JOHN C. GIBBS
KEITH F. WIDAMAN

In collaboration with Anne Colby

SOCIAL INTELLIGENCE

Measuring the Development of Sociomoral Reflection

With a foreword by

LAWRENCE KOHLBERG

SOCIAL INTELLIGENCE

Measuring the Development of Sociomoral Reflection

JOHN C. GIBBS
Department of Psychology
The Ohio State University

KEITH F. WIDAMAN
Department of Psychology
University of California at Riverside

—————— *In collaboration with ANNE COLBY* ——————

—————— *With a foreword by* ——————

LAWRENCE KOHLBERG

Editorial/production supervision
  and Interior design by Virginia Livsey
Cover design by 20/20 Services, Inc.; Mark Berghash
Manufacturing buyer: Edmund W. Leone

ISBN 0-13-815910-6

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

SOCIAL INTELLIGENCE

Measuring the Development of Sociomoral Reflection

Dedicated to

Larry Kohlberg

FOREWORD

Lawrence Kohlberg

John Gibbs, the first author of this volume, was a valued member of a Harvard team which developed a method of assessing stages of spontaneous moral reasoning (Standard Issue Scoring) and validated this method against 20-year American longitudinal data and 10-year longitudinal data from Turkey and Israel (Colby, Kohlberg, Gibbs, & Lieberman, in press). Building from this Standard Issue Scoring method, John saw the possibility of developing an alternative open-ended instrument which would be easier to administer and score than the Harvard method, and yet would retain the qualitative nature of responses lost in Jim Rest's (1979) multiple-choice recognition test (the Defining Issues test). In part, John was inspired by the format of the Loevinger & Wessler (1970) group-administerable Sentence Completion Test. Like the Loevinger test, Gibbs and Widaman's Sociomoral Reflection Measure (SRM) has a defined stem (in this case, moral norms) to which the subject responds. The value content is precoded, and the rater need only attend to the stage structure of the response. This simplification of test administration and scoring seems to retain much of that captured by the more complex Standard Issue Scoring method. This volume reports a correlation of .85 with the Standard method in an age-heterogeneous sample, and a correlation of .50 with age controlled. Given the test-retest reliability of both instruments (higher for the Standard Issue method), these correlations represent considerable concurrent validity of the SRM. To the extent that the correlation between the two measures is moderately good, the SRM can claim the basic validity (e.g., invariant sequentiality, structured wholeness) of the Standard Issue method of assessment. Eventually, factor-analytic and longitudinal studies by the authors should establish the construct validity of the measure in its own right.

While in general agreement with the rationale and usefulness of the SRM, I differ with John Gibbs on one major point, the existence and inclusion of a post-conventional fifth stage of moral judgment. Adapting from distinctions made by Gibbs (1977, 1979a), Kohlberg, Hewer, and Levine (Note 1) distinguish between "hard" Piagetian structural stages and "soft" qualitative stages, such as the Perry (1968), Loevinger (Loevinger & Wessler, 1970) and Fowler (1981) stages. Both "hard" and "soft" stages represent qualitative changes in development and define developmental sequences. In addition, "hard" stages represent: (1) new thought operations; (2) internalized schemata of action; (3) logically or philosophically formalizable systems in equilibrium; and (4) culturally universal patterns. Soft stages are new levels of reflection on

the self and the world, rather than new operations of thought and judgment. I believe that Stage 5 meets all the criteria of "hard stages"; Gibbs does not. Further research is clearly needed on this point. I personally would recommend, in using the SRM, that one treat Gibbs's "theory-defining level" (specifically, Theoretical Principles) responses as indicating a fifth stage.

In summary, this book and the measure it provides represent a useful introduction to moral stages and their measurement--even for those who will wish to go on to master Standard Issue Scoring for more intensive research purposes (requiring exactness in assigning stage scores to individual subjects in longitudinal and clinical work). For those who want a measure of group means, educational effects, and correlations with other variables, the greater efficiency of the SRM may make it the method of choice. It is a pleasure to recommend this volume to anyone interested in research method in moral development.

June, 1981
Cambridge, Massachusetts

PREFACE

This book is dedicated to Lawrence Kohlberg, one of the most influential developmental psychologists of our era. Although Kohlberg has made important contributions in areas such as sex-role development and early childhood education, his major work has been in moral development. Indeed, were it not for Kohlberg's work over the past several decades, it is doubtful that moral reasoning would have attained its distinct and major status in developmental psychology.

When, in the late 1950s and early 1960s, Lawrence Kohlberg undertook to study moral reasoning and conduct, his choice of topic made him something of an "odd duck" within American psychology . . . . Social scientists were not nostalgic for that period when morality seemed to be mainly a bludgeon for controlling sex and, possibly, swearing. No up-to-date social scientist, acquainted with psychoanalysis, behaviorism, and cultural anthropology, used such words at all. To appreciate what Kohlberg has done to deepen the intellectual interest in what is, after all, a very substantial aspect of human psychology, one must have some sense of the tide he swam against. Moral reasoning, as a process, was something of which behavioral scientists were at least professionally unaware. Certain aspects of the process entered into behavior research under such concepts as "attitude," "norm," "custom," and "value." But in a very muddled way that usually failed to take account of even the most elementary distinction of moral philosophy . . . . The study of attitudes, customs, norms, and values had the effect, in the past, of "trivializing" the subject of morality. In general, attitudes were just "pro" or "con" some object of social concern (a politician, a policy, an ideology); one toted up the pros and cons, rather like a preelection poll, and that was thought to be the end of it. Customs and norms were described as relatively concrete guides to behavior, and a culture was likely to seem a kind of bundle of these, like so many unrelated sticks tied together with a string. People were not often asked to think about cases in which norms were in conflict--as norms, in fact, often are--or to try to reason their way to a resolution of such conflict. When people are asked to deal with ethical dilemmas, as in Kohlberg's research, a complex and hard-to-understand process emerges. (Brown & Herrnstein, 1975, pp. 307-308)

Kohlberg found, then, that moral thinking is a complex process not reducible to the expression of moral attitudes, norms, or values. He has claimed that morality is not totally relative, that individual, social-class, and cultural differences in moral reasoning permit--upon sufficient probing--the discernment of common <u>structures</u>. Further, these structures are claimed to evolve in a standard sequence of developmental stages, with individual and cultural variation accounting mainly for differences in rate of development through the sequence. Kohlberg's theory of moral stages has attracted attention from extraordinarily diverse areas in the contemporary intellectual community. Kohlberg's Center for Moral Development and Education at Harvard has been visited not only by psychologists and educators, but also by philosophers, theologians, counselors, and others interested in ethics from numerous countries. Kohlberg's theory is controversial: in the social science literature, Kohlberg's claims have attracted at least as many adversaries as proponents (e.g., Kurtines & Greif, 1974, vs. Broughton, 1978).

Kohlberg's reaction to this theoretical controversy has generally consisted of pointing out that his claims are, after all, essentially empirical. While I (Gibbs) was a research associate and part of the Kohlberg group at the Center from 1975 to 1979, I frequently heard Kohlberg explain that his stages were more a matter of discovery than of invention, that moral reasoning structures could be discerned by anyone willing to take the time to investigate. Nonetheless, those of us at the Center who were helping to construct a systematic assessment manual based on Kohlberg's 20-year longitudinal moral judgment study were also mindful of the truth of another Kohlberg comment: namely, that social cognition does not emerge from subjects with moral stage tags already attached. Constructing a manual, in other words, was hard work.

That work has, however, provided the basis for an empirical answer to Kohlberg's critics. Blind scoring of that portion of the longitudinal data <u>not</u> used in the initial manual construction has yielded a clear picture of a stage-sequential development in moral reasoning (Kohlberg, 1981). Further, the manual itself (Colby, Kohlberg, Candee, Gibbs, Hewer, Kaufman, Power, & Speicher-Dubin, in press) has been found to provide a reliable and valid instrument for the stage assessment of moral judgment (Colby, Kohlberg, Gibbs, & Lieberman, in press).

As Kohlberg explains in the Foreword, this book provides a simplified and group-administerable equivalent to the Harvard assessment instrument. The prominent developmental psychologist James Rest was a research associate at the Harvard Center during the late 1960s. Over the course of his involvement as part of the Kohlberg group, Rest (1979) gained the impression that Kohlberg's primary aim in guiding manual construction "was not to put together a handy instrument, but to devise a theoretical system to represent the logic of moral thinking, analogous to Chomsky's work on syntactical structures" (p. xviii). I believe Rest's inference was--and continues to be--essentially correct. The Standard Issue Manual is a brilliant systematization of the content and structure of moral thought--but its intricacies limit the readiness and ease with which it can be used as a research tool. A "handy instrument" it isn't.

Rest recollects that, during his years at the Center, "alternative schemes for assessing moral judgment became a compelling interest" (p. xviii) for him. This "compelling interest" led Rest to develop his Defining Issues Test (DIT), which assesses moral judgment (specifically, moral evaluation) through a multiple-choice format. During my work at the Center in the late 1970s, I

came to be captured by the same "compelling interest" in the possibility of "alternative schemes" referred to by Rest. Yet my departure has not been as radical as Rest's. Rest (1975) has rejected interpretations of the DIT as simply a more "handy" way of indexing moral reasoning (the correlations are low anyway), and has instead championed the referent for the DIT--moral <u>evaluation</u>--as a form of moral judgment worthy of study in its own right.

Our Sociomoral Reflection Measure (SRM), in contrast, remains within the realm of the justification or reasoning sense of moral judgment. The SRM unabashedly rides piggyback on the parent Standard Issue manual. Like its parent, the SRM is a production-task measure of moral reasoning whereby subjects must express their thinking with respect to moral dilemmas and associated normative values. Whereas sociomoral norms constitute only one classificatory feature of the Harvard manual, however, norms become the crucial feature defining the questionnaire and scoring format used in the SRM. Systematic use of the norms has enabled us to amplify the strength of the dilemma questions and to simplify the procedures for stage assessment of the responses (see Chapter 2). These innovations have in turn made possible the convenience of group administration (rather than individual interviewing) for data collection and of self-training (rather than workshop participation) for learning reliable stage assessment.

The extensive assistance of the second author, Keith Widaman, in the construction and validation of the SRM was, of course, crucial. As I was making a career move from Harvard to Ohio State in 1979, Keith (then a graduate student at Ohio State) introduced himself and expressed his wish to work with me in the moral development area. I was soon delighted to discover how well my project was supported not ony by Keith's quantitative expertise, but as well by his ability to make keen inferences as to the qualitative meanings of sociomoral responses. Keith's highly successful completion of Kohlberg's June 1979 Moral Judgment Scoring Workshop provided the foundation for Keith's substantive participation in the new reference manual construction, in preliminary work on scorability and other assessment rules or procedures, and in the training of our vanguard group of SRM raters: Helen Ahlborn, Kevin Arnold, and Miriam Galevi (these raters are also thanked for their subsequent contributions to the project). Keith's participation was also major in the data collection phase, and in the psychometric analyses of the new instrument's properties. His contributions to the SRM project continued during 1980 and early 1981 despite his busy first year as assistant professor at the University of California, Riverside.

Many persons (and in some cases, the institutions they represent) made crucial contributions to the success of our instrumentation project. Especially crucial on a practical level was the funding for the project provided by the Small Grant Program of the National Institute of Mental Health. Larry Kohlberg consistently encouraged this work from its very start in January of 1979 and provided an invaluable early critique of the manuscript draft. Anne Colby (Radcliffe), who is the foremost authority on Kohlberg's theory and stage constructs (apart from Kohlberg himself), served as a consultant on the project. The importance of her contributions to the sharpening of the reference criteria is reflected in her status as collaborator. Marvin Berkowitz (now at Marquette University) conducted for us the first psychometric study of the SRM, specifically, of its concurrent validity with the Harvard instrument. We owe considerable thanks to Ted Fenton (Carnegie-Mellon University) for his early belief in the SRM and his inclusion of it as one of the evaluative measures used in connection with his Civic Education Project. Various personnel

associated with the Ohio Youth Commission, the Columbus Public Schools, and the Ohio State University also provided crucial assistance in other aspects of the data collection. We thank Janina Jolley, Scott Mullarky, Catherine O'Connor, Mark Tappan, and Bedonna Weiss for their diligent work as self-trainees. Most recently, we thank Phil Clark (Ohio State), Bill Damon (Clark University), and Fred Damarin (Ohio State) for their constructive comments regarding portions of the manuscript and Mark Tappan (again) and Kevin Arnold for extensive proofreading assistance. Our gratitude also goes to our families for their unfailing sympathy and support throughout this project.

The "reference manual" in this book is comprised of Chapters 5 through 12. These chapters provide the criteria to which the scorer refers in making developmental assessments of the responses provided by subjects on the SRM questionnaire. Other portions of the book should make possible the informed and effective use of the reference manual. Chapter 1 introduces the developmental approach to the assessment of intelligence by contrasting it with the individual differences approach, and then makes a corresponding--and more specifically pertinent--contrast within the framework of social intelligence. In Chapter 2, we elaborate on the developmental approach to the measurement of social intelligence, specifically with reference to the measurement of moral judgment (or what we call sociomoral reflection). Chapter 2 concludes by overviewing our psychometric evaluations of the SRM. Chapter 3 discusses at length the sociomoral stages themselves and their relation to situational action as well as to the concept of maturity. The discussion in this chapter represents a consolidation of the revised view of the stages previously articulated by Gibbs (1977, 1979a). Chapter 4 provides guidance for self-training and for scoring by reference to Chapters 5-12. Appendices B and C provide the materials for the self-training program outlined in Chapter 4; Appendix A contains the SRM questionnaire itself (Forms A and B), plus the standard rating form. With its arious parts, this book should be sufficient to make possible research use of the Sociomoral Reflection Measure.

September, 1981
Columbus, Ohio

SOCIAL
INTELLIGENCE
Measuring the Development
of Sociomoral Reflection

CONTENTS

1

TWO APPROACHES

TO

INTELLIGENCE

The precise focus of this book is on the measurement of reflective sociomoral thought. Since sociomoral reflection is a specified form of social intelligence, however, our work more fundamentally relates to intelligence. Elkind (1969) has compared and contrasted the "psychometric" with the "Piagetian" approach to the conceptualization and measurement of intelligence. It is important for an understanding of our assessment work to articulate this distinction. The terms we will use for the two approaches, however, are "individual difference" (or differential) and "developmental": psychometric techniques are not the exclusive possession of the former approach, nor is Piaget the only theorist important in the latter. Thanks to works by Elkind and others (Cowan, 1978; DeVries, 1974; Furth, 1973; Stephens, McLaughlin, Miller, & Glass, 1972; Tuddenham, 1971), psychologists have come to grasp the fundamentally distinct conceptualization of intelligence provided by the developmental approach; yet the point has been made primarily with respect to nonsocial intelligence. A parallel point remains to be grasped with respect to social intelligence. In this chapter we will articulate our developmental approach to reflective sociomoral thought against the backdrop of a discussion of the individual-differences approach.

"I.Q." is a quintessential symbol of the individual differences approach. Tests such as the Stanford-Binet provide an "Intelligence Quotient" (I.Q.), which represents the intellectual brightness, power, or agility of an individual relative to that of his or her same-aged peers. Assessment of such individual differences is based on the relative numbers of items correctly answered on a test typically containing a large number of questions tapping simple as well as complex cognitive abilities. These items derive from diverse sources, and the criterion for their inclusion is essentially empirical: those questions which are correctly answered by a gradually increasing percentage of persons in successively older age brackets are retained as test items --and those which show no such age trends are eliminated. As Tuddenham (1971) observed, "Current tests consist of items chosen more for their statistical properties than for their content" (p. 65). Large cross-sectional samples are then typically used to establish the "normal" or average percentage of items passed at specified ages. An individual's I.Q. is defined and evaluated as bright or dull relative to these age norms. Since one's degree of intellectual brightness is considered by many differential theorists to be a stable and possibly inherited trait, one's relative position as one gets older is expected to remain fairly constant. Hence, although intelligence tests do entail use of age-related (and therefore to some extent developmentally related)

differences, the aim in doing so is to establish age-constant individual differences.

A psychometrically acceptable intelligence test--or any psychometrically serious test--must meet certain standards, chiefly those pertaining to reliability and validity. The test must be demonstrated to be a valid measure of that which it purports to measure, e.g., by correlating highly with an already accepted measure of the variable or by appropriately discriminating among persons who can already be described on other grounds as possessing high or low levels of that variable. The test must also show some stability or reliability, e.g., by correlating highly with the results of a second testing administration or by showing good consistency across its component parts. Reliability is thought to be promoted by standard procedures of test administration, perhaps even by using tape-recorded instructions to guard against subtle variations.

The developmental approach is not concerned with intellectual brightness so much as with intellectual maturity. An individual who can infer or work out the correct answers to a relatively large quantity of miscellaneous problems may be bright but is not necessarily intellectually mature. Criteria for maturity must be derived from studies of normal human cognitive development, with an eye to identifying the basic and maximally generalizable patterns of the individual's mental evolution--in other words, to identifying the character of intelligence at different ages. Although complexity is not necessarily to be equated with maturity, there is a theoretical assumption that development from birth to maturity fundamentally entails progressive differentiations and integrations in, for example, thought processes. This assumption is common to Piaget's cognitive-structural perspective as well as Werner's organismic perspective, and for that matter can be found in the nineteenth-century writings of J. M. Baldwin. For all these theorists, the experiential differentiations and integrations of intelligence represent active and responsive modes of adaptation which extend beyond relatively "closed" processes of instinctual and other regulations (see especially Piaget, 1969/1971).

Further, there is an assumption that these differentiations and integrations lead to a sequence of hierarchically arranged cognitive structures, or "stages," that has maturity as a normal end-state. Indeed, the structuralist view--which we adopt in our work--is that one cannot even properly speak of the development of a psychological variable unless a progressive organizational complexity (usually entailing a hierarchical sequence) can be discerned. (This does not mean that developmental psychologists necessarily should confine their interests to "development" defined in this narrow sense. We accept Wohlwill's 1973 position that any robust age-related change in a psychological function over time is fair game for the developmental psychologist--although we do feel that those functions of greatest significance for the developmentalist are those which undergo an evolving complexity.) In this strict sense of development, an age-related increase in the percentage of correct problem solutions (as reflected in mental age) is viewed as epiphenomenal to developmental processes--not as development itself.

The development of progressively mature mental structures cannot be discerned at the surface level of problem-solving solutions; it must instead be studied at the level of the processes which underlie and generate those solutions. For example, psycholinguists find transformational patterns not at the surface level of literal syntactical combinations but rather in terms of the

deep or generic functional level of language constructions and operations. To investigate the cognitive dynamics which underlie a task performance, the developmental researcher may probe by introducing task variations and extensions at opportune points during the subject's task activity. When one is working with verbally fluent subjects, it is helpful to elicit justifications or explanations of their various task decisions. These justifications often provide significant indications of how the subject is approaching and thinking about the problem. In this phase of the developmental approach, the completely "standard" procedures associated with the differential approach would actually block rather than facilitate research progress. As Tuddenham (1971) noted, rigidly standardized procedures "would never have provided the insights which have led Piaget, Inhelder, and their co-workers to their theoretical formulations" (p. 74). Although the task format must be standard enough to allow for comparisons across subjects (particularly comparably aged subjects), the experimenter must be procedurally flexible and alert to unexpected features of the subject's reasoning that may yield new structural discoveries. As Cowan (1978) argues, "standardized" does not necessarily mean "scientific." It is an instructive point that some of the most replicable findings in developmental psychology have come not from perfectly standardized investigations but instead from Piaget's nonstandard but penetratingly flexible and inquisitive méthode clinique.

Individualized inquiry need not characterize the full saga of developmental research, however. In the context of moral development, Kohlberg (1979) has suggested that the trajectory of developmental research entails three phases. In the first phase, the emphasis is on exploration. One works with cross-sectional, longitudinal, and cross-cultural samples to identify the "broad outlines" (p. x) of sequential, structural development, typically using the méthode clinique for this purpose. Eventually, however, one becomes reasonably confident that the "transcontextual validity" (Weisz, 1978) of the basic sequence has been sufficiently explored, and one perceives diminishing returns from further exploratory investigations. At this point, a second, primarily methodological phase becomes appropriate, and emphasis shifts from exploration to assessment. Thanks to one's phase-one work, one can bring to the assessment phase a considerable knowledgeability of the relevant structural sequences and of the most efficacious techniques for eliciting them. With the phase-one experience as a foundation, then, effective assessment methodology can be developed. Standard procedures can now be introduced with reasonable confidence that their use will not preclude the discovery of important new structures. Indeed, in the second phase it is entirely appropriate, scientific, and beneficial to develop systematically structured and standard tests and to apply the traditional psychometric criteria of validity and reliability to those tests.

Kohlberg then sees a third phase, not yet fully attained in moral development, whereby the assessment advances achieved in the second phase enable one to return to the basic theoretical questions not fully answered in the first phase (e.g., questions as to the developmental relationship of sociomoral intelligence to other relevant variables or as to the degree of unitariness to the stages) with one's new, scientifically powerful assessment tools at one's disposal. It is to the advent of this third phase that we trust the assessment tool presented in this book will contribute.

The distinctions noted between the differential and the developmental approaches can be elaborated specifically with reference to social intelligence

(cf. Greenspan, 1979). As Walker and Foley (1973) note, most investigators of social intelligence have from the start "approached their problems with an individual difference orientation" and with a "definite interest in psychometrics" (p. 840). The psychology of social intelligence has a research history, beginning with Thorndike's (1920) conceptualization of the construct as "the ability to understand and manage men and women, boys and girls--to act wisely in human relations" (p. 228). Although Thorndike viewed social intelligence as an ability distinct from those of "abstract" and "mechanical" intelligence, that view has become controversial. Guilford (1967) shares the Thorndike view, reserving for social intelligence distinct factor combinations in his complex multifactorial theory of human intelligence. Wechsler (1958), on the other hand, although using the term "social intelligence," views it as merely the application of general intelligence to social contexts. This controversy continues to be reflected in recent literature (e.g., Keating, 1978; Osipow & Walsh, 1973). Most researchers on both sides, however, have conceptualized social intelligence in individual-difference rather than in developmental terms.

Among the most widely used tests of social intelligence are those innovated by Guilford and associates. Guilford's social intelligence tests include tasks eliciting performances such as: choosing the most appropriate interpersonal episode; identifying the drawing which does not belong in a set of drawings of socioemotional facial expressions, gestures, and postures; and identifying the type of interpersonal situation in which a given verbal statement will have appropriate meaning. On the basis of the age-normative data compiled for these tests (O'Sullivan & Guilford, 1975), one can assess an individual's degree of interpersonal brightness, in much the same spirit as one determines an individual's I.Q.

Although most investigators have been concerned with social intelligence as a fairly stable trait or facility, some have been interested in social intelligence as an evolving, progressively mature capacity. These researchers have typically used more open-ended techniques, consistent with the first-phase concern in the developmental approach with flexible inquiry into more spontaneous modes of behavior that may reveal features of organization or structure. Piaget's chief method in his early work (1932/1965) on the topic was to use children's comparative evaluations of moral concepts such as naughtiness or fairness as a vehicle for probing their reasoning. After a subject made a comparative evaluation with respect to a given pair of stories differing in certain morally relevant respects, the subject would then be asked to explain or justify that evaluation. Feffer and colleagues (e.g., Feffer & Suchotliff, 1966) studied social cognitive capacities at diferent ages by assessing whether subjects retain thematic coherence as they retell TAT-inspired stories from the viewpoint of each of the story protagonists. Similarly, Flavell, Botkin, Fry, Wright, and Jarvis (1968) used adaptability in story re-telling performance as a method for studying the development of role-taking and communication skills. Kohlberg (1969), followed by Selman (1980) and Damon (1979), retained and intensified the Piagetian focus on justification by presenting moral or social problems and then probing the ways in which subjects justify their decisions and evaluations with respect to their proposed solutions to the problems.

All of the above researchers have used cross-sectional (and in Kohlberg's case, longitudinal and cross-cultural) samples for studying social intelligence. This commonality is not accidental, of course, since by definition the developmentalist is concerned with the ways in which a