort and ability. First graders did not refer to it at all, third graders only rarely (M=.12), but sixth graders used it more than any other kind of justification (M=.68).

HA HE LO CONDITION

In this condition, there was high ability, high effort, and yet a poor outcome. As in all the other conditions, first graders used the outcome justification most, with decreasing emphasis as children got older (Ms = 1.68, 1.33 and .70 for grades 1, 3 and 6, respectively). There was a slight developmental increase in the use of the effort justification (Ms = .32, .48 and .60 in grades 1, 3 and 6 respectively). First and third graders utilized this criterion significantly less than did sixth graders.

HA LE LO CONDITION

In the condition where there was a high ability but low effort and a poor outcome, the following results were found: There was a developmental decrease in the use of outcome to justify rewarding behavior (*M*s=1.65, 1.03 and .58 for grades 1, 3 and 6, respectively). All grades made a similar number of references to effort (*M*=.41 for all grades combined). Only third and sixth graders referred to the Actor's ability (*M*=.36 for grades 3 and 6 combined) There was a developmental trend in the use of the compensatory relationship between effort and ability found. First graders did not acknowledge it at all. Third graders utilized it only .11 times, on average, and sixth graders utilized it .63 times, on average.

LA HE LO CONDITION

In the condition where there was low ability, high effort but a poor outcome, it was found that, as in other conditions, first graders justified their rewarding behavior with the outcome criterion more often than any other group (*M*s=1.68, 1.44 and .52 respectively for grades 1, 3 and 6). The criterion of effort was used .32 times by first graders, .53 times by third graders, and .43 times by sixth graders, on average. Reference to ability was only made by sixth graders (*M*=.22). But the compensatory relationship between effort and ability was also noted by sixth graders and was overwhelmingly preferred by them (*M*=.78). This is a strong developmental finding.

LA LE LO CONDITION

In the condition where there was low effort and ability, and a poor outcome as well, the following was found: First graders used the outcome justification most, with progressively older children using it progressively less (Ms=1.83, 1.44 and .72 respectively in grades 1, 3 and 6). There was also a marked developmental trend in the use of ability to justify awarding behavior. First graders did not use it at all, third graders used it only .05 times, on average, and sixth graders, .58 times, on average. The compensatory relationship between effort and ability was used only by sixth graders, and even then, only rarely (M=.20).

In sum, only sixth graders acknowledged the ability of greater effort compensating for less ability, and the converse. They utilized it mainly in situations where one was high and the other was low. First graders focused largely on the outcome quality, with some minor adjustments for effort. Third graders focused on outcome predominantly, but put more emphasis on the effort of the actor, often emphasizing conversationally that effort in this case was a reflection of the Actor's obedience to parental authority.

Chapter IV

Discussion

It was found that the development of children's evaluations of a peer regarding a concrete achievement task was quite similar to development of their thinking in other domains, such as in moral judgment. Many tasks which utilize hypothetical, theoretical techniques, most notable among them the classical moral judgment procedures of Piaget et al. (1962) have repeatedly demonstrated that young children focus on the outcome as the most important judgment criterion. Saltzstein & Goldhammer (1990) found the outcome to be the major criterion utilized by first graders. In the same studies we found that as children mature, they become cognizant of the relevance of the Actor's effort, and its implications for evaluation.

The current study was undertaken partially as a demonstration of children's developing criteria for evaluation in a concrete, achievement task. Basically, the development was found to progress from a focus on outcome to one on intention, as expected. Additionally, the question of how children determine whether an event was intended or not was an issue. Two possible explanations are the Matching Rule, as described by Shultz and his colleagues and/or a pre-determination of possible controllability of the outcome by the actor. Support was found for both of these possible mechanisms, although younger children were found more likely

to use the matching rule, while older ones spoke more about the controllability of outcome. And finally, developmental changes in children's perception of what constitutes skill was also a theoretical question.

It was the assumption that an output of effort implies an intention for the outcome to occur. Effort is the physical manifestation of intention in a concrete task situation such as this one. In the main procedure, it was predicted that older children would acknowledge effort as a variable, while the younger ones who would not acknowledge intentions in other domains do not acknowledge effort as a criterion here. As expected, a developmental trend was found in the current task, wherein children's criteria for evaluations changed from an immature one focusing on outcome to a more mature one, focusing on intention, or in this case, effort. The findings in the auxiliary procedure, while more complex, were also supportive of these findings. Evidence supporting this finding was as follows:

While all children awarded more stars for a positive outcome than for a negative one, the first graders rewarded more heavily for success, and less for failure than either the third or sixth graders. Clearly, first graders rewarded heavily if the actor succeeded in pulling out a chip with the color that he intended, and did not reward if the other color was retrieved. This shows a distinct emphasis on outcome by the younger children. Additionally, the rewarding patterns of

first graders was unique in that this distinction between matches and misses (between stated intention and outcome) was the only criterion they appeared to utilize. Their rewarding behavior was extremely consistent across all other experimental manipulations. The youngest group's awarding pattern was so consistently outcome-anchored that it can only be described as resistant to the effects of either the luck-skill distinction or the differentiation between easy and hard task conditions.

Additional evidence in support of this hypothesis consists of children's use of verbal justifications regarding their rate of star awarding. Figures 8, 9, 10 and 11 all illustrate that regardless of whether children were responding in the easy or difficult condition, luck or skill condition, match or miss outcome conditions, first graders justified their awarding overwhelmingly by mentioning the outcome as the criterion. The number of references to all the other criteria was negligible by comparison. Third graders also mentioned the outcome as justification most often. However, in certain conditions, such as in bad outcome conditions (both skill and luck dependent) and the skill-dependent good outcome condition, more third graders tended to describe the activity necessary to create the outcome. By contrast, sixth graders engaged in justification by using the outcome far less. At the same time, sixth graders turned to use of difficulty level and controllability of the outcome rewards much more often.

This was true of their reward patterns, where a hard match was rewarded more than an easy match and skill-based good outcomes were well rewarded but luck-based good outcomes were not. It was also true of their pattern of verbal justifications, where the number of references to the outcome diminished drastically (compared to younger children), while being replaced with more references to the actor's skill, or lack of it, and to the controllability of outcome, as reflected in the luck/skill and easy/hard distinctions. Actually, it can not be concluded that sixth graders did not utilize outcome as criterion, only that their thinking was more flexible, and utilized a wider array of criteria. Difficulty level and controllability (luck-skill distinction) were at the forefront of these newly developed criteria. Effort, the visible evidence of intention, is necessary for control, and even more so in the more difficult level of the task.

In the Auxiliary Procedure (the replication of Wimmer, Wachter & Perner's study), similar evidence was found. Here too, first graders rewarded with cookies predominantly based on whether the child in the story managed to paint the complete fence or only half of it. This is essentially a successful/unsuccessful outcome criterion, similar to the one found in the main procedure, and consistent with the hypotheses. The number of cookies they chose to reward the actor did not change appreciably whether there was evidence of greater or less ability, or greater or less effort on the

part of the actor.

In the Auxiliary Procedure, third graders actually drew attention to the factor of effort, among other factors, although they too continued awarding cookies largely based on the positive/negative outcome distinction, much like the first graders. The only additional criterion which they utilized was that of effort. Children who tried hard were given more cookies than children who had not, even when the outcome was held constant. Such a finding was possible here because there were explicitly stated high and low effort conditions for the observers to focus on, in contrast to the main procedure, where the Actor's effort had to be inferred by the Observer. Therefore, this is not a contradiction to the findings in the main procedure, but an elaboration, or enhancement of them. One explanation for this finding is that third graders acknowledge only concrete, physically apparent features of an actor, such as how hard he worked, and not yet the internal features such as personality traits or abilities. Another possible explanation is that in this story, children were told to paint the fence by a parent. Therefore, in this case, expending effort was synonymous with obedience and failing to do so was synonymous with disobedience. In the casual, open-ended de-briefings which followed some of the trials, third graders were indeed found to focus on the issue of obedience more than either first or sixth graders.

Sixth graders were also influenced by the presence or

absence of a positive outcome in the Auxiliary Procedure. They rewarded with more cookies in the Auxiliary Procedure if the fence was fully painted, as opposed to half painted. However, their responses were tempered by the intention, as demonstrated by effort, as well as all of the variables available. In short, their criteria for evaluation were far more sensitive to condition effects. Generally, this allows more flexibility and complexity of response. Descriptions of the other criteria used by sixth graders will follow below.

This study incorporated several other variables available for consideration by the children beyond outcome and intention. It was indeed confirmed that children at different ages and in different experimental conditions clearly verbally described and used each of the criteria which were enumerated in the introduction. Obviously, not all children under all circumstances used all of the criteria to evaluate an actor's performance. The criteria which were listed were: outcome, controllability (the skill-luck distinction), level of task difficulty (and different levels of skill needed to accomplish tasks of different difficulty levels), effort and ability, as they pair up to formulate skill, and finally, the compensatory relationship between effort and ability. This is the understanding that a person with less ability can compensate for it with more effort to achieve the same level of skill, and thereby, to accomplish the same concrete physical task, and vice versa.

In the main procedure, first, there were large developmental differences found in the use of different criteria as justification for rewarding. First graders used outcome, whether good or bad, almost to the exclusion of all else. Third graders also focused on the outcome a great deal. They did, however, repeatedly refer to the Actor's effort. This was apparent in two ways. In the main procedure, in the luck-dependent conditions, some third graders awarded more stars than expected for a miss. When a justification was elicited, they frequently offered responses such as: "He tried hard. It's not his fault." There was even an occasional explicit response, such as "He tried hard, but had bad luck." The latter response not only acknowledged the actor's effort, but also implied at least an intuitive level of understanding of the lack of controllability here.

In the Auxiliary Procedure, the outcome, effort and ability were all alluded to by the Experimenter in the instructions. In this context, third graders focused a great deal on the outcome, just as in the main procedure, but put even more emphasis on the effort. In addition to the fact that trial conditions which included high effort yielded more cookies for the Actor, the Observer children also verbally justified it by talking about the Actor's high level of effort. As was mentioned previously, since the request to paint the fence came from a parent, high effort was synonymous with obedience. Many third graders emphasized the aspect of

obedience, apparently a strong issue between parents and children at this age. Another possibility, as mentioned previously, is that effort is an external, visible, tangible feature, and therefore, that children become conscious of it at an earlier age than ability, which is a less tangible, internal trait.

Sixth graders both in their awarding behavior and in their justifications demonstrated an understanding of all of the available variables. First, they too focused on the outcome, but to a far smaller degree than the other factors and other age groups. In their verbal justifications on the main procedure, many did mention the actor's success or failure, but only in passing, as they galloped along to other issues. They acknowledged the level of difficulty of the skill task when it was more difficult, and did indeed reward more heavily for it. They also acknowledged the lack of control over outcome in the luck-dependent (random) condition, and did not withhold rewards as much for a poor outcome (mismatch) in these situations. Their understanding of controllability was further demonstrated by the fact that in the skill-dependent condition, they rewarded more heavily in the difficult than in the easy condition, and mentioned the different levels of skill needed. This was not the case among younger children. However, in the luck-dependent condition, they did not reward more heavily for a less probable successful outcome than for a more probable successful

outcome, understanding that both of the outcomes were completely random.

In conclusion, the rationale for having children reward a peer for achieving or not achieving on a skill or luck version of a task was an interest in testing their understanding of the skill-luck distinction as it pertains to controllability of outcome, using a non-verbal evaluative measure, thereby eliminating the potentially confounding effect of differential verbal fluency. Also, because children have been shown to go through an intuitive phase before they can reason clearly about a mental operation, there was reason to believe that here, too, children might award for achievement, using intuitively more mature criteria than they could actually describe.

The developmental hypothesis was confirmed using this primarily nonverbal measure. Younger children awarded stars based mainly according to the outcome of the act. That is, a positive outcome was strongly valued regardless of whether it was easy or difficult, or based on luck or skill. In contrast, older children, especially sixth graders put great emphasis on the luck-skill distinction. That is, a match in the luck-dependent condition had much less value for them than did a match in the skill-dependent condition, especially where the necessary tactile discrimination was made more difficult. These findings were further supported by the verbal measures, in the form of justifications for the size of rewards. First-

graders used the outcome to justify their rewards, whether large or small, more than any other factor. This was irrespective of the skill-luck distinction, or the difficult-easy task distinction. Sixth-graders, on the other hand, used the justifications which focused on the difficulty of the task, and on the availability of control over it. Generally, the development of children's actual rewarding patterns matched the development of their verbal reasoning about what the important features in an achievement are. For example, the sixth-graders frequently used controllability- or difficulty-based justifications for their reward to the peer; in contrast, first graders relied heavily on outcome to justify their awarding behavior.

It was surprising, however, to find that first-graders often made more correct statements about the actor's intention than the third-graders, and almost as many as sixth-graders. There are several possible explanations for this. The most likely explanation is one which supports the original hypothesis. That is, causal attribution was hypothesized to be the functional underpinning for the ability to evaluate an achievement. It is necessary to understand the cause of an event, including whether to attribute it internally (to the actor) or externally (to an accident or divine intervention, etc.) before one can correctly evaluate an achievement. It can be argued that young children (first-graders) already have some causal attribution skills, but that they are not yet

ready to apply their understanding to a concrete task such as the one being discussed here.

Another possibility is that the coding scheme made it difficult for sixth-graders to demonstrate their optimum performance. The coding scheme was designed to use a match between intention and outcome as criterion of intentionality. Often, sixth graders mentioned that wanting is not the same as intending when there is no full control. If some sixth-graders' performances were undermined somewhat by the inflexibility of the coding scheme, first-graders might have benefitted from the comparison. For example, in a luck-dependent/match condition, when a child in grade 1 was asked whether the actor meant to pick that chip, he would generally respond with "yes", relying on the match between stated intention and outcome. This was the correct response, according to the coding scheme. On the other hand, a significant number of sixth-graders responded with "no" in the same scenario, bypassing the 'match' criterion, and elaborating that one can not intend to do something which is random, and not within his control. Technically, this was coded as an incorrect answer, since it does not conform to the expectation, built into the coding scheme, regarding the match which children try to find between outcomes and preceding events, as Shultz and his colleagues have demonstrated. Yet, these sixth-graders' responses indicate a more mature understanding of the luck/skill distinction. In fact, it

should be noted that there seems to be a developmental trend in use of these rules. That is, that younger children attempt to determine the intendedness of an outcome by looking for a match between previous events and the outcome, but older children try to ascertain the controllability of outcome to determine its intendedness.

Any future investigation will have to include a refinement of the coding scheme, which should have added flexibility in order to accommodate this development in types of responses. Further investigation is warranted as well, to determine the direction of development-whether it proceeds from an intuitive, non-verbal understanding of achievement evaluation criteria, to full, verbal reasoning about it, and whether it proceeds from a development of causal attribution skills generally to their application in the domain of achievement evaluations.

This study was designed to be a non-threatening game task which deliberately bears no resemblance to the more typical academic or sports achievement tasks which children routinely encounter. The latter are more heavily loaded with emotional baggage dealing with past performance and expectations of future performance, as well as self-esteem issues. Farmer, Vispoel & Maehr (1991) found that among children in several grades of school, successful achievement in a school context enhanced what they referred to as "Mastery Values". However, achievement in a sports context was more related to an

increase in "Social Approval Values". Therefore, the context of the achievement determines what type of self-esteem gain is accomplished. However, they also found that across contexts, Mastery Values were associated more with effort attributions than with luck attributions. So, while context is important, the basic cognitive reasoning which is detailed in the current study is very relevant.

Gender effects were not found in the current study, despite their prominence in other kinds of studies involving achievement. For example, Stipek & Gralinski (1991) asked Junior High School students to complete a questionnaire measuring achievement-related beliefs after taking a math exam. Girls rated their ability lower, expected to do less well, were less likely than boys to attribute their success to high ability or failure to luck. They also tended to over-attribute failure to low ability. These findings indicate that achievement in school is often attributed differently by boys than by girls. In the context of sports, White (1993) examined gender and age effects on children's attributions along the dimensions of causality, stability and controllability. After finishing a softball game, they completed the Causal Dimension Scale (CDS). Males perceived the reason behind their successful outcome to be more internal than females. Females perceived the reason behind their victory to be more controllable (the implication is that less skill was required) than males. Thus, in more familiar tasks,

such as in the domains of sports and academic achievement, attributions of the two sexes show opposite biases. Yamauchi (1990) also examined attributions for success and failure in a non-competitive task (jigsaw puzzle assembly). Subjects rated outcomes to four causes: ability, effort, task difficulty and luck. Mainly, Subjects attributed failure to lack of ability and greater task difficulty, not to lack of effort. However, there were gender differences found in attribution of success. Males attributed success to internal factors more than females did.

The failure to find gender differences in the current study may be due to the fact that the Subjects had no personal experiences with the task, unlike those in real life. This would support the theory that girls may reason about these issues in a similar manner to boys, unless their experience with the type of task or with evaluative feedback about similar tasks has altered their own causal reasoning.

The current research has important implications for the interaction between children's cognitive development and their choice of which goals to pursue, and how to pursue them. Their self-esteem and sense of efficacy in play and work settings relies on appropriate evaluations of their achievements. Apparently, children, both males and females develop in a parallel fashion in the elementary school years with regard to the cognitive capability to evaluate achievement. Yet, slightly older children, within their more typical work and

play environments, already display differences in how they attribute success and failure. This phenomenon is probably due to social experiences with the task and with other evaluations, rather than to some fundamental change in reasoning.

The children who participated in the current study represent a sub-culture of the American population. They are Orthodox Jewish children. However, they are all thoroughly embedded in American culture, speaking English as their primary language, they are exposed to the same school curriculum, television shows, and pop culture of their times. Therefore, this researcher feels confident that the results reported here are not different than they would be if they had been based on another population of children. However, these findings may not be as applicable to children from other cultures. for example, Mizokawa & Rykman (1990) examined causal attributions of Japanese children for success and failure in language and social studies as well as in mathematics and sciences. Unlike the current findings, they found that effort was the overwhelmingly popular attributional choice in explaining academic performance. Ability was a more distant second.

The rationale for including Wimmer & Wachter's study as an auxiliary procedure was as follows: It focused on similar theoretical issues as were broached in the main procedure, such as the opportunity for children to reward and evaluate a

hypothetical peer, on an achievement task, utilizing criteria such as outcome and issues of controllability and effort, in achievement situations which include the factor of intention. These aspects super-impose themselves very efficiently onto the main procedure, allowing the added confidence of similar results which had been achieved with different operations. But beyond this, the most unique contribution of the Auxiliary Procedure is the contribution of one factor which could not be built into the main procedure: In addition to making the distinction between outcomes which were uncontrollable (luck-dependent) and those which were controllable (skill-dependent), this procedure allowed for a finer-grained analysis of how children reason about skill than the main procedure afforded. In the main procedure, children had the opportunity to notice that easy tasks take less skill, and harder tasks require more skill to complete. Additionally, the occurrence of luck-based outcomes is independent of skill level. The Auxiliary Procedure goes one step further. As Heider's theory stipulates, skill is a combination of ability and effort. Each of these features was available to children alone, in the auxiliary study. Therefore, the opportunity for children to demonstrate a developing understanding of <u>how</u> effort and ability interact to form skill, was the most significant contribution that the auxiliary study provided. As expected, young children valued either effort or ability, or both <u>independently</u>. Older children indicated an understanding that the two factors of effort and ability can compensate for each other as they each contribute to skill.

As a final point, in the current research, there was no correlation found between children's ability to conserve or at what level they balanced a beam on a fulcrum, and the kinds of causal reasoning which they utilized when evaluating a peer during the described games. In other words, there was no correlation found between children's nonsocial general reasoning and their social cognitive reasoning. Perhaps this is because a social reasoning task such as this one is more complex than nonsocial reasoning, such as in traditional Piagetian tasks. The additional social, cultural and environmental influence may create a decalage between reasoning levels in these different domains.

The findings of the current study demonstrate a very clear picture of social cognitive development. Children change the criteria which they use in evaluating the performance of a peer. Young children were demonstrated to be exclusively outcome anchored. The next older group of children demonstrated an additional emphasis on effort. The oldest children studied succeeded in making the distinction between luck and skill dependent tasks, presented an understanding of controllability of outcome. Perhaps this indicates a fundamental change in their causal reasoning or in the degree to which it is relevant to evaluation of performance. They also made the distinction between easier and more difficult tasks. Finally, they also showed evidence of an understanding of the compensatory relationship between effort and ability as they combine to form skill.

Table 1.0 Means of stars allocated by match/miss outcome conditions, by grade.

Grade	1	3	6
Match	8.22	7.69	5.78
Miss	2.67	2.56	1.80

Table 2.0 Mean frequency of action description (Reason II) justifications in skill and luck conditions, by grade.

Grade	1	3	6
Skill	.22	.34	.75
Luck	.13	.11	.05

Table 3.0 Mean use of difficulty level-type of justifications across conditions, by grade.

Grade	1	3	6
Combined conditions	.02	.15	.76

Table 4.0 Mean use of controllability (luck/skill distinction) justification with luck and skill conditions by grade.

Grade	1	3	6
Skill	.00	.05a	.00b
Luck	.08	.23a	.97b

Note: Significant differences within columns are indicated by the same letter.

Table 5.0 Mean number of correct intention statements within skill, luck, match and miss conditions, by grade.

Grade	1	3	6
Skill	2.33	2.72	3.00a
Luck	2.33	2.66	2.39a
Match	2.81b	2.88c	2.61
Miss	1.84b	2.50c	2.78

Note: Maximum number of correct intention statements - 3.0

Note: Statistically significant differences are indicated by same letters within column.

Table 6.0 - Mean number of internal attributions by skill and luck, match and miss conditions, by grade.

Grade	1	3	6
Skill	2.06a	2.44	2.67b
Luck	2.61a	2.11	.83b
Match	2.44	2.56c	1.86
Miss	2.23	1.98c	1.64

Note: Maximum number of internal attributions = 3.00

Note: Significant differences indicated by same letters.

Table 1A - List of conditions in auxiliary study

1	HA	HE	HO
2	HA	LE	HO
3	LA	LE	HO
4	LA	HE	HO
5	HA	HE	LO
6	HA	LE	LO
7	LA	HE	LO
8	LA	LE	LO

*H = High *L = Low

*A = Ability *E = Effort

*O = Outcome

Table 2A - Mean number of each kind of justification (outcome, effort, ability and compensatory relationship) by condition and grade.

CONDITION	OUTCOME	EFFORT	ABILITY	COMP. RELAT.
AAA				
GRADE 1	1.78	.22	0	0
GRADE 3	1.50	.44	.06	0
GRADE 6	1.04	.46	.50	.08
ABA				
GRADE 1	1.77	.23	0	0
GRADE 3	1.40	.60	0	0
GRADE 6	.20	.22	.21	1.60
BBA				
GRADE 1	1.70	.21	.09	0
GRADE 3	1.21	.45	.25	.07
GRADE 6	.80	.72	.36	.10
BAA				
GRADE 1	1.76	.24	0	0
GRADE 3	1.22	.45	.21	.12
GRADE 6	.20	.31	.29	1.68

Table 2A-Continued

AAB				
GRADE 1	1.68	.32	0	0
GRADE 3	1.33	.48	.29	.68
GRADE 6	.70	.60	.70	0
ABB				
GRADE 1	1.65	.35	0	0
GRADE 3	1.03	.46	.40	.11
GRADE 6	.58	.47	.32	1.63
BAB				
GRADE 1	1.68	.32	0	0
GRADE 3	1.44	.53	.03	0
GRADE 6	.52	.43	.22	1.78
BBB				
GRADE 1	1.83	.17	0	0
GRADE 3	1.44	.51	.05	0
GRADE 6	.72	.50	.58	.20
AAB				
GRADE 1	1.68	.32	0	0
GRADE 3	1.33	.48	.29	.68
GRADE 6	.70	.60	.70	0

Table 2A - Continued

ABB				
GRADE 1	1.65	.35	0	0
GRADE 3	1.03	.46	.40	.11
GRADE 6	.58	.47	.32	1.63
BAB				
GRADE 1	1.68	.32	0	0
GRADE 3	1.44	.53	.03	0
GRADE 6	.52	.43	.22	1.78
BBB				
GRADE 1	1.83	.17	0	0
GRADE 3	1.44	.51	.05	0
GRADE 6	.72	.50	.58	.20

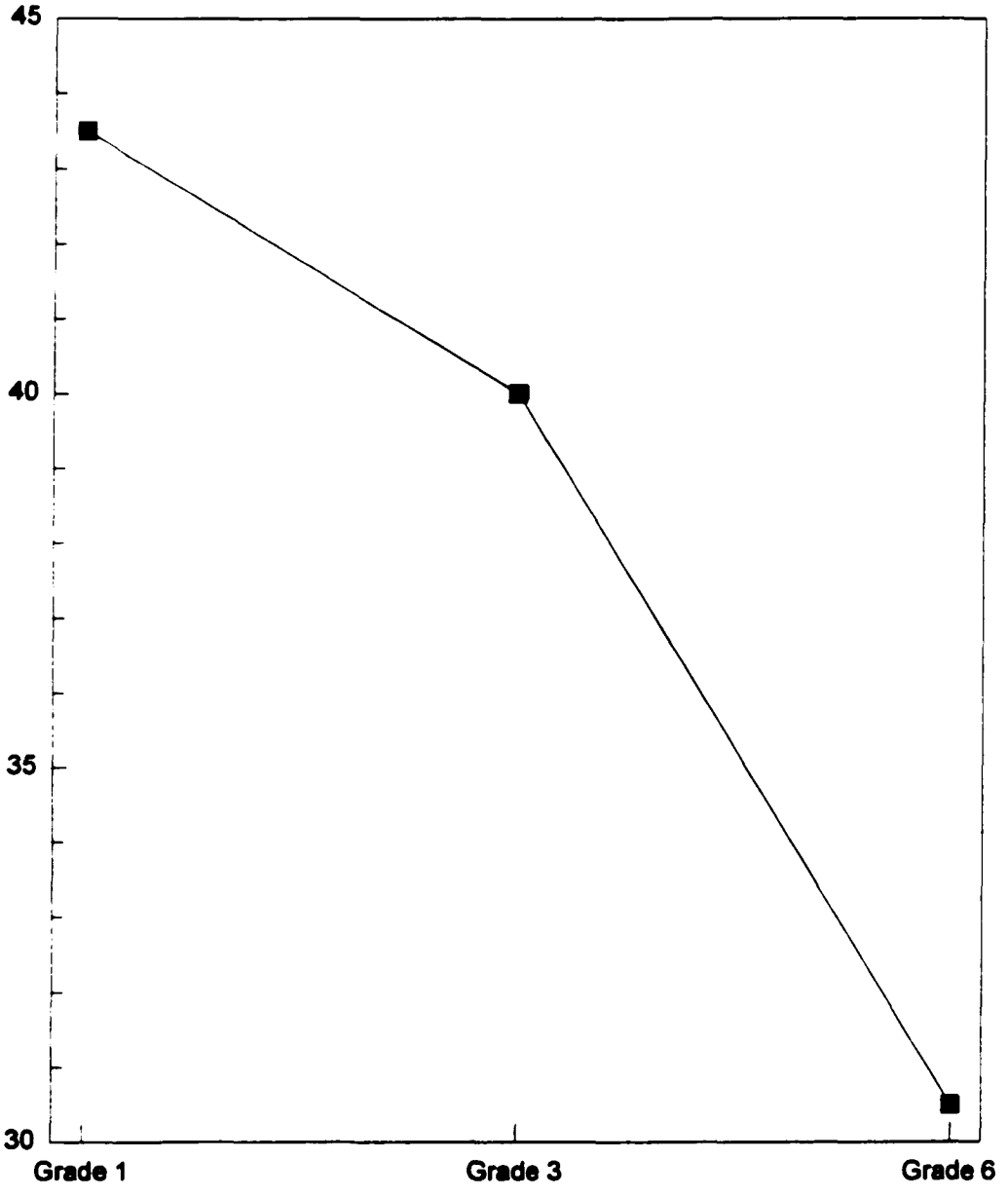


Figure 1.0. Average number of stars awarded, across conditions, grades 1, 3 and 6.

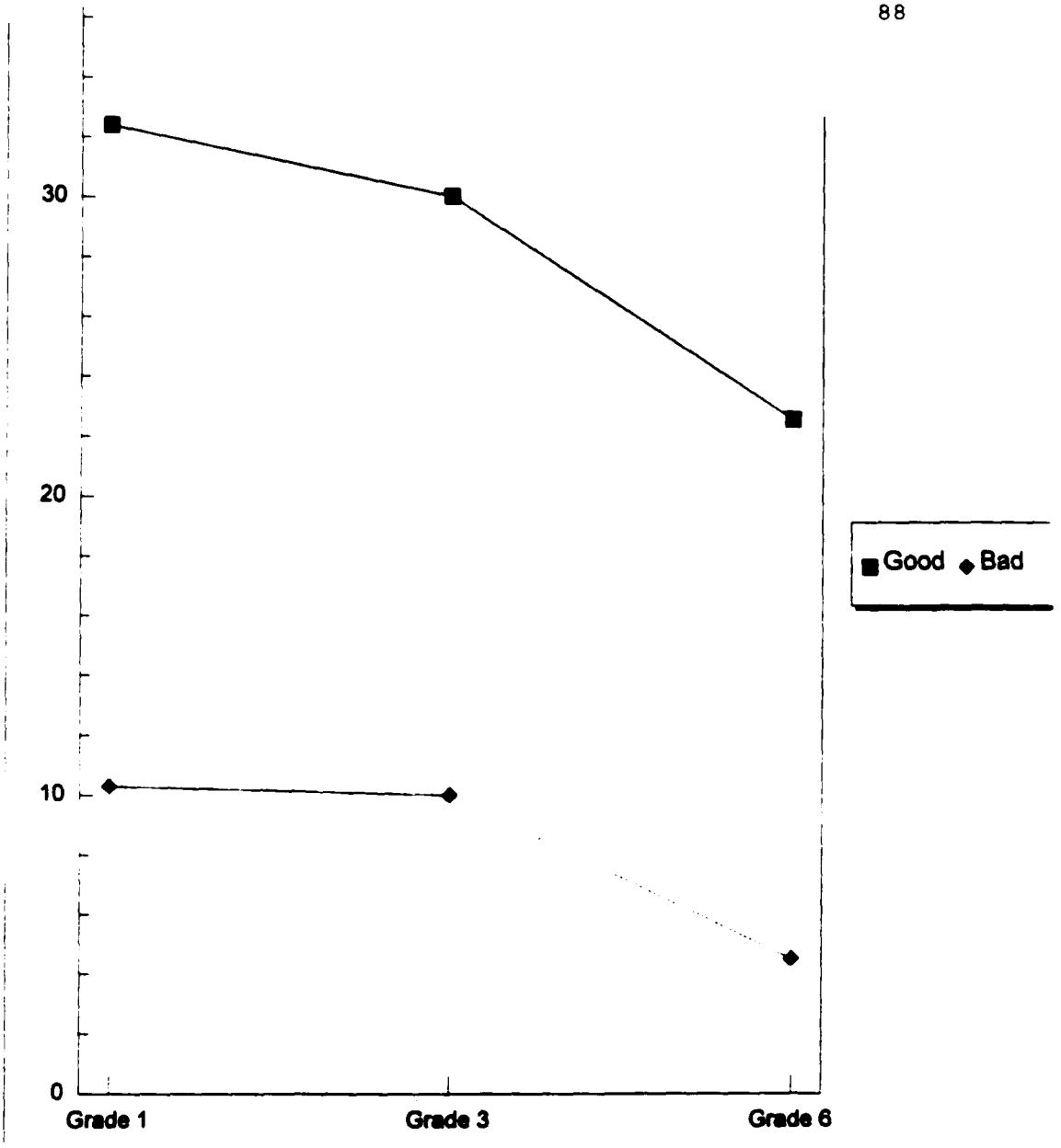


Figure 2.0. Average number of stars awarded in good outcome conditions and bad outcome conditions, grades 1, 3 & 6.

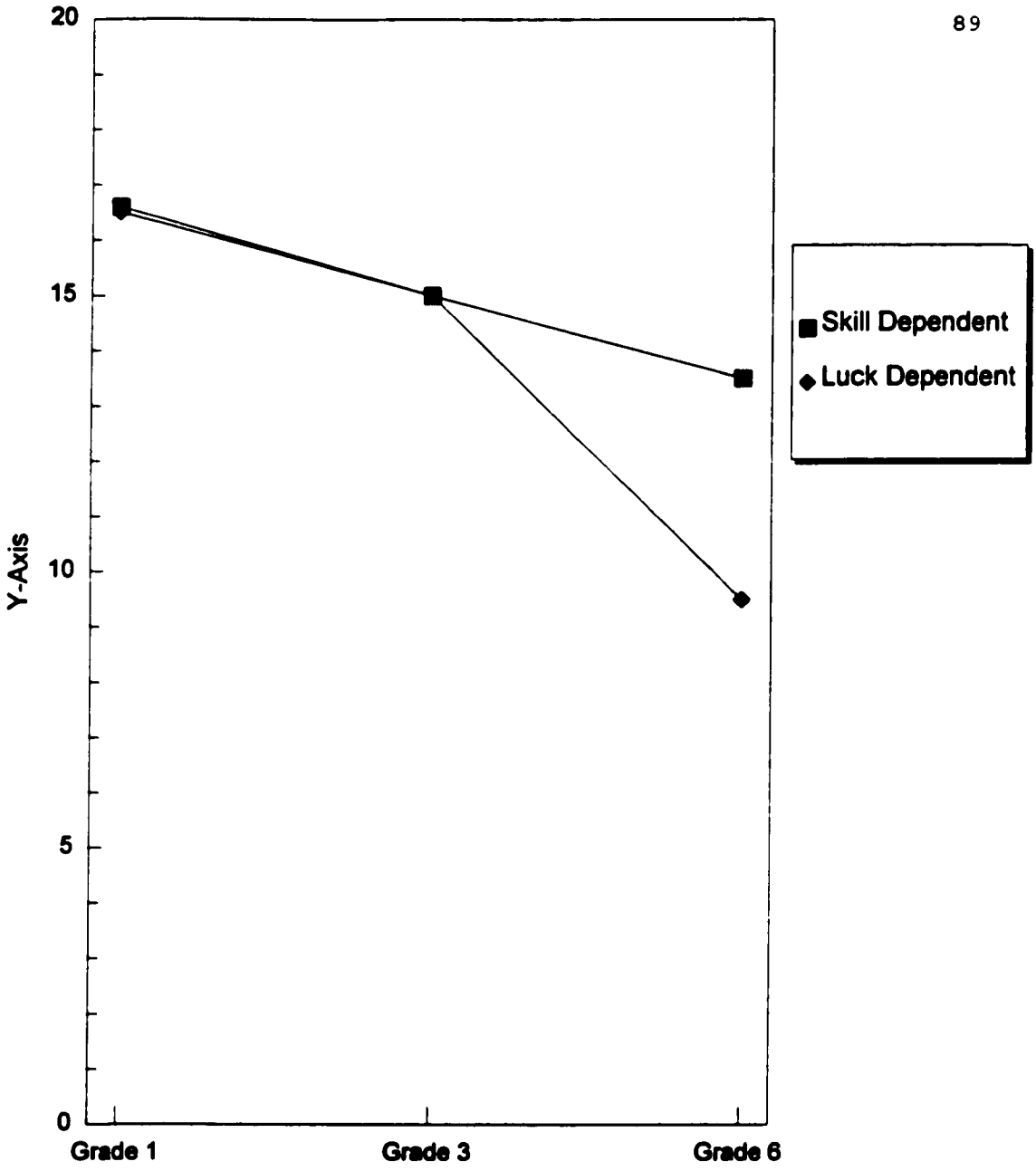


Figure 3.0. Average number of stars awarded in luck dependent and skill dependent good outcome conditions.

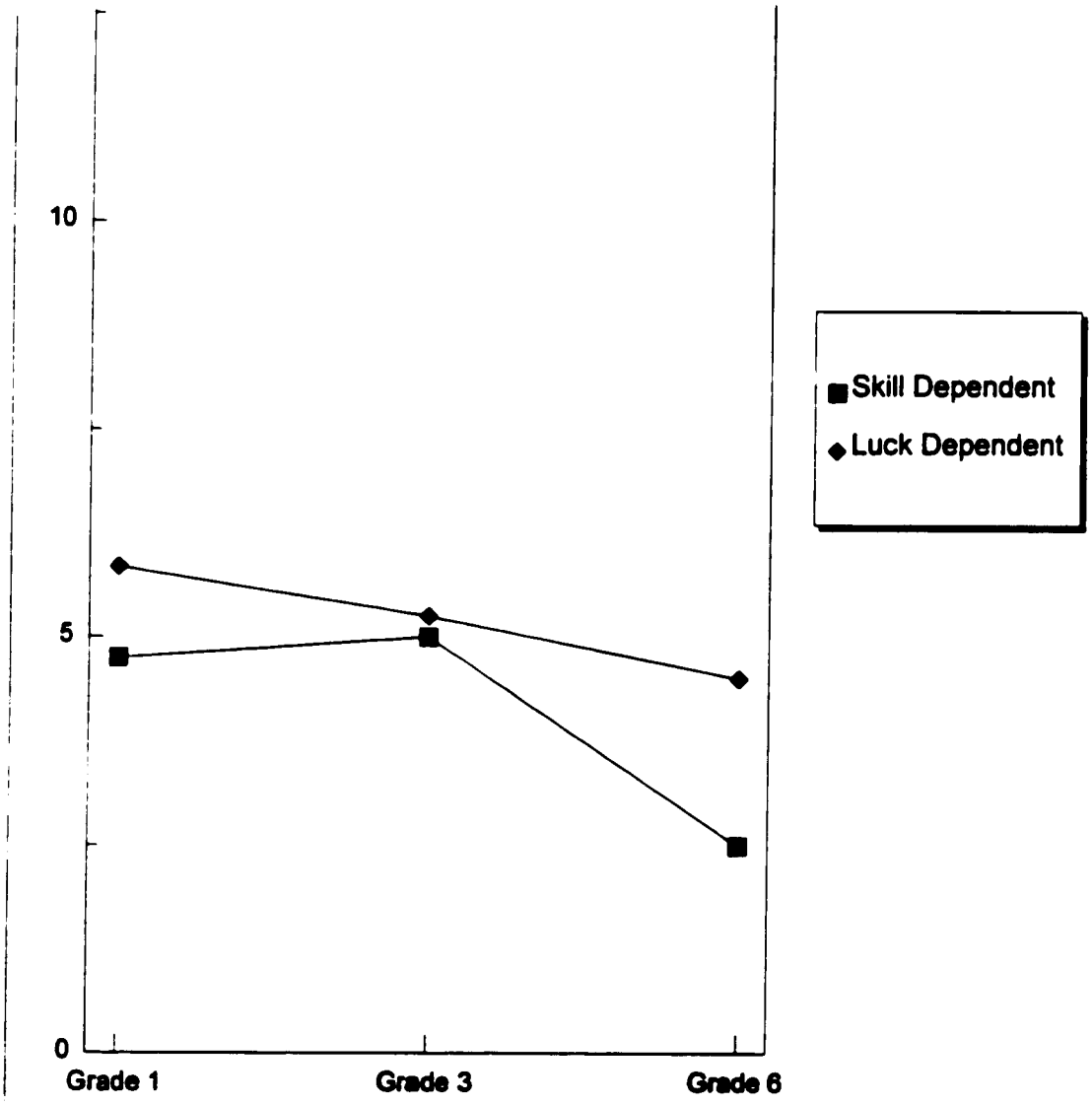


Figure 4.0. Average number of stars awarded in skill dependent and luck dependent bad outcome conditions

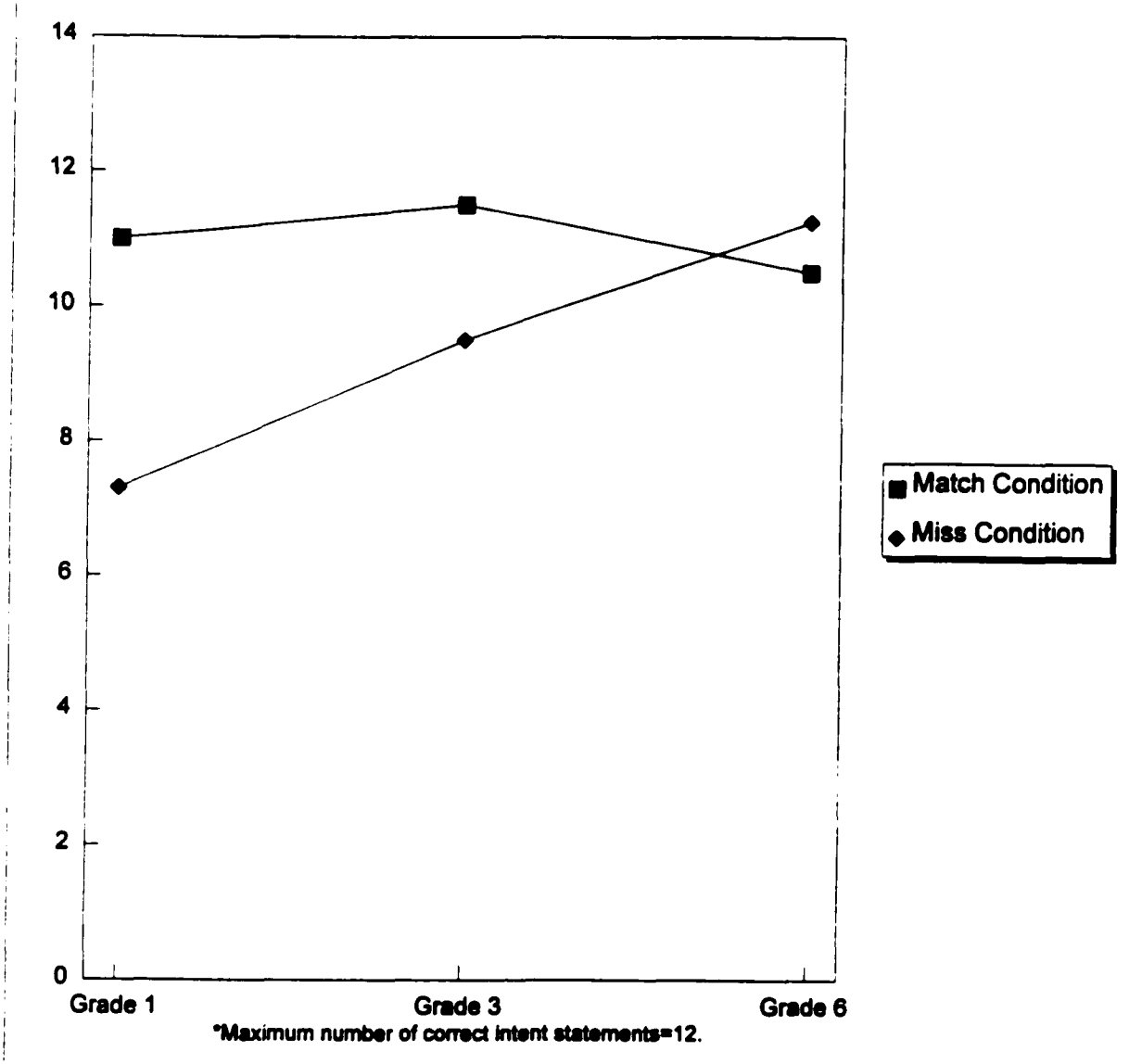


Figure 5.0. Average number of correct intention statements in good and bad outcome conditions.

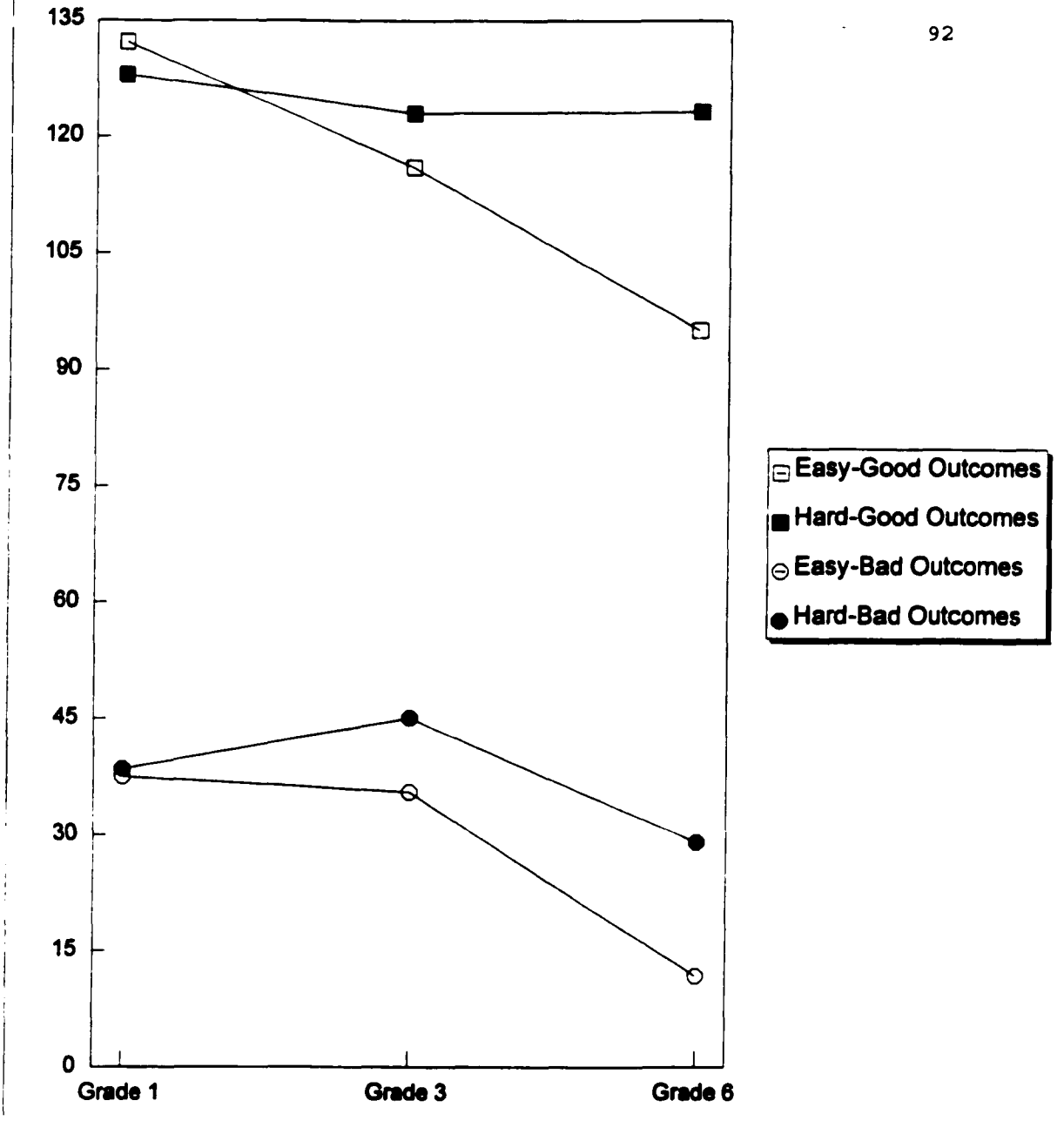


Figure 6.0. Average number of stars awarded in skill-dependent conditions (easy and hard, good and bad outcomes).

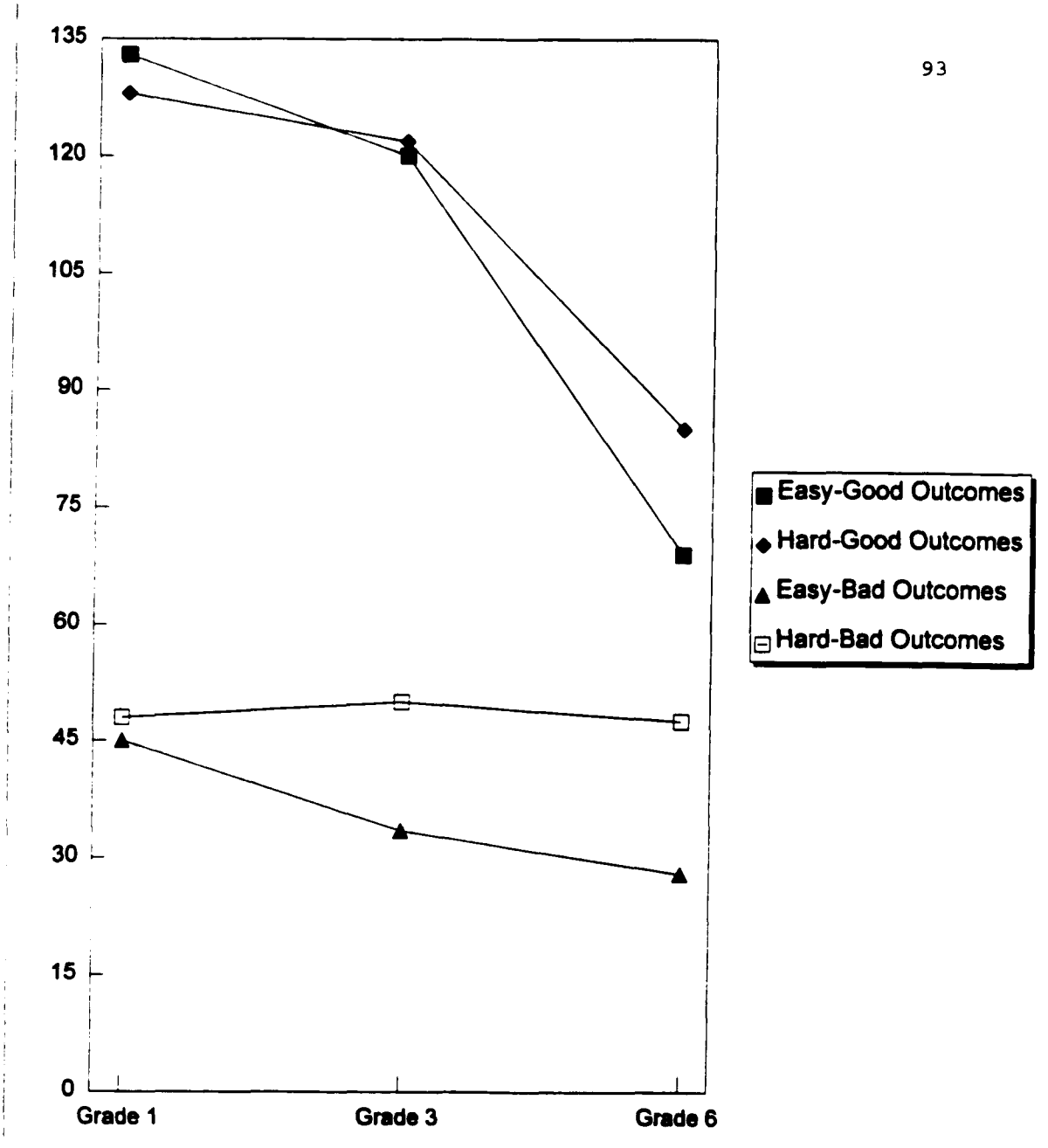


Figure 7.0. Average number of stars awarded in luck dependent conditions (easy and hard, good and bad outcomes).

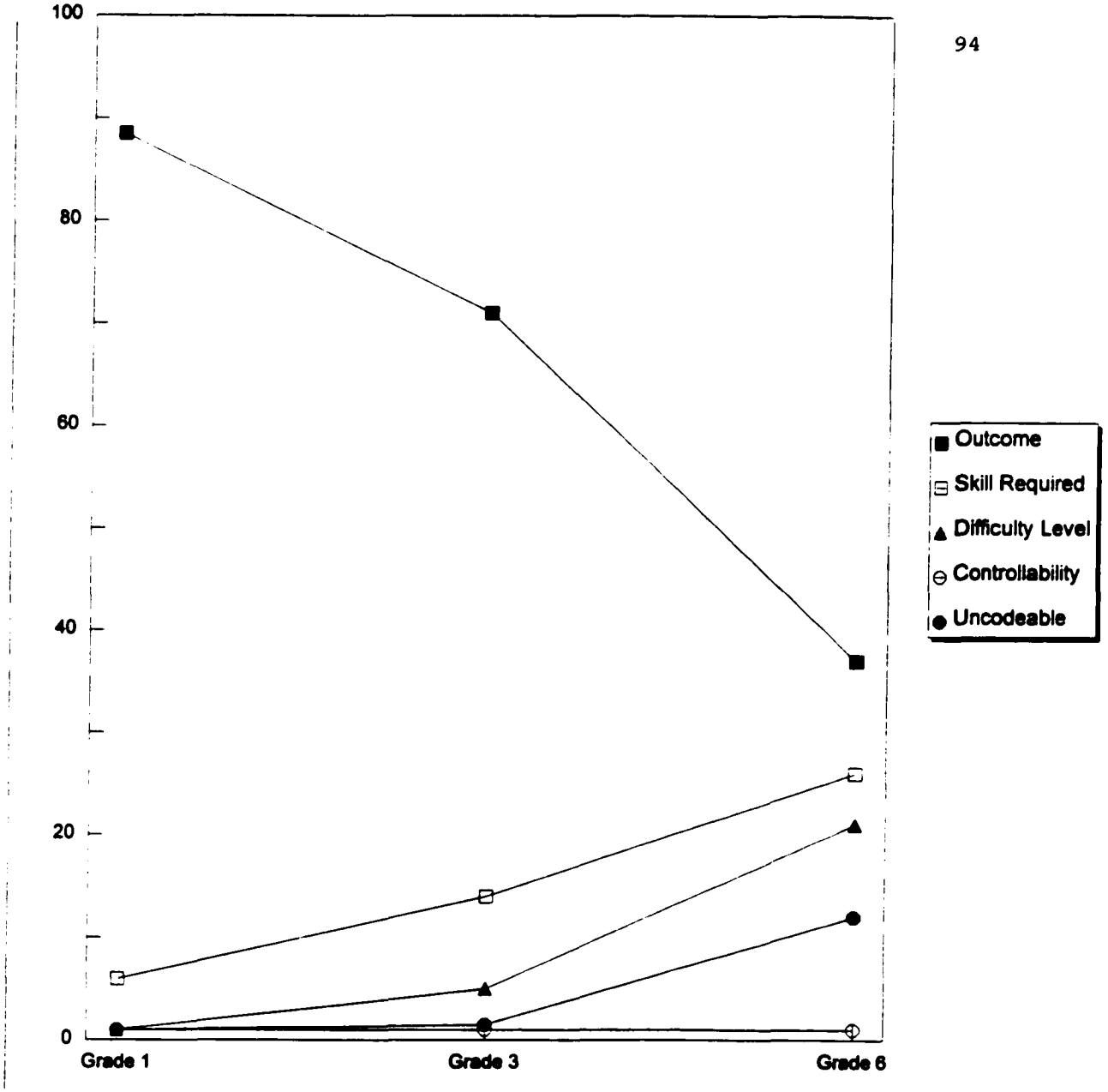


Figure 8.0 Total number of each kind of justification in skill dependent, good outcome conditions.

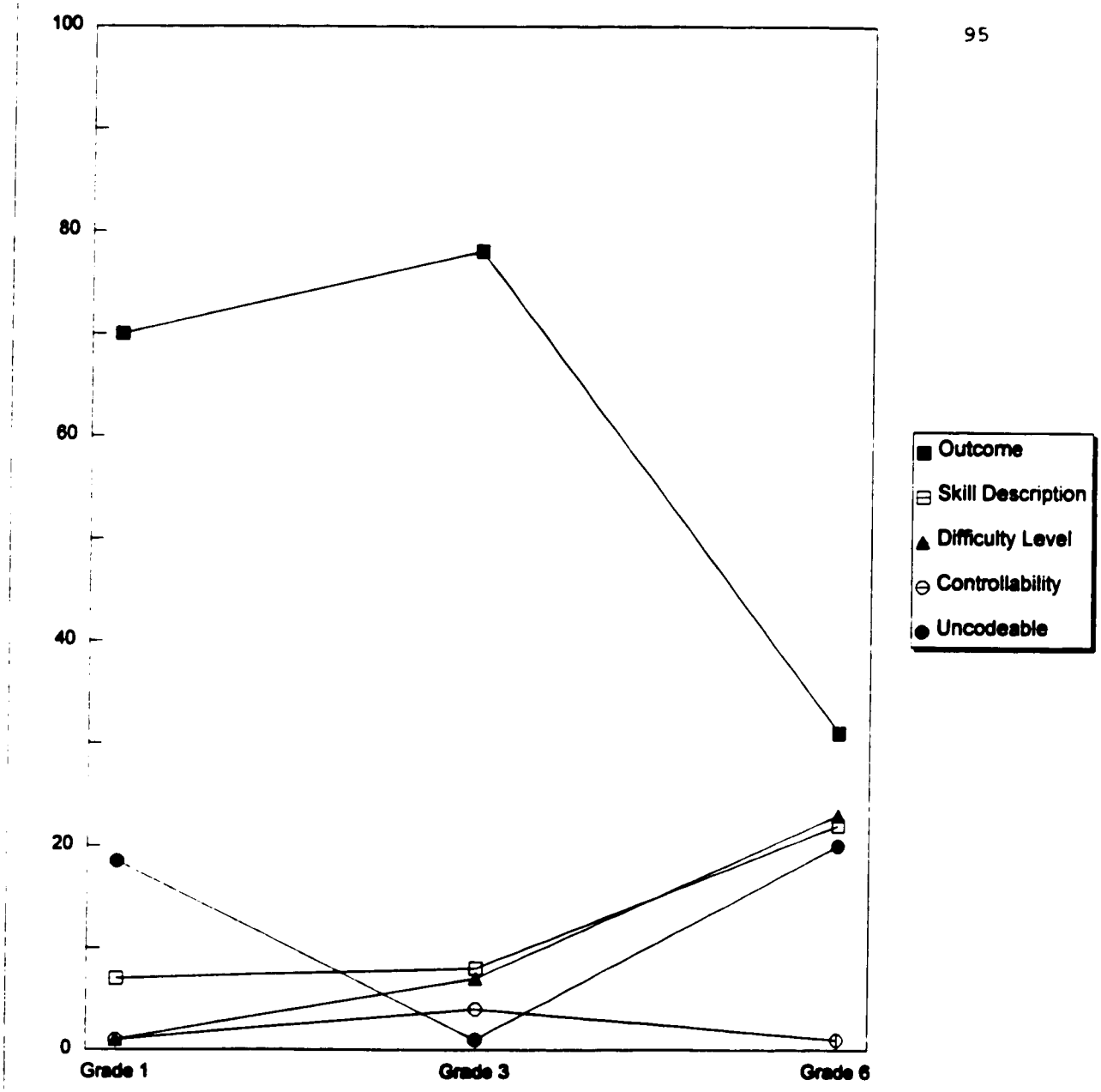


Figure 9.0. Total number of each kind of justification in skill-dependent, bad outcome conditions.

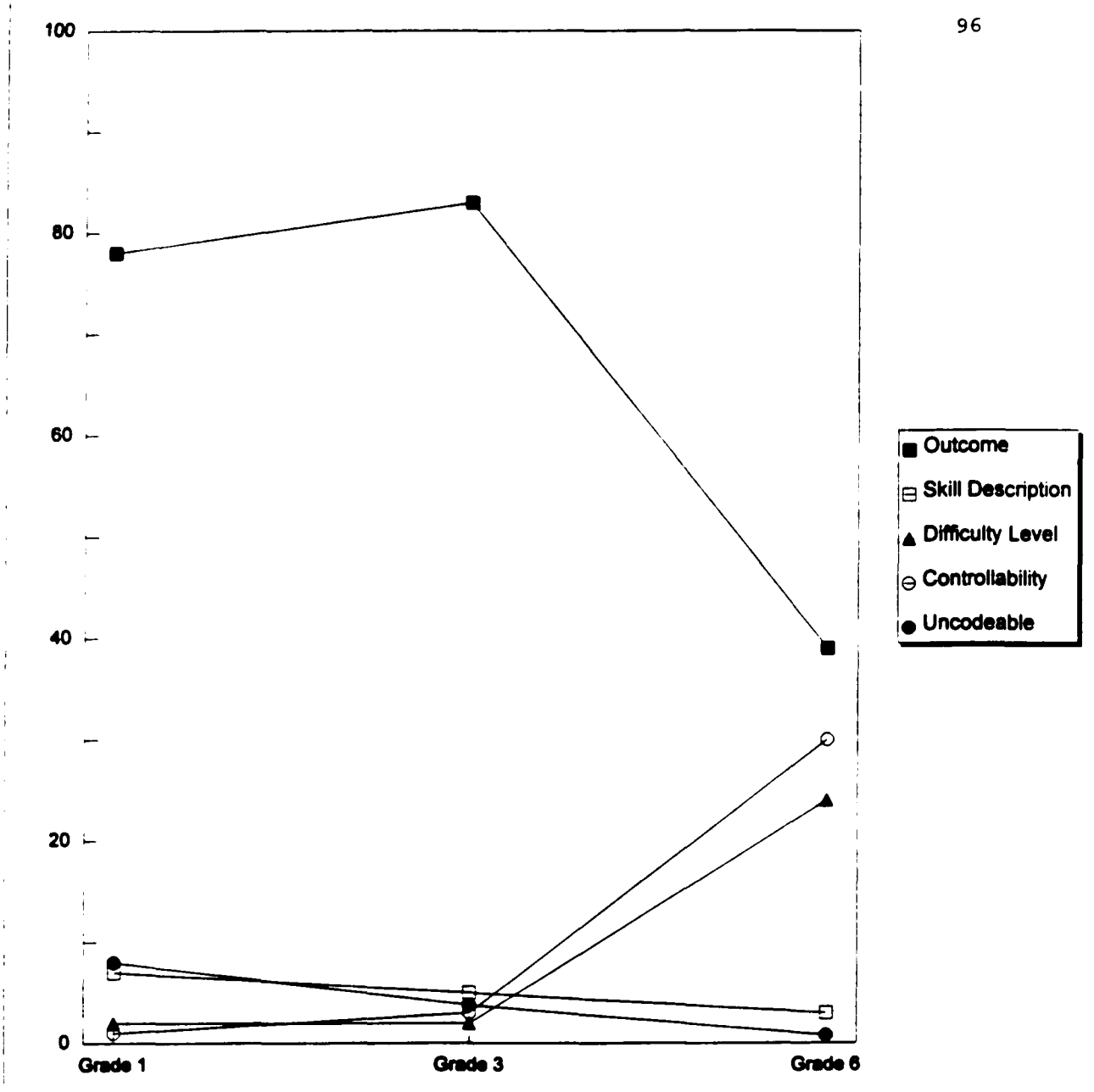


Figure 10.0. Total number of each kind of justification in luck dependent, bad outcome conditions.

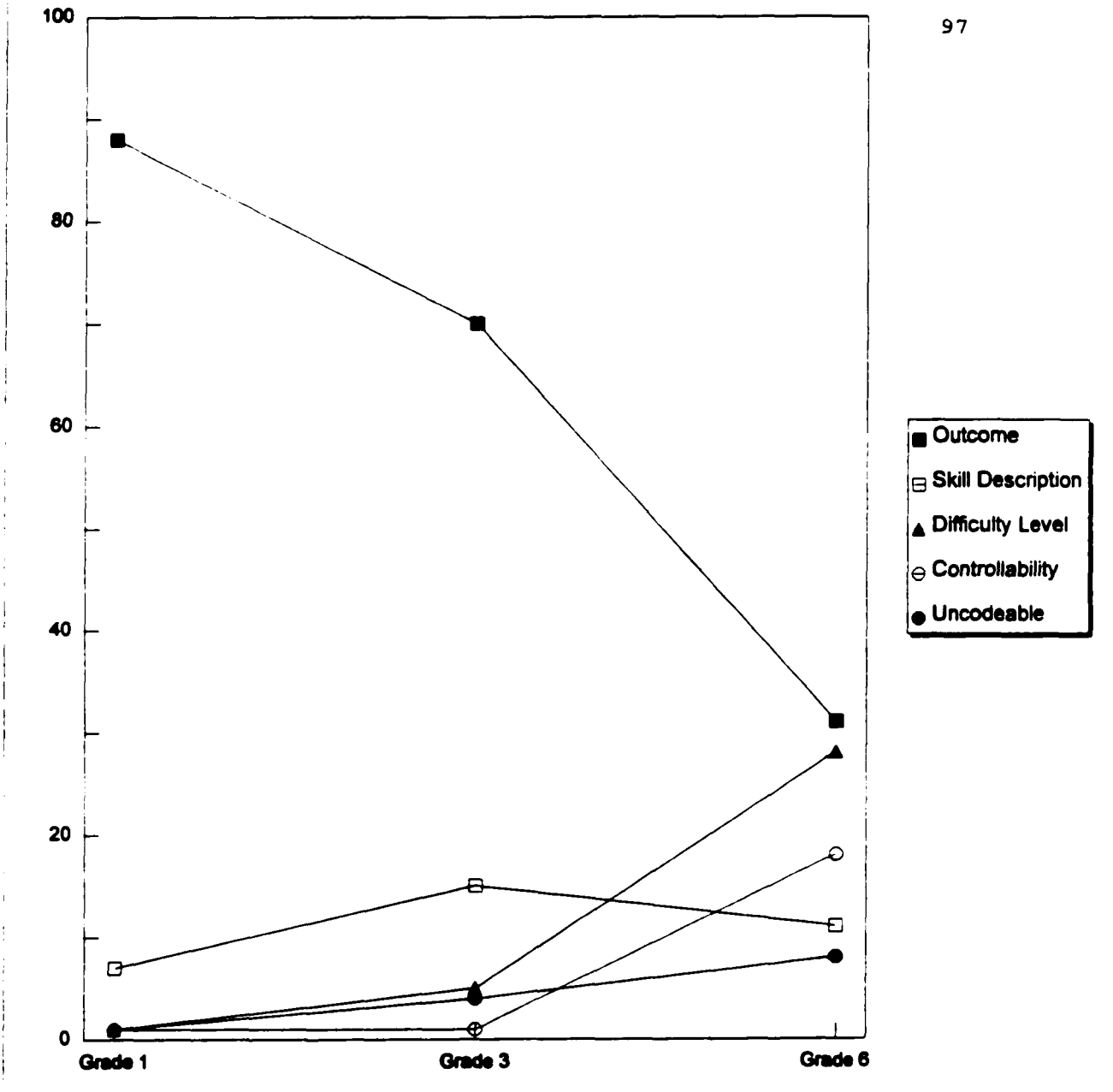


Figure 11.0. Total number of each kind of justification in luck dependent, good outcome conditions.

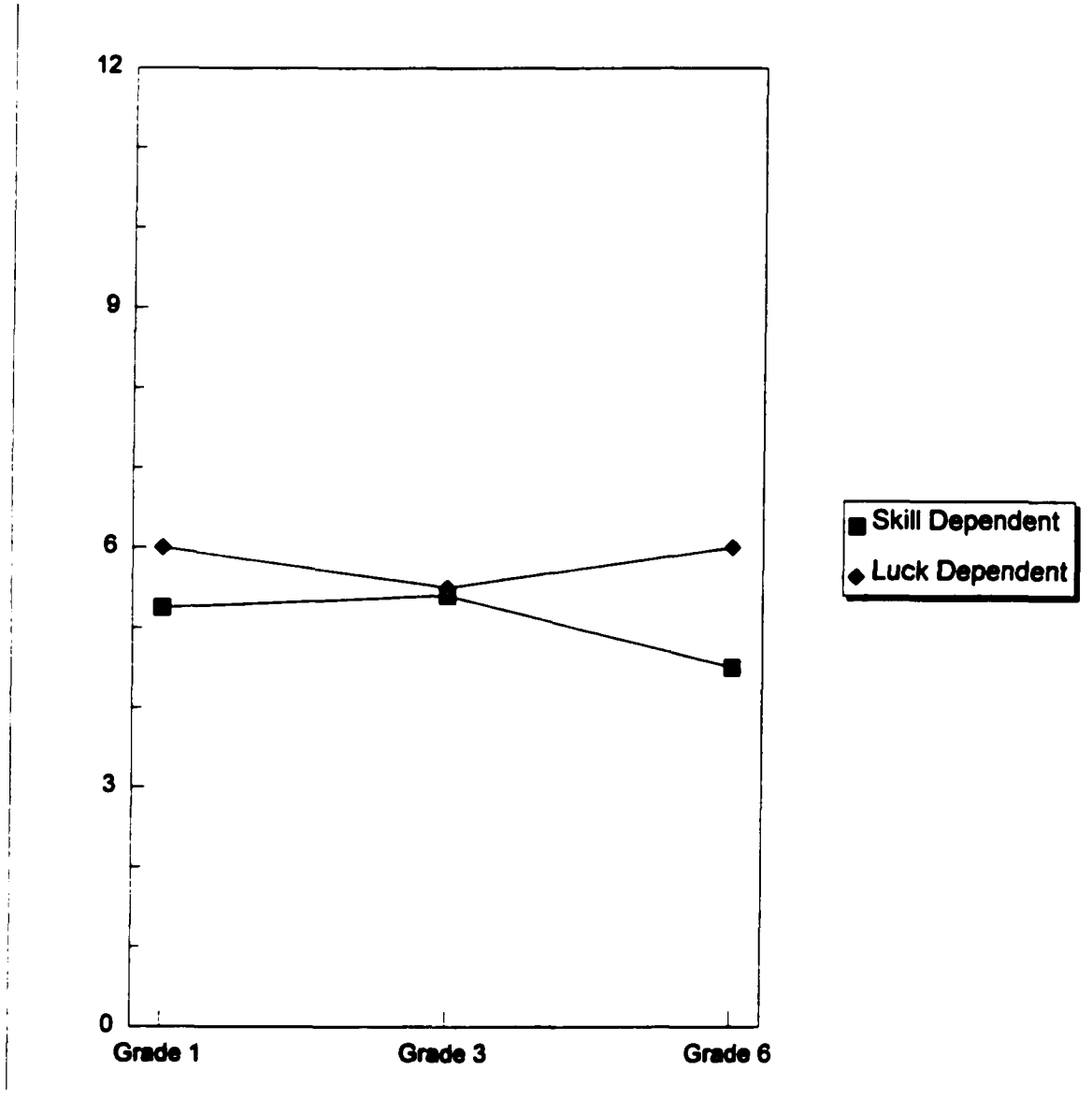


Figure 12.0. Average number of correct intention statements in luck and skill dependent good outcome conditions.

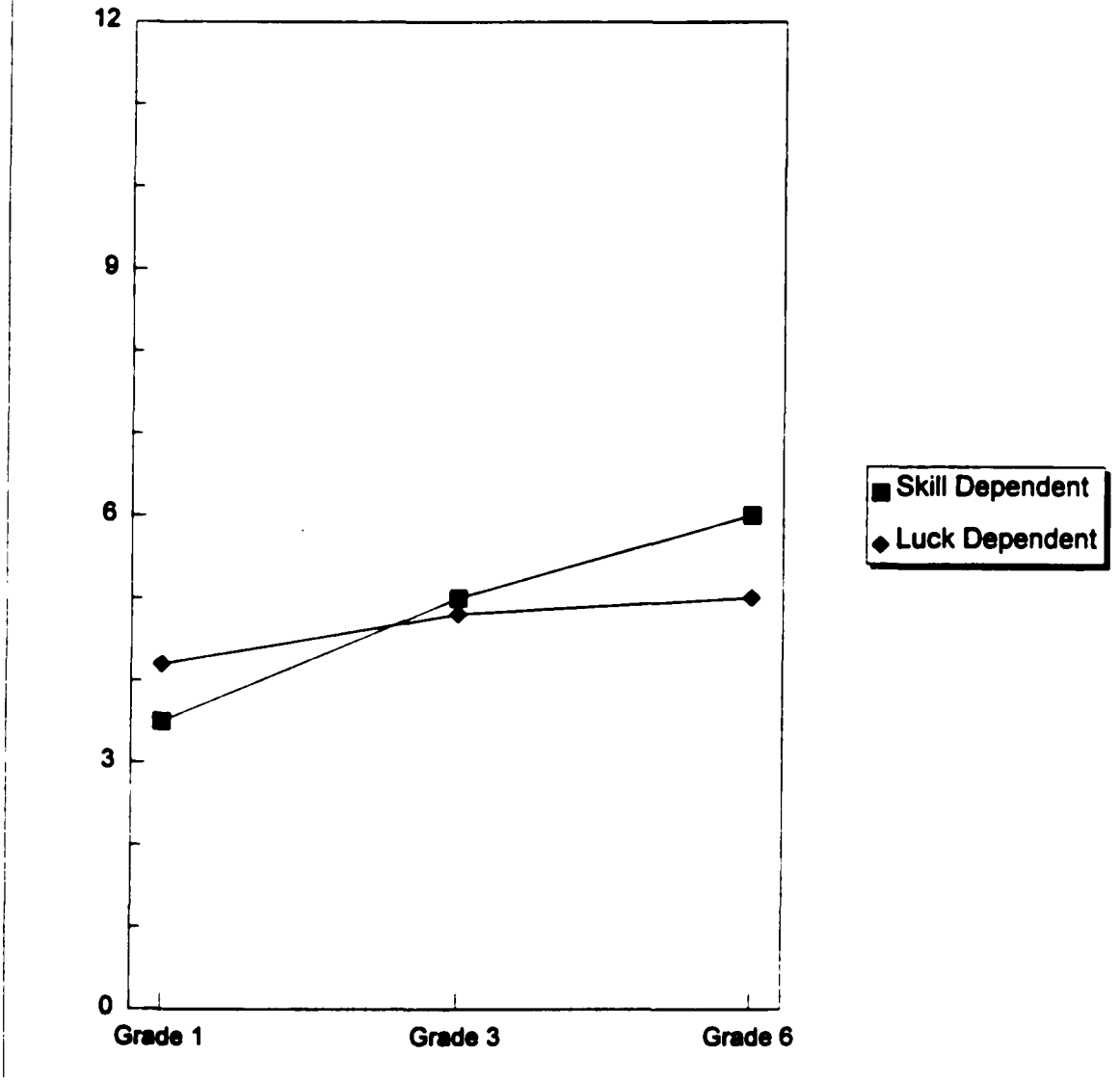


Figure 13.0. Average number of correct intention statements in luck and skill dependent bad outcome conditions.

APPENDIX A

Protocol and Verbal Instructions in Main Procedure - Skill Condition

Two children of the same sex and age group were brought by the Experimenter to the lunch area or other deserted comfortable area on the camp grounds. The children were introduced to each other, and Experimenter introduced herself again. Instructions began as follows:

"One of you (the Actor) will play a game, which I will show you both, and then the other of you (the Observer) will reward him or her by giving him or her 0, 1, 2 or 3 stars, depending on how many you think he or she has earned, how much you think he or she deserved for how well he or she did. We will flip a coin to decide who will play, and who will give out stars."

"Here we have a bag with two colors of chips - yellow and blue. The yellow ones have a smoother texture and the blue ones have a coarser texture. (Both children were given the opportunity at this point to look at and manually inspect the chips). For now, we will play with these chips, which are quite different from each other, and later, we will try it with chips which are a little more similar to each other. (Again, both children were given the opportunity to look at and manually inspect the alternative pair of chips, with the

more similar textures). We need exactly 10 yellow/fine textured and 10 blue/coarse textured chips in the bag. Please count them and put them in the bag now".

For alternate pairs of children, the hard condition was presented first, and the easy condition, second.

"Here is how we will play: Actor (using child's name) will tell us which type of chip he or she would like to pull from the bag. Then, he or she can reach in and feel around. How it feels, how smooth or rough it is should help him or her to select a chip. Then, we will all look at which color it is. Then, Observer (using child's name) will reward Actor (name) with 0-3 stars, and I will ask her a few questions about why he or she thinks that was the right number of stars. I am not asking to argue with you, but only because I am interested in how you think. We will play this game several times. If either of you feel that you would rather not play this game, you can go back to your group, and participate in the usual activities (soccer ball, baseball, arts & crafts, etc., mentioned as appropriate)"

(Not to Subjects) The game was played until at least 3 examples of matches and mismatches between the stated intention and outcome had occurred. Data was recorded regarding the number of trials necessary to be exposed to criterion numbers of both matches and mismatches. This was done in order to have information regarding the proportion of yellow to blue chips to place in the bag for each child's

yoked partner in the subsequent luck-dependent condition.

This also provided information to ascertain whether selecting the intended chip was indeed more difficult for younger children than it was for older ones, which would have implications for how they determine whether skill is necessary, and how much skill is necessary to carry out this task successfully. This eventuality had been adjusted for by interviewing pilot Subjects in all three age groups regarding the ease or difficulty of discriminating the two textures from each other. The first graders were exposed to chips with one grade more textural difference between them, since it had already been found in pilot studies that they have more difficulty making distinctions between the same two textures as the older children. In other words, first graders technically had an easier discrimination to make than older children, but functionally, it was the equivalent discrimination. After the chip had been selected, and the Actor child had received his reward, Observers were interviewed about the following issues, in the following manner:

1) The reasons why they allocated the number of stars they did (providing matching criterion responses). The exact query was: "Why did you give him or her __ number of stars?"

2)Whether the Actor was exhibiting sufficient skill (eliciting a skill assessment). The exact query was: "Was he or she playing well?"

3)Whether the Actor intended that particular outcome (eliciting an intention attribution). The exact query was: "Did he or she mean for the blue, or yellow chip to come out?" 4)When the answer was negative, a follow-up question eliciting a causal attribution statement. The exact query was: "What made that color chip come out?"

All of the above constituted one trial. As stated previously, each child was interviewed in this fashion after three trials where the intention and action had matched, and after three trials where the intention and action had not matched. Then, children who had begun with the more difficult discrimination were switched to the easier discrimination. Children who had been exposed to the easier discrimination first were switched at this point to the more difficult discrimination. The same children also participated in the Auxiliary Procedure, to be described, as well as in one of two Piagetian tasks to help assess their general reasoning level.

APPENDIX B

Protocol and Verbal Instructions in Main Procedure - Luck Condition

Two children of the same sex and age group were brought by the Experimenter to the lunch area or other deserted comfortable area on the camp grounds. The children were introduced to each other, and Experimenter introduced herself again. Instructions began as follows:

"One of you (the Actor) will play a game, which I will show you both, and then the other of you (the Observer) will reward him or her by giving him or her 0, 1, 2 or 3 stars, depending on how many you think he or she had earned, how much you think he or she deserved for how well he or she did. We will flip a coin to decide who will play, and who will give out stars."

"Here we have a bag with two colors of chips - yellow and blue." (Both children were given the opportunity at this point to look at the chips). " There will be twelve yellow and eight blue." (These numbers are given as an example. In reality, the number of chips of each color was determined by the yoking procedure with the previous skill condition. That is, each child in this condition was yoked to a particular child who had participated in the skill condition. That child's proportion of successes and failures in both the easy

and hard versions of the task were noted. Later, when the child who was to be yoked with that one came to participate in the luck task, the proportion of yellow to blue chips which was loaded into the bag was made equivalent to the proportion of successes and failures in the original skill task. This ensured that proportion of successes and failures remained comparable, yet enabled children in the skill condition to control the outcome, while children in the luck condition could not). "For now, we will play with these chips, and later we will change them, so there will be 10 yellow and ten blue ones" (again, an example, indicating the change in proportion, representing findings from the change in difficulty level in the original skill task). "I will ask you (the Observer) to give him or her (the Actor) stars- anywhere from 0 through 3 each time, depending on how well you think he or she did. I will also ask you some questions, not to argue with you, but to get information about how children of different ages feel about certain things. If you are not enjoying the game, you can stop and are free to go back to your activity (soccer, arts & crafts, etc., mentioned as appropriate). As with the skill condition, the game continued until at least three instances of match between intention and outcome and three instances of mismatch between intention and outcome had occurred. On those occasions where enough of one kind of trial had occurred, but not of the other kind of trial, the game was continued, rewards were given, but the

interview about justifications, etc. was eliminated, until the necessary type of trial had been generated.

For alternate pairs of children, the 'hard' condition, or the one with less likelihood of success, was presented first, and the 'easy' condition, the one with greater likelihood of success, was presented second. For the other pairs, the order was reversed.

After each trial instance of intention statement and chip retrieval, the Observer was interviewed about the following issues, in the following manner:

1) The reasons why they allocated the number of stars they did (providing matching criterion responses). The exact query was: "Why did you give him or her ___number of stars?"

2) Whether the Actor was exhibiting sufficient skill (eliciting a skill assessment). The exact query was: "Was he or she playing well?"

3) Whether the Actor intended that particular outcome (eliciting an intention attribution). The exact query was: "Did he or she mean for the blue, or yellow chip to come out?"

4)When the answer was negative, a follow-up question eliciting a causal attribution statement. The exact query was: "What made that color chip come out?"

All of the above constituted one trial. As stated previously, each child was interviewed in this fashion after three trials where the intention and action had matched, and after three trials where they had.

APPENDIX C

Auxiliary Procedure Description and Protocol

The fence study was an auxiliary procedure, selected to complement the actual dissertation experiment. It is a replication of a study conducted by Wimmer & Wachter (1985) It was chosen because the study addresses itself to at least some, though not all of the experimental questions inherent in the dissertation experiment. Just as the main study, it is a developmental study, looking at children of different ages. It deals with children's evaluations of the concrete achievements of a peer, and what criteria are important to them at several ages. The significant factors under investigation are similar to the ones in the main study. They are: the younger children's emphasis on outcome (whether positive or negative), and older children's emphasis on the skill exhibited by the actor. In this theoretical formulation, skill is composed of effort and ability. It is generally found that younger children see effort as the equivalent of ability, eg, "if he tried hard, he is smart". Somewhat older children recognize that effort and ability are independent factors, each of which contributes to skill. Yet older children acknowledge that there exists a compensatory relationship between effort and ability. That is to say, more effort can compensate for less ability, and if one has more

ability, one can expend less effort to achieve the same outcome. (Naturally, the difficulty of the task determines how much effort and ability cumulatively one needs to achieve it. This issue has been addressed in the main study, and not here, however).

DESCRIPTION OF PROCEDURE-There were 56 Subjects in each of grades 1, 3 and 6. For some of the Subjects (32 per grade in the original study), their capability to conserve was another dimension in their eligibility to participate in the study. No first graders succeeded in conserving mass, all sixth graders succeeded, and among the third graders conservation of mass was inconsistent. In the subsequent replication, 24 more children participated at each grade level. Here, their eligibility to participate was determined by appropriate age and grade ranking as well as the ability to balance a beam on a fulcrum at an age-appropriate level, according to Siegler's descriptions of Piaget's balance beam task.. All participating first graders were classified as Level I in capability to balance the beam, all third graders who participated were classified as Level II, and all sixth graders who participated were classified as Level III. Once they had been pre-selected in this fashion, they were exposed to the following procedure:

There was a basic story, told to each child, in all age groups. The story was as follows: A father asked his son to paint a fence. The factors added to the story from this point

are as follows: Ability-less or more, represented by level of maturity of the boy as shown in one of two drawings, either of a young or older boy, shown to each Subject. Effort-the child was either shown to have put in much effort, represented by a picture of a paint bucket and brush, or little effort represented by a picture of a baseball, bat and glove. Outcome-a positive outcome was represented by a picture of a fully painted fence. A negative one was depicted by a picture of a partially painted fence. When the three factors are combined in every possible way, there are 8 story types . Each child in each age group was exposed to each story type two times, and the data collected represents the sum of each child's two responses for that story type. The story types were as follows:

ABILITY	EFFORT	OUTCOME
HIGH	HIGH	HIGH
HIGH	LOW	HIGH
HIGH	HIGH	LOW
HIGH	LOW	LOW
LOW	LOW	LOW
LOW	LOW	HIGH
LOW	HIGH	HIGH
LOW	HIGH	LOW

The dependent measures were as follows: 1) number of cookies awarded (0-8), to the fictional character for his achievement on the previously described task.

2) Justification for the number of cookies given, to be coded as outcome-oriented, effort-oriented, ability-oriented, or compensatory relationship-oriented. Multiply coded answers were possible. For example, if a child justified the number of cookies awarded by saying "he painted all of the fence, but he did not try very hard so I only gave him 5 cookies", it was coded as 1/2 an outcome-oriented answer and 1/2 an effort-oriented answer.

All children who participated in the main procedure also participated in the auxiliary study. There were twice as many participants per grade here because all children in skill-based conditions and all children in luck-based conditions participated in the auxiliary procedure. By random selection, half of the Subjects were exposed to the main procedure first and the auxiliary procedure next; the other half were exposed to the auxiliary procedure first, and the main procedure next. Subjects were introduced to this task as follows:

"I am going to show you a picture of a boy, John. His father has asked him to paint their fence. In this game, sometimes, he will look like this (showing a flash card of a young, approximately 4 year old boy). At other times, he will look like this (Showing a flash card of an older boy, approximately 10 years old). I suspect that this big John can do more work around his house than this little John can."

"I will also be showing you a picture of what he actually spent the afternoon doing. That picture will sometimes look

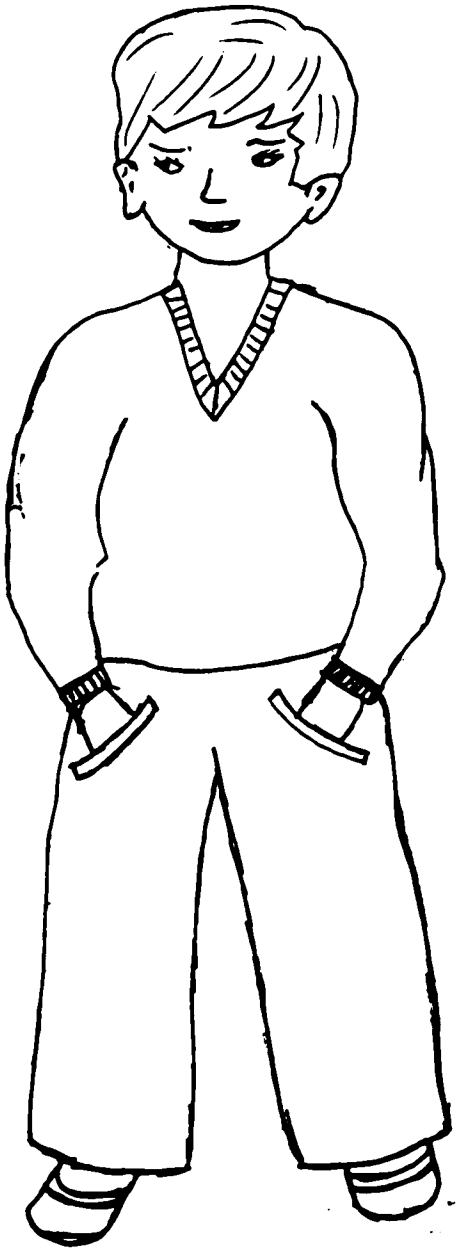
like this (showing a flash card of a bucket of paint and a wet paintbrush), or sometimes like this (showing a flash card of a baseball, bat and mitt). It certainly looks like he was trying to paint that fence more when he was using these things (the paint and brush) than when he was using these things (showing a baseball, bat and mitt)."

"I will also show you what the fence ended up looking like. That picture will sometimes look like this (a flash card of a fully painted fence) or like this (a flash card of a partially painted fence). This one looks much better than this one, doesn't it?'

In this way, all of the visual stimuli were introduced before any trials occurred. Their implications for ability, effort and outcome were also elucidated in this fashion. During the two examples of each kind of trial, as listed above, children were asked to award cookies, (0-8) to the hypothetical actor, and to justify their choice of the extent of rewards given. Justifications were coded as outcome-oriented, ability-oriented, effort-oriented, or reflecting a compensatory relationship between effort and ability.

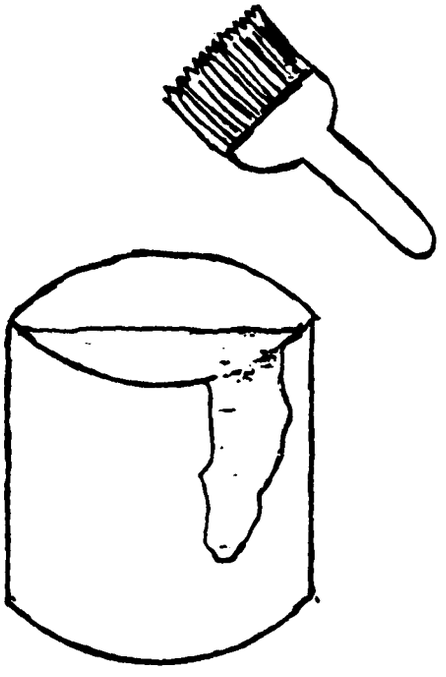
APPENDIX D

Flash Card Stimuli Utilized in Auxiliary Study

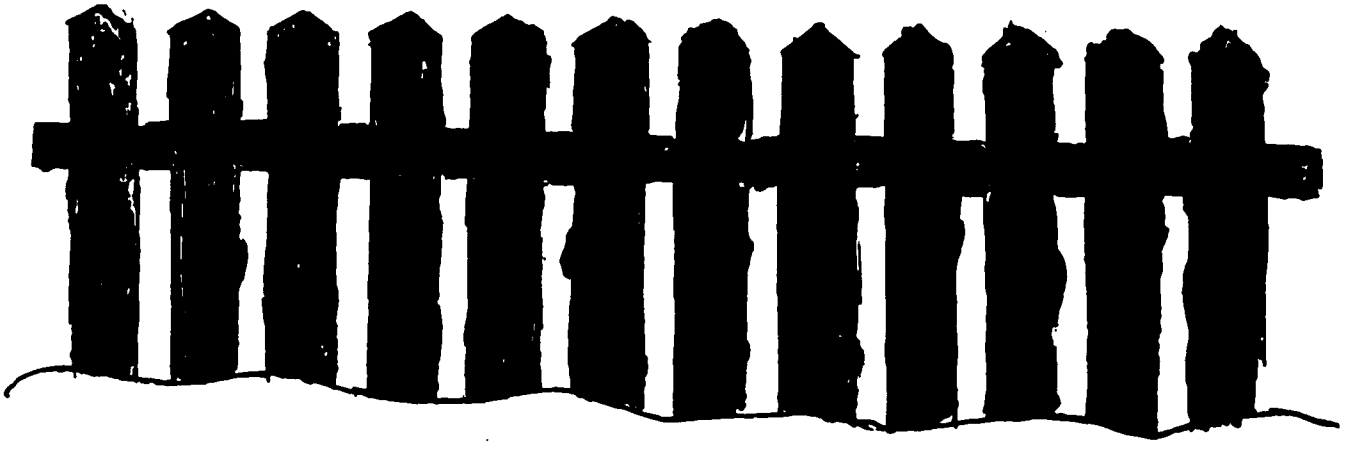


High Ability, represented by older boy.

Low Ability, represented by younger boy.

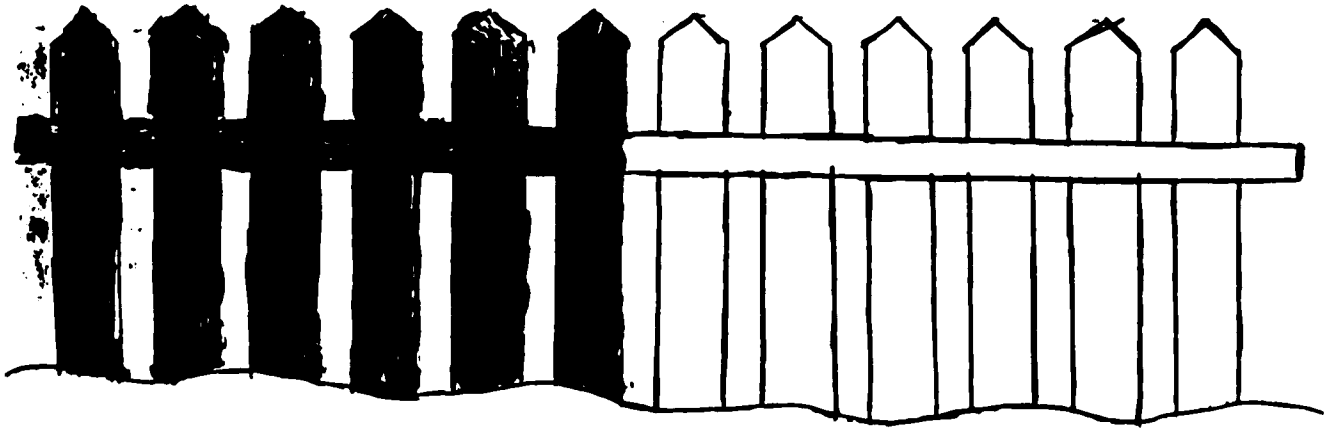


High Effort, represented by paint and brush.



Good Outcome, represented by fully painted fence

Low Effort, represented by baseball and bat.



Poor Outcome, represented by partially painted fence.

REFERENCES

Farmer, H. Virpoel, W. & Maehr, M. (1991). Achievement contexts: Effect on achievement values and causal attributions. Journal of Educational Research, 85(1), 26-38.

Goldhammer, E. & Saltzstein, H.D. (1992). Development of criteria for achievement evaluations. Paper presented at the Annual Convention of the American Psychological Association, Washington, D.C.

Heider, F. (1958). The Psychology of Interpersonal Relations. New York: Wiley.

Keasey, C.B. (1978). Children's developing awareness and usage of intentionality and motives. In C.B. Keasey (Ed.), Nebraska Symposium on Motivation (Vol. 25). Lincoln: University of Nebraska Press.

King, M. (1971). The development of some intention concepts in young children. Child Development, 42, 1145-1152.

Kun, A. (1977). Development of the magnitude-covariation and compensation schemata in ability and effort

attributions of performance. Child Development, 48, 862-873.

Mizokawa, D. & Ryckman, D. (1990). Attributions of academic success and failure: A comparison of six Asian-American ethnic groups. Journal of Cross-Cultural Psychology, 21(4),434-451.

Nicholls, J.G. (1978). The development of the concepts of effort and ability, perceptions of academic attainment and the understanding that difficult tasks require more ability. Child Development, 49, 800-814.

Nicholls, J.G. & Miller, A. (1984). Reasoning about the ability of self and others: A developmental study. Child Development, 55, 1990-1999.

Nicholls, J.G. & Miller, A.T. (1985). Differentiation of the concepts of luck and skill. Developmental Psychology, 21, 76-82.

Piaget, J.The Origins of Intelligence in Children. New York: Norton, 1963.

Piaget, J. & Inhelder, B. (1975). The Origin of the Idea of Chance in Children. New York: Norton.

Piaget, J. (1932). The Moral Judgment of the Child. London: Kegan-Paul.

Ryckman, D. & Mizokawa, D. (1991). Cross-situation variability of attributions for success and failure: A cross-sectional study. Journal of

Adolescent Research, 6(2), 197-211.

Saltzstein, H.D. (1987). The representation of other persons' moral judgments. Presented at Moral Judgment Symposium: Moral Responsibility and Causal Reasoning. Society for Research in Child Development, Baltimore, Maryland.

Saltzstein, H.D. & Goldhammer, E. (1986). Children's use of intentionality and outcome as criteria in rewarding peer performance at three age levels. Paper presented at The Sixteenth Annual Convention of the Piaget Society, Philadelphia, Pennsylvania.

Saltzstein, H.D. & Goldhammer, E. (1990). Developmental changes in children's criteria for rewarding a peer in an experimental game. Merrill-Palmer Quarterly, (36), 557-572.

Shultz, T.R. & Wells, D.(1985). Judging the intentionality of action-outcomes. Developmental Psychology, 21, 83-89.

Siegler, R.S. (Ed.), Children's Thinking: What Develops?, Laurence Erlbaum Associates, Inc. 1978.

Stipek, D. & Gralinski, J. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. Journal of Educational Psychology, 83(3), 361-371.

Weiner, B. & Peter, N.V. (1973). A cognitive-developmental analysis of achievement and moral judgments. Developmental Psychology, 9, 290-309.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. Psychological Review, 92, 548-573.

White, S. (1993). The effect of gender and age on causal attribution in softball players. International Journal of Sport Psychology, 24(1), 49-58.

Wimmer, H. Wachter, J. & Perner, J.(1982). Cognitive autonomy of the development of moral evaluation of achievement. Child Development, 53, 668-676.

Yamauchi, H. (1990). Actor and observer attributions by Japanese subjects for success and failure in non-competitive situations. Psychologia: An International Journal of Psychology in the Orient, 33(4), 212-219.