DOCUMENT RESUME

ED 048 341 TM 000 414

AUTHOR Lieberman, Marcus
TITLE Estimation of a Moral Judgment Level Using Items Whose Alternatives Form a Graded Scale.

PUB DATE Feb 71
NOTE 118p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$6.58
DESCRIPTORS Age Differences, Behavior Theories, Correlation, Cultural Differences, Decision Making Skills, *Developmental Psychology, Factor Analysis, Item Analysis, Learning Theories, *Moral Values, Psychological Patterns, *Psychometrics, *Research Methodology, Socioeconomic Status, *Statistical Analysis
IDENTIFIERS Great Britain, *Latent Trait Theory, United States

ABSTRACT

The limitations in the methods which are usually used in stage theory test analysis to determine the stage of development of an individual are outlined. Piaget's and Kohlberg's conceptions of moral judgment are presented along with one method of estimating a subject's level of development. The concept of a latent trait is developed. Its relationship to a stage sequence, through stating the probability of being in each stage, is illustrated in a non-technical manner. Using nine situations or stories pointing up a moral dilemma, the stage boundaries within stories, the discriminating powers of the stories, and the latent trait value parameters are derived using maximum likelihood procedures. The calculation of discriminating power requires the derivation of a new form of polychric correlation coefficient. Various studies reported in the literature concerning differences in moral judgment among ages, sexes, socioeconomic classes and culture are replicated and the procedures developed here applied to the data. Finally, the tests of the assumptions required by the model, the results of the above replications, and the generalizability of the method of analysis are discussed in some detail. (Author/DG)

Estimation of a Moral Judgment Level Using Items Whose Alternatives Form a Graded Scale

Marcus Lieberman

Institute for Educational Research
Downers Grove, Illinois

Paper presented at the annual meeting of the American Educational Research Association, February, 1971

INTRODUCTION

Developmental psychologists have proposed many stage theories which are hierarchical in nature and supported by empirical studies. These include, among others, concepts of dreams, realism, animism and life, liquid and solid conservation and moral judgment.

Determination of the stage of development of a subject, in general, involves presenting the subject with a series of tasks or situations. Through responses made to questions about these tasks or situations, a stage "score" is determined for each and an arithmetic average represents the degree of development.

However, an arithmetic average does not take into account all the information contained in the small amounts of data extant, specifically (1) boundaries between stages are typically non-equidistant, indicating that it may be more difficult to pass from one stage to the next than to pass from the second to a third, or, that subjects remain in some stages longer than others; (2) some tasks or situations may be simpler or more complex than others, resulting in responses at a higher or lower level than the subject's actual stage of development; (3) some tasks or situations may be more accurate indicators of the stage of development, i.e., more discriminating than others.

In addition to the above data, an underlying or latent trait value, which is a continuous variable assumed normally distributed, can be derived from the task or situation characteristics and a subject's pattern of stage scores. If the characteristics are determined for different ages, subjects who advance through the stages while they grow older will be judged at a nearly constant moral judgment level not unlike the concept of intelligence quotient. If the characteristics are developed from an arbitrary norm group or average of different ages, the same story characteristics would apply to subjects of all ages and the moral judgment level would increase as a subject passed through the stages in a manner analogous to the concept of mental age. In either case, this continuous variable is related to the theoretical stages by stating the probability of a subject with a given latent trait value being in each stage.

It is the purpose of this paper to propose an analysis that will quantitatively describe the four values discussed above. The first three represent task or situation differences, while the fourth represents subject differences.

The method of analysis is applied to data gathered in studies concerned with estimating a level of moral judgment development. Subjects in these studies differed in age, sex, socioeconomic status and culture. In the moral judgment stage concept, the situations are stories representing moral dilemmas. Subjects' responses to questions about the stories are rated from one to six corresponding to the six stages in moral judgment development first described by Piaget (1932) and later refined by Kohlberg (1963).

The organization of this paper is as follows. First a discussion of Piaget's and Kohlberg's conceptions of moral judgment is presented along with one method of estimating a subject's level of development. Then, the concept of a latent trait is developed. It's relationship to a stage sequence through stating the probability of being in each stage is illustrated in a non-technical manner. Next, the stage boundaries within stories, the discriminating powers of the stories and the latent trait value parameters are derived using maximum likelihood procedures. The calculation of discriminating power requires the derivation of a new form of polychoric correlation coefficient. These procedures are then applied to the data and various studies reported in the literature concerning differences among ages, sexes, socioeconomic classes and cultures are replicated. Finally, in the closing chapter, the tests of assumptions required by the model, the results of the above replications and the generalizability of the method of analysis is discussed.

CHAPTER I

THE DEVELOPMENT OF MORAL JUDGMENT

This paper is concerned with the quantifying of stage-type data from one concept of developmental psychology. Thus, it is useful at the onset, to describe the nature of developmental psychology by contrasting it with learning theory.

While learning theory is concerned with the acquisition of behavior, the construction of mathematical models and the analysis of learning processes, developmental psychology emphasizes the description of stages defined by structural wholes, forming an invariant order of succession where passage from an inferior to superior stage is equivalent to an integration. Learning theorists' views are essentially reductionistic and exposed to strict experimentation, usually on lower species than man, while developmentalists are content working with broader concepts underlying the stage sequencies.

The developmental theory discussed in this paper is that which underlies the concept of moral judgment. The main difference between moral judgment and moral knowledge is that the former involves the child's use and interpretation of rules in conflict situations and his reasons for moral action, while the latter indicates only a familiarity with the basic rules and conventions of our society, learned by first grade as shown by Hartshorne and May (1928-1930). Moral knowledge scores appear to indicate intelligence, cultural background and desire to make a good impression, rather than level of moral development. Originally formulated by Piaget (1932), the theory of moral judgment has been refined by Kohlberg (1963) yielding a quantitative method to estimate the level of development of that trait.

In Piaget's (1932) developmental theory, the child first moves from an amoral stage to a stage of respect for sacred rules. This transition, usually taking place between three and eight years, coupled with the cognitive limitations of that span, results in a confusion of moral rules with physical laws. Rules are viewed as fixed external things rather than instruments of human purposes and values. The young child's inability to distinguish subjective and objective aspects of his experience ("realism") along with his inability to distinguish his own perspective on events from that of others ("egocentrism") leads him to view rules as things, sacred and unchangeable.

Intellectual growth transforms the child's perceptions of rules from external authoritarian commands to internal principles. Piaget views these internal norms as a sense of justice; i.e., a concern for reciprocity and equality between individuals. At about eight to ten years an "autonomous" justice replaces the earlier "heteronomous" morality based on unquestioning acceptance of adult authority.

Piaget assumes a culturally universal age development of a sense of justice involving increasing concern for the needs and feelings of others along with the conceptions of reciprocity and equality. Respect for authority and the rules of adult society are reenforced, though, since adult institutions are founded on reciprocity, equality of treatment, service to human needs, etc.

Kohlberg (1963) has defined six stages in the development of moral judgment where in each, the way of viewing or solving a moral dilemma or conflict situation is characterized differently. Pairs of stages are grouped beneath levels with the following characteristics:

Level I: Value resides in external quasi-physical happenings, in bad acts, or in quasi-physical needs rather than in persons or standards.

Stage 1. Obedience and punishment orientation. Egocentric deference to superior power or prestige, or a trouble avoiding set. Objective responsibility.

Stage 2. Naively egoistic orientation. Right action is that instrumentally satisfying the self's needs and occasionally, others'. Awareness of relativism of value to each actor's needs and perspective. Naive egalitariansim and orientation to exchange and reciprocity.

Level II: Moral value resides in performing good or right roles in maintaining the conventional order and the expectations of others.

Stage 3. Good boy orientation. Orientation to approval and to pleasing and helping others. Conformity to stereotypical images of majority or natural role, behavior, and judgment by intentions.

Stage 4. Authority and social order maintaining orientation. Orientation to "doing duty" and to showing respect for authority and maintaining the given social order for its own sake. Regard for earned expectations of others.

Level III: Moral value resides in conformity by the self to shared or shareable standards, rights or duties.

Stage 5. Contractual legalistic orientation. Recognition of an arbitrary element or starting point in rules or expectations for the sake of agreement. Duty defined in terms of contract, general avoidance of violation of the will or rights of others, and majority will and welfare.

Stage 6. Conscience or principle orientation. Orientation not only to actually ordained social rules but to principles of choice involving appeal to logical universality and consistency. Orientation to conscience as a directing agent and to mutual respect and trust.

These characteristics may also be defined with respect to various aspects of morality such as "Concept of Rights" and "Motivation for Rule Obedience or Moral Action."

Kohlberg's six stages form a universal sequence such that movement from one stage to the next is dependent on attainment of each of the preceding stages. A more advanced stage is not merely an addition to a less advanced stage but represents a reorganization of thinking, displacing, or controlling less advanced stages.

Kohlberg (1964a) demonstrated that middle class children follow the same sequence in the moral judgment graded scale scheme but at a faster rate than lower class children. Little data has been gathered to explore sex differences, but pilot studies done by Kohlberg (1964b) indicate similar results with girls reaching a constant level earlier. In addition, Grahman and Simpson (unpublished) found that English children also follow the sequence but at a slower rate than American children. Age trends in various cultures and social classes consistent with Kohlberg's stage orderings have been found by Kohlberg (1968). To support the stage structure hypothesis, a Guttman quasi-simplex pattern in the correlations between the stages is reported by Kohlberg (1963). Further support of the stages in moral judgment development comes from studies of individual development by Kohlberg (1968).

Kohlberg has written nine stories or social dilemmas which he uses to estimate a subject's level of moral judgment. These are given in the Appendix. Each story is presented to the subject and probing questions are asked about each. Either or both the story and questions may be in written or oral form.

Responses are judged as to the stage in moral development they most closely reflect. If a very large proportion of the responses are at the same stage, that score is assigned and weighted with 100 units. If the subject's responses are mixed, two stage scores are assigned, e.g., 3(1), where the major stage receives a weight of 66 2/3, and the minor stage a weight of 33 1/3. The sum of the scores is divided by the number of stories employed to obtain the global score or moral quotient.

An example will serve to illustrate this procedure.

Story No.	I	II	III	IV	V	VI
Score	2	2(1)	3(2)	3	4(1)	3(1)

$$2(100) + (2(66.67) + 1(33.33)) + (3(66.67) + 2(33.33)) + 3(100) + (4(66.67) + 1(33.33)) + (3(66.67) + 1(33.33)) = 1{,}467$$

$$\text{Moral Quotient or Global Score} = \frac{1{,}467}{6} = 244$$

There are four major assumptions in the above method of estimation. One is that the stories are of equal complexity. For example, it is as "easy" to be rated at stage four on one story as it is on another. A second involves the equal "distances" between stages; i.e., it is as easy to pass from any stage to the next highest one. Third, all stories discriminate equally; i.e., each could serve as a measure of a subject's moral judgment level. The fourth is concerned with the interrelationships among the stories. The correlation matrix among the item ought to remain constant regardless of age, race, culture, or sex.

This paper will use psychometric techniques to estimate a level of moral judgment and to test these assumptions.

CHAPTER II

MORAL JUDGEMENT AS A LATENT TRAIT

In any latent trait theory, it is assumed that behavior can be accounted for, to a great degree, by isolating certain consistent and stable human characteristics called traits. If we know a subject's values on these traits, we may predict or explain his performance in relevant situations.

Lord and Novick (1968) discuss latent traits in their concluding chapter:

> The problem of identifying and defining such traits in terms of observable variables and determining that traits are important in a given behavioral context is the major problem in any theory of psychology. In studying any particular behavioral context, the psychologist will bring to bear the relevant theoretical formulation and will perhaps employ one or more models within that theory as a basis for designing experimental studies and for analyzing the data from these studies. It is this interplay of theory, model and data that adds to scientific knowledge.

One relationship between a latent trait value and stages can be described in terms of the probability of crossing a "boundary" between stages. This relationship is illustrated by a graph called the item (story) characteristic curve.

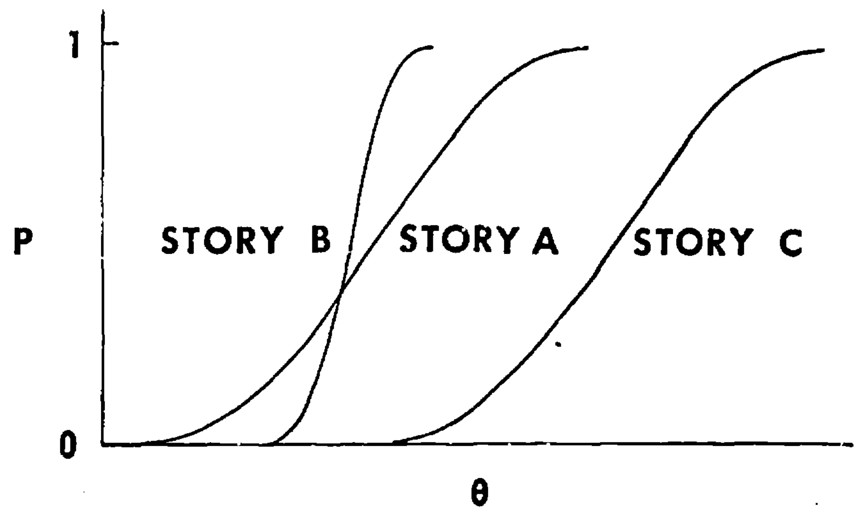


Fig. 1.--Characteristic Curves for Three Stories

In the above diagram, three story characteristic curves are represented. P is the probability of crossing the boundary between two adjacent stages and θ is the latent trait value of a randomly chosen subject.

We may compare the stories on two characteristics; the latent trait value necessary for a given probability to cross to the next highest stage and the precision in being able to discriminate people who have crossed from those who have not.

In the first case, we may say story C is more "difficult" than stories A or B. For, in the diagram, we can see that a subject must have a higher latent trait value to achieve the same probability of crossing into a higher stage than either story A or story B. However, with respect to the second characteristic, story B is more discriminating at this stage boundary than stories A and C. Below a certain value of Θ, for story B, the boundary has no probability of being crossed. Above that value the boundary is certainly crossed. In stories A and C, a wide range of Θ, the latent trait, results in finite probabilities of crossing and not crossing.

Thus, for the six stage concept of moral judgment there are five boundaries for each of the stories. These boundaries will be determined from the numbers of subjects found at each stage and from the assumption that the underlying latent trait has a normal distribution.

In addition, each story will have a discriminating power represented by its loading resulting from a factor analysis of the assumed unifactoral matrix of correlations between the stories. It is this matrix which deserves special interest, since the items, taken pairwise, are scored in unequal interval values and thus require the correlation between them to be of a polychoric nature.

To determine the value of the latent trait for a randomly chosen subject with a given pattern of stage scores, the story parameters and the pattern yield the probability of such a pattern occurring. Then the value of θ which maximizes the probability of the pattern occurring is the level of the latent trait for an individual whose responses result in that pattern.

To determine the probability of being in each stage once the level of the latent trait is known, it is necessary to refer to a diagram such as the one depicted below.

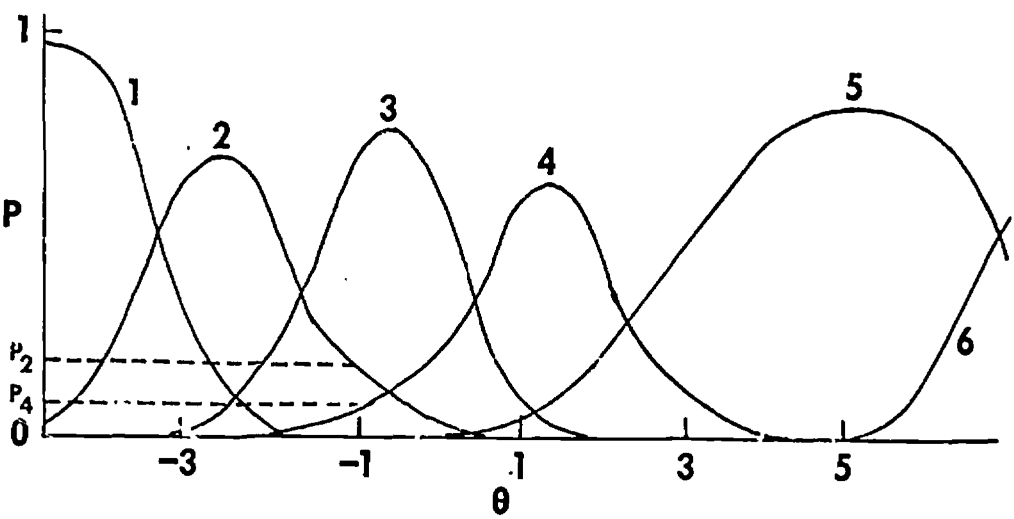


Fig. 2.--Probability of Being at Each Stage, Given a Moral Judgment Level

The graphs numbered 1 through 6 represent the probability (P) that a subject with latent trait value θ would answer with a "pure" pattern of stage scores for the tasks or situations, i.e., 1 corresponds to 1 1 1 1, 2 corresponds to 2 2 2 2, etc.

Now the occurrence of such "pure" patterns is extremely rare and subject's scores are bound to vary somewhat from task to task. We can see from the diagram that a subject with a latent trait value between -1.5 and +0.5 has a higher probability of being in stage 3 than any other stage. However, a subject with a value of -1.0 has a finite probability of being in stage 2 and stage 4 also, which can be read at the left of the diagram at probabilities P_2 and P_4 respectively.

In the next chapter, formulas are developed for the boundary levels, the discriminating powers of the stories, the polychoric correlation coefficient and the estimate of the latent trait. The non-mathematically oriented reader is urged to skim these pages and turn to Chapter IV. There, the formulas are applied to analyse the data at hand to estimate the stories' characteristics for different populations of subjects and to replicate some of the moral judgment studies described in the review of the literature.

CHAPTER III

PSYCHOMETRIC ESTIMATION OF A LATENT TRAIT

Review of Previous Research

Estimation of a latent trait using item parameters and an assumed underlying distribution has been discussed widely and more recently, with increasing frequency. However, with few exceptions, estimations are based on dichotomous scoring and the moral judgment scheme necessitates a polychotomous model.

Item Parameters

Among the more well known discussants of dichotomously scored item parameter estimation are Lawley (1943,1944), Tucker (1948), and Lord (1952), all of whom applied heuristic methods. A maximum likelihood solution for such item parameters is presented by Bock and Lieberman (1967). One polychotomous system is discussed by Bock (1968) who uses a multinomial model in analyzing items with categorical responses. Another was suggested by Michael and Perry (1955) who derive a theory of item analysis based on items scored at three levels of appropriateness of response using weights of 2c, c, and zero with c an arbitrary constant. Srivastava and Webster (1967) go beyond polychotomy and estimate a true score where items are scored on a continuous scale, but parameters are estimated for the distribution of true and observed scores, not items.

One parameter, item difficulty, is traditionally defined as the proportion of the population correctly answering the item. In the moral judgment items, where degree of "correctness" is considered, proportions of subjects responding to each level must be considered. For dichotomously scored items, improvements on the estimates of difficulty have been suggested, e.g., Feder (1947), who proposed averaging the proportions of the lower and upper 27 per cent score groups correctly answering a given item. Lord (1952), assuming an underlying normal distribution of the latent trait uses the normal deviate producing the proportion that correctly answer an item as the difficulty.

The other item parameter considered here, discriminating power, generally measures the relationship between correctly or incorrectly answering one item of a group and scoring high or low on the total group of items. In the present polychotomous case, item score - total score product moment correlations might be suggested, but this would not take into account the unequal distances between the levels in an item. In the dichotomous score case, two methods are frequently applied. Guilford (1954) and others use point biserial correlations, while Findley (1956) suggests the difference of the proportions of the upper and lower 27 per cent score groups correctly answering an item as the discriminating power. The method which lends itself to polychotomously scored items is most recently discussed in Lord and Novick (1968) who use the first loadings of the assumed unifactoral matrix of interitem correlations as estimates of discriminating power. Tucker (1948) and Samejima (1968) propose using the inverse of the loading to produce an invariant estimate.

The interitem correlation matrix is traditionally tetrachoric in nature when items are dichotomously scored. Estimates of these correlations can be calculated by either inverting a complicated polynomial or interpolating tables of the normal bivariate distribution. Froemel (1969) discusses a method of calculating tetrachoric correlations using an approximation to the normal bivariate distribution evaluated with Gaussian quadrature techniques. However, the items discussed in this paper have a number of hierchical levels and a form of polychoric correlation must be calculated. Richie-Scott (1919) describes a method of estimating a polychoric correlation by dividing the frequency surface into columns and rows at each point into four quadrants, and for each quadrant, a tetrachoric correlation is calculated. These approximate the value of the true correlation and their weighted mean is found where the weights are determined so that the probable error of the mean r calculated is minimal. Richie-Scott compares this method to that of computing an enneachoric (3 × 3) correlation coefficient by reducing a frequency table to a central block, four marginal and four corner block portions. The latter method involves calculating tetrachoric coefficients and results in solving a polynomial in r. Pearson and Pearson (1922) describe a lengthy arithmetic procedure also employing tetrachoric coefficients. Lancaster and Hamdan (1964) describe a method to produce a polychoric correlation that is more easily calculated but still involves solving a complicated set of simultaneous equations. Tallis (1962) discusses a maximum likelihood procedure for the case of 2 × 2 and 3 × 3 frequency tables requiring the assumption of normal bivariate distribution in the pair of variables considered. In addition, since category boundaries between the levels are to be estimated, each time one item would be paired with others, different boundaries would result.

Estimation of a Latent Trait

Measures of ability or latent trait have been estimated either simultaneously with item parameters or as a function of known parameters. Lord (1968) analyzed the Verbal Scholastic Aptitude Test using Birnbaum's three parameter model, employing a guessing correction. Lord assumed a unidimensional trait and dichotomous scoring and found that initial estimates must be good to avoid having ability estimates go to plus or minus infinity. However, divergence may still occur because of one item or one person, either or both of which can be removed. The process required the solving of N(examinees) + 2n(items) - 2 equations in as many unknowns. Samejima (1968) derives both Bayes modal estimates and mean value estimates of ability for an individual based on a function of difficulty and discriminating power for each item and the entire pattern of responses. This method is applied to both dichotomously scored items and those whose alternatives form a graded scale structure given that the necessary parameters have been determined.

Item Characteristic Curves and Distribution of the Latent Trait

Togerson (1958) shows that the proportion of subjects in the population who respond correctly to an item can be predicted using a function of two factors; the assumed distribution of subjects on the underlying variable and the characteristic curves of the items. The item characteristic curve or the graph relating the probability that the subject will correctly answer an item, given his ability, has been described mainly by two models, a logistic function and normal ogive. Among those who have assumed the normal ogive model include Guilford (1936), Lawley (1943,1944), Lord (1952) and Lord and Novick (1968). Logistic models have been employed by Maxwell (1959) and Lord and Novick (1968). Baker (1961) computed maximum likelihood estimates of item parameters of a scholastic aptitude test using both normal and logistic models. The goodness of fit of ogives specified by the pairs of item parameters to observed data was determined for each item. While negligible differences were found for the item difficulties, differences in item discriminating power indicated that interpretation of these indices requires separate frames of reference. Empirical results showed the logistic model to be a useful alternate to the normal model in item analyses. The underlying distribution of a moral judgment score will be empirically shown to approximate the normal curve centered at some stage for a particular sex, social class, culture or age.

Estimation of the Item (Story) Parameters and the Latent Trait (Moral Judgment)

Introduction

There are two types of parameters to be estimated for each item, a common discriminating power and a difficulty level for each alternative. The alternative or stage difficulties will provide some measure of "distance" between each of the hierarchical levels and discriminating power will indicate the relative strength of the item to predict the total score received on all items considered.

Discriminating Power - the Polychoric Correlation Coefficient

Since the graded scale structure in the alternative will not, in all probability represent categories which are equidistant between levels, any measure of interitem correlation, which will lead to the calculation of discriminating power must be of a polychoric nature. The method described in Tallis (1962) will be employed, extending the equations for the 2×2 and 3×3 cases to $m(j)$ categories in each item (j) considered.

If two variables U and V have a joint bivariate normal density, and have values of a graded scale nature where U has r levels and V, s levels, then two variables X and Y may be defined such that

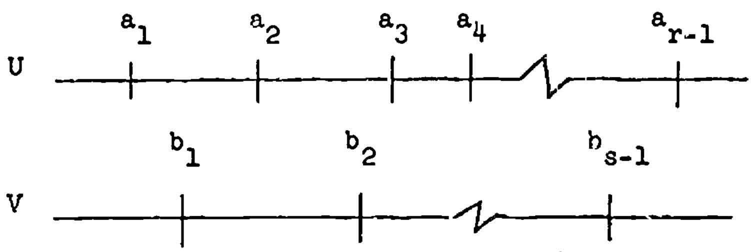


Fig. 3.--Two graded scale variables

$$Pr\ (X=x_1) = Pr\ (U < a_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_1} e^{-t^2/2} dt = \int_{-\infty}^{a_1} \varphi\ (u) = \Phi(a_1) = P_{1.}$$

$$Pr\ (X=x_2) = Pr\ (a_1 < U < a_2) = \Phi(a_2) - \Phi(a_1) = P_{2.}$$

down to

$$Pr\ (X=x_r) = Pr\ (U > a_{r-1}) = 1 - \Phi(a_{r-1}) = P_{r.}$$

Likewise,

$$Pr\ (Y=y_1) = Pr\ (V < b_1) = \Phi(b_1) = P_{.1}$$

$$Pr\ (Y=y_2) = Pr\ (b_1 < V < b_2) = \Phi(b_2) - \Phi(b_1) = P_{.2}$$

down to

$$Pr\ (Y=y_s) = Pr\ (V > y_{s-1}) = 1 - \Phi(b_{s-1}) = P_{.s}$$

It should be noted that x_i and y_j refer here to discrete classifications. The joint densities are of the following form:

$$P_{ij} = \Phi(a_i, b_j, \rho) = \Phi(a_i, b_j)$$

and

$$P_{11} = \Pr\,(X=x_1, Y=y_1) = \Pr\,(U < a_1, V < b_1) = \Phi(a_1, b_1)$$

$$P_{12} = \Pr\,(X=x_1, Y=y_2) = \Pr\,(U < a_1, b_1 < V < b_2) = \Phi(a_1, b_2) - \Phi(a_1, b_1)$$

down to

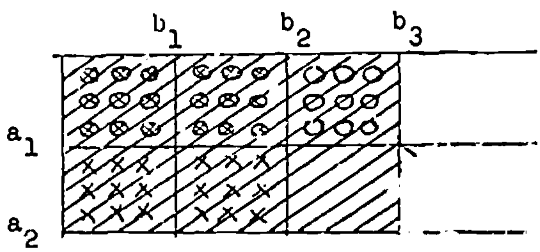
$$P_{15} = \Pr\,(X=x_1, Y=y_2) = \Pr\,(U < a_1, V > b_{s-1}) = \Phi(a_1) - \Phi(a_1, b_{s-1})$$

and for cells not along the edges of the table, e.g.

$$P_{23} = \Pr\,(X=x_2, Y=y_3) = \Pr\,(a_1 < U < a_2, b_2 < V < b_3)$$

$$= \Phi(a_2, b_3) - \Phi(a_1, b_3) - \Phi(a_2, b_2) + \Phi(a_1, b_2)$$

A general form for the probabilities can be derived from consideration of the following figure:



$$P_{23} = \Phi(a_2, b_3) - \Phi(a_1, b_3) - \Phi(a_2, b_2) + \Phi(a_1, b_2)$$

Fig. 4.--Diagram for Calculating P_{ij}

In general,

$$P_{ij} = \Phi(a_i, b_j) - \Phi(a_i, b_{j-1}) - \Phi(a_{i-1}, b_j) + \Phi(a_{i-1}, b_{j-1}).$$

If $j=s$

$$P_{is} = \Phi(a_i) - \Phi(a_i, b_{s-1}) - \Phi(a_{i-1}) + \Phi(a_{i-1}, b_{s-1})$$

if i=r

$$P_{rj} = \Phi(b_j) - \Phi(a_{r-1},b_j) - \Phi(b_{j-1}) + \Phi(a_{r-1},b_{j-1})$$

If i=r and j=s

$$P_{rs} = 1 - \Phi(a_{r-1}) - \Phi(b_{s-1}) + \Phi(a_{r-1},b_{s-1})$$

and if either or both subscripts of the arguments for any term are zero, the term vanishes.

We wish to estimate ρ using a random sample of size $n_{..}$ whose responses to two variables are classified along the two dimensions, X and Y.

A frequency table of the following nature might appear:

	y_1	y_2	y_3		y_s	
x_1	n_{11}	n_{12}	$n_{1'}$	...	n_{1s}	$n_{1.}$
x_2	n_{21}	n_{22}	n_{23}	...	n_{2s}	$n_{2.}$
x_3	n_{21}	n_{32}	n_{33}	...	n_{3s}	$n_{3.}$
	⋮	⋮	⋮		⋮	
x_r	n_{r1}	n_{r2}	n_{r3}	...	n_{rs}	$n_{r.}$
	$n_{.1}$	$n_{.2}$	$n_{.3}$		$n_{.s}$	$n_{..}$

Fig. 5.--Distribution of subjects for two graded scale variables

The likelihood, given such a cell distribution is

$$L = K \prod_{\substack{i=1 \\ j=1}}^{\substack{i=r \\ j=s}} P_{ij}^{n_{ij}}$$

where K is a constant, independent of the parameter to be estimated, ρ.

One necessary differentiation is that of $\frac{\partial\Phi(a,b)}{\partial\rho}$.

Tallis (1962) shows that this expression is merely equal to $\varphi(a,b)$ which can be written in the form

$$\varphi(a,b) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \frac{-1}{2(1-\rho^2)}[a^2-2\rho ab+b^2]$$

The derivatives of P_{ij} with respect to ρ all vanish, with the exception of those terms of the form $\Phi(a_i,b_j)$, the only function of ρ in the expressions.

Let

$$\ell = \log L = K' + \sum_{\substack{i=1 \\ j=1}}^{\substack{i=r \\ j=s}} n_{ij} \log P_{ij}$$

then

$$\frac{\partial\ell}{\partial\rho} = \sum_{\substack{i=1 \\ j=1}}^{\substack{i=r \\ j=s}} \frac{n_{ij}}{P_{ij}} \frac{\partial P_{ij}}{\partial\rho}$$

and the $\frac{\partial P_{ij}}{\partial\rho}$ are a linear combination of the form $k\varphi(a,b)$.

Now,

$$\frac{\partial^2 \ell}{\partial \rho^2} = \sum_{i,j} n_{ij}[(\frac{-1}{P_{ij}^2} \frac{\partial P_{ij}}{\partial \rho} \frac{\partial P_{ij}}{\partial \rho} + (\frac{1}{P_{ij}} \frac{\partial^2 P_{ij}}{\partial \rho^2})].$$

From Kendall and Stewart (1961)

$$\mathcal{E} \frac{\partial^2 \ell}{\partial u \partial v} = - \mathcal{E} \frac{\partial \ell}{\partial u} \frac{\partial \ell}{\partial v}$$

where, in the situation, $u=v=\rho$ so

$$- \mathcal{E} \frac{\partial^2 \ell}{\partial \rho^2} = - n_{..} \sum_{i,j} \mathcal{E}[\frac{-1}{P_{ij}} (\frac{\partial P_{ij}}{\partial \rho})^2 + (\frac{\partial P_{ij}}{\partial \rho})^2]$$

$$= n_{..} \sum_{i,j} [\frac{1}{P_{ij}}(\frac{\partial P_{ij}}{\partial \rho})^2 - (\frac{\partial P_{ij}}{\partial \rho})^2].$$

In the case of multiple cells, the second term in the brackets becomes much smaller than the first due to P_{ij} being small.

Thus, we can state

$$- \mathcal{E} \frac{\partial^2 \ell}{\partial \rho^2} = n_{..} \sum_{i,j} \frac{1}{P_{ij}} (\frac{\partial P_{ij}}{\partial \rho})^2$$

Parameter estimation by the use of the Newton-Raphson method of iteration reduces to the following for one parameter:

$$\rho^{(2)} = \rho^{(1)} + I^{-1} \delta^{(1)}$$

where $\rho^{(2)}$, $\rho^{(1)}$, I^{-1}, and $\delta^{(1)}$ are all scalar quantities.

and

$\rho^{(2)}$ is the new estimate of ρ

$\rho^{(1)}$ is the old estimate of ρ

I^{-1} is the negative of the inverse of the expected second derivative of the log likelihood function with respect to ρ at $\rho=\rho^{(1)}$

$\delta^{(1)}$ is the first derivative of the log likelihood function with respect to ρ at $\rho=\rho^{(1)}$.

Here,

$$I = -\mathcal{E}\left(\frac{\partial^2 \ell}{\partial \rho^2}\right) = n_{..} \sum_{i,j} P_{ij}^{-1}\left(\frac{\partial P_{ij}}{\partial \rho}\right)^2$$

and

$$\delta^{(1)} = \sum_{i,j} \frac{n_{ij}}{P_{ij}} \frac{\partial P_{ij}}{\partial \rho}\bigg|_{\rho=\rho^{(1)}}$$

Calculation of the volume function $\Phi(a,b)$ and the area function $\Phi(c)$ involve the technique of Gaussian Quadrature for the former and a series approximation for the latter.

The threshold values for the categories in U and V are estimated in the following manner:

$$a_1 = \Phi^{-1}\left(\frac{n_{1.}}{n_{..}}\right) \qquad b_1 = \Phi^{-1}\left(\frac{n_{.1}}{n_{..}}\right)$$

$$a_2 = \Phi^{-1}\left(\frac{n_{1.}+n_{2.}}{n_{..}}\right) \qquad b_2 = \Phi^{-1}\left(\frac{n_{.1}+n_{.2}}{n_{..}}\right)$$

$$\vdots \qquad\qquad \vdots$$

$$a_{r-1} = \Phi^{-1}\left(\frac{\sum_{i=1}^{r-1} n_{i.}}{n_{..}}\right) \qquad b_{s-1} = \Phi^{-1}\left(\frac{\sum_{j=1}^{s-1} n_{.j}}{n_{..}}\right)$$

If I is thought of as the information "matrix", then $\frac{I^{-1}}{n_{..}}$ is an estimate of $V(\hat{\rho})$ so that $\sqrt{\frac{I^{-1}}{n_{..}}}$ estimates the standard error of $\hat{\rho}$.

Tallis (1962) states that for an r×s table, whenever r-s-1 is larger than the number of parameters to be estimated, a chi square test of the assumptions of the underlying bivariate normal distribution has the form.

$$\chi^2_{rs-1-t} = \sum_{i,j} \frac{(n_{ij}-n_{..}\hat{P}_{ij})^2}{n_{..}\hat{P}_{ij}}$$

where t is the number of parameters to be estimated and $\hat{P}_{ij}$ is the maximum likelihood estimate of P_{ij}.

An unrestricted maximum likelihood factor analysis described in Jöreskog (1967), assuming a single factor, will be applied to the matrix of polychoric correlations, producing loadings for each item corresponding to an estimate of item discriminating power.

Item Difficulty

The difficulties ass- ated with each item will be determined by the proportion of subjects responding at each level. Here, the hierarchical structure necessitates the scoring of each level below the one chosen as also having been selected.

As the distribution of the latent trait in the population is assumed to be normal, the deviate producing that cumulative proportion of the normal curve will be the measure of each alternative's difficulty.

Latent Trait

Estimation of the latent trait, a moral judgment level, will be accomplished by the method of maximum likelihood estimation. Samejima (1968) discusses sufficient conditions for the existence of this estimator. The procedure requires the probability of a given subject's response pattern, which is a function of known item parameters, and the assumed distribution of the underlying trait.

P_{jk} = the probability that a randomly chosen person with ability level Θ answers at the k^{th} level of the j^{th} item.

$j = 1,2,\ldots,n;\ k = 1,2,\ldots,\max(j)$

$$P_{jk} = \Phi(g_{j,k}(\Theta)) - \Phi(g_{j,k-1}(\Theta))$$

where

$$g_{jk}(\Theta) = \frac{\gamma_{jk} - \alpha_j \Theta}{\sqrt{1-\alpha_j^2}}$$

and

Θ = ability level to be estimated

γ_{jk} = the difficulty of the k^{th} level of the j^{th} item - the standard normal deviate corresponding to the proportion of the population below the k^{th} level.

α_j = the discriminating power of the j^{th} item $(0 \le \alpha_j \le 1.00)$

For convenience of notation, let $\sigma_j = \sqrt{1-\alpha_j^2}$

Let L = the likelihood that a randomly chosen person with ability level Θ will answer in a given pattern

$$L = \prod_{j=1}^{n} P_{jk}$$

Let $\ell = \log L = \sum_{j=1}^{n} \log P_{jk}$.

Then

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{n} \frac{1}{P_{jk}} \frac{\partial P_{jk}}{\partial \theta} = \sum_{j=1}^{n} \frac{1}{P_{jk}} \frac{\partial P_{jk}}{\partial g_{jk}} \frac{\partial g_{jk}}{\partial \theta}$$

Now, if $1 < k < \max(j) - k$, a middle level of item j

$$\frac{\partial P_{jk}}{\partial g_{jk}} = \frac{\partial}{\partial g_{jk}} [\Phi(g_{jk}) - \Phi(g_{j,k-1})]$$

$$= \varphi(g_{jk}) - \varphi(g_{j,k-1}).$$

Also

$$\frac{\partial g_{jk}}{\partial \theta} = \frac{-\alpha_j}{\sigma_j}.$$

Therefore,

$$\frac{\partial P_{jk}}{\partial \theta} = -\frac{\alpha_j}{\sigma_j} [\varphi(g_{jk}) - \varphi(g_{j,k-1})]$$

and

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{n} - \frac{\alpha_j}{\sigma_j} \frac{\varphi(g_{jk}) - \varphi(g_{j,k-1})}{\Phi(g_{jk}) - \Phi(g_{j,k-1})}$$

If $k = 1$, the lowest level of item j,

$$P_{jk} = \Phi(g_{jk})$$

$$\frac{\partial P_{jk}}{\partial \theta} = -\frac{\alpha_j}{\sigma_j}\varphi(g_{jk})$$

and

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{n} -\frac{\alpha_j}{\sigma_j}\frac{\varphi(g_{jk})}{\Phi(g_{jk})}, \quad \text{for all } j.$$

If $k = \max(j)$, the highest level of item j,

$$P_{jk} = 1 - \Phi(g_{j,k-1})$$

$$\frac{\partial P_{jk}}{\partial \theta} = \frac{\alpha_j}{\sigma_j}\varphi(g_{j,k-1})$$

and

$$\frac{\partial \ell}{\partial \theta} = \sum_{j=1}^{n} \frac{\alpha_j}{\sigma_j}\frac{\varphi(g_{j,k-1})}{1-\Phi(g_{j,k-1})}, \quad \text{for all } j.$$

For a pattern including the lowest, highest and middle levels of items, the appropriate term, selected from those listed above, would be entered in the sum forming the first derivative of the log likelihood function.

The second derivative of the log likelihood function must also be computed.

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{n} \frac{\partial}{\partial \theta}\frac{1}{P_{jk}}\frac{\partial P_{jk}}{\partial \theta}$$

$$= \sum_{j=1}^{n} \frac{-1}{P_{jk}^2}\frac{\partial P_{jk}}{\partial \theta}\frac{\partial P_{jk}}{\partial \theta} + \frac{1}{P_{jk}}\frac{\partial^2 P_{jk}}{\partial \theta^2}$$

where

$$\frac{\partial^2 P_{jk}}{\partial\theta^2} = \frac{\partial}{\partial\theta}\frac{\partial P_{jk}}{\partial\theta} = \frac{\partial}{\partial\theta}\left[-\frac{\alpha_j}{\sigma_j}(\varphi(g_{jk}) - \varphi(g_{j,k-1}))\right]$$

$$= -\frac{\alpha_j}{\sigma_j}\left(\frac{\partial\varphi(g_{jk})}{\partial\theta} - \frac{\partial\varphi(g_{j,k-1})}{\partial\theta}\right).$$

Now,

$$\frac{\partial\varphi(g_{jk})}{\partial\theta} = \frac{\partial}{\partial\theta}\ \frac{1}{\sqrt{2\pi}}\ e^{\frac{-g_{jk}^2(\theta)}{2}}.$$

$$= \frac{1}{\sqrt{2\pi}}\ e^{\frac{-g_{jk}^2(\theta)}{}}\quad \frac{-2g_{jk}(\theta)}{2}\quad \frac{\partial g_{jk}}{\partial\theta}$$

$$= -\frac{\alpha_j}{\sigma_j}\ [\varphi(g_{jk}(\theta))][-g_{jk}(\theta)].$$

For notational convenience, let

$$\varphi(g_{jk}(\theta)) = \varphi(g_{jk})$$

and

$$g_{jk}(\theta) = g_{jk}.$$

Then

$$\frac{\partial^2 P_{jk}}{\partial\theta^2} = -\frac{\alpha_j}{\sigma_j}\ [(-g_{j,k})(-\frac{\alpha_j}{\sigma_j})\ \varphi(g_{jk}) + \frac{\alpha_j}{\sigma_j}\ (-g_{j,k-1})\ \varphi(g_{j,k-1}]$$

and

$$\frac{\partial^2 \ell}{\partial\theta^2} = \sum_{j=1}^{n}\left\{\frac{-1}{[\Phi(g_{jk})-\Phi(g_{j,k-1})]^2}\ [-\frac{\alpha_j}{\sigma_j}\ (\varphi(g_{j,k})-\varphi(g_{j,k-1}))]^2\right.$$

$$+ \frac{\alpha_j^2}{\sigma_j^2} \quad \frac{g_{j,k-1}\varphi(g_{j,k-1})-g_{jk}\varphi(g_{jk})}{\Phi(g_{jk})-\Phi(g_{j,k-1})} \Bigg\}$$

or

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{n} \frac{\alpha_j^2}{\sigma_j^2} - \frac{\varphi(g_{j,k})-\varphi(g_{j,k-1})}{\Phi(g_{j,k})-\Phi(g_{j,k-1})}^2 \frac{g_{j,k-1}\varphi(g_{j,k-1})-g_{jk}\varphi(g_{jk})}{\Phi(g_{jk}) - \Phi(g_{j,k-1})}$$

If $k = 1$, the lowest level of item j,

$$P_{jk} = \Phi(g_{jk})$$

$$\frac{\partial^2 P_{jk}}{\partial \theta^2} = \frac{\partial}{\partial \theta} \frac{\partial P_{jk}}{\partial \theta} = \frac{\partial}{\partial \theta} [- \frac{\alpha_j}{\sigma_j} \varphi(g_{jk})] = \frac{\alpha_j^2}{\sigma_j^2} g_{jk} \varphi(g_{jk}) .$$

Thus, for an answer pattern where all items are answered at the lowest level,

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{n} [\frac{-1}{\Phi^2(g_{jk})} \frac{\alpha_j^2}{\sigma_j^2} \varphi^2(g_{jk}) + \frac{-1}{\Phi(g_{jk})} (\frac{\alpha_j^2}{\sigma_j^2} g_{jk} \varphi(g_{jk}))]$$

$$= - \sum_{j=1}^{n} [\frac{\alpha_j^2}{\sigma_j^2} \frac{\varphi(g_{jk})}{\Phi(g_{jk})} (\frac{\varphi(g_{jk})}{\Phi(g_{jk})} + g_{jk})]$$

If $k = \max(j)$, the highest level of item j,

$$P_{jk} = 1 - \Phi(g_{j,k-1})$$

$$\frac{\partial^2 P_{jk}}{\partial \theta^2} = \frac{\partial}{\partial \theta} \frac{\partial P_{jk}}{\partial \theta} = \frac{\partial}{\partial \theta} [\frac{\alpha_j}{\sigma_j} \varphi(g_{jk})] = \frac{\alpha_j^2}{\sigma_j^2} g_{j,k-1} \varphi(g_{j,k-1})$$

Thus, for an answer pattern where all items are answered at the highest level,

$$\frac{\partial^2 \ell}{\partial \theta^2} = \sum_{j=1}^{n} \left[\frac{-1}{[1-\Phi(g_{j,k-1})]^2} \frac{\alpha_j^{\,2}}{\sigma_j^{\,2}} \varphi^2(g_{j,k-1}) + \frac{1}{[1-\Phi(g_{j,k-1})]^2} \frac{\alpha_j^{\,2}}{\sigma_j^{\,2}} g_{j,k-1}\varphi(g_{j,k-1}) \right]$$

$$= \sum_{j=1}^{n} \left[\frac{\alpha_j^{\,2}}{\sigma_j^{\,2}} \frac{\varphi(g_{j,k-1})}{[1-\Phi(g_{j,k-1})]} \left(\frac{-\varphi(g_{j,k-1})}{[1-\Phi(g_{j,k-1})]} + g_{j,k-1} \right) \right]$$

Here again, for a pattern including the lowest, highest and middle levels of items, the appropriate term, selected from those listed above would be entered in the sum forming the second derivative of the log likelihood function.

Employing the Newton-Raphson technique as in the polychoric correlation derivation described previously, the following iteration procedure is

$$\theta^{(2)} = \theta^{(1)} - \left. \frac{\frac{\partial \ell}{\partial \theta}}{\frac{\partial^2 \ell}{\partial \theta^2}} \right|_{\theta=\theta^{(1)}}$$

where

$\theta^{(2)}$ = the new estimate of ability level

$\theta^{(1)}$ = the old estimate of ability level

and the first and second derivatives of the log likelihood function are evaluated at the old estimate of ability level.

The standard error of this estimate of ability level is given by

$$S.E.: = \frac{1}{\sqrt{\left|\frac{\partial^2 \ell}{\partial \theta^2}\right|_{\theta=\theta^{(1)}}}}$$

Application to Moral Judgment

As an illustrative example of the previous derivation, consider the Great Britain sample of 14 year olds. Each subject was presented with five stories and was "graded" on each with a "score" of 1 to 6 corresponding to the level of moral judgment attained as evidenced by his responses to probing questions.

In calculating the polychoric correlations among the stories, the category boundaries between the six levels were also computed yielding five "difficulty" levels for each story. When the matrix of interitem correlations was subjected to unrestricted maximum likelihood factor analysis, the loadings for each story on the resulting single factor serve as discriminating powers of the stories.

Various patterns of responses are listed below along with the maximum likelihood estimate of the latent trait or moral judgment level.

TABLE 1

STORY SCORE PATTERNS AND
THEIR CORRESPONDING MORAL JUDGMENT LEVELS

Answer Pattern I II III VII VIII	Moral Judgment Level
1 1 1 2 2	-3.2698
2 2 2 2 2	-1.8676
2 2 3 3 3	-0.5737
2 3 5 4 2	0.1261
3 3 3 3 3	0.1399
3 3 3 4 4	1.1360
4 4 4 4 4	1.8734
4 4 4 5 5	2.6270
5 5 5 5 5	3.2619

The moral judgment level is distributed as $N(0,1)$ and is far more sensitive to story differences than the global scoring method. This is immediately apparent when patterns that would result in the same global score reveal a range of moral judgment levels when differences in category boundaries and discriminating powers are taken into account.

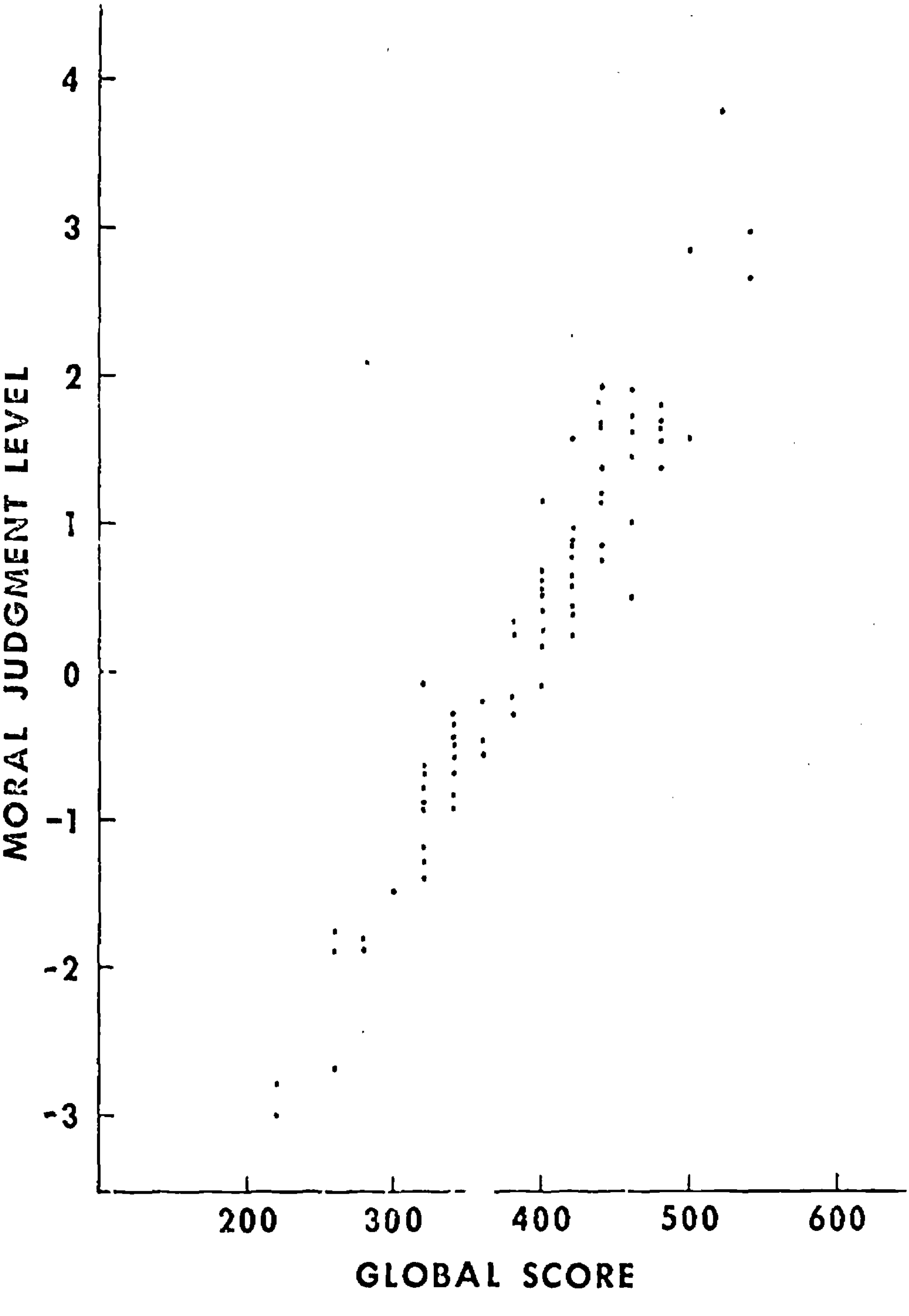
TABLE 2

EXAMPLES OF THE SAME GLOBAL SCORE PRODUCING VARIOUS MORAL JUDGMENT LEVELS

Global Score	Pattern					M.J. Level
	I	II	III	VII	VIII	
	3	2	2	3	2	-1.2103
240	2	3	2	2	3	-0.8742
	2	2	2	3	3	-0.8694
	3	2	3	2	3	-0.6950
260	2	2	3	3	3	-0.5737
	3	2	2	3	3	-0.5616
280	2	3	3	3	3	-0.1612
	3	3	2	3	3	-0.1468

Figure 6 illustrates the relationship between the global score and the moral judgment level for a number of answer patterns. In one instance, a global score of 320 represents a range of almost 1.5 in moral judgment level. In a later section of this paper, we will see that this may make a difference of two stages.

Fig. 6.--Moral Judgment Level
vs
Global Score for U.S.
19 year olds


4
3
2
1
0
-1
-2
-3
MORAL JUDGMENT LEVEL
200
300
400
500
600
GLOBAL SCORE

CHAPTER IV

DATA ANALYSIS AND RESULTS

The data presented here is the form of story parameters. Each story has five category boundaries which separate the six stages in the moral judgment scheme and one discriminating power, representing an estimate of the correlation of a story score and the moral judgment level.

These story parameters have been calculated for four major subgroups. The first is culture with populations of subjects from the United States and Great Britain. The second is sex, with both males and females included. Age is the third subgroup and includes 11, 13, 14 and 19 year olds. The 11 and 14 year old groups are British subjects and the 13 and 19 year olds are American. The last subgroup is social class with the British subjects encompassing both lower and middle classes while the American subjects are all rated as middle class.

In addition, the British subpopulation responded to seven stories, the 19 year old U.S. group responded to five, (a subset of the British group's seven), and the 13 year old U.S. group responded to all nine. The stories presented to each group are listed below.

TABLE 3

STORIES PRESENTED TO VARIOUS SUBGROUPS

G.B. 11 and 14 year olds	I,II,III,IV,VII,VIII,IX
U.S. 19 year olds	I,III,IV,VII,VIII
U.S. 13 year olds	I,II,III,IV,V,VI,VII,VIII,IX

Category Boundaries

The category boundaries for the stories which were presented to all four age groups are given in two forms. The first form is a series of tables with the boundaries between the six stages for each subgroup, e.g., British 11 year old middle class.

The second form is five graphs, one for each story, given to all age groups, illustrating the differences in stage or category boundaries among the four age groups. The middle class subjects are used throughout to eliminate socio-economic differences among the different age populations.

The culture-age confounding is evidenced by the appearance of more U.S. 13 year olds at higher stages (producing lower category boundaries) than British 14 year olds.

TABLE 4

CATEGORY BOUNDARIES BY AGE AND CULTURE

		GB		US	
		11	14	13	19
I	1	- .61730	-1.26890	-1.26708	-2.01341
	2	.38033	.07248	- .36120	- .88983
	3	1.12814	.82585	- .22689	- .38122
	4	1.68188	1.07039	.91732	.57923
	5	5.41998	2.61630	1.94911	1.80322
II	1	- .76471	-1.57148	- .91732	
	2	.18676	.11214	.22689	
	3	1.17314	1.17231	.82344	
	4	1.84653	1.43325	1.63255	
	5	2.35508	2.10016	2.23161	
III	1	- .87533	-1.46757	- .72979	-2.04038
	2	.04082	- .26487	- .04323	-1.31647
	3	1.04143	.65021	.59518	- .34876
	4	1.84446	1.15305	2.38152	.35151
	5	2.60074	1.65572	5.47119	.86247
IV	1	-1.06464	-1.15305	- .98784	-1.92910
	2	.57570	- .21898	- .06334	- .56465
	3	1.27107	.70630	.73122	.07776
	4	2.60233	.96444	1.74602	.77555
	5	2.60233	.98222	2.32257	1.52164
V	1			-2.05786	
	2			- .65130	
	3			- .18722	
	4			1.34626	
	5			5.41998	

TABLE 4-Continued

		GB		US	
		11	14	13	19
	1			-2.05786	
	2			- .65130	
VI	3			- .18722	
	4			1.34626	
	5			5.41998	
	1	- .44350	- .90693	- .64723	-1.88905
	2	- .11631	- .163651	- .16730	- .96113
VII	3	1.00517	.67799	.80715	- .04350
	4	1.68188	1.22064	1.65511	.61845
	5	2.60233	1.93221	2.3394	1.07157
	1	- .40540	-1.22064	- .59882	-2.34735
	2	- .06968	- .60946	- .21334	-1.31958
VIII	3	1.00517	.55663	.73098	- .28009
	4	1.68188	1.09036	1.72511	.16398
	5	5.37995	2.21636	2.19492	1.37844
	1	-1.73166	-1.40507	-1.19980	
	2	- .10463	- .25335	- .62439	
IX	3	.19859	.08365	- .15389	
	4	1.6464	1.05084	1.64138	
	5	2.60233	5.41998	5.41998	

TABLE 5

CATEGORY BOUNDARIES
BY AGE AND SEX

		GB 11		GB 14		US 19	
		Male	Female	Male	Female	Male	Female
I	1	- .76802	- .46661	-1.2815	-1.25212	-2.37103	-1.82681
	2	.30409	.46661	.02279	.13231	- .96890	- .82326
	3	.96154	1.35708	.66025	1.04020	- .40244	- .36230
	4	1.53863	1.89379	.90846	1.25212	.57155	.58616
	5	5.66270	5.47119	2.36189	5.37995	1.75172	1.85396
II	1	- .70972	- .82783	-1.52495	-1.61986		
	2	.28093	.08529	.09128	.13231		
	3	1.29683	1.05536	.97962	1.41219		
	4	1.93413	1.76437	1.33518	1.54310		
	5	2.37185	2.33742	2.09284	2.10734		
III	1	- .96154	- .78685	-5.19934	-1.07874	-2.11000	-1.98494
	2	.52212	.02458	.02458	- .30974	- .22168	-1.20579
	3	.80531	1.29281	.63946	.66075	- .25109	- .43975
	4	1.80678	1.88951	.94955	1.41219	.43673	.27681
	5	5.66270	2.33377	1.46053	1.93793	.94585	.79212
IV	1	- .89333	-1.29833	-1.34020	-1.00315	-1.94076	-1.91880
	2	.57340	.57823	.21620	.22168	- .57974	- .55115
	3	1.20013	1.35708	.66742	.74522	.12869	.03202
	4	2.37185	5.47119	1.79866	1.81078	.83232	.72642
	5	2.37185	5.47119	2.09652	2.10734	1.58358	1.47040
VII	1	- .44701	- .43965	- .78504	-1.04020	-2.01081	-1.79912
	2	- .16714	- .06088	- .53219	- .74522	- .97783	- .94629
	3	.89333	1.14447	.43073	.96742	.04726	- .12566
	4	1.70339	1.65914	1.06188	1.41219	.71517	.53594
	5	5.66270	2.33742	1.69492	2.37511	1.12252	1.02789
VIII	1	- .47166	- .33468	- .98561	-1.54310	-2.28739	-2.40964
	2	- .09999	- .03651	- .48084	- .74522	-1.40627	-1.24924
	3	.99737	1.01380	.38167	.74522	- .24417	- .31272
	4	1.93413	1.49115	.84807	1.41219	.22703	.10778
	5	5.54259	5.41998	1.92640	5.34703	1.43680	1.32965

TABLE 5-Continued

	GB 11		GB 14		US 19	
	Male	Female	Male	Female	Male	Female
1	-1.70339	-1.76437	-1.46053	-1.35493		
2	- .12231	- .08528	- .43072	- .08807		
3	.07772	.33468	- .05649	.22168		
4	1.20013	.93553	.88089	1.25212		
5	5.66270	2.33742	5.41998	5.37995		

TABLE 6

CATEGORY BOUNDARIES BY CLASS

		Middle	Lower
	1	- .69741	-1.05343
	2	.22650	.22244
I	3	1.08605	.90203
	4	1.57348	1.17574
	5	2.52528	5.37995
	1	-1.48054	- .90203
	2	.02174	.23208
II	3	1.08605	1.23393
	4	1.48054	1.69599
	5	2.27142	2.17058
	1	- .94077	-1.27518
	2	- .19687	- .06106
III	3	.64299	.95995
	4	1.25883	1.52467
	5	1.68243	2.17058
	1	-1.16808	-1.07231
	2	.21166	.50731
IV	3	.85407	1.00804
	4	1.89719	2.17206
	5	2.27142	2.28372
	1	-1.03519	- .46520
	2	- .59041	- .23600
VII	3	.87513	.79713
	4	1.36133	1.44217
	5	1.99317	2.28372
	1	- .89658	- .67449
	2	- .37831	- .30391
VIII	3	.69741	.79713
	4	1.36133	1.29875
	5	2.27142	2.67546
	1	-1.52537	-1.55728
	2	- .37831	- .05615
IX	3	- .03623	.25528
	4	.91845	1.15946
	5	2.52528	5.47119

TABLE 7

CATEGORY BOUNDARIES
BY SEX AND CLASS

		Male		Female	
		Middle	Lower	Middle	Lower
I	1	- .81577	-1.10421	- .59858	- .99969
	2	.12258	.18776	.32226	.26157
	3	.85918	.77267	1.35372	1.06757
	4	1.36986	1.07207	1.83887	1.30917
	5	2.25092	5.47119	5.41998	5.47119
II	1	-1.54664	- .82152	-1.42608	- .99969
	2	- .09183	.35391	.12427	.09963
	3	.95128	1.24600	1.22711	1.22064
	4	1.36986	1.72180	1.59869	1.66839
	5	1.97050	2.45256	5.47119	1.98075
III	1	-1.29557	-1.42233	- .70066	-1.13590
	2	- .24704	- .05298	- .15208	- .07024
	3	.47590	.93050	.81062	.99446
	4	1.05225	1.42233	1.50678	1.66456
	5	1.54664	2.03101	1.83887	2.40891
IV	1	-1.16529	-1.04554	-1.17060	-1.10343
	2	.09183	.57783	.32226	.43073
	3	.65542	1.07653	1.06757	.93613
	4	1.79176	2.19492	2.01451	2.14759
	5	2.25092	2.19492	2.29075	2.41182
VII	1	-1.22783	- .34194	- .88970	- .61329
	2	- .58139	- .21344	- .59858	- .26157
	3	.69380	.62007	1.06757	1.03304
	4	1.16529	1.42233	1.59869	1.46523
	5	1.65679	2.45510	5.47119	2.14759
VIII	1	- .95128	- .57783	- .84950	- .79164
	2	- .40847	- .21344	- .35142	- .40900
	3	.47590	.77791	.93140	.81915
	4	1.16529	1.25009	1.59869	1.35756
	5	1.97050	2.45510	5.41998	5.41998

TABLE 7-Continued

		Male		Female	
		Middle	Lower	Middle	Lower
	1	-1.54664	-1.58628	-1.50678	-1.52610
	2	- .73324	- .03531	- .09656	- .07966
IX	3	- .24704	.15955	.15208	.36611
	4	.81577	1.17640	1.02008	1.14076
	5	5.47119	5.41998	2.29075	5.47119

TABLE 8

CATEGORY BOUNDARIES
BY AGE AND CLASS

		11		14	
		Middle	Low	Middle	Low
I	1	- .38846	- .78788	-1.08966	-1.40311
	2	.35722	.39573	.10101	.06408
	3	1.13634	1.12276	1.03890	.72712
	4	1.81342	1.60865	1.40199	.90595
	5	5.54259	5.41998	2.27363	5.41998
II	1	-1.13634	- .56918	-2.27363	-1.35574
	2	- .11685	.39573	.15913	.08243
	3	1.19379	1.15974	.99069	1.31123
	4	1.99072	1.76882	1.20065	1.63433
	5	5.47119	2.16004	1.99560	2.13081
III	1	- .76906	- .95202	-1.14340	-1.76919
	2	- .05833	.10708	- .33764	- .21972
	3	.89256	1.15507	.43073	.80592
	4	1.67966	1.99072	.99069	1.27334
	5	5.66270	2.42039	1.32882	2.01909
IV	1	-1.13634	-1.02008	-1.20065	-1.12434
	2	.42009	.68664	.01441	.35227
	3	1.08241	1.42608	.66547	.73266
	4	2.26921	5.47995	1.68618	1.89592
	5	2.26921	5.37995	2.27363	2.01909
VII	1	- .93685	- .17442	-1.14340	- .78103
	2	.35772	.03857	- .85816	- .51194
	3	1.13634	.92715	.66547	.68594
	4	1.99072	1.54199	1.3890	1.35974
	5	5.47119	2.42319	1.68518	2.18368
VIII	1	- .51772	- .33389	-1.48354	-1.09080
	2	- .02915	- .09656	- .77764	- .51194
	3	1.08241	.95721	.39932	.66313
	4	1.99072	1.54199	1.3890	1.12434
	5	5.54259	5.41998	1.99560	2.44481

TABLE 8-Continued

		11		14	
		Middle	Low	Middle	Low
IX	1	-1.47753	-199398	-1.57636	-1.31531
	2	- .32632	.03857	- .43073	- .14583
	3	- .05833	.37496	- .01441	.14583
	4	.84996	1.23889	.99069	1.09080
	5	2.26921	5.41998	5.66270	5.41998

Fig. 7.--Story I Category Boundaries for Middle Class Subjects

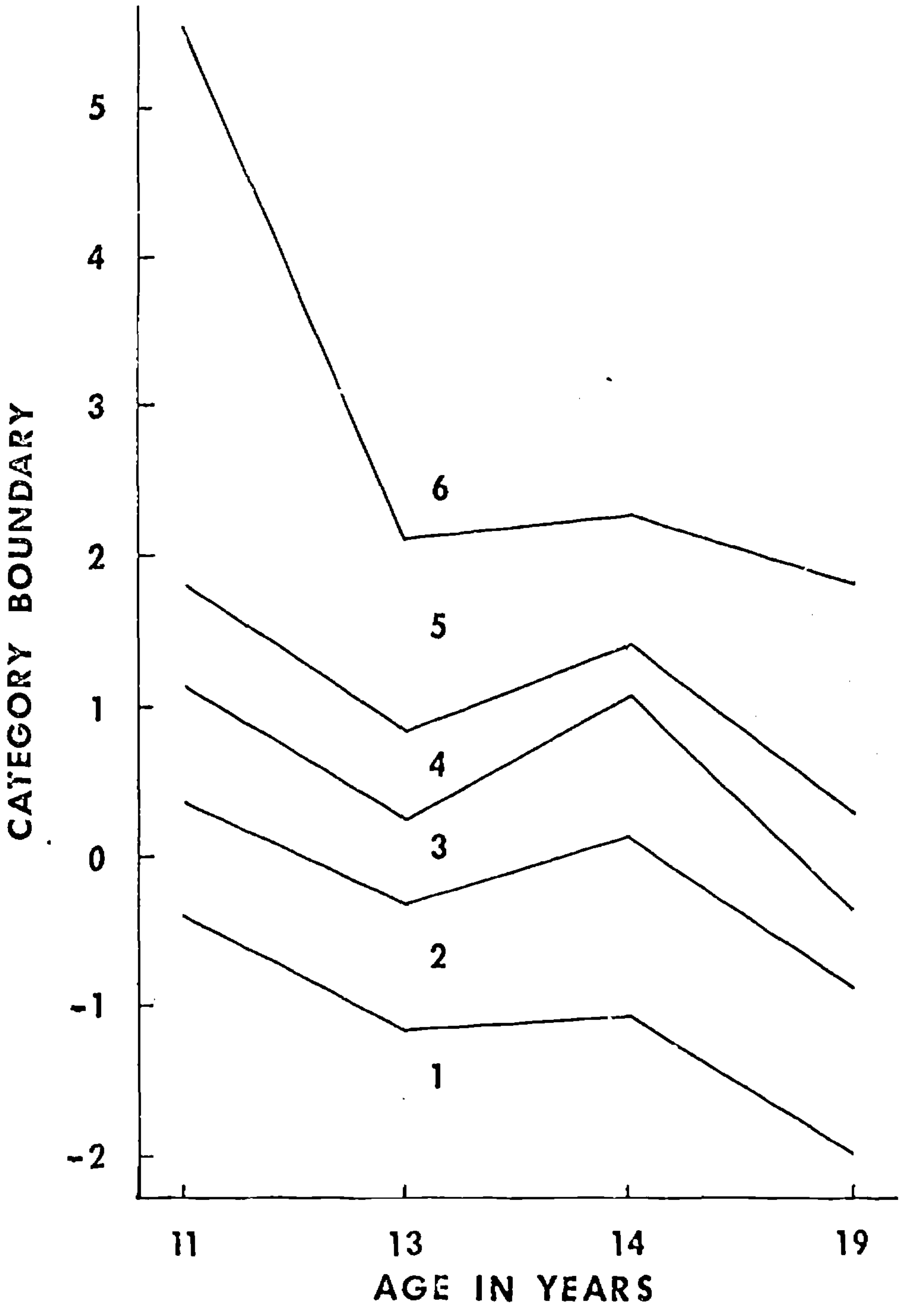

CATEGORY BOUNDARY
5
4
3
2
1
0
-1
-2
6
5
4
3
2
1
11
13
14
19
AGE IN YEARS

Fig. 8.--Story III Category Boundaries for Middle Class Subjects

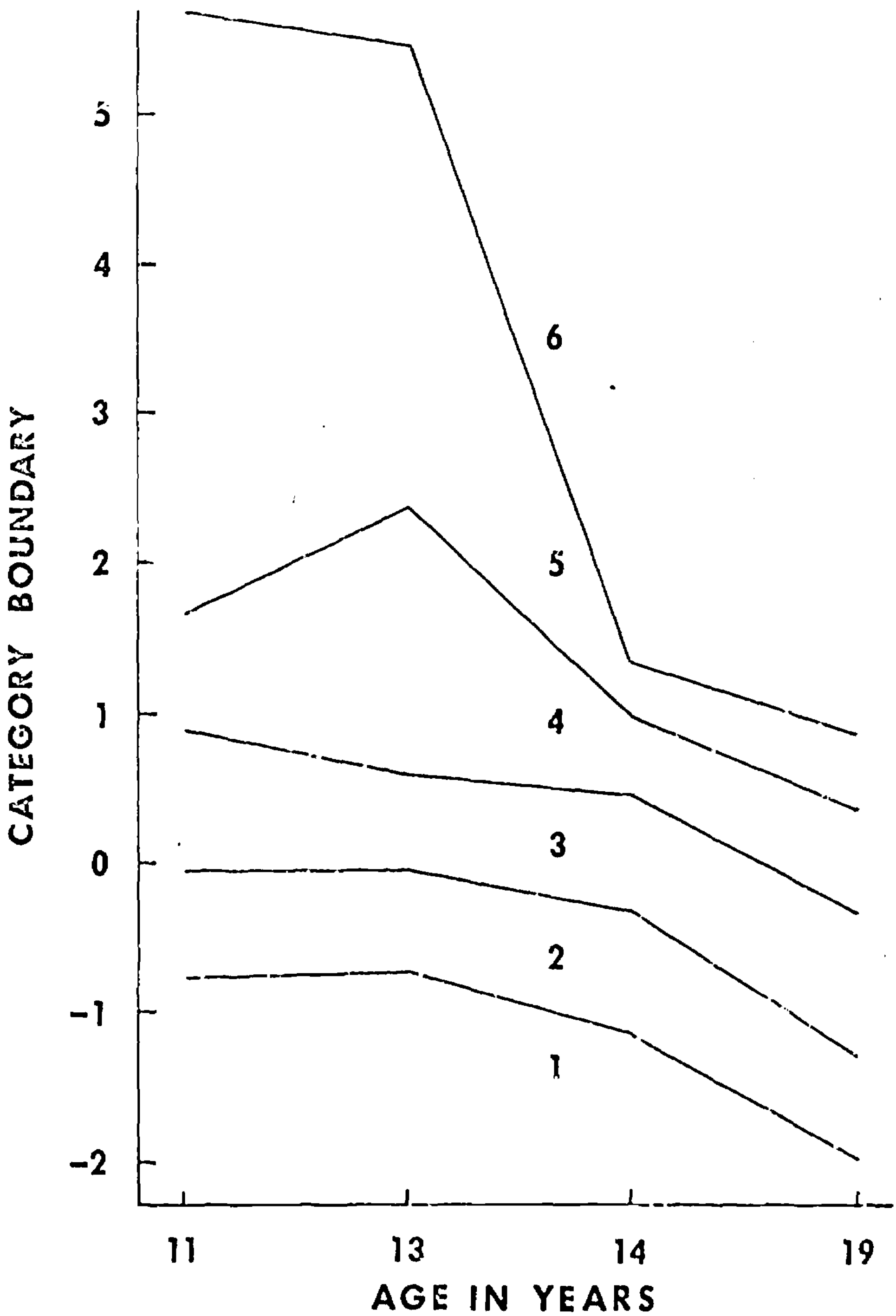
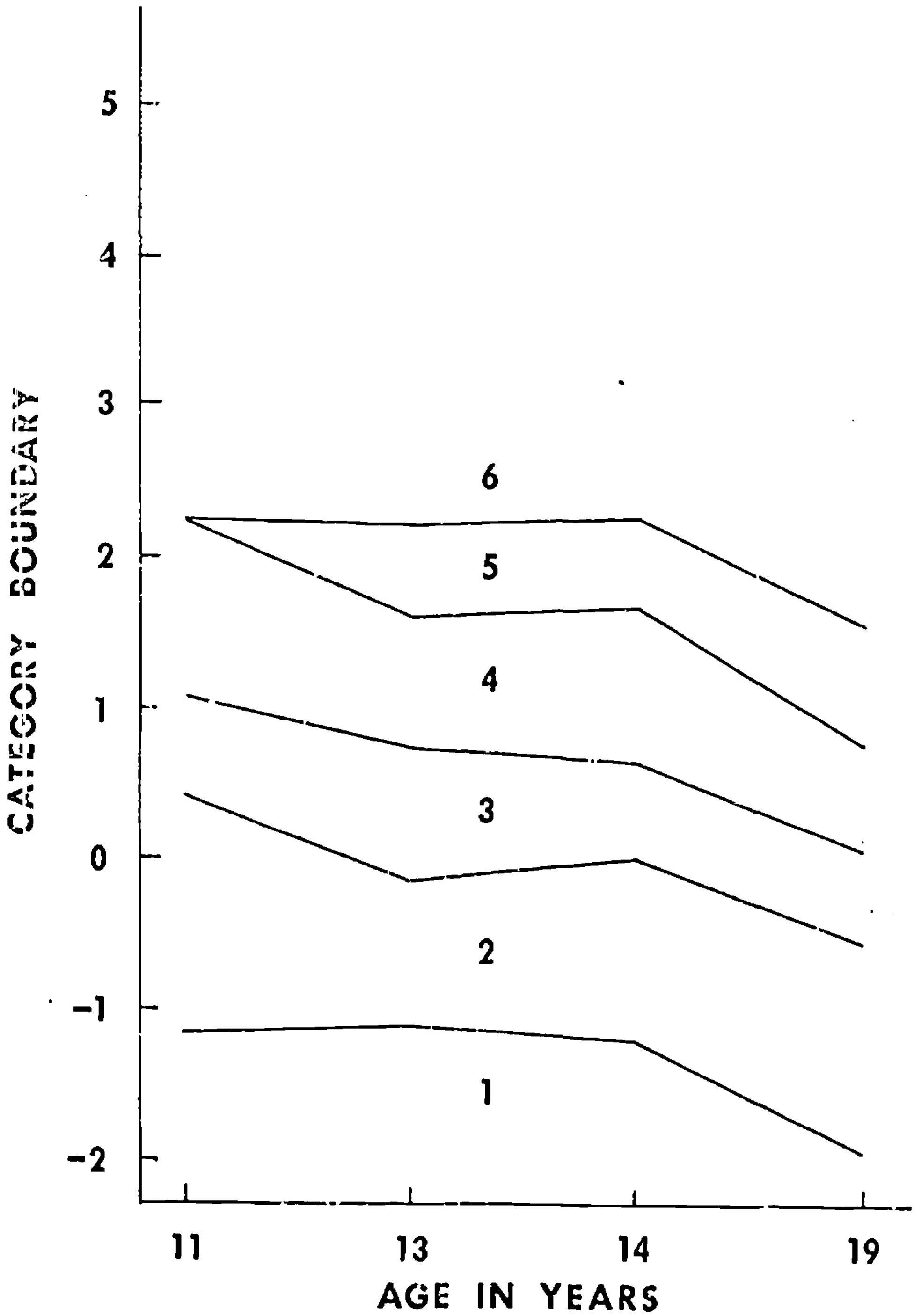

CATEGORY BOUNDARY
5
4
3
2
1
0
-1
-2
6
5
4
3
2
1
11
13
14
19
AGE IN YEARS

Fig. 9.--Story IV Category Boundaries for Middle Class Subjects



CATEGORY BOUNDARY
5
4
3
2
1
0
-1
-2
6
5
4
3
2
1
11
13
14
19
AGE IN YEARS

Fig. 10.--Story VII Category Boundaries for Middle Class Subjects

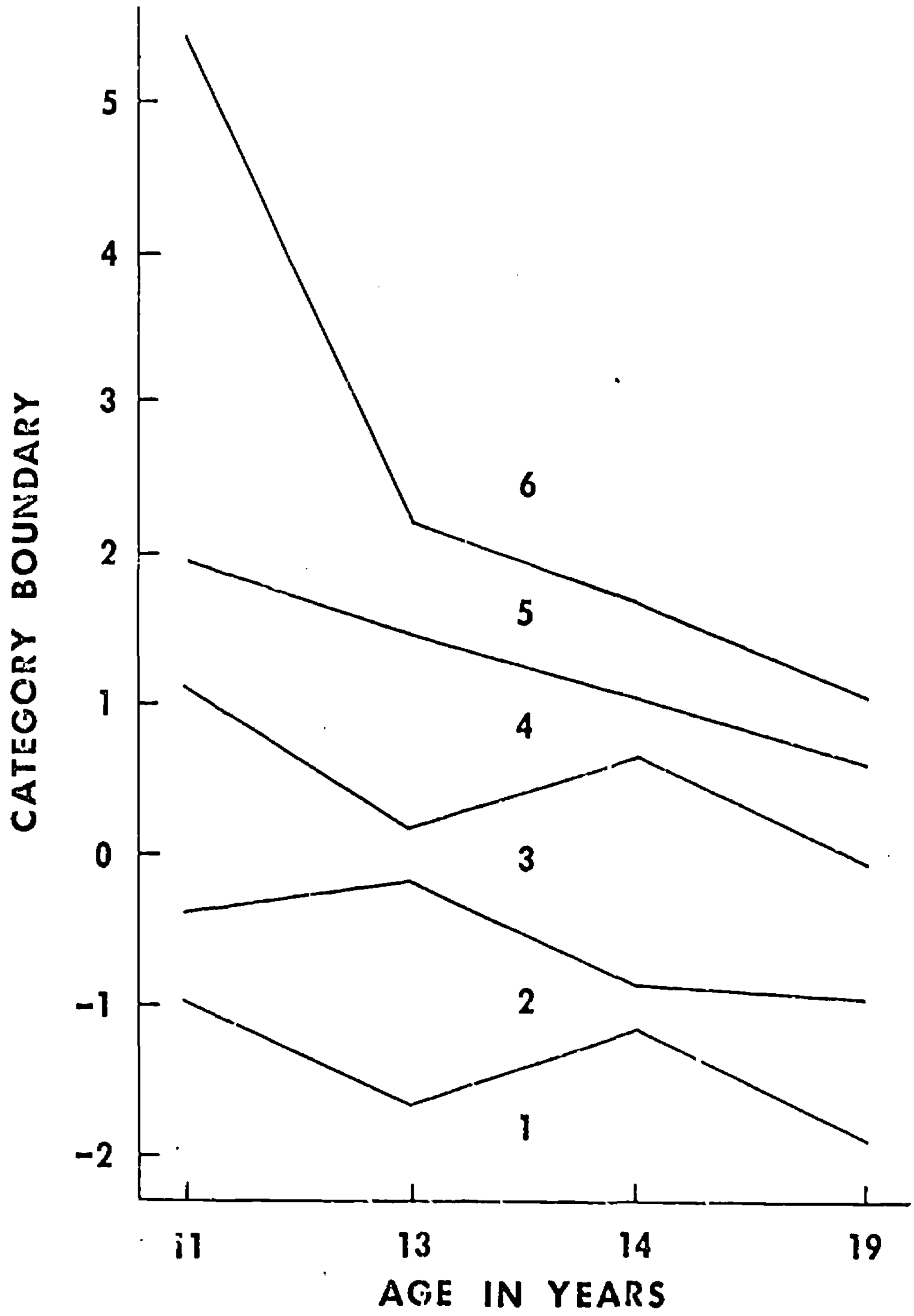
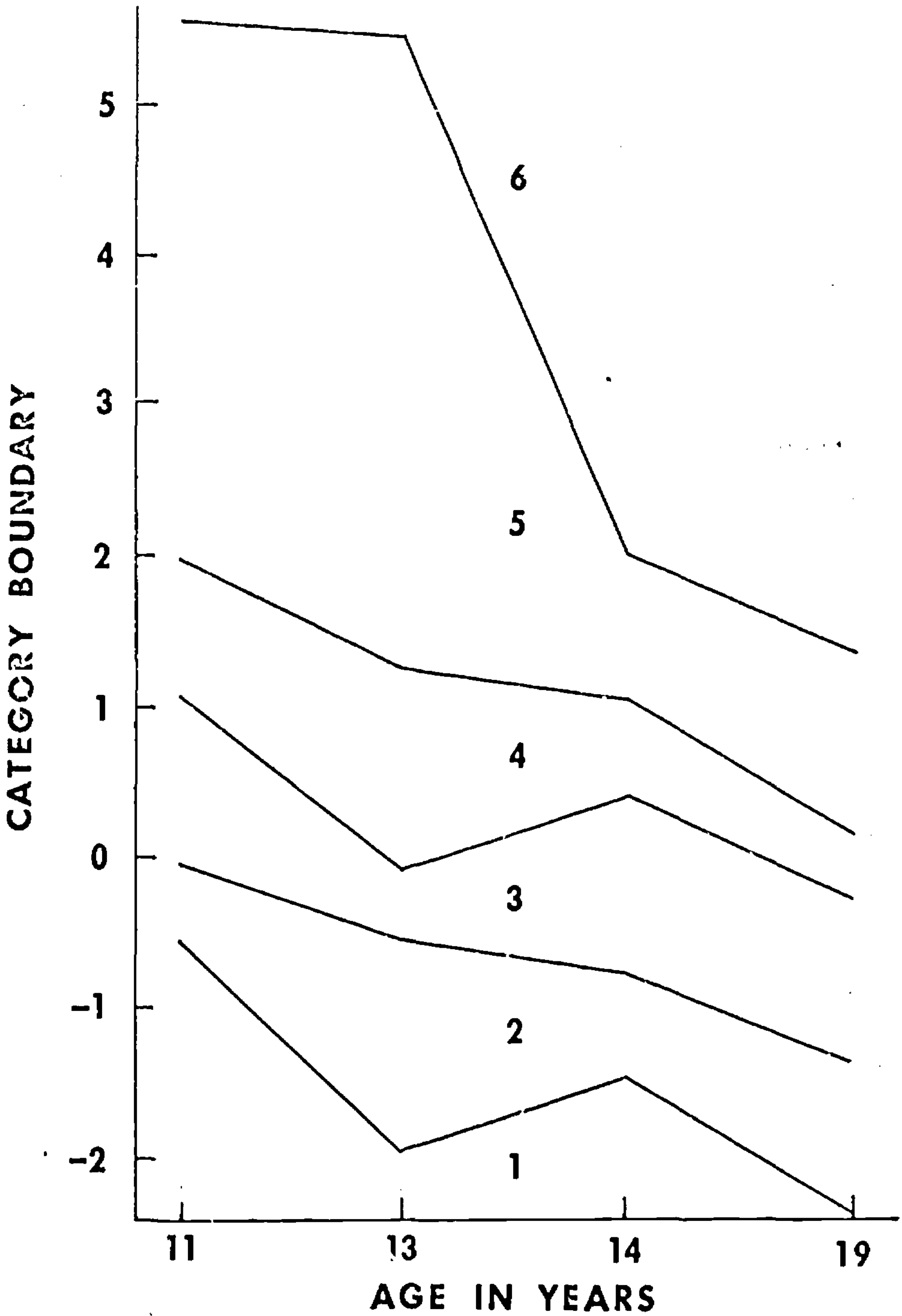

CATEGORY BOUNDARY
5
4
3
2
1
0
-1
-2
6
5
4
3
2
1
11
13
14
19
AGE IN YEARS

Fig. 11.--Story VIII Category Boundaries for Middle Class Subjects



CATEGORY BOUNDARY
5
4
3
2
1
0
-1
-2
6
5
4
3
2
1
11
13
14
19
AGE IN YEARS

Comparison of Groups (Category Boundaries)

As stated before, category boundaries are determined by the cumulative proportion of subjects answering (scored) at a particular stage or below. The boundary value is the normal deviate corresponding to that proportion.

For example, in the following situation, children's responses to a story are scored at stages one through six with the following result.

TABLE 9

CUMULATIVE NUMBERS OF SUBJECTS OF EACH STAGE

Stage	Number at that stage	Cumulative Proportion	Normal Deviate
1	18	0.16216	-0.98561
2	17	0.31532	-0.48084
3	37	0.64865	0.38167
4	17	0.80180	0.84807
5	19	0.97297	1.92640
6	3	1.00000	

Two groups may be compared by studying the relative positions of their stage boundaries.

TABLE 10

COMPARISON OF CATEGORY BOUNDARIES
BETWEEN TWO GROUPS

	Boundary	Stg.	11 yrs	14 yrs
	1	1	-0.87533	-1.46757
	2	2	0.04082	-0.26487
Story III	3	3	1.04143	0.65021
	4		1.84446	1.15305
	5		2.60074	1.65572

A few of the contrasts considered have been represented in graphical form. In this manner, the category boundaries across stories may be compared between two subgroups.

For example, in the British 14 yr vs. 11 yr graph, the 11 year olds' category boundaries are displaced about the value of one stage upward, indicating more children at lower levels than the 14 year old group.

In the U.S. 19 year old group, the boundaries for the male and female subjects are almost congruent for the five stories administered.

Fig. 12.--Comparison of Category Boundaries Between 11 and 14 year old British Subjects

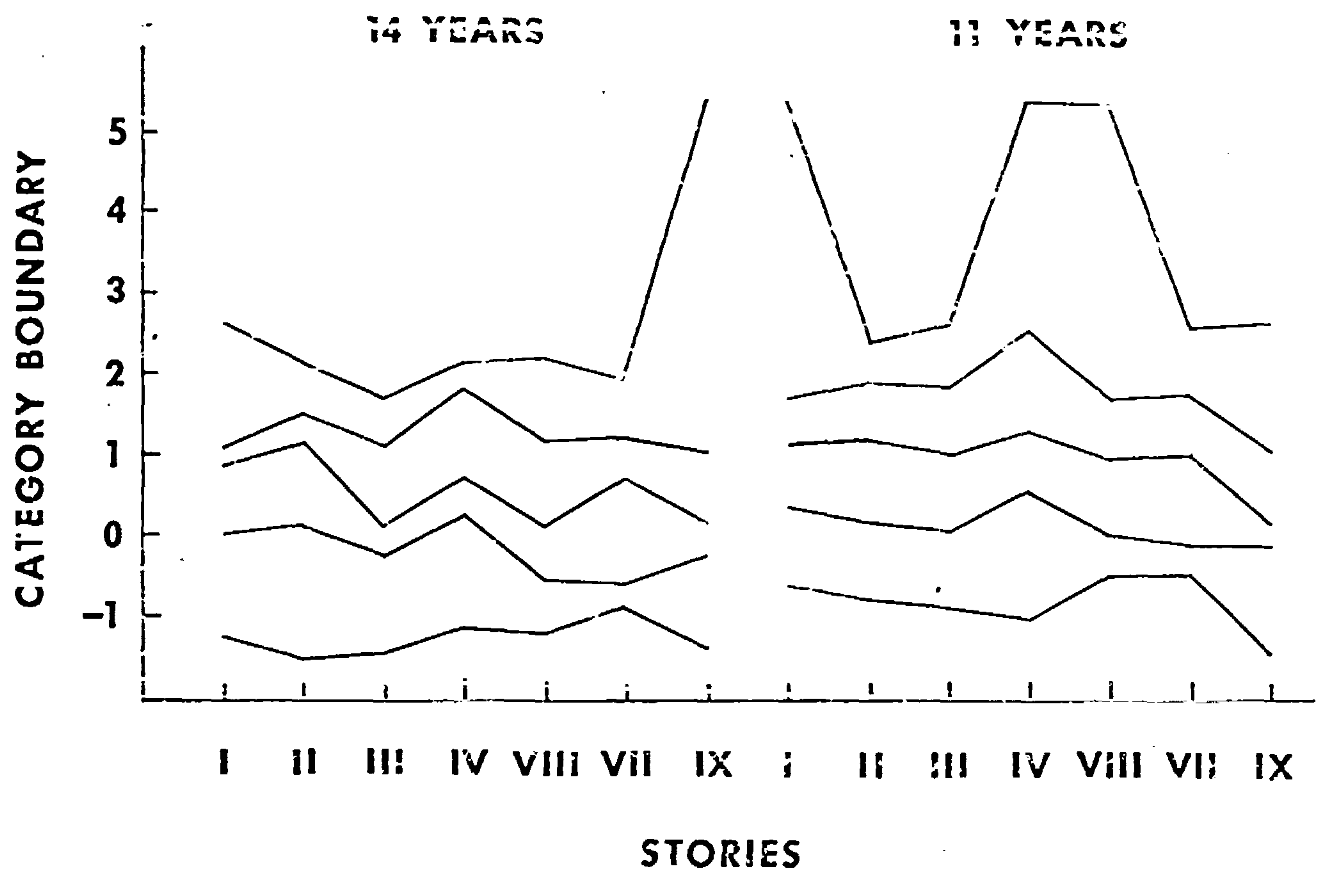
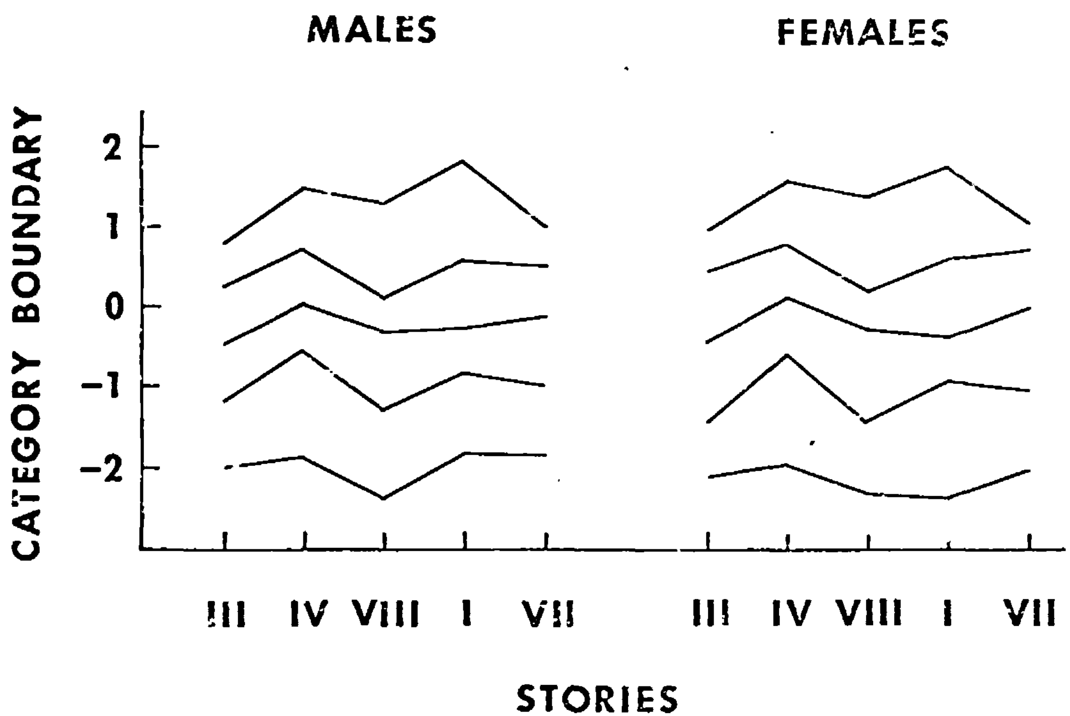

14 YEARS
11 YEARS
CATEGORY BOUNDARY
5
4
3
2
1
0
−1
I II III IV VIII VII IX I II III IV VIII VII IX
STORIES

Fig. 13.--Comparison of Category Boundaries Between Male and Female 19 year old U.S. Subjects



MALES
FEMALES
CATEGORY BOUNDARY
2
1
0
−1
−2
III IV VIII I VII
III IV VIII I VII
STORIES

Fig. 14.--Comparison of Category Boundaries Between Low and Middle Class British Subjects

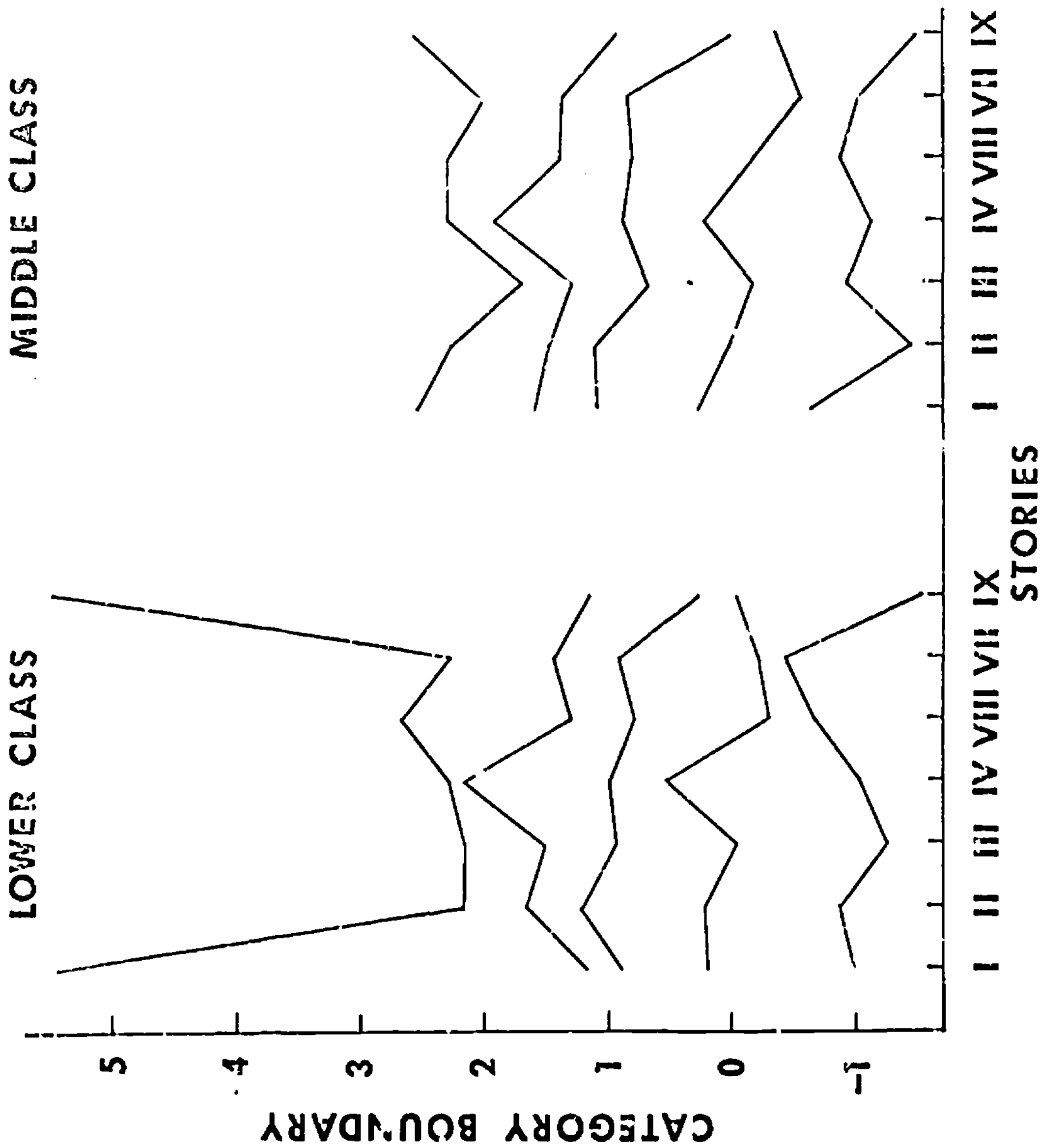
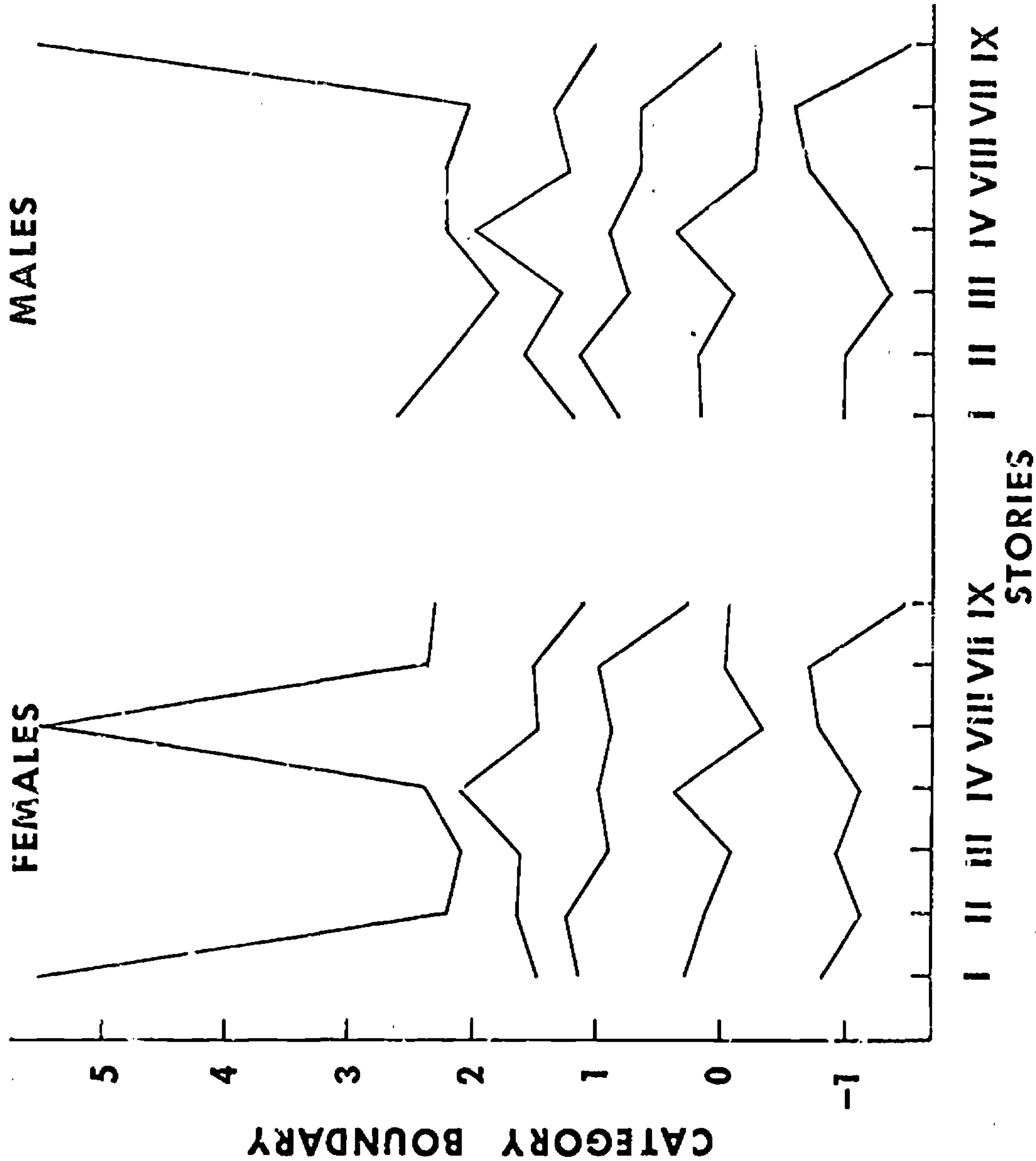

LOWER CLASS
MIDDLE CLASS
CATEGORY BOUNDARY
5
4
3
2
1
0
-1
I II III IV VIII VII IX
I II III IV VIII VII IX
STORIES

Fig. 15.--Comparison of Category Boundaries Between Male and Female British Subjects


FEMALES
MALES
CATEGORY BOUNDARY
5
4
3
2
1
0
-1
I II III IV VIII VII IX
I II III IV VIII VII IX
STORIES

Significance Tests Among the Subgroups

The stage boundary between the first and second stages is lower for 14 year olds than for 11 year olds. This corresponds to a lower proportion of 14 year olds being at the lowest moral judgment stage than 11 year olds.

The upper boundaries are lower for 14 year olds indicating a larger proportion above those boundaries in the upper stages than in 11 year olds.

Significance tests may be accomplished by squaring the difference between two groups' values for each category boundary, dividing by the sum of the variances for each boundary and summing the quotients, producing a variable distributed as χ^2 with 5 degrees of freedom.

Consider the following example:

TABLE 11

DATA FOR SIGNIFICANCE TEST BETWEEN TWO SETS OF CATEGORY BOUNDARIES

No.	C.B.11	Var.	C.B.14	Var.	ΔC.B.	$(\Delta C.B.)^2$	ΣVar	$\chi^2_{(1)}$
1	-.87533	.00970	-1.46757	.01589	.59224	.35075	.02559	13.70652
2	.04082	.00731	- .26487	.00716	.30569	.09345	.01447	6.45818
3	1.04143	.01095	.65021	.00815	.39122	.15305	.01910	8.01308
4	1.84446	.02763	1.15305	.01149	.69141	.47805	.03912	12.22009
5	2.60074	.11715	1.65572	.02013	.94502	.89306	.13728	6.50539

The sum of the final column = 46.90326, and whereas $\chi^2_{.001;5} = 20.5$, $p < .001$.

Note: The variance of a normal deviate is given by $\sigma^2 = \frac{pq}{Nh^2}$

where p is the proportion producing the deviate

q is 1. less that proportion

N is the number of subjects

h^2 is the square of the ordinate corresponding to the deviate

Sex Effects in the Category Boundaries among the Stories

In the Great Britain sample (the 11 and 14 year olds), three stories showed a significant difference between male and female subjects.

Story I: All category boundaries were lower for males than females. ($p < .05$)

Story III: All category boundaries were lower for males than females. ($p < .025$)

Story VII: The lowest two category boundaries were lower for females, the upper three, lower for males. ($p < .05$)

In the U.S. sample, (the 19 year olds), two stories showed a significant difference between males and females.

Story III: The lowest two category boundaries were lower for females, the upper three were lower for males. $(p < .005)$

Story VII: The lowest two category boundaries were lower for females, the upper three were lower for males. $(p < .05)$

Thus, three of the seven stories administered to the British subjects showed more boys in higher stages than girls. Two of the five stories presented to the American sample showed more girls than boys at the middle stages and more boys than girls at the higher stages.

Kohlberg had postulated that girls proceed through the stages faster than boys, so that at the younger ages, more girls should be found at the higher stages. Without taking age into account, no strong evidence is contributed by the data considered here to refute or support that hypothesis.

Social Class Effects in the Category Boundaries among the Stories

In the Great Britain sample, four stories showed a significant difference between lower and middle class subjects.

Story I: All category boundaries were lower for lower class than for middle class. As no subjects scored in the sixth stage for this story, the sixth stage for this story, the fifth category boundary is arbitrarily set at 5.54000. ($p < .05$)

Story II: The highest category boundary was lower for lower class subjects than for those in the middle class, but the remaining four boundaries were lower for middle class subjects than for lower class subjects. ($p < .01$)

Story III: The lowest category boundary was lower for lower class subjects than for those in the middle class, but the remaining four boundaries were lower for middle class subjects than for the lower class group. ($p < .005$)

Story VII: All category boundaries were lower for middle class subjects with the exception of the third which was lower for lower class subjects. ($p < .001$)

These four stories out of the seven presented show an indication, but no firm support for the hypothesis that middle class children are at higher stages than lower class children.

Age by Social Class Interactions in the Category Boundaries among the Stories

In the Great Britain sample, one story exhibited a significant difference between middle and lower classes of 14 year olds.

Story III: All category boundaries were lower for middle class 14 year olds than for lower class 14 year olds. ($p < .005$)

Two stories showed a significant difference between 11 year old middle and lower class subjects.

Story II: The lowest two category boundaries are lower for middle class 11 year olds than for lower class 11 year old subjects, while the upper three boundaries were lower for lower class 11 year olds. ($p < .01$)

Story VII: The lowest category boundary was lower for middle class 11 year olds, but the other four boundaries were lower for lower class 11 year old subjects. ($p < .001$)

Since seven stories were presented, with only one or two showing significant differences, there appears to be no support for the hypothesis that middle and lower classes differ within 11 and 14 year old age groups.

Age by Sex Interactions in the
Category Boundaries among the Stories

In the Great Britain sample, two stories showed significant differences among the boundaries between 14 year old male and female subjects.

Story VII: The lowest two boundaries were lower for females while the upper three boundaries were lower for male 14 year olds. ($p < .01$)

Story VIII: The lowest two category boundaries were lower for females, while the upper three boundaries were lower for male 14 year olds. ($p < .05$)

One story exhibited a significant difference among the category boundaries for the G.B. sample of male and female 11 year olds.

Story III: The highest and second lowest boundaries were lower for female 11 year olds while the remaining boundaries were lower for male 11 year old subjects. ($p < .025$)

It was expected that the younger girls would be at higher stages than the younger boys with random variation at the older ages, but only two stories given to the 14 year olds and one given to the 11 year olds detected any differences, failing to support this hypothesis.

Sex by Social Class Interactions in the
Category Boundaries among the Stories

In the Great Britain sample, five stories showed significant differences between middle and lower classes for males.

Story II: All category boundaries were lower for middle class males than for lower class males. ($p < .005$)

Story III: The lowest category boundary was lower for lower class males, while the remaining four categories were lower for middle class males. ($p < .05$)

Story IV: Except for the middle boundary, all category boundaries were lower for middle class males than for lower class males. ($p < .001$)

Story IX: While the lowest category boundary was lower for lower class males, the remaining boundaries were lower for middle class males. ($p < .001$)

In the G.B. sample no stories showed significant differences among the category boundaries between female lower class subjects and female middle class subjects.

In the G.B. sample, among the lower class subjects, no stories exhibited any significant differences among the category boundaries between male and female subjects.

Among the Great Britain middle class subjects, two stories showed significant differences between male and female subjects in the category boundaries.

Story III: All category boundaries were lower for middle class males than for females. ($p < .05$)

Story IX: With the exceptions of the highest boundary, all the category boundaries were lower for middle class males than for females. ($p < .01$)

Though the British male middle class subjects showed higher numbers at the more advanced stages in four of the stories than the lower class males, no such difference occured among the females.

Within the middle and lower classes there was no systematic difference between the male and female subjects.

TABLE 12

SUMMARY OF SIGNIFICANT DIFFERENCES AMONG THE CATEGORY BOUNDARIES OF THE STORIES

	I	II	III	IV	VII	VIII	IX
Age Effect	x	x	x	x	x	x	x
Social Class Effect	x	x	x		x		
Sex Effect	x		x		x		
Age by Class Interaction		x	x		x		
Age by Sex Interaction			x		x	x	
Sex by Class Interaction		x	x	x	x		x

Story Discriminating Power

For each subgroup, a polychoric correlation matrix among the stories administered was calculated. Then, a maximum likelihood factor analysis procedure was applied which resulted in a single factor in every subgroup considered but one, which was assumed to be a chance occurrence. Loadings were computed for each story on that single factor. These loadings, displayed in Table 13, will serve as the estimate of the correlation between story score and the latent trait, moral judgment level, i.e. discriminating power.

This analysis also produces standard errors of the loadings which will be employed in significance tests among the age subgroups.

Comparison of Age Groups - Story Loadings - Discriminating Power

Even though the category boundaries for any story may change from one group to another, it was expected that the relationships among the stories ought to remain constant. Since a single factor model fit the data in the subgroups, the loadings were compared using two-way analysis of variance, stories by age level. Factor loadings are not normally distributed, so a transformation was employed and a weighted means analysis performed.

In place of the loadings, the log of the reciprocal of the loadings were entered. These values were weighted with reciprocal of the variance of this transformed variable.

since

$$\mathcal{V}[f(x)] = \left(\frac{\partial f}{\partial x}\right)^2 \mathcal{V}[x]$$

where

$$f(x) = \log \frac{1}{x},$$

TABLE 13

FACTOR LOADINGS OF THE STORIES

	I	II	III	IV	V	VI	VII	VIII	IX
US Male, 19, Middle	.330		.467	.212			.314	.430	
US Female, 19, Middle	.233		.333	.180			.206	.519	
US 19, Middle	.305		.400	.138			.298	.444	
US Male, 13, Middle	.523	.472	.522	.653	.625	.694	.582	.635	.571
GB 11	.264	.302	.552	.256			.588	.656	.354
GB 14	.239	.471	.467	.559			.393	.599	.411
GB Middle	.226	.373	.579	.456			.332	.614	.445
GB Lower	.413	.405	.509	.453			.609	.679	.284
GB Male	.358	.546	.599	.487			.485	.648	.335
GB Female	.257	.248	.496	.397			.497	.701	.364
GB 14 Middle	-.055	.305	.548	.631			.257	.589	.625
GB 14 Lower	.479	.478	.373	.532			.511	.599	.293
GB 11 Middle	.437	.326	.609	.242			.307	.457	.270
GB 11 Lower	.228	.260	.596	.295			.682	.702	.364
GB Male Middle	.355	.653	.763	.503			.246	.507	.356
GB Male Lower	.432	.422	.508	.410			.597	.697	.273
GB Female Middle	.071	.009	.396	.304			.256	.730	.519
GB Female Lower	.393	.378	.538	.497			.608	.682	.294
GB Male 11	.399	.403	.681	.342			.473	.601	.189
GB Male 14	.178	.695	.601	.530			.355	.518	.429
GB Female 11	.143	.212	.420	.179			.638	.711	.534
GB Female 14	.178	.260	.540	.572			.261	.667	.328

$$f' = \frac{\partial}{\partial x}[\log \frac{1}{x}] = [\frac{1}{1/x}][-\frac{1}{x^2}] = -\frac{1}{x}$$

and

$$(f')^2 = \frac{1}{x^2}$$

$$\therefore \mathcal{V}[\log \frac{1}{x}] = \frac{\mathcal{V}[x]}{x^2}$$

and the weights are

$$w_i = \frac{[\log \frac{1}{\text{loading}_i}]^2}{[\text{s.e. of loading}_i]^2}$$

The two factors were stories (I,III,IV,VII and VIII) and age (11,13,14 and 19) and simple contrasts were used to determine significant differences. The interaction was used as the error variance.

Null Hypothesis One:

The stories will discriminate equally, i.e., the loadings of the stories will be equal; all will correlate equally with the single factor.

Between Mean Square = 169.0108

Degrees of Freedom = 4

P - value less than .01 (hypothesis rejected)

Null Hypothesis Two:

The relationship among the stories will remain constant regardless of age differences, i.e., the loadings for the

TABLE 14

DATA FOR TWO-WAY ANALYSIS OF VARIANCE

Story	Age	Loading	log $\frac{1}{\text{loading}}$	s.e.	weight
I	11	.264	1.332	.081	253
	13	.523	.648	.083	60
	14	.239	1.431	.080	341
	19	.305	1.187	.054	470
III	11	.552	.594	.079	59
	13	.522	.650	.083	60
	14	.467	.761	.078	97
	19	.400	.916	.059	280
IV	11	.256	1.363	.082	265
	13	.653	.426	.079	30
	14	.559	.582	.077	56
	19	.138	1.981	.051	1308
VII	11	.588	.531	.079	47
	13	.582	.541	.082	42
	14	.393	.934	.079	145
	19	.298	1.211	.054	489
VIII	11	.656	.422	.080	30
	13	.635	.454	.080	34
	14	.599	.512	.077	44
	19	.444	.812	.062	165

stories will remain the same for each of the four ages.

Between Mean Square = 47.0005

Degrees of Freedom = 3

P - value less than .01 (hypothesis rejected)

Relation of Moral Judgment Level to Age

To eliminate socioeconomic effects, only the middle class British subjects were used so that all four age groups are middle class children.

It was expected that the only parameter which would change from one age group to another would be the category boundaries. Therefore a log reciprocal average among the discriminating powers (one-factor loadings) of the four age groups was computed for each story.

A level of moral judgment was calculated for various patterns and slopes were computed from the four age values. Table 16 gives the moral judgment level for each pattern using the category boundaries for each age, the average loading among the age groups, and the resulting slopes. Graphs representing the relationship between moral judgment level and age for each of the five selected patterns are given in Figure 16.

The moral judgment level pictured here has an analogue with the intelligence quotient. Given an answer pattern, a subject of lower age is judged at a higher level than a subject of higher age. So, as a subject gorws older, he should also answer at increasingly higher patterns causing his moral judgment level to be relatively constant.

The category boundaries for the four ages among the middle class subjects was given earlier in Table 4 and Table 8. The following log reciprocal average of the discriminating powers among the age groups was used for each story.

TABLE 15

AVERAGE DISCRIMINATING POWERS OF THE FIVE STORIES CONSIDERED

Story	I	III	IV	VII	VIII
Discriminating Power	.3167	.4817	.3370	.4474	.5769

Age, Class, Culture and Sex Comparisons
Using the Moral Judgment Level

To make comparisons among the age, sex, class and culture groups, it is necessary to define story parameters which would be used for all subgroups. The discriminating powers will be the log reciprocal averages discussed previously and the category (stage) boundaries will be the arithmetic average of each boundary across the four ages. This procedure results in the following parameters, given in Table 17.

TABLE 16

MORAL JUDGMENT LEVEL FOR A GIVEN PATTERN FOR EACH AGE

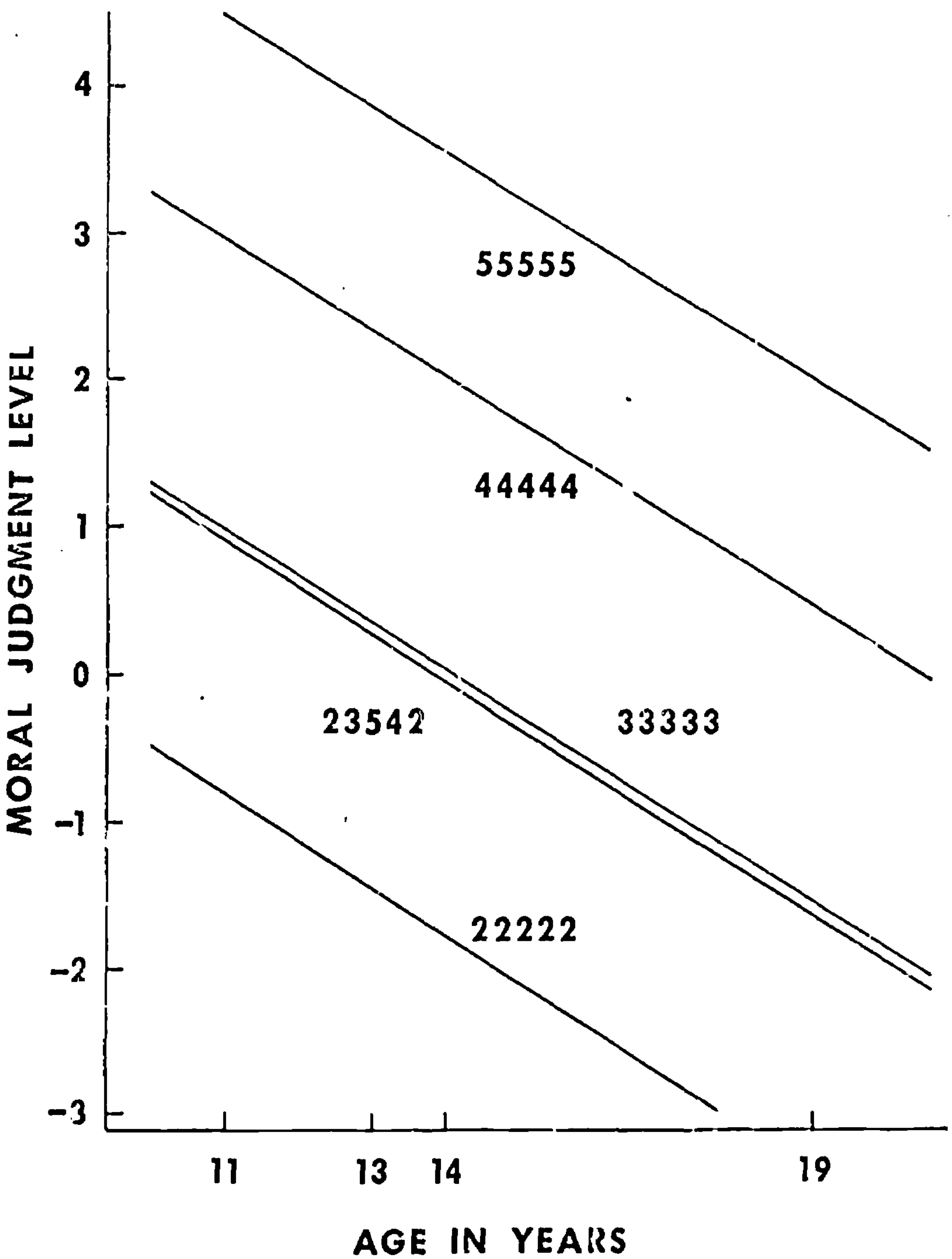
	Age				
Pattern	11	13	14	19	Slope
11122	-2.1053	-2.9861	-3.2698	-4.7894	-0.3721
22222	-0.7691	-1.5377	-1.8676	-3.4078	-0.3523
22333	0.5132	-0.5088	-0.5737	-2.0329	-0.3143
23542	0.7427	0.1411	0.1261	-1.4913	-0.2830
33333	1.1730	-0.0041	0.1399	-1.3979	-0.3043
33344	2.3380	0.9400	1.1360	-0.5433	-0.3252
44444	3.2270	1.9568	1.8734	0.2378	-0.3322
44455	4.7544	3.5976	2.6270	1.0306	-0.4166
55555	7.8895	5.4441	3.2619	1.8504	

TABLE 17

STORY PARAMETERS

Story		I	III	IV	VII	VIII
Disc Power		.3167	.4817	.3370	.4474	.5769
Cat. Bound.	1	-1.16577	-1.15902	-1.34213	-1.15567	-1.57095
	2	-0.18472	-0.43518	-0.06652	-0.58667	-0.67087
	3	0.50850	0.39344	0.64134	0.49074	0.27614
	4	1.5297	1.35264	1.58417	1.28556	1.11006
	5	2.93346	3.33138	2.06890	2.60977	3.58486

Fig. 16.--Moral Judgment Level as a Function of Age Given an Answer Pattern of Story Scores


MORAL JUDGMENT LEVEL
4
3
2
1
0
-1
-2
-3
55555
44444
23542
33333
22222
11
13
14
19
AGE IN YEARS

This resulting moral judgment level can be viewed as an analogue to mental age, for as children pass through the stages, their moral judgment level increases. The same story parameters are used regardless of the subject's age.

Sex by Age Comparisons

Since all the 13 year olds in the available sample were boys, comparisons will be made among the 11, 14 and 19 year old subjects.

H_0: There will be no significant difference among the three ages on the moral judgment level.

$F = 313.5488$ d.f. = 2,1416 P .0001

(hypothesis rejected)

H_0: There will be no significant difference among the sexes on the moral judgment level.

$F = 0.5433$ d.f. = 1,1416 P .4612

(hypothesis accepted)

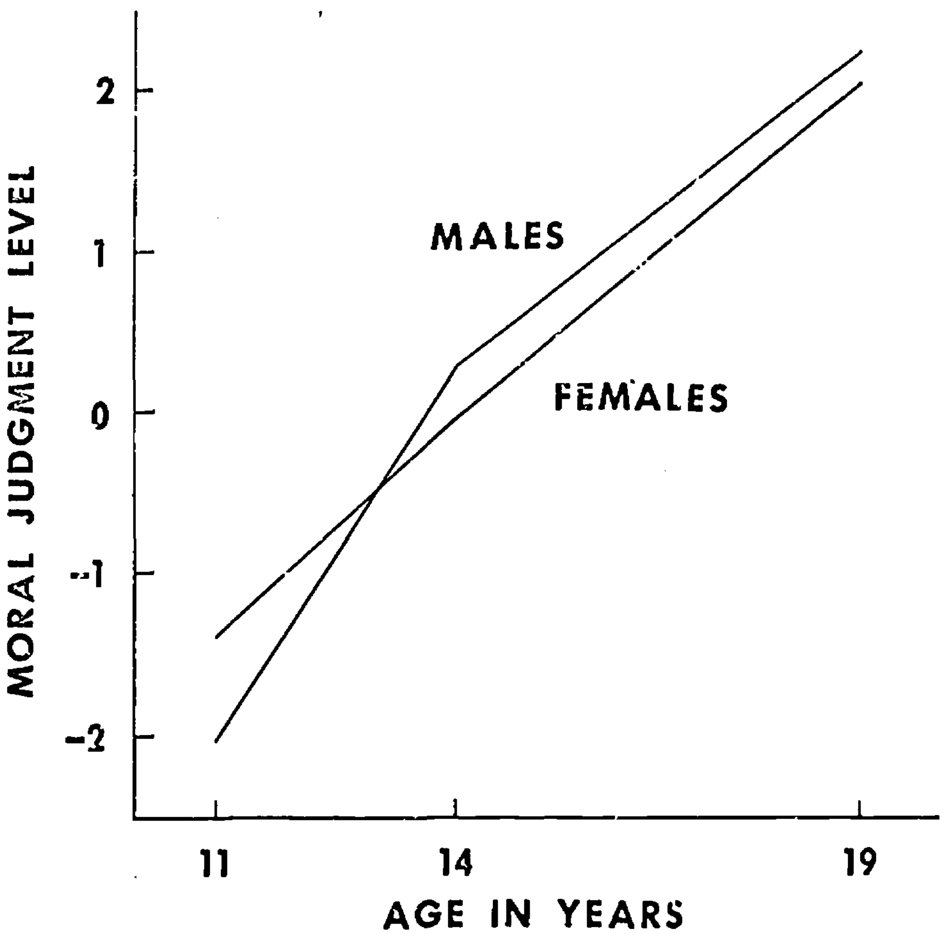
H_0: There will be no significant interaction between age and sex groups.

$F = 3.6284$ d.f. = 2,1416 P .0269

(hypothesis rejected)

This interaction is illustrated in Figure 17.

Fig. 17.--Illustration of Sex by Age Interaction


MORAL JUDGMENT LEVEL
2
1
0
-1
-2
MALES
FEMALES
11
14
19
AGE IN YEARS

As previously discussed, Kohlberg hypothesized that, as in other traits, younger girls are found at higher stages than younger boys, but older groups of subjects had nearly equal levels of development. The significant interaction displayed above supports this hypothesis.

Social Class Comparisons

Only in the 11 and 14 year old Great Britain sample were there both lower and middle class subjects, therefore, the class comparison will be made across these two age groups.

H_0: There will be no significant difference in moral judgment level between lower and middle class subjects.

$F = 0.0308$ d.f. = 1,437 P .8609

(hypothesis accepted)

Age by Social Class Comparison

H_0: There will be no significant interaction in moral judgment level between age and social class subgroups.

$F = 2.5188$ d.f. = 1,437 P .1133

(hypothesis accepted)

Sex by Social Class Comparison

H_0: There will be no significant interaction between sex and social class groups on moral judgment level

$F = 0.0322$ d.f. = 1,437 P .8578

(hypothesis accepted)

Sex by Age by Social Class Comparison

H_0: There will be no significant interaction between sex, age and social class groups on moral judgment level.

$F = 4.2256$ d.f. = 1,437 P .0405

(hypothesis rejected)

This interaction is illustrated in Figure 18.

Culture Comparison Using the Moral Judgment Level

Since there were no similar age groups in both the U.S. and Great Britain samples, the closest ages are the U.S. 13 year olds and the Great Britain 14 year olds.

H_0: There will be no significant difference between the 13 year old group and the 14 year old group on moral judgment level.

$F = 9.285$ d.f. = 147,224 $P < .01$

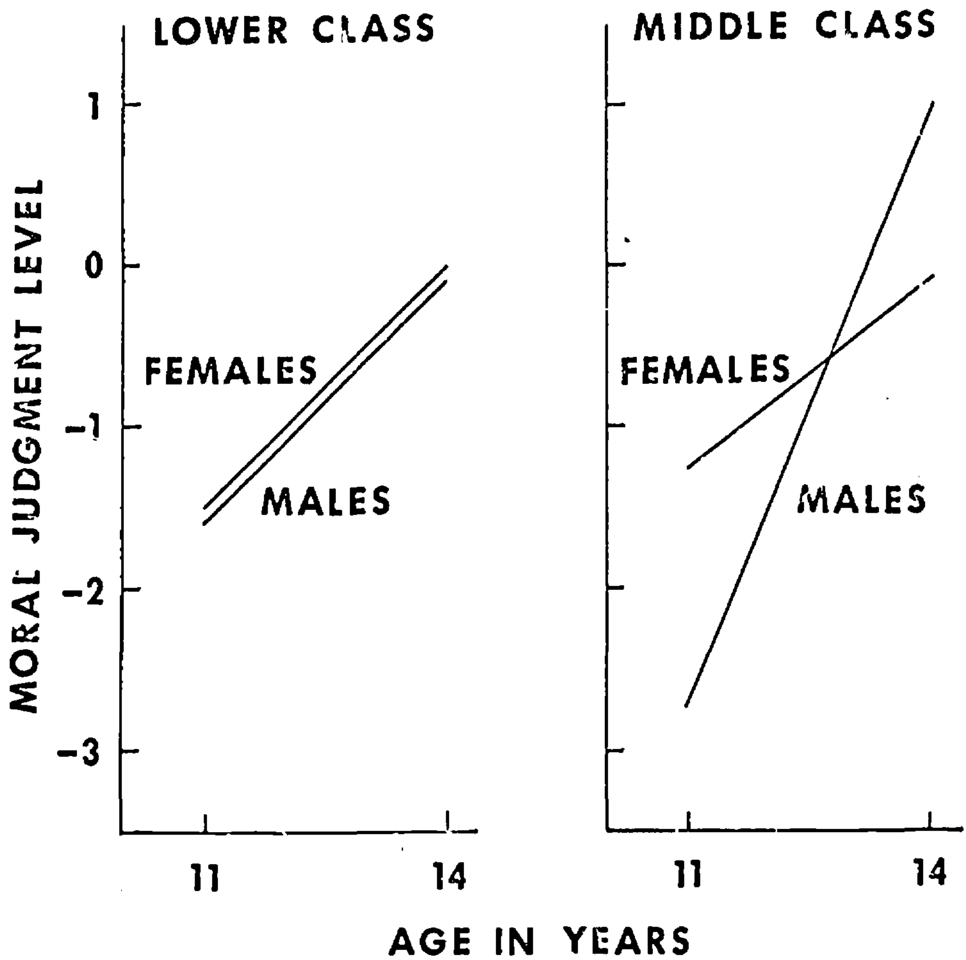
(hypothesis rejected)

Although Kohlberg has postulated that U.S. children proceed through the stages faster than Great Britain children, the Great Britain 14 year olds were significantly higher in moral judgment level than the U.S. 13 year olds.

Relation Between Moral Judgment Level and the Six Stages

The probability of a randomly selected subject having a moral judgment level θ answering with a pattern of scores

Fig. 18.--Illustration of Age by Sex by Class Interaction



LOWER CLASS
MIDDLE CLASS
MORAL JUDGMENT LEVEL
1
0
-1
-2
-3
FEMALES
MALES
FEMALES
MALES
11
14
11
14
AGE IN YEARS

$(k_1, k_2, \ldots k_n)$ was presented earlier in the chpater discussing the estimation of the moral judgment level.

If one assumes that an answer pattern consisting of the same stage response for each story represents a subject at a pure stage (e.g., a 2 2 2 2 2 response for five stages indicates a subject at stage two of Kohlberg's six stages.) then the probability of being at a pure stage can be computed for a large range of moral judgment levels.

It has already been shown that two patterns pr. ucing the same global score result in different moral judgment levels. Thus, one answer pattern with the same number of 2's and 3's as another may result in a higher probability of being at stage three than the second. This is the reason that the moral judgment level is used as the basis for determining the probability of being in each stage.

Table 18 gives the results of a few patterns with their respective moral judgment levels and probabilities of a subject with that level being in each stage.

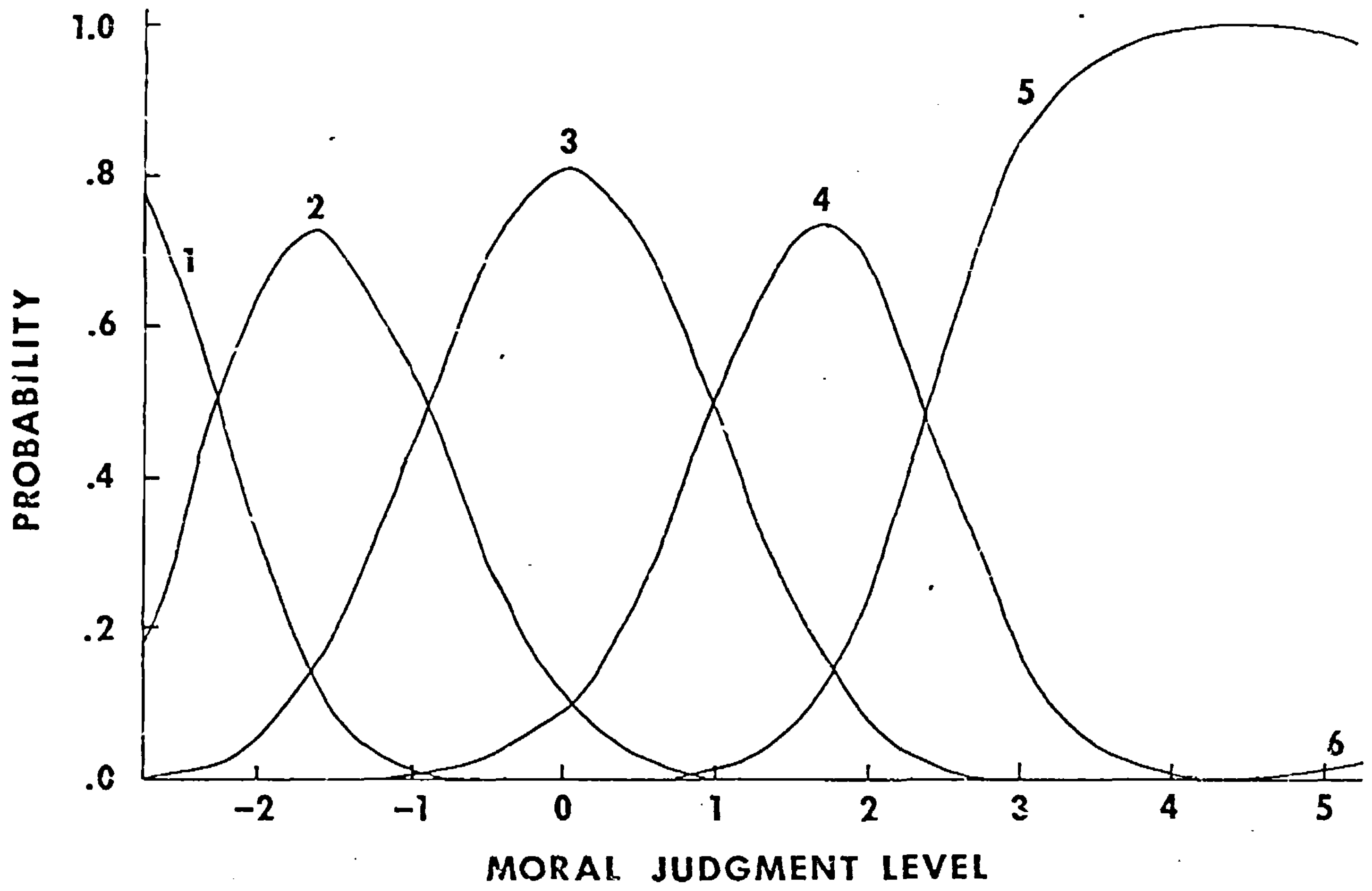
Figure 19 is a representation of the relationship between a subject's moral judgment level and the probability of that subject being in each of Kohlberg's six stages.

TABLE 18

PROBABILITY OF BEING IN EACH STAGE
AS A FUNCTION OF MORAL JUDGMENT LEVEL

Story					Latent Trait	Prob. of Being at Each Stage					
I	III	IV	VII	VIII	(M.J.Level)	1	2	3	4	5	6
4	4	6	6	6	5.7799	0.00	0.00	0.00	0.00	0.89	0.11
6	2	4	4	3	1.5506	0.00	0.22	0.71	0.07	0.07	0.00
5	3	2	4	3	0.9828	0.00	0.01	0.51	0.48	0.00	0.00
4	x	3	2	4	0.4992	0.00	0.03	0.73	0.24	0.00	0.00
3	1	3	4	3	0.0350	0.00	0.11	0.80	0.09	0.00	0.00
5	1	2	2	4	-0.4954	0.00	0.29	0.68	0.03	0.00	0.00
5	3	4	2	1	-0.9965	0.02	0.55	0.42	0.01	0.00	0.00
1	3	1	1	3	-1.4951	0.09	0.72	0.19	0.00	0.00	0.00
3	1	4	3	1	-2.0155	0.32	0.63	0.05	0.00	0.00	0.00
1	2	1	1	2	-2.9895	0.90	0.10	0.00	0.00	0.00	0.00

Fig. 19.--Probability of Being in Each Stage Given a Level of Moral Judgment


1.0
.8
.6
.4
.2
.0
PROBABILITY
1
2
3
4
5
6
−2
−1
0
1
2
3
4
5
MORAL JUDGMENT LEVEL

CHAPTER V

DISCUSSION AND CONCLUSIONS

In discussing the technique proposed in this paper, the following questions are considered. Did the data fit the model, i.e., are the assumptions of the model met? Can item (story) and latent trait (moral judgment level) parameters be derived using accepted statistical methods? When these parameters are used with real data are results produced consistent with the theory? Does the proposed procedure result in more information than the traditional technique? Is the graded scale model widely generalizable? The answers to all these questions are definitely affirmative.

Assumptions about the Model

One assumption for the parameter estimation procedure described was that the latent trait, moral judgment level, be normally distributed in the population. From the data considered here, the developmental level of moral judgment in subjects below 14 years of age had some degree of dependence on sex or social class, resulting in multimodal distributions. By the age of 14 and still continuing at 19, the latent trait appeared normally distributed regardless of sex or social class.

In this analysis, it was also necessary to assume that the stories were distributed multivariate normally in a pairwise bivariate normal manner and represented a unifactoral model. Polychoric correlations were calculated for each pair of items independently and in each case, an expected contingency table was produced, assuming a bivariate normal distribution of the stage scored for the two items using th[illegible]ed category boundaries and the resultant correlation coe[illegible]nt. A chi-square was produced to determine whether the observ[illegible]d table differed significantly from this expected bivari[illegible] [illegible]ormal table. Since there were nineteen subgroups with [illegible] items, one subgroup with nine items and three subgro[illegible] five items, there were 19 × 21 + 1 × 36 + 10 = 465 c[illegible]ency tables produced and polychoric correlations calculat[illegible]. [illegible] these, 397 or 86 per cent did not differ significant[illegible] [illegible] > .05) from the expected bivariate normal distribution. Of [illegible]l polychoric correlation matrices produced for the subgroups, 23 or 74 per cent represented a single factor model ($p > .05$). Therefore, it can be concluded that the moral judgment stage score data used in this analysis do fulfill the assumptions described above.

Characteristics of the Items (Stories)

Story parameters were derived using the method of maximum likelihood estimation, resulting in five stage boundaries and a discriminating power for each story. Using the same estimation procedure, and employing the known story characteristics, a moral judgment level was derived for a given pattern of story scores.

Final story parameters were calculated using a log reciprocal average across age for discriminating power and an arithmetic average across age for category boundaries. From most highly discriminating to least discriminating, the stories were VIII, III, VII, IV, and I. Since the category boundaries were not consistently lower or higher from one story to the next, only an approximate estimate of difficulty can be determined. The stories, in order from easiest to most difficult were VIII, VII, I, III, and IV. For this situation, difficulty is intended to mean less subjects at the higher stages.

Moral judgment values were calculated using age-linked story characteristics as well as norm group story characteristics.

Using the norm group parameters, the probabilities of a subject with a given answer pattern being at each stage presented.

Hypothesis Tests Among the Subgroups

Two methods of testing differences among the subgroups were employed. The first contrasted the category boundaries as determined by one group against those determined by the other group. In the second method, the moral judgment level of each group was compared to the other using the averaged story parameters.

Both the category boundary comparisons and the latent trait value comparisons yielded significantly different means between each of the four age groups.

There were not significant sex effects found, either among the British 11 and 14 year olds or the American 19 year olds on the two tests.

There was only an indication of social class difference when four of the seven stories presented to the British 11 and 14 year olds produced significantly different category boundaries but one of these actually showed more lower class children at higher stages than middle class children. The social class comparison using the moral judgment level resulted in no significant difference between middle and lower class subjects.

With no similar age groups in both British and American subgroups, the closest pure culture comparison was that between U.S. 13 year olds and G.B. 14 year olds. The moral judgment level contrast between these groups resulted in a significant difference in favor of the 14 year old Britishers indicating a stronger age than culture effect in the trait. The category boundary test substantiated this result.

Since all American subjects were judged to be middle class subjects, age by social class interactions were investigated only in the British 11 and 14 year old groups of middle and low class subjects. Neither the category boundary comparisons nor the moral judgment level comparison revealed any such interaction.

The category boundary contrast revealed no clear significant age by sex interaction, but the moral judgment level comparison yielded a significant interaction with younger girls at higher stages than boys, but with the sexes being indistinguishable in moral judgment level at later ages.

Investigation of sex by social class interactions among the British subjects revealed three stories showing middle class males at higher stages than lower class males, one story with the reverse occurring and no stories exhibiting differences between the classes among the females. The moral judgment level contrast also indicated the lack of any sex by social class interaction.

There was a significant triple interaction among sex, age and social class in a comparison using the moral judgment level. Lower class males and females were indistinguishable at both 11 and 14 years, but 11 year old middle class males, who were lower than 11 year old middle class females, were above their female counterparts by the age of 14.

Addition Information Produced by the Graded Scale Model

As described above, story differences in difficulty and discriminating power are determined from the model. Also, a continuous, normally distributed, moral judgment level can be estimated with its standard error for group comparisons. Apart from these data, the method provides a description of the subjects probability of being in each stage, given his response pattern to the stories.

These probabilities can be used in a variety of ways. For example, it may be interesting to know which children in a given group are "nearly ready" to change to the next highest stage. This occurs when the probabilities of a subject being in two adjacent stages are nearly equal. This information will be valuable to experimenters attempting to change the moral judgment level of subjects with a series of discussions, readings or experiences.

Generalizability of the Graded Scale Model

There are many concepts in developmental psychology which are being viewed as sequences of stages where each additional stage attained implies not only a substitition or addition but an integration with earlier stages. Some of these are concepts of dreams, liquid and solid conservation, realism, animism and life. In each of these, stories are discussed or tasks attempted. The experimenter then rates the subject at a particular stage of development for each task or story. Using the technique described in this paper, story or task parameters may be estimated and used to determine subjects' latent ability levels for contrasts among groups of varying characteristics.

In the area of market research, stages have been defined to describe consumer behavior towards a product. First, there is an awareness of the new product, then interest accompanied by questions or requests for information, then trial, acceptance, use and repurchase. Various market indicators could be tested as to their predictive capabilities. It would also be very useful to have an underlying continuous variable to compare different geographical or socioeconomic groups' reaction to new products.

Economists describe stages in nations' economic growth. Various indicators are used to determine whether a country is in the amassing capital stage, the production oriented stage, or the industrially mature stage. These indicators could also be judged as to predictive strength and scaled to allow for differences in complexity.

The physiological stages of development in the human being; i.e., infant, child, adolescent, adult, elder, may be more rigorously defined using characteristics which change through the stages. Hair color, height or weight, bone ossification, coordination, etc. would be the "items" or characteristics.

The psychological reaction to knowledge that one has a terminal disease has been shown to follow a predictable pattern, including rejection, fear, anger and finally acceptance. The indicators are in the form of reactions to questions or other stimuli.

Other stage-like structures that may be tested with this model are taxonomies of cognitive and affective behaviors. Each level adds, substitutes and integrates some defining properties to those below it. It may be possible to determine interlevel distances and the position of a subject's latent ability on some continuum describing the depth and scope of his response to some stimulus in the form of a question or event.

APPENDIX

MORAL JUDGMENT SITUATIONS

(Probing questions not included)

I. Joe is a 14-year-old boy who wanted to go to camp very much. His father promised him he could go if he saved up the money for it himself. So Joe worked hard at his paper route and saved up the $40 it cost to go to camp and a little more besides. But just before camp was going to start, his father changed his mind. Some of his friends decided to go on a special fishing trip, and Joe's father was short of the money it would cost. So he told Joe to give him the money he had saved from the paper route. Joe didn't want to give up going to camp, so he thought of refusing to give his father the money.

Should Joe refuse to give his father the money or should he give it to him? Why?

II. Joe lied and said he only made $10, and went to camp with the other $40 he made. Joe had an older brother Bob. Before Joe went to camp, he told Bob about the money and about lying to their father.

Should Bob tell their father? Why?

III. In Europe, a woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost him to make. He paid $200 for the radium and charged $2,000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about $1,000 which is half of what it cost. He told the druggist that his wife was dying and asked hime to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So Heinz got desparate and broke into the man's store to steal the drug for his wife.

Should the husband have done that? Why?

IV. The Dr. finally got some of the radium drug for Heinz's wife. But it didn't work, and there was no other treatment known to medicine which could save her. So the Dr. knew that she had only about six months to live. She was in terrible pain, but she was so weak that a good dose of a pain-killer like ether or morphine would make her die sooner. She was delirious and almost crazy with pain, and in her calm periods, she would ask the Dr. to give her enough ether to kill her. She said she couldn't stand the pain and she was going to die in a few months anyway.

Should the Dr. do what she asks and make her die to put her out of her terrible pain? Why?

V. In Korea, a company of Marines was way outnumbered and was retreating before the enemy. The company had crossed a bridge over a river, but the enemy were mostly still on the other side. If someone went back to the bridge and blew it up as the enemy were coming over it, it would weaken the enemy. With the head start the rest of the men in the company would have, they could probably then escape. But the man who stayed back to blow up the bridge would probably not be able to escape alive; there would be about a 4 to 1 chance he would be killed. The captain of the company has to decide who should go back and do the job. The captain himself is the man who knows best how to lead the retreat. He asks for volunteers, but no one will volunteer.

Should the captain order a man to stay behind, or stay behind himself, or leave nobody behind? Why would that be best?

VI. The captain finally decided to order one of the men to stay behind. One of the men he thought of was one who had a lot of strength and courage but he was a bad trouble maker. He was always stealing things from the other men, beating them up and wouldn't do his work. The second man he though of had gotten a bad disease in Korea and was likely to die in a short time anyway, though he was strong enough to do the job.

If the captain was going to send one of the two men, should he send the trouble maker or the sick man? Why?

VII. Two young men were in trouble. They were secretly leaving town in a hurry and needed money. Al, the older one, broke into a store and stole $500. John, the younger one, went to a man who was known to help people in town. John told the man that he was very sick and he needed $500 to pay for the operation. Really he wasn't sick at all, and he had no intention of paying the man back. Although the man didn't know John very well, he loaned him the money. So John and Al skipped town, each with $500.

VIII. While all was happening, Heinz was in jail for breaking in and trying to steal the medicine. He had been sentenced for ten years. But after a couple of years, he escaped from the prison and went to live in another part of the country under a new name. He saved money and slowly built up a big factory. He gave his workers the highest wages and used most of his profits to build a hospital for work in curing cancer. Twenty years had passed when a tailor recognized the factory owner as being Heinz, the escaped convict whom the police had been looking for back in his home town.

Should the tailor report Heinz to the police? Why should(n't) he?

IX. During the war in Europe, a city was often heavily bombed. All the men in the city were assigned to different fire-fighting and rescue stations all over the city. A man named Diesing was in charge of one fire engine station station near where he worked. One day after an especially heavy bombing, Diesing left the shelter to go to his station. But on the way, he decided he had to see whether his family was safe. His home was quite far away, but he went there first.

Was it right or wrong for him to leave the station to protect his family?

BIBLIOGRAPHY

Baker, Frank (1961) Empirical comparison of item parameters based on the logistic and normal functions. Psychometrika 26 239-246.

Bock, R. Darrell (1969) Fitting a response model for n items when item j is in m_j nominal categories. (in preparation)

Bock, R. Darrell and Lieberman, Marcus (1967) Fitting a response model for n dichotomous items. Statistical Laboratory Research Memorandum No. 8. Department of Education, University of Chicago.

Feder, Daniel D. (1947) Estimation of test item difficulty by averaging highest and lowest quarter performance compared with total population count. Educational and Psychological Measurement 7, 133-134.

Findley, Warren G. (1956) A rational for evaluation of item discrimination statistics. Educational and Psychological Measurement 16, 175-180.

Froemel, Ernest (1969) Calculation of the tetrachoric correlation coefficient. (in preparation)

Graham, Douglas and Simpson, Anna Louise, Moral Development of School Children, unpublished.

Guilford, J. P. (1936) Psychometric Methods New York, N. Y. McGraw-Hill.

Guilford, J. P. (1954) Psychometric Methods New York, N. Y. McGraw-Hill.

Hartshorne, H. and May, M. A. (1928-1930) Studies in the nature of character: Vol. I, Studies in deceit; Vol. II, Studies in self-control; Vol. III, Studies in the organization of character. New York, Macmillan.

Joreskog, Karl G. (1967) UMLFA: A computer program for unrestricted maximum likelihood factor analysis. Education Testing Service Research Memorandum RM-66-20. Princeton, N. J.

Kendall, M. G. and Stuart, A. The Advanced Theory of Statistics, Vol. II, London: Griffen, 1961.

Kohlberg, Lawrence (1968) Moral Development (in preparation).

Lancaster, H. O. and Handan, M. A. (1964) Estimation of the correlation coefficient in contingency tables with possible non-metrical characters. Psychometrika 29, 383-391.

Lawley, D. N. (1943) On problems connected with item selection and test construction. Proceedings of the Royal Society of Edinburgh, 61, 273-287.

Lawley, D. N. (1944) The factorial analysis of multiple item tests. Proceedings of the Royal Society of Edinburgh, 62-A, 74-82.

Lieberman, Marcus (1969) Calculation of a polychoric correlation coefficient. Paper presented at Psychometric Society Spring Conference, Princeton, N. J., April, 1969.

Lord, Frederic M. (1952) A theory of test scores. Psychometric Monograph No. 7.

Lord, Frederic M. (1968) An analysis of the verbal Scholastic Aptitude Test using Birnbaum's three parameter logistic model. Educational and Psychological Measurement 28, 989-1020.

Lord, Frederic M. and Novick, Melvin R. (1968) Statistical Theories of Mental Test Scores, Reading, Massachusetts: Addison-Wesley.

Maxwell, A. E. (1959) Maximum likelihood estimation of item parameters using the logistic model. Psychometrika 24, 221-228.

Michael, William G. and Perry, Normal C. (1955) A Theory of item analysis based on the scoring of items at three levels of appropriateness of response. Education and Psychological Measurement 15, 404-415.

Pearson, K. and Pearson, E. S. (1922) On polychoric coefficients of correlation. Biometrika 14, 127-156.

Piaget, Jean (1932) The Moral Judgment of a Child, Glencoe, Illinois: Free Press (1948) (orig. 1932).

Richie-Scott, A. (1919) The correlation coefficient of a polychoric table. Biometrika 12, 93-133.

Samejima, Fumiko (1968) Estimating latent ability using a response pattern of graded scores. Educational Testing Service Service, Research Bulletin RB-68-2, Princeton, N. J.

Srivastava, A. B. L. and Webster, H. (1967) An estimate of true score in the case of items scored on a continuous scale. Psychometrika 32, 327-338.

Tallis, L. R. (1962) The maximum likelihood estimation of correlation from contingency tables. Biometrics 18, 342-353.

Torgerson, Warren S. (1967) Theory and Methods of Scaling, New York, N. Y. John Wiley and Sons.

Tucker, Ledyard R. (1948) A method for scaling ability items in difficulty taking item unreliability into account (abstract). American Psychologist 3, 309-310.