Theory and Research in Social Education

Statement of Purposes and Style for Manuscripts

Theory and Research in Social Education is designed to stimulate and communicate systematic research and thinking in social education. The purpose is to foster the creation and exchange of ideas and research findings that will expand knowledge about purposes, conditions, and effects of schooling and education about society and social relations.

Conceptualizations and research from all of the social sciences, philosophy, history and the arts are needed in clarifying thinking and practice in social education. Manuscripts are welcomed on topics such as those that follow:

- Purposes of social education;
- Models, theories, and related frameworks concerning the development, diffusion, and adoption of curricular materials;
- Instructional strategies;
- The relation of the social sciences, philosophy, history and/or the arts to social education;
- The politics, economics, sociology, social psychology, psychology, anthropology, philosophy, and/or the history of social education;
- Alternative social organizations and utilizations of the school for social education;
- Comparative studies of alternative models of social education;
- Models of and research on alternative schemas for student participation and social action;
- Relationship of different pre- and in-service patterns of teacher training to social education;
- Models of the utilization of objectives in social education and related research findings;
- Implications of learning theory, child development research, socialization and political socialization research for the purposes and practice of social education;
- The relationship of different independent, explanatory variables to educational achievements in the area of learning about society and social relations;
- The social organization, climate, cohesion of schools and other school characteristics as independent, explanatory variables predicting to general educational achievement.

Form for Submission of Manuscripts

In order to facilitate the processing of manuscripts, authors are asked to follow the procedures noted below:

1. Manuscripts should be typed with a dark black ribbon, clearly mimeographed, or multilithed. Authors should avoid submitting ditto copies of articles unless clearly legible. Some corrections in dark ink will be accepted. Copies containing numerous corrections will be returned for retyping.

2. Four copies of each manuscript should be submitted. This will speed up the reviewing process and guard against loss of manuscripts.
3. Everything should be double-spaced including footnotes and references.

4. Since manuscripts will be sent out anonymously for reviewing and due to the fact that the abstracts will be published, the author’s name and affiliations along with an abstract of approximately 150 words in length not exceeding 200 words should appear on a separate covering page. Information identifying the author, position, and institutional affiliation should appear on a separate page.

5. No responsibility is assumed for loss or injury to manuscripts submitted for publication.

Manuscript Style

1. When citations are made, the author’s name, publication date, and page (where necessary) should be enclosed in parentheses and located directly in the text. The complete reference will be included in a “References” section at the end of the article. For example, “Another problem arises if inductive methods are used to teach a generalization. The generalization may be reified, treated as a fact, when all generalizations, empirical or theoretical, are, as Popper argues, only corroborated for the time being (Popper, 1959).”

2. Do not cite references by means of footnotes.

3. Only substantive footnotes should be sequentially numbered within the text and located at the end of the manuscript.

4. References should be alphabetized and located at the end of the manuscript. They should take one of the following forms:


5. Each table should be placed on a separate page and placed in a separate section at the end of the manuscript. Arabic numbers should be used for numbering tables; they should be numbered consecutively throughout the manuscript. Show where they belong in the text by the following note:

Table One About Here

6. Figures should be submitted in their final form. Use India ink and place them on separate pages in a separate section at the end of the manuscript. Number them and locate them in the text in the same way as tables.

7. Send Manuscripts To:

Professor Lee Ehman, Editor
Theory and Research in Social Education
School of Education 341
Indiana University
Bloomington, Indiana 47401
CONTENTS

In Memorial

Articles

Foshay with Burton

Citizenship as the Aim of the Social Studies 1

Larkins and Oldham

Patterns of Racial Separation in a Desegregated High School 23

Napier

The Ability of Elementary School Teachers to State Moral Thought Statements 39

Angrist, Mickelsen and Penna

Development and Evaluation of Family Life Courses 57

Allen

Children's Associations with Their Own and Other Countries 80

Theory and Research in Social Education is the official journal of the College and University Faculty Assembly of the National Council for the Social Studies. It is a general review open to all social studies educators, social scientists, historians, and philosophers. A general statement of purposes and style for manuscripts may be found on the inside front and back covers.

A subscription to Theory and Research in Social Education may be obtained by membership in the College and University Faculty Assembly of the National Council for the Social Studies. Back issues may be obtained for $3.50 each and institutional subscriptions are $7.50 per year. Write the editor for these orders.

Copyright 1976 by the College and University Faculty Assembly of the National Council for the Social Studies. All rights reserved.
Officers of the National Council for the Social Studies, 1976-1977:

*President:* Howard Mehlinger
Indiana University

*President-Elect:* Anna Ochoa
Indiana University

*Publications Board Chairperson:* Charles L. Mitsakos
Chelmsford (Massachusetts) Public Schools

**Reviewers for this issue:**

Mel Arnoff
Robert Barr
Shirley Engle
Karen Fox
Richard Jantz
Benita Jorkasky
Robin McKeown
Jack Nelson
William Norris
Anna Ochoa
Thomas Popkewitz
Donald Schnieder
Jan Tucker
Mary Jane Turner
In Memorial

JOHN R. LEE

JONATHON C. McLENDON

Lovely and pleasant in their lives; and in their death they were not divided; they were swifter than eagles, they were stronger than lions.
CITIZENSHIP AS THE AIM OF THE SOCIAL STUDIES

Arthur W. Foshay
Assisted by
William W. Burton
Teachers College
Columbia University

"...Social scientists are faced by an enormous and growing number of students who don't really want to be social scientists or theorists; they want to be practical; they want to understand our system so that they can work in it, change it, improve it. These two aspirations—the scientific, theoretical, scholarly interest, and the real practical interest of many of the students—are in implicit conflict with each other."

—Gerald F. Eise, in Humanities, Sept. 1972

There are three main themes offered here. First, we argue that the Social Studies needs a single governing concept—citizenship—to give it programmatic meaning, or coherence. Second, we examine certain widely used secondary school texts for their treatment of citizenship, and find them unsatisfactory. We also examine some "new Social Studies" materials and find them more promising. Third, we make broad suggestions for a reconstituted Social Studies program, organized around the concept of citizenship.

Any instructional program must meet at least the following criteria if it is to have meaning:

1. It must be coherent. It must be organized around some theoretically sound set of ideas or intentions.
2. It must be systematic. One part of the program must lead to another in some logical way.
3. It must be appropriate for students. It must be within their life-space as they develop in maturity, understanding, and knowledge. It must also be appropriate for the society within which it is offered—consistent with societal beliefs and customs.
4. It must be teachable. It must be within the grasp of some conceivable teaching staff, and it must be thought of in instructional terms.
5. It must be important in a human sense. It must deal with some significant aspect of the human condition.
6. It must be focused. It should be possible to distinguish its aims from those of education as a whole.

The Social Studies meet some of these criteria better than others. One can
argue that the various learning activities typically included are teachable, often appropriate for students, and usually socially acceptable. It would be much harder to argue that the program as a whole is coherent or focused, that it is systematic, or that it seeks specifically to deal with Man the social being (taken as an aspect of the human condition).

These difficulties arise from the heritage of the field. This relatively young field has History and Geography as its ancestors. During the Nineteenth Century, the schools had offered a somewhat chauvinistic American History, some Ancient History (a field left over from the Renaissance and Eighteenth Century), and Geography—principally the memorization of the locations of the land masses and the political boundaries. The field was academic, by the lights of those times. That is, it took erudition to be knowledge, and treated the student as a passive recipient of information.

The field continues to be academic in this sense, in the main. American History continues to celebrate the American Achievement; the Civics course presents the American governmental forms; Economics, sometimes taught, seeks to be true to its parent discipline. "World History," which is not a discipline in its own right, is typically European History with a bow toward the Middle East and Asia. The only part of the usual high school Social Studies offering that does not reflect this view of the "academic" is Problems of Democracy, which more or less deals with current affairs and current events.

Well, one asks, what of that? Could not these courses be well done, and a sound Social Studies program be offered at the secondary level? The difficulty has been suggested by Prof. Eise, quoted above. Students cannot but consider such courses, even when well done, as "merely academic." To say this is not to damn the academic; at its best, it offers ways of grasping reality that rise far above common sense and the school of hard knocks. But to be informed is not to practice grasping the unstructured reality of daily life. Information is necessary, but not sufficient.

There is another difficulty with the Social Studies offering as it exists at present: it is not a program. It is a more or less consistent array of subjects, each separate from the others, all dealing in their various ways with Man the Social Creature, but disunited. The difficulty is that there is no unifying aim to the offering; it has no comprehensive purpose; it is scattered. There are as many aims as there are subjects.

This whole matter has become urgent in our times. Our disadvantaged populations have found their voices. They respond, often unwittingly, to antique public ideas: anarchy, oligarchy, aristocracy. What they do not respond to is American Democracy. Anyone who listened to the college screamers of the Sixties must have wondered how they could have become so alienated from all those American institutions we thought they had learned about in school—those institutions which, if imperfect in specific
areas, could nevertheless be improved by well-tested procedures.

It is the best educated portion of the population that has expressed the deepest suspicion of our institutions. In their various ways, they say the institutions are unjust; that our expressed ideals are violated so often and so widely that (some say) the whole nation is hypocritical.

As a matter of tradition, we have expected the schools to bring about affiliation with the nation. In school, we have supposed, children learn the basic ideals and traditions—and legends and myths—of the country. To know these, we have supposed, is to love them. What a shock to find that this is not so! What has been happening?

What has been happening has been, in large measure, the same old thing. What we have to get used to is that children have many sources of information now, not only one, concerning the Republic and its affairs. If what they read and hear in school doesn't square with what they see and hear in the mass media, they believe the media, not the textbook. If both sources are biased, so are the children.

The schools have always been supposed to build citizenship among our youth. The Founding Fathers intended it so when they encouraged universal education, and we have supposed that citizenship was being fostered by the schools, and especially by the Social Studies.

We propose here that this old aim be examined afresh. Where the Social Studies fall short of its demands, we propose that the field be reformed. We have implied that the field of the Social Studies suffers from two technical shortcomings now: it is too narrowly academic, and it has no coherent purpose. If, in addition to this, the field fails to promote citizenship, we have three fundamental criticisms to make of it.

CITIZENSHIP AS THE AIM OF THE SOCIAL STUDIES

The term, "citizenship," has been bent and tortured through our history until it has almost lost its meaning. At various times, it has meant patriotism, conformity with the business ethic, abiding by the law, supporting the New Deal, and participating in war. Given this lineage, it is not surprising that some would abandon the term altogether, perhaps by reducing it to a technical maneuver—making it into behavioral objectives that avoid it.

The term has often been replaced by such technical maneuvers. Any curriculum plan, as is often pointed out in educational literature, has to respond to three forces: the student, society, and the nature of the knowledge to be learned. During its earliest period, the Social Studies responded to the apparent demands of its fields of knowledge, History and Geography, almost to the exclusion of the demands of child development. In doing this, it merely extended its Nineteenth Century heritage, which had been subject-centered, though (as has been seen) responsive to the needs of
society as interpreted at the time. The Progressive Educationists, in rebellion against the subject-centered curriculum of the early Twentieth Century, devised systems that called for overt activity by the students. They also called for the Social Studies to be responsive to the social needs of the day. The "Core Curriculum" grew out of these responses to the student and society, partial as the responses were. As Hertzberg points out, the Social Studies have alternated between a Core approach and a subject-centered approach. At present, the Core is in a decline; the curriculum development of the Fifties and Sixties is subject-centered, though with a considerable reliance on what little is known about how students grow into such knowledge. All these moves are technical in character.

Citizenship, we have said traditionally, is the main aim of schooling in the United States. Certainly, we say here, it should be the main aim of the Social Studies. What are the prospects that the term can overcome its past, and compete successfully with the technical maneuvers and the academic tradition of the field? Can this amorphous field be brought together around such a term?

Popularly, the term "citizenship" refers to feelings of affiliation with the country—loyalty, patriotism—and also to the disposition to take an active part in governmental affairs, at least through voting. On what basis shall affiliation be built? Why should people take part in government? Is participation an expression of loyalty? These questions are no longer rhetorical. As we have all seen, there are many people who do not feel affiliated, and therefore take no part. To know us is not necessarily to love us. Too many people have, as we say, lost the faith. They do not trust the established processes to be just.

There is the key term—justice. Fortunately for us, John Rawls has put together the fruits of ten years' work in his *A Theory of Justice*, and we are in a position now to reconsider our term "citizenship" because of his efforts. Influenced by his ideas, we can offer a definition of citizenship powerful enough, we think, to sustain a Social Studies program in the schools.

Here is the definition, based on Rawls: *citizenship is all those activities that seek just relations between individuals and social institutions.* Such a definition suggests what should be emphasized, and what subordinated, in the social studies program. It emphasizes the need for knowledge of our institutions, including their evolution from the deep past. It emphasizes, too, the need for direct practice in making and modification of social institutions of all kinds, including the mini-institutions within schools as well as the larger institutions in the world.

Rawls' book is close-knit and carefully constructed. Because of this quality, we can sketch the argument by quoting from its opening, for he builds the entire argument on these foundations. Institutions are to be assessed according to their justice, he says:
Justice is the first virtue of social institutions, as truth is of systems of thought...laws and institutions no matter how efficient and well-arranged must be reformed or abolished if they are unjust. (p. 3)

Among individuals with disparate aims and purposes a shared conception of justice establishes the bonds of civic friendship; the general desire for justice limits the pursuit of other ends. One may think of public conception of justice as constituting the fundamental charter of a well-ordered human association. (p. 5)

...the primary subject of justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation. (p. 7)

Any notion of democracy is based on a notion of justice; any notion of freedom includes justice as an assumption. Social institutions are viewed, according to this definition of citizenship, as processes; the principal skills to be developed among students are those necessary to the analysis and criticism of such processes, according to the justice of their consequences.

To leave it at analysis, however, would be insufficient; we would merely have shifted to a new subject-centeredness. A citizen is one who participates in social decision-making and therefore in social change. A program that sought such participation as part of its major purpose would perforce offer direct, first-hand practice in such decision-making and social change. Students would not stop with book knowledge; they would be drawn by degrees into full participation in the world beyond the school, to the end that their first vote, at age eighteen, would be a climactic event in their education—an event placed in a setting of active political and social participation, as well as knowledge.

A Social Studies program so conceived would, perhaps, contribute to the remedy for some of our contemporary social ills: the disposition to see the Establishment as remote from one's self; the disposition to resort to violent action; the naive failure to count votes that characterizes some youthful political groups; in general, the blind thrashing-out that our young people indulge in.

The need for such a conception can be argued on other groups, as well. The regulatory agencies in the federal government require constant monitoring and supervision by the populace. In order for the system to work, one must know it in depth. For example, the coal industry has been ridden by tragedy for generations; the miners don't know in depth how the system works, and they die. In 1950, Centralia No. 5 explored. As John Bartlow Martin pointed out in Harper’s at the time, the men knew the mine was dusty and dangerous, and had appealed to their local union, the national union, the federal Bureau of Mines, and the Governor of Illinois,
all without effect. They thought they had exhausted the possibilities. They hadn’t, of course. They could also have appealed to the political opposition, and to the nearby newspaper, the St. Louis Post Dispatch. They didn’t, though, and the results were lethal. One might ask, could they have learned what to do in school? What is required is that the whole system be learned. Those who rail against the system don’t usually know it in depth. If we are to reduce the amount of fatuous rhetoric about our political system, we had better be about the business of mass education. And the most available place to undertake this remedial action is in the Social Studies.

Criteria to be Applied to Social Programs

Two substantive criteria arise from the foregoing. If a Social Studies Program is to build citizenship among students, the following questions must be answered affirmatively:

1. Is the program organized around a conception of justice—citizenship—broad enough to be of equal value to all people?
2. Does the program provide for direct practice of citizenship?

The first of these has already been argued, above. The second grows both from the requirements of the first, and also from some elementary principles of pedagogy.

It is elementary that any effective learning must meet four requirements, else learning does not take place. The student must:

1. Want something (drive)
2. Notice something (cue)
3. Do something (response)
4. Get something (reward)

The wanting, noticing, doing, and getting are all bound up in action. Many academic programs, including most Social Studies programs, are short of action, emphasizing only noticing (i.e., didactic instruction, the presented subject-matter), and getting something (a surrogate for real reward—the course grade). Kilpatrick, following Dewey, used to say, “you learn your response. That is what you learn, and that is all you learn.” Our second criterion asks that explicit provision be made for students in the Social Studies to practice the desired response—i.e., that they undertake direct practice of citizenship.

THE EXISTING PROGRAMS

Programs in the Social Studies

It is impossible to attend to all the “existing programs.” At the present
rate of production, two men could not hope to keep pace, let alone catch up. But any selection implies assumptions which must be made explicit, if the critique is to be rational.

Hertzberg, in attempting to make sense of the recent history of Social Studies, has suggested that there are two basic models of reform in the field.

Source study represents one type, the core curriculum another. The first is oriented to the disciplines, to cognitive skills, to the acquisition of knowledge, and to an alliance with the liberal arts colleges and universities. It sees in the student a future academic scholar. It is only tangentially concerned with the present, with affect, with social problems and social reform. The other is oriented to the fusion or the disregard of disciplines, to affective skills, to establishing a connection between the world inside and outside the school, and to an alliance with the schools of education. It sees in the student the future good citizen and—often—social reformer. (p. 20, "Historical Parallels for the Sixties and Seventies..."

An inspection of recent proposals or a review of Social Education for the last ten years tends to support the Hertzberg thesis. However, we should remember that she is discussing models of reform, not citizenship in the sense we have discussed above.

Hertzberg also notes that, at its height, the Core movement was reported in about 11 per cent of public junior and senior high schools enrolling over 500 students, and that the "overwhelming majority" of the programs were in the junior high schools. Further, she suggests that the new Social Studies movement of the 60's is only recently beginning to have significant effect on secondary schools, and that a counter movement seems to be gathering strength.

We suggest that reform models exist as a ripple on the ocean of traditional Social Studies—History, Civics, World History, and Problems of Democracy.

An executive from a major publishing company provides support for this conclusion. According to him, the most widely used texts in the Social Studies at present are the following:

8th grade: Henry F. Graff, The Free and the Brave, Rand McNally
9th grade: Frank Magruder, American Government, Allyn and Bacon
10th grade: Leften Stavrianos et al., A Global History of Man, Allyn and Bacon
11th grade: Henry Bragdon and Samuel McCutchen, History of a Free People, Macmillan Co.
The all-time best seller is Magruder, which has dominated its market since prior to World War II. The odds are that the readers of this paper themselves read Magruder, as did the senior writer in the 20's.

The Holt, Rinehart program authored by Edwin Fenton et al. for grades 9-12 is reported to be dropping off in sales since it first appeared, but it is still one of the best sellers of a reform nature. Other new materials, with good but not overwhelming distribution, are the Houghton Mifflin Law in America program, Stavrianos et al. A Global History of Man (Allyn and Bacon), Oliver and Newman, Public Issues Series, and Katz, Eyewitness: The Negro in American History.

These lists are open to question. We have been unable to obtain hard information concerning the distribution of all materials.

We shall examine Graff, Magruder, Bragdon and McCutchen, and Todd and Curti as "traditional" materials; Fenton's program as a "discipline-oriented" program, and Katz, Oliver, and Newman as "problems-oriented" programs. Though the categories are not mutually exclusive, and the selections perhaps too narrow, the attempt seems preferable to a critique of some idealized abstraction.

The "Traditional" Materials

The materials selected for examination here seem to resist our two main criteria. Without exception, the texts examined fail to list "justice" in their indexes. "Citizenship" fares but little better. The Todd and Curti text lists one entry for citizenship, which refers us to the definition of citizenship given in the Fourteenth Amendment to the Constitution:

Section 1. Citizenship defined. All persons born or naturalized in the United States and subject to the jurisdiction thereof, are citizens of the United States and of the state wherein they reside.

Important as is the Fourteenth Amendment, it defines nothing, from the point of view taken here.

The Magruder text deals with citizenship at somewhat greater length, but with the same position. The definition of citizenship contained in the founding documents is explained legalistically.

If we cannot pose our first question directly to the materials, perhaps we can intuit answers from the materials' manifest intentions. The traditional program (grades 8-12) includes two years of American History, one year of American Government, and one year of World History, plus Problems of Democracy. Other options offered include geography, sociology, and social psychology in the form of a course on Marriage and the Family. The minimum requirement is usually two years of American History and one year of American Government.

Only courses dealing with the United States are required. Even if a
five-year program of Social Studies is assumed, only one year need include concern with any social organization other than ours—the year of "World History." Geography, unless it is Political Geography, does not deal with our main questions. Problems of Democracy, a frequent elective, is really Problems of American Democracy. In fact, the usual electives—Anthropology, Sociology, and Marriage and Family Life—have a strong nationalistic coloration, and they tend to gloss over the problems we in fact face.

The most pervasive example of glossing over is the notion of "The United States as a Free Society." Two of the texts proclaim this belief in their titles: *The Free and the Brave* and *History of a Free People*. All the texts make frequent use of the word "free"—free world, free people, free society, etc. But the term "freedom" is rarely if ever confronted as if it were a complex concept. It is used vaguely; it is often "defined" implicitly as the opposite of the communist or fascist "way of life." Graff and Todd and Curti fail to list "free" or "freedom" in their indexes. Magruder and Bragdon and McCutchen refer to the Constitution, the Bill of Rights, and a 1944 speech by Judge Learned Hand. The effect of the materials is to project a notion of binary freedom—either off or on, especially through comparisons of the free world and the communist bloc. There is stout support of the Free United States, with little or no attention to the freedoms that are or are not enjoyed, and by whom.

There is a kind of absolutism, or perhaps oversimplification, in these materials that throws into question their academic respectability, though they seek academic accomplishment as their principal goal. Communism is treated as an absolute evil, and by implication the United States is treated as an absolute good. The justice of the systems is not treated. Consider this, from Bragdon and McCutchen: "Communism is more dangerous than fascism because its expressed ideals are higher." Or, "The strength of communism rests not merely on Russian armies and the attractions of the Communist message, but also on Communist agents all over the world. Unquestioning party members, and their dupes, are found everywhere." As if political propaganda were carried on only by the Communists! As if propaganda, or persuasion, were somehow evil!

Magruder is guilty of carelessness:

*The Dictatorship of the Proletariat.* The communists do not foresee a proletariat able to govern themselves after a revolution. Rather, the proletariat would need "guidance and education"—from the communist party, of course. Hence, the dogma calls for a *dictatorship of the proletariat*, a totalitarian regime to lead the people to the theoretical goal of communism: a "free classless society."

This passage overlooks the debate concerning what Marx meant by the dictatorship of the proletariat; this formulation is at variance with most of
Marx’s interpreters, and almost exactly at variance with what Marx himself said. Or let us consider the following quotation from de Tocqueville:

There are...two great nations...which seem to tend toward the same end, although they started from different points: I allude to the Russians and the Americans...The American struggles against the natural obstacles...the adversaries of the Russian are men...the conquests of the one are therefore gained by the plowshare; those of the other by the sword...the principal instrument of the former is freedom; of the latter servitude...

De Tocqueville wrote, of course, before the Civil War or the final slaughter of the Indians. But Magruder (or the editors who followed him) obviously liked the mischievous good-evil texture of the quotation. We may agree that no adversary has a monopoly of good or evil; certainly our children should not be hampered by such flat dichotomizing.

Another pervasive theme in the texts is what we will call technological salvationism. Hear Graff: “The rate of technological change is one important measure of human progress.” The book tends to give the impression that cultural advance can be measured almost completely on the scale of technological advance.

Todd and Curti do address one of the social problems of technology: technological unemployment. But the small space devoted to this matter leaves the impression that it was typical only of the turn of the century.

Yet another pervasive theme is the heroics of war. Consider this, from Graff:

The defense of freedom has always been the chief business of American patriots. In the twentieth century serious threats from abroad have made this task far heavier than in the past. The perils have had to be met by the sacrifices and often the lifeblood of patriots in uniform—GI’s.

There is no single typical GI, for GI’s come from every race and from every part of the country. They are the sons of poor parents and those who are well-to-do. They were born in the city as well as on the farm.

The consistent difficulty with such passages is that the inherent problems—those that call for decisions by citizens—are glossed over. The question, “Who is the GI?” which has plagued us for ten years, is ignored here—buried in a kind of platitudinous glue.

This is a century of war. Hear what Graff has to say to the teacher in the Teacher’s Edition of his text: “The human side of the (Civil) war is kept uppermost through details given about colorful and heroic individuals.”
Whatever war is, it is not in the main colorful and heroic. So to portray it is to do the children a disservice. The killing of people becomes "to wipe out Japanese resistance," or "wiping out the St. Mihiel salient." Pentagonese. The basic lack here is that the texts examined do not deal with the development of critical attention to the justification of war.

The irresistible progress of democracy in the United States is nowhere more apparent, according to these authors, than in the brief and successful struggle for civil rights. The idea that there is an ongoing struggle to preserve, protect, and extend the civil liberties of people under all governments simply does not appear. Civil rights and "Negro" rights are considered to be the same. Again Graff, the most idealistic of the writers:

Using Force. Step by step the courts and the office of the United States Attorney General tried to carry out the changes made necessary by (Brown vs. Board of Education). Too often force was required. President Eisenhower, for instance, called out federal troops to integrate...Central High School in Little Rock, Arkansas, in 1957. In 1962 President Kennedy also called on federal power—to bring about the enrollment of a Negro student at the previously all-white University of Mississippi.

The complexities, the question of equitable (or just) distribution of federal force and injunctive power, the value conflicts are all ignored.

The two characterizations we have offered here, that of dangerous over-simplification and the Irresistible March of Progress in Our Democracy, are brought together in the following, from Bragdon and McCutchen:

Meanwhile, through private action, taboos which kept Negroes out of certain employments were removed. Places were open to them on the basis of ability in fields as diverse as the professions, white collar jobs in business, and big league baseball. It has been estimated that the average earnings of Negroes in the United States tripled between 1940 and 1953.

These beginnings of progress for Blacks are portrayed as if the solution had been found; the passage as a whole conveys untruth, though in particular it is accurate enough.

Another constant theme in these texts is that capitalism serves all the people equally. This claim is usually put in the democracy-communism dichotomy, and it takes note of the possibility of mistakes, such as our overlooking the Black man's plight.

Not only has the productivity of America run way ahead of its increase in population, but available wealth has become more divided. During the "golden 20's" the top 5 per cent of income
gained 35 per cent of the wealth. By 1952 this share had sunk to 16 cent. Meanwhile, income levels had risen so much at the bottom of the scale that poverty was now generally limited to handicapped individuals, to social groups moving into a new environment (such as the Puerto Ricans in New York), or to special areas such as towns where a mine or factory had closed. (Stavrianos)

Both government reports and the reports of independent economists deny this statement; one wonders what its basis was. Students should, of course, be brought into direct contact with the problem of poverty in the midst of plenty—but this statement, like so many others, glosses the problem over.

Now, this evaluation of the “traditional” program in the Social Studies is biased. It does not include the account of the history of the country or its social institutions as presented in the materials of instruction. The bias, however, grows out of the questions we have raised. A program of nationalistic political socialization, which is what the materials offer, is defensible—but not from the point of view of its contribution to citizenship.

The thrust of the traditional program does not deal with citizenship because it does not deal with justice. In attempting to socialize students to an idealized society, the program glosses over existing injustice. The mass media deal with injustice every day and the students hear about it. A school program that fails to make the connections between our ideals and the daily injustices may well promote the juvenile cynicism that is now so commonplace. In suggesting that the progress of democracy is inevitable, the programs do not imply that any action is necessary.

The program we have would be unacceptable to people in the “original position,” as described by Rawls, who argues for a return to the Social Contract. To be acceptable, the program would need to emphasize strongly the existing just institutions of society, and the progress made toward relieving injustices, so that the just institutions might be preserved and protected. As things stand, large numbers of young people, having learned no better in school, damn the entire Establishment.

The Reform Movements

We have selected for examination two of the most widely recognized current reform curriculum programs: the Holt, Rinehart and Winston materials stemming from the work of Edwin Fenton and his associates at Carnegie Tech, and the Public Issues Series published by American Education Publications (AEP) and developed by Donald Oliver and Fred Newmann.

The two programs represent the two kinds of reform mentioned by Hertzberg—the subject-centered and the core. While the Holt materials are
centered around an inquiry method, they are strongly discipline-oriented: i.e., they seek to teach History, Political Science, Economics. The AEP materials also embody a method—rational discussion—but their theme is taken from persistent social questions, without regard for traditional discipline lines.

In one sense, however, the two programs are not comparable. The Holt materials are a complete secondary school Social Studies curriculum. The AEP materials are limited to 28 pamphlets of approximately 60 pages each. They would probably be used either as the basis for a Problems of Democracy course, or as "enrichment" materials for other aspects of the usual Social Studies program. However, the AEP materials could probably be expanded to cover the entire program, if they were judged valuable enough.

Holt Materials

We selected four texts and the accompanying teachers’ manuals for examination, from the wide range of Holt materials. These were: A New History of the United States (Bartlett, Fenton, Fowler, and Mandelbaum, 1969), The Shaping of Western Society (Good, 1968), Comparative Economic Systems (Coleman, 1968), and Comparative Political Systems (Schultz, 1967). Each of these volumes (as do all in the series) carries the subtitle, An Inquiry Approach.

In general, it must first be said that the texts are a pleasant relief from the traditional texts examined. The coverage seems thorough, though impressionistic in a lifelike manner, rather than systematic in a “scholarly” manner. There is much less glossing over of significant issues than is true of the traditional materials, although even in these materials, the treatment does not always develop the full adversary positions it suggests.

For example, the linking of “command economies” with “totalitarian governments” (p. 140 of the Teachers’ Manual, pp. 328-332 Shaping of Western Society) perhaps oversimplifies things. The “code” words in Comparative Economic Systems are “democratic-totalitarian” and “market-command.”

Unlike the traditional texts, however, these materials do deal with the more obvious areas in which U.S. reality does not correspond with U.S. ideals. There is a treatment of “How America Modifies the Market,” and a full treatment of civil liberties under “The Protection of the Individual.” Similarly, A New History of the United States deals fairly completely with the issues of political franchise, labor’s right to organize, the justice of war, etc.

The treatment is not faultless, however. Although the notion of the Absolute Good of the United States is qualified, there is little qualification of the Absolute Evil of the USSR. The series is limited pedagogically by its
original intent, to be *A High School Social Studies Curriculum for Able Students* (Bureau of Research, April 1969, Project Nos. HS-041 and H 292, with M. Lichtenberg). The inquiry approach calls for a high level of facility with print media, abstract concepts, and verbal expression. It seems likely that the non-‘able’ 75% of the population, which includes large proportions of the ethnically and economically disprivileged, would be handicapped in using these materials. The materials provided for the “slow learners” in the Holt materials seem, by comparison, hastily constructed and poorly thought through.

The Holt materials do not meet the first criterion we have put forward—that the materials place a sound concept of justice at the center of their attempts to deal with citizenship. The second criterion—that students enter the process actively—therefore does not apply.

The American Education Publications Materials

While the American Education Publications materials are not a full secondary curriculum, they are, taken as a whole, a rather remarkable set of texts in social problems. The problems they deal with are both national and international; the materials offer instruction in U.S. history, the history of other countries, political science, economics, sociology, and a range of other areas. The treatment is, of necessity, given the format, rather spotty. But the issues presented are considered in the most forthright and unbiased manner of any of the materials we examined. Each topic is considered in reasonable sub-topic form, and open-ended discussion questions are framed for both the sub-topics and the broad topic. The questions are usually directed toward ethical and moral judgments, as opposed to the cognitive-analytical questions formulated by the Holt materials, and the recall-indoctrination questions too often found in the traditional materials.

In general, the materials are evocative, engaging, and probing. They call for moral judgments, and thus raise questions of justice. The justice embodied in the materials seems generally consistent with our first criterion, as does the sense of justice they seem to seek to develop. Within their limits, the materials meet criterion one.

Since the first criterion is met, the second (that the student actively participate in citizenship) becomes relevant. This criterion poses difficult problems for these materials. For one thing, the problems and issues raised are very big. They do not lend themselves to meaningful action by students (as would, for example, a section on student rights). Moreover, the materials do little to provide an interface between the students and the issues, nor are the big issues related to root causes in such a way that students could see connections. For example, the authors missed an opportunity, in dealing with the labor movement, to connect the high school student with the big issue via recognition that youth are a large,
unorganized, often exploited labor group, often excepted from minimum wages, lacking fringe benefits, and lacking upward mobility.

We are critical of the American Education Publications materials because of their view of rationality. They are trapped by it: they project a view of rational discussion that would finally hold all problems up to the clear light of reason, as if we still lived in the eighteenth century. They overlook what Crane Brinton called the "anti-intellectual" aspect of modern thought—that thought represented, as he points out, by Rousseau, Newman, Pavlov, and Freud, not to mention more recent thinkers. It is the thought of this century, not the Enlightenment, that is needed if people are to carry ideas into action.

While these materials go a long way in purging idealism from the discussion of problems, they maintain a highly idealistic position concerning solutions: i.e., if only we think clearly about problems, they will be solved.

PROPOSALS FOR A SOCIAL STUDIES PROGRAM FOR CITIZENSHIP

We have urged that citizenship be defined as the pursuit of justice as between men and their institutions. This definition is not the popular one, but we do not think it is in conflict with what is generally meant by the term. The popular definition of citizenship combines loyalty (or patriotism) with limited action. We believe that what inspires loyalty to American institutions is our belief in their openness to improvement, and in particular to their openness to the improvement of justice. Where they are only partially just, or are unjust, we believe we have the power to change them. Where they are just, we believe we have the power to conserve them. Loyalty to American institutions is directed at a constant process of amelioration.

Dissent and criticism have always been a part of the process. It seems likely that there is less cynicism about our institutions of government than there was a century ago, if one is to believe the testimony of the great American humorists of the mid-nineteenth century, such as Artemus Ward, Josh Billings, Petroleum V. Nasby, Mark Twain, Eli Perkins, or the scathing cartoons of Thomas Nast. The popular scorn of those days gave support to the great social legislation of the turn of the century, and the more idealistically-oriented dissent and criticism of these days may have the same effect during the years ahead.

Meanwhile, where are the schools? According to Langton and Jennings, in an important study, loyalty, participation, use of media for political information, and the disposition to discuss politics show "scant differences...as a consequence of whether the student had taken a more traditional American Government course or the more topically oriented,
wider ranging American Problems course.” (p. 857) They conclude that “If the educational system continues to invest sizable resources in government and civics courses at the secondary level—as seems most probable—there must be a radical restructuring of these courses in order for them to have any appreciable pay-off.”

It is such a radical restructuring that we wish to sketch here. We suggest that the Social Studies curriculum, all of it, including the curriculum of the elementary school, be designed to promote citizenship. In order that this be done, major changes in content and intent would have to be made. Currently, the program seeks to teach History, in the main, with occasional excursions into other disciplines: Geography, Political Science, Economics.

We propose that the program be oriented around the nature of citizenship, here and elsewhere in the world, with our two criteria (a concept of citizenship, and active participation in the making and remaking of institutions) constantly observed. Such a program would not be organized around a single discipline, as the present one seems to be organized around History. It would seek to equip students with the processes of inquiry peculiar to specific disciplines, only as these are instrumental to the clarification of issues and the pursuit of action. The program would be organized according to the main issues that arise as men seek the ideal of just institutions. If one discipline appeared more than the others, it would be the discipline of political science. We repeat, however, that the purpose of the program would be to build citizenship, not to make junior-grade political scientists. It would be infused with political behavior by students.

A program for the development of citizenship would begin at the elementary school level. “By the time students reach high school many of their political orientations have crystallized or have reached a temporary plateau.” (Langton and Jennings, p. 854) Children arrive at grade 1 with rudimentary notions of justice (some of them seem not to progress beyond this level). They have learned at home the idea of “fairness,” which they interpret as even-handedness by those who have power over them; they have learned to “be fair” in dividing things between themselves. As the senior writer found in a study in 1953, the children have a special version of the Golden Rule: “Whatever you do, you get it right back.” By age 6 or 7, children have acquired a considerable vocabulary of blame and recrimination, but only a meager vocabulary of praise.

As things stand, the elementary school deals with all of this, and with the idea of citizenship, by insisting on “good manners,” and (in some schools) by seeking children’s cooperation in making rules for their joint conduct. In all schools, children are confronted with rules to be obeyed; such obedience is often called “good school citizenship.” As a kind of gentle overlay, elementary children are offered a somewhat fictionalized version of American History (grade 5) and some Geography.

Adelson and O’Neil suggest the conditions that apply at the late
elementary and secondary levels, from the point of view of child development:

Eleven-year-olds: “We might say that the 11-year-old has not achieved the capacity for formal operations. His thinking is concrete, ego-centric, tied to the present; he is unable to envision long-range social consequences; he cannot comfortably reason from premises; he has not attained hypothetico-deductive modes of analysis.”

Thirteen-year-olds: “The 13-year-old has achieved these capacities some (much?) of the time, but is unable to display them with any consistent effectiveness.”

Fifteen-year-olds: “The 15-year-old has an assured grasp of formal thought. He neither hesitates nor falters in dealing with the abstract.”

Eighteen-year-olds: “Taking our data as a whole we usually find only moderate differences between 15 and 18...The 18-year-old is, in other words, the 15-year-old, only more so. He knows more; he speaks from a more extended apperceptive mass; he is more facile; he can elaborate his ideas more fluently. Above all, he is more philosophical, more ideological in his perspective on the political order.”

A citizenship program for the elementary grades would need to be present-oriented, concrete, and first-hand. Such a program, to meet our two criteria, would need to build on the notions of justice (“fairness”) the children already have, while at the same time giving them constant direct practice in the making and changing of institutions, within their classrooms at first, later within the school, still later in the larger community. Elementary teachers already know something of how this may be done, but they do it on a sporadic, ad hoc basis. They have not considered our two criteria.

First, concerning justice: children would need to be presented with moral dilemmas for discussion and resolution. Such materials have been developed by Foshay and F. Shaftel of Stanford and more recently by Kohlberg and others. The problems would need to be close to the daily lives of the children, of course, though at the later elementary school ages, the problems might become increasingly complex. (E.g., at age six, the problem might have to do with whether someone ought to be excluded from a birthday party to “get even”; at age ten or eleven, the problem might have to do with a contest with a class in another school, in which the other class is strongly suspected of cheating.)

Second, concerning action: it is a commonplace for teachers of little children to lead them to the making of their own classroom rules. The teacher
could lead the children to apply the criterion of fairness to the rules, as well as the criterion of consensus. The teacher could begin building the concept of law, sanctions, and power if, as the rules were formulated, the matter of penalties for infractions were raised, together with the administration of such "justice." Obvious as it is, this is rarely done.

Ever since the days of Progressive Education, it has been common for children to work on projects in groups. Many children make faulty learnings out of this experience, such as that groups are not productive, or that one person usually takes over and does most of the work. Such committees are a natural example of an ad hoc social institutions; teachers could (some few do) help the children to look at the committee work structure objectively, and to develop the necessary skills, such as delegation of work, correction for faulty planning, the application of appropriate criteria to the product of the committee, and attention to the social process within the committee. Knowledge about these matters has been in the professional literature since the Thirties, though in the climate of these times it tends to be overlooked.

The U.S. History offered to elementary children tends to "come off" as a collection of myth and fable about people who lived in high and far-off times. It need not be so. Again, we can re-learn some of the lessons from the Progressive period. When children re-enact historical episodes, and especially when they construct the episodes themselves, they tend to internalize whatever the episodes contain. The broad aim of the Social Studies, being the development of citizenship, requires that in the upper elementary grades children enact episodes pertaining to the evolution of our political system, with an emphasis on the debating of the historical issues we have faced as a nation. This kind of material would tend to replace the anthropologically-oriented materials that dominated the Thirties and Forties, with the endless re-enactments of Pioneer Life and the Plains Indians.

The program should also include first-hand contacts with the broader community, both through the use of local people in the classroom and through visits and interviews. We used to think that such contacts should proceed from what was immediate to the children, such as the local policeman and the mailman, to what was more remote, such as local civic officials. With the advent of television, especially, it has become apparent that such an "expanding communities" approach does not correspond to the actual world children live in. The local policeman should not be presented as a "community helper," but as a law enforcement official. The mailman should be seen as part of a necessary communications system. The local politician (who should be a part of the lives of elementary school children) should be seen as one who works with conflicting interests, with the "art of the possible," and so on.

The secondary program, continuing from the elementary program,
should also deal with political reality. Like the elementary program, it should seek to deepen the concept of justice, and provide first-hand experience with political action.

This approach requires a reconception of the subject matter. History would be taught as the evolution of political systems; cultural history would be included as it enlightened citizenship, otherwise it would be offered as a part of a parallel Humanities program. Many of the subjects offered now would continue to be offered, but with the consistent attempt to discern their influence on our political action. We would not have to present the American story as an account of constant progress, if this were done. The rise of the labor movement, the development of the corporate systems, the Westward Movement, all these and the other familiar topics would be reinterpreted.

The Civics course would break from its Nineteenth Century tradition of structuralism, taking advantage of the current state of Political Science, which is empirical. The special emphasis in the Civics course, however, would be on attempted political action. Students would learn to analyze elections and campaigns as they occurred, taking part in them in appropriate ways. They would attempt to influence the actions of their local political bodies; they would learn through first-hand experiences how citizens may appeal to the political system for redress of grievances. In short, they would learn directly all the alternatives to rioting and street demonstrations—but these, too, they would study as directly as possible.

They would be taught propaganda analysis through historical as well as current examples. They might well practice writing propaganda, to the end that they recognize it when they see it.

Student government, from the point of view adopted in the revised Civics course, would become a laboratory for experimentation with governmental processes and forms. In one of the alternative schools, for example, the students rejected the usual forms of student government in favor of what they called a “community.” Faced with the necessity to make community decisions, the students went through the whole process of reinventing the vote. They discovered for themselves the concept of minority rights, having voted and found that some did not agree with the majority. They invented an oligarchy to lead them, and eventually rejected it. And so on. Their teachers were offered one rich possibility after another to bring alive the basic political concepts.

Problems of Democracy would remain, but would be transformed into yet another (more complex) set of action opportunities. Current events would be taken as current problems, as the students, now almost old enough to vote, would be brought in contact with the possibilities and limitations of our current version of American political citizenship.

Such a program would emphasize political knowledge and political behavior. Its broad purpose would be to deepen and energize the concept of
justice. In pursuing these purposes, it would be necessary to subordinate or alter some fields now offered.

In order that the student’s school program as a whole not be out of balance, it would be necessary to build a Humanities offering as strong as this Social Studies offering. Literature and the Arts, Cultural Anthropology, and Cultural History would be offered in the Humanities program. Economics would appear in this proposed Social Studies program as it pertained to political behavior. It would not be offered in its own right as a discipline.

Nor would the other subjects currently offered be offered as disciplines. The modes of inquiry appropriate to the Social Sciences would be developed with the students as they were instrumental to the principal aim of the program. Where they were not instrumental to this aim, they would not be offered.

What, then, of disciplined knowledge? The proposal we offer here is that the teaching of formal disciplines be made elective in high school, especially for the college-bound, and that the full-scale treatment of the disciplines be done at the college level. What this implies is that the principal aim of the lower schools be to equip the majority of non-college-bound students to act, now. The more formal kinds of knowledge would not be required of them, except as these kinds of knowledge contribute directly to present action.

We do not believe that such a policy is elitist, because we do not agree that academic knowledge is elite. Formal knowledge is not a more advanced form of practical knowledge; it is an abstract form of it. We do not seek to make social science theorists out of high school students. We seek to make citizens of them, within the meaning adopted in this paper. What we seek, as a consequence of twelve years of education in the lower schools, is an effective participant in society who will help us all in our pursuit of justice.

REFERENCES


**Bibliographical Note:**

Other materials consulted for background information but not specifically cited in the paper are too numerous to list in full. The following were especially helpful, however.

Additional critical attention to texts was provided by Social Studies in the United States: A General appraisal, by Benjamin Cox and Byron Massialas (Harcourt, Brace and World, 1967); “Reflective Thinking, Values, and Social Studies Textbooks,” by James P. Shaver (The School Review, 73 [Autumn, 1965], pp. 226 ff.); and by a doctoral dissertation in progress by Peter Dublin at Columbia University, Teachers College.
PATTERNS OF RACIAL SEPARATION
IN A DESEGREGATED HIGH SCHOOL

A. Guy Larkins
Sally E. Oldham
University of Georgia

Georgia public high schools have been desegregated since 1970: Black students and white students attend school in the same building. Students of each race can be found in the majority of courses and classrooms. But integration—the unrestricted association of races—is rare. Numerous patterns of racial separation are the rule in most schools with which the authors are familiar.

The purpose of this study was to determine whether racial separation in a desegregated high school was associated with ability grouping, student selection of elective courses, student selection of seat location in classrooms, verbal interaction between students and between students and teachers during instruction, and adult authority roles and student roles in extracurricular activities such as football, band and cheerleading.

METHOD

Research Posture

The statement of research purpose should not mislead the reader into believing that we entered this study with a neat package of objectives and predetermined procedures for meeting them. In fact, our method was similar in spirit to Skinner’s (1956) first principle of scientific research: When you run into something interesting, drop everything else and study it. Based on previous experience in visiting numerous north Georgia high schools, we had some fairly fixed ideas of what we would look for, but we were determined to remain sensitive to the unexpected, to immerse ourselves in the setting, and to adjust our observations and analyses to fit whatever interesting problems we encountered (Shaver and Larkins, 1973, pp. 1254-1257).

Our method was also influenced by Smith and Geoffrey (1968), who argued that it is legitimate to enter a study without fixed hypotheses, and who alerted us to the need for multiple methods of verifying claims, as have also Denzin (1970) and others (Webb, Campbell, Schwartz and Sechrest, 1966).
Research Setting

The school selected for this study serves students in Grades 8-12 from a community of 5,000 people, and from a surrounding county of approximately the same population. The building was used by black students from 1956 until desegregation in 1970, when it became the only public high school in the area. At the time of the study, the school enrolled 825 students. The black to white ratio was 65 percent to 35 percent, but the proportion of whites would have been higher had there not been a small private academy less than a mile away from the public high school.

Within the school population, this study focused on the approximately 200 students enrolled in American History. For instance, data reported concerning the number of students by race enrolled in Remedial Reading refer not to the total school population, but to the set of students who were both taking American History and Remedial Reading. Similarly, data reporting voluntary seating patterns by race are based on observations made in American History and other social studies classrooms.

There were several reasons for focusing on the set of students taking American History: They were a representative sample by race of the total school population—it was a required course. They were a theoretical sample (Denzin, 1970) which made it possible to study whether ability grouping in some courses affected the racial composition of courses which were not ability grouped. And, being social studies specialists, the investigators were particularly interested in the effects of desegregation on social studies courses.

Data Collection

Data were gathered over a three month period, beginning with approximately a week of unstructured observation of social studies classes by the junior author. The senior author made occasional observations in some of the same classrooms.

Impressions formed during the first week of unstructured observation became the foci for several weeks of systematic observation of ten American History classes, plus elective courses in the same department. Frequency counts of seating patterns and of student-teacher interaction were made, as were anecdotal notes.

Classroom observations raised questions about patterns of racial imbalance in other departments in the school, particularly whether ability grouping, remedial programs and electives were related to racial imbalance in various courses and programs of study. Information relevant to some of those questions was obtained from permanent records and interviews. Several weeks were spent compiling California Achievement Test scores and rosters of students in various courses. In addition, scheduled interviews...
were conducted with eight department heads, the principal, superintendent of the school system, two counselors and a school secretary. Casual interviews were also held with an available sample of teachers from all departments. Those interviewed were very cooperative and seemed candid.

Reliability and Validity

Inter-rater reliabilities could not be estimated for the classroom observations—a single observer was used—but it is unlikely that gross errors were made in counting the number of students in a room, or in categorizing them by race. Interviews with faculty and administrators were the most likely source of error for the information reported here. Information which they provided, however, agreed with data obtained through direct observation, California Achievement Test scores, class enrollment figures, our previous experience and the experience of our colleagues who were familiar with other southern high schools.

Reliability was estimated for the California Achievement Test scores by using KR 21. Although surprisingly low—.61—it was high enough for classifying students by groups (Borg and Gall, 1971, p. 142).

Data Analysis

Much of the data analysis was by inspection; the data were primarily descriptive. In those cases where estimates of statistical significance were appropriate, chi-square was used. The level of significance was set at .05. Yates correction was calculated when cell entries were five or less with one degree of freedom.

FINDINGS: SOCIAL STUDIES

Voluntary Seating and Verbal Interaction

Systematic observations of intra-classroom patterns of racial interaction were made in five American History classes and four social studies elective classes. Data were obtained by recording student seating patterns by race, counting the number of questions which black students and white students were asked by their social studies teachers, and by noting instances of verbal interaction by race between students. Casual observation in numerous other classrooms were consistent with the frequency-count data.

Without exception, students segregated themselves into seating groups by race. Moreover, voluntary segregation occurred regardless of the racial composition of the class. For instance, one class contained 16 black students and 10 whites, which was similar to the racial division in the school population. Although this class would be considered desegregated, and
racially balanced, it was not integrated; integration refers to alterations in the relations between people, rather than mere removal of legal barriers.

Previous experience led us to expect that certain seats in classrooms would be "owned" by whites or blacks; that the same race would tend to occupy the same seats from one class period to the next. This expectation was confirmed in an interesting way. In four out of five American History classes taught by one teacher, four seats on the front left side of the room near the instructor were always occupied by white middle class students. This pattern was especially noticeable because the remaining seats on that side of the room were occupied by black students in every class. The rest of the white students sat on the opposite side of the room. Hence, there appeared to be some stigma attached to the seats which regulated who could be afforded their status.

That the observed social studies classes were not integrated was also evident in the lack of verbal interaction between students across race; voluntary verbal interaction appeared to be nonexistent. Exceptions to this finding were occasionally noted during casual observations in other settings, but lack of free verbal interaction between students across race appears to be the rule in those schools with which the authors are familiar.

Patterns of verbal interaction between teachers and students by race were also noted. Three male, white social studies teachers who exhibited marked differences in teaching style were observed for this purpose. In inquiry based lessons, one teacher directed all questions to black students. The second teacher asked all but two questions of black students. In contrast, a third teacher reserved all of his questions for whites. In each of these classes, the proportion of students by race was the same as in the school population. This tendency to teach to either one race or the other is consistent with our casual observations elsewhere. It appears that teachers' verbal behavior may reinforce voluntary segregation within classrooms.

Choice of Social Studies Electives by Race

Experience prior to this study led the investigators to expect that students would not only segregate by race through voluntary seating patterns within classes, but also through the selection of social studies electives. For example, during the initial desegregation year of 1970, the senior author taught a high school black studies course which contained 21 black students and 4 whites. Directly across the hall, a sociology-psychology course contained no black students. Table 1 indicates a similar pattern of racial imbalance among some social studies electives in the present study.

Of the 81 white students enrolled in social studies electives, 42 percent selected Psychology, but only 13 percent of the black students made that choice. Of 139 black students enrolled in electives, 37 percent chose the course which had a unit on Black History, but only 15 percent of the whites
Table 1: Choice of Electives by Race

<table>
<thead>
<tr>
<th>Course</th>
<th>Percent of White Students</th>
<th>Percent of Black Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociology, You and the Law</td>
<td>12.35</td>
<td>11.51</td>
</tr>
<tr>
<td>Economics, Careers, World Leaders, Black History</td>
<td>14.81</td>
<td>36.69</td>
</tr>
<tr>
<td>Family Relations, Criminology, Communism, History of War</td>
<td>19.75</td>
<td>12.23</td>
</tr>
<tr>
<td>Psychology</td>
<td>41.97</td>
<td>12.95</td>
</tr>
<tr>
<td>Consumer Education</td>
<td>11.11</td>
<td>26.62</td>
</tr>
<tr>
<td>Totals</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>

\[^{a}\text{Multiple course titles indicate that some electives were combinations of several mini-courses.}\]

\[^{b}\text{Chi-square values were computed using frequency data, not proportions. } X^2 = 34.5, df = 4, p < .001, N \text{ for whites } = 81, N \text{ for blacks } = 139.\]

made that choice. Enrollments in Consumer Education followed the pattern in Black history: the black to white ratio was more than four to one, even though blacks outnumbered whites by less than two to one in the school population. Sociology was the only elective in which the black to white ratio was similar to the school population ratio.

FINDINGS: EFFECTS OF ABILITY GROUPING

That we found evidence of racial separation in voluntary seating patterns, patterns of voluntary verbal interaction and choice of social studies electives was not surprising; on the whole, this study reinforced our previous experience with desegregated schools. We were puzzled, however, to note that some American History classes appeared to have an obvious racial imbalance in enrollment. For example, one class had 19 black students and only three white students, but another class had nine black students and 18 white students.

Several possible explanations for the racial imbalance in enrollment in
in American History were explored and rejected. According to teachers, counselors and administrators whom we interviewed, American History was not an elective, it was not ability grouped, students were not allowed to select the instructor or the period during which they would take the course, and counselors made no attempt to fit students to instructors. It occurred to us, therefore, that ability grouping and tracking in other areas of the curriculum may have had a "spillover effect" in determining the racial balance in American History classes. Having found an interesting problem, we followed Skinner's (1956) advice and turned our attention to two questions: To what extent does this school practice ability grouping and tracking? Are ability grouping and tracking related to patterns of racial separation in the school?

Interviews with the high school faculty and administrators, plus examination of the composition of various classes by race and academic achievement indicated that ability grouping, tracking and racial separation within the school were related to several curricular innovations which were adopted during the three years prior to this study. Those innovations involved courses of study which prepared students to receive one of three types of high school diplomas. Innovations also involved remedial reading, English, and remedial and advanced math. Each is explained briefly below.

Specialized Diplomas

In addition to the standard high school diploma, two special diplomas were created; one was college preparatory, the other was career development. Compared to the standard diploma, the college preparatory track required additional hours in science and math. Career development required six to nine units in one of the six areas in vocational education; such as agriculture, woodworking or cosmetology.

How did tracking for special diplomas affect racial separation in the school? Interviews with faculty and administrators indicated that very few black students graduated with the college prep diploma, and that the majority of students in the career development track were blacks or lower class whites. Furthermore, instructors reported that the small handful of college prep students who signed up for an occasional course in the vocational track, such as woodworking or agriculture, formed clear-cut cliques within the classroom and did not interact with the career development students.

Remedial Reading

All students unable to read at the 8.0 level on the California Achievement Tests were required to enroll in remedial reading and remain there until they either graduated or overcame their reading deficit. More than half of the
total student body could not meet the 8.0 standard. Classroom observations and interviews with faculty and administrators indicated that a disproportionate number of the remedial reading students were black. The following data illustrate the problem.

Reading scores on the California Achievement Tests were obtained for the students enrolled in American History. Those students whose scores were more than one standard deviation below the mean for the total group were classified “low.” Those who scored more than one standard deviation above the mean were classified “high.” The rest were classified “average.” Table 2 presents the results of that classification.

Table 2: Percent of Students by Race at Three Reading Levels

<table>
<thead>
<tr>
<th>Reading Level</th>
<th>Low</th>
<th>Average</th>
<th>High</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>28.24</td>
<td>66.41</td>
<td>5.34</td>
<td>100%</td>
</tr>
<tr>
<td>White</td>
<td>1.49</td>
<td>41.79</td>
<td>56.72</td>
<td>100%</td>
</tr>
</tbody>
</table>

aChi-square values were computed using frequency data, not proportions. $X^2 = 72.87$, df = 2, $p < .001$, N for whites = 67, N for blacks = 131.

Visual inspection of the categories in Table 2 is impressive. When students in this sample were segregated by reading scores they were also separated by race: the proportion of blacks to whites in the low category was 19 to 1. In contrast, the proportion of blacks to whites in the high category was 1 to 11.

English and Math

One of the recent innovations in this high school was the addition of 13 mini-courses on subjects ranging from Shakespeare to The Supernatural in Literature. Remedial reading students were not allowed to take these courses. Interviews with faculty and administrators indicated that most of the students in these courses were white.

Ten math courses were offered: Developmental (remedial) Math; General Math; Business Math; Vocational Math; Elementary Algebra Part
I; Elementary Algebra Part II; Algebra I; Algebra II; Geometry; Trigonometry and Functions.

All students were required to take math unless they had completed Algebra I or Elementary Algebra I and II. Course placement was determined by the student’s score on the math section of the California Achievement Tests. One function of the test was to segregate students into remedial or nonremedial math courses. As can be seen from Table 3, course placement also tended to separate students by race.

Table 3: Relationship between Race and Math Courses

<table>
<thead>
<tr>
<th>Courses</th>
<th>Black</th>
<th>White</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trigonometry</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Geometry</td>
<td>7</td>
<td>12</td>
<td>19</td>
</tr>
<tr>
<td>Algebra II</td>
<td>4</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Algebra I</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Elementary Algebra II</td>
<td>11</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Elementary Algebra I</td>
<td>12</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Vocational Math</td>
<td>29</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>Business Math</td>
<td>35</td>
<td>7</td>
<td>42</td>
</tr>
<tr>
<td>General Math</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Remedial Math</td>
<td>27</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td><strong>132</strong></td>
<td><strong>60</strong></td>
<td><strong>192</strong></td>
</tr>
</tbody>
</table>

\[ X^2 = 58.59, \ df = 9, \ p < .001 \]

\[ ^a \text{Includes all students enrolled in American History for whom CAT math scores were available.} \]

When the data in Table 3 were converted to percentages and the courses were classified as college preparatory—Trigonometry, Geometry, Algebra I and II—and remedial or vocational—Vocational, Business, General and Remedial Math—the racial imbalance within the math program was striking. Only 10 percent of the black students in our sample were enrolled in the college preparatory courses, compared to 53 percent of the whites. Similarly, 73 percent of the black students were enrolled in remedial or vocational math compared to only 23 percent of the whites. The rest of the students were enrolled in Elementary Algebra I or II, which are difficult to classify.
The data reported so far illustrate an important dilemma in American education; the conflict between attempts to meet academic needs of students and attempts to provide integrated schooling. For our sample, it appears that curricular arrangements such as providing college preparatory programs of study, vocational programs, remediation in basic subjects, and student choice of social studies electives separated students into racial groups—a college preparatory track populated primarily by whites, and a remedial-vocational track populated primarily by blacks.

Some educators believe that the apparent conflict between attempts to provide a quality education and attempts to integrate schooling can be resolved by requiring all students to take some courses for which there are no academic prerequisites (Westby-Gibson, 1966; Findley, 1970). Under this plan, students would be separated by academic interest and ability during some courses and integrated during others. Our review of literature uncovered no assessment of the effects of this procedure. On the basis of our experience in public schools, however, the present investigators hypothesized that ability grouping in some courses would cause racial imbalance in non-grouped classes. Fortunately a partial test of this hypothesis was available in the target school; all students were required to take American History, and the ten classes offered in this subject were not ability-grouped.

The first step in testing our hypothesis was to determine whether the distribution of students by race among the ten American History classes differed significantly from a random distribution.

Data presented in Table 4 indicate that taken as a whole the ten American History classes were racially imbalanced; the chi-square value of 26.46 was significant at the .01 level. Furthermore, inspection of Table 4 indicates that white students were much less evenly distributed across classes than were blacks. Whites appear to have been virtually isolated in Classes 1, 5, 6, and 9, which contained only two or three white students each. In contrast, though whites were the minority race in this school, they outnumbered blacks by two to one in Class 10, which contained more than 25 percent of the total sample of white students. Class 10 contributed so heavily to the racial imbalance among the American History courses, that when it was removed from the analysis the chi-square value was not statistically significant.

Having determined that there was a statistically significant racial imbalance in the American History classes taken as a whole, and especially
Table 4: Racial Composition of American History Classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Black</th>
<th>White</th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.15</td>
<td>4.41</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>14.07</td>
<td>7.35</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>9.63</td>
<td>16.18</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>8.89</td>
<td>8.82</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>14.07</td>
<td>4.41</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>7.41</td>
<td>4.41</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>10.37</td>
<td>14.71</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>9.63</td>
<td>10.29</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>11.11</td>
<td>2.94</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>6.67</td>
<td>26.46</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>Totals</td>
<td>100.00</td>
<td>100.00</td>
<td>135</td>
<td>68</td>
</tr>
</tbody>
</table>

*a This column contains the percent of the total sample of black students or white students in each class.

*b This column contains the number of students by race in each class. Chi-square values were computed using the data in this column. $X^2 = 26.46, df = 9, p < .01$.

in Class 10, the next step in the *ex post facto* testing of our hypothesis was to examine rival hypotheses. Several were explored and tentatively rejected: Teachers, counselors and administrators claimed that students could not elect whether to take American History, they were not assigned to American History on the basis of ability or race, students were not allowed to select a favorite instructor or choose a period when they could be with friends, and school personnel made no attempt to match students with compatible instructors. Furthermore, the ten classes were divided evenly between two instructors, and some of the above claims were checked by computing a 2 x 2 chi-square, the results of which indicated that race of student was independent of instructor for the American History classes ($X^2 = 2.71, df = 1, p < .05$).

The final step in testing the hypothesis that tracking and ability grouping in other academic subjects can affect the composition of non-grouped classes involved interviewing students and instructors concerning the unusually large number of white students in Class 10. Responses indicated that most of these students were in the college preparatory track and played in the school band. Their advanced academic courses were taught during the
in the school band. Their advanced academic courses were taught during the first four periods of the day. Band was taught sixth period. The only time open for American History was fifth period.

The investigators tentatively concluded that the conflict between offering an enriched curriculum on the one hand and integrated schooling on the other may not always be resolved by requiring students to take some courses, such as American History, which are not ability-grouped. Tracking students for a large portion of the day may remove the degrees of freedom needed to “untrack” them for one or two periods.

FINDINGS: DESEGREGATION AND EXTRACURRICULAR ACTIVITIES

Social studies is frequently seen as a vehicle for the democratic mixing of students who are otherwise kept apart by ability grouping and tracking (Westby-Gibson, 1966; Findley, 1970). Extracurricular activities such as football, band and cheerleading are sometimes seen as fulfilling the same purpose.

The following data were collected at a home football game. Particular attention was paid to the race of persons holding positions of high status or authority, and to the racial composition and racial interaction patterns of the football team, cheerleaders, band and the hometown crowd.

At the game, several of the high status or authority roles were evenly partitioned by race. Tickets were sold by one white and one black female counselor. Tickets were taken at the gate by one white and one black male, the principal and the head of the math department. There were five referees, only one of which was black, but the coin toss was attended by one referee of each race. Similarly, each team sent co-captains to the coin toss, one black and one white. The head coach of each team was white, but each had one white and one black assistant.

Although there was an obvious attempt to have each race represented in visible roles, several examples of race stereotyping were evident. At the game, there were: five white policemen, who tended to stay close together; two white team physicians; four whites who held the sideline markers; one white student photographer; and four black waterboys. The home team had eight cheerleaders, five white and three black, selected by popular vote. The captain of the cheerleaders was white. The band was entirely white except for two flag girls, one female saxophone player, one male drummer and the drum majorette. The band master was white.

The most unusual example of race separation was the visiting football team. With one exception, their offensive team was all white, and with one exception, their defensive team was all black. The lone exception on the offense was the black quarterback, who was obviously their superstar. Although racially desegregated on the playing field, the home team tended
to cluster and interact by race when on the sidelines before and during the game.

Seating patterns by race were obvious among the fans whether adults or students. Interaction across race lines was absent with one exception; black students would occasionally make requests of white cheerleaders. Generally the crowd was good natured and supportive of the whole team. It was observed, however, that black adults tended to be particularly supportive of black star players. They seemed to know them personally, or at least spoke as though they did. No cheers or jeers from anyone in the crowd were heard that might be construed as having racist overtones.

DISCUSSION

External Validity

There have been advantages in focusing our attention on a single desegregated school: We were able to spend far more time "living with" our data source than would have been possible had we studied several schools; we were able to adjust our inquiry to interesting and important problems which came to our attention during the course of the study; and we were able to return to our data source to double check claims.

But there were also disadvantages. One of the most obvious was external validity: Would our findings replicate elsewhere? One reason for thinking that they might not is that our target school was atypical in several ways: Black Americans are in the minority nationally, but formed a two to one majority in this school. The majority of Americans live in or near large urban centers, but our sample was from the semi-rural South. And some aspects of the target school's curriculum were probably atypical: for instance, the requirement that all students must either demonstrate competence in reading and math or take remedial courses in those subjects may not be widespread.

Nevertheless, we think that the general patterns of racial separation reported in this study would replicate. Our findings are not only consistent with our professional experiences in other desegregated schools, and with the experiences of our colleagues, but also with other types of research on ability grouping and desegregation. For instance, ability group placement, based on standardized aptitude or achievement scores, almost without exception tends to separate students by race (Svensson, 1962; United States Civil Rights Commission, 1962; Coleman, 1966; Findley and Bryan, 1971; Shafer and Olexa, 1971). That there is a relation between race and placement in ability grouped classes is also predictable from the considerable research which indicates that black students consistently perform lower on the average than whites on standardized tests (Klineberg, 1934; Dreger and Miller, 1960; Coleman, 1966; Pettigrew, 1971).
Despite the congruence between the findings of this study and those of other types of research, the question of replicability needs to be settled through additional studies similar to this one but conducted in a variety of settings and by investigators with a variety of frames of reference.

One aspect of the present study which particularly needs replication is the hypothesis that ability grouping in some courses affects racial balance in non-grouped classes. This issue probably cannot be addressed experimentally; a number of *ex post facto* studies will likely be required to provide a reliable answer.

**Other Research Questions**

Additional studies are needed to answer questions about student perception of racial integration and ability grouping. Data gathered for the present study throw only indirect light on that issue. For instance, how might the perceptions of school differ for each of the following four students: two blacks and two whites, one of each race in the college prep track and the other in the vocational-remedial track?

From the evidence at hand, we can reasonably assume that students in the college prep track would be part of an elite who take their advanced academic courses together and may even share the same non-grouped classes. What we do not know is how perception of their schooling experiences within their elite status may differ by race. For instance, although whites were a minority in our target school, we can infer that white college prep students were almost always in classrooms in which whites were the majority. During the hour to hour business of schooling, they probably were seldom forced to cope with minority status, except perhaps in the lunchroom and hallways. In contrast, black college prep students, though part of the school majority, appeared to be set off from their racial peers and may have confronted the hour to hour business of schooling as part of a distinct minority.

Similar questions could be asked about the perceptions of schooling by students of each race in the vocational-remedial track. How do black students react to their race being overrepresented in the remedial classes? Do white students in those classes feel stigmatized? Are the remedial courses seen as a desirable way to acquire essential academic skills, or as a deadly bore to be endured? How do students of each race and in each of the two tracks view schooling in terms of their chances for social mobility? We suspect, but lack evidence, that student responses to many of these questions would be disturbing.
A Moral Dilemma

Patterns of within-school racial separation pose a potential moral dilemma for those responsible for educational policy. Meeting the individual needs of students through an enriched curriculum containing electives, remedial courses in basic subjects, advanced courses for able students and career preparation programs for those who do not aspire to a college education often seem desirable. But these practices may also unwittingly contribute to racial separation.

No easy solution to the dilemma is apparent: Students cannot be placed in advanced math and science courses for which they have inadequate preparation. Restriction of students' freedom of choice in electives, seating patterns, and informal social interaction is also questionable. Requiring all students to take the same courses, regardless of interest, ability, or career aspiration would be seen as a step backward by many citizens and educators.

Regardless of whether solutions to the general dilemma are found, social studies educators can take specific measures which may help. One step might be to become informed about research on the effects of different patterns of teacher-pupil interaction. Teacher-Student Relationships: Causes and Consequences by Brophy and Good (1974) contains an extensive review of that subject.

After reviewing much of the relevant research, Brophy and Good (1974, p. v) concluded that "teachers behave quite differently toward different students in their classroom" and "the treatment of some students is so consistently inappropriate as to erode their general confidence and aspirations." Generally, the less appropriate teacher behaviors are reserved for students who are perceived as low achievers or as coming from families of low socioeconomic status. Given the well known relation between race, SES and pupil achievement, the relevance of Brophy and Good's review to the present research is obvious. Social studies educators may help counter some of the tendencies toward, and effects of, racial separation by learning about and adopting appropriate patterns of verbal interaction with their students. A good beginning would be to avoid behaving as though desegregated classrooms were populated only by students of one race.

A second step involves the assumption that social studies can be used as a vehicle for the democratic mixing of students who are otherwise separated by tracking and ability grouping in other courses. The present research seems to indicate that if racial balance in social studies is desired, educators may need to deliberately assign students to classes. Social studies teachers cannot assume that lack of ability grouping in their courses will result in a desirable racial balance.

A third step toward avoiding racial separation is to use ability grouping only when it is essential. Grouping for teacher convenience is a questionable
practice given the lack of consistent evidence that it improves pupil achievement (Findley, 1970).

Educators should also be more cautious about using standardized tests as the primary means of assigning students to ability grouped courses; our review of the literature has raised serious questions about their predictive validity when used with black students (Oldham, 1974). We suggest that multiple criteria be used for ability group placement, and that students be trained in test taking skills. Students should also be acquainted with the test administrator, the testing environment, and the test format before an instrument is used to help assign students to courses or programs of study.

Perhaps the most important step is for individual social studies teachers to decide whether they are committed to integration as desirable social and educational policy. Unless that decision is faced by teachers on the firing line, mere desegregation will likely perpetuate through default.

REFERENCES

Shafer, W. E. and Olexa, C. Tracking and Opportunity. Scranton:
Shaver, J. P. and Larkins, A. G. "Research on Teaching Social Studies."
In R. M. W. Travers (Ed.), *Second Handbook of Research on Teaching.*
THE ABILITY OF ELEMENTARY SCHOOL TEACHERS TO STAGE SCORE MORAL THOUGHT STATEMENTS

John D. Napier
University of Georgia

The work of Lawrence Kohlberg and his associates has provided a description of moral thought development (Figure 1) as well as instructional procedures for developing moral thought (Kohlberg & Turiel, 1971; Kohlberg, 1975). This construct of moral thought development and program for moral education has been discussed as a possible medium for value education in social studies curriculum (e.g., Fraenkel, 1973, Ferguson & Friesen, 1974; Kohlberg, 1973, Shaver & Strong, 1976). One of the instructional procedures used in most programs employing Kohlberg's ideas was stage scoring of moral thought statements prior to or during instruction (Rest, 1974). Stage scoring of moral thought statements was necessary in order to fulfill the instructional procedure of exposing students to the next higher stage of moral thought (Kohlberg & Turiel, 1971, p. 455; Kohlberg, 1975, p. 675).

The stage scoring of moral thought statements is a complex task. If teachers cannot stage score the moral thoughts of students before instruction then the successful stage scoring of the moral thoughts of students during instruction is doubtful. Further, if teachers cannot stage score moral thought statements then the successful use of the instructional procedures advocated by Kohlberg and his associates for moral education is also doubtful.

In 1972, Porter and Taylor published a self training global rater manual to aid teachers in stage scoring moral thought statements (Porter and Taylor, 1972). At the time of this study, this rater guide was the only completed manual available to aid teachers. The manual was published by the Ontario Institute for Studies in Education. Also, a version of this same manual is available from the Harvard Center for Moral Education. This rater guide has been used by Kohlberg and his associates in studies of moral development and moral education (e.g., Blatt & Kohlberg, 1973). The purpose of this study was to investigate whether teachers could use the Porter and Taylor global rater manual to correctly stage score different moral thought statements, and what factors were related to their stage scoring abilities.

STAGE SCORING MORAL THOUGHT STATEMENTS

The measurement system used to determine the stage of moral thought statements involved the qualitative coding of responses to hypothetical
Figure 1
Stages of Moral Development*

PRE-CONVENTIONAL LEVEL

Moral value resides in external quasi-physical happenings, in bad acts, not in persons and standards.
Stage 1: *Punishment and Obedience Orientation.* Egocentric deference to superior power or prestige or a trouble-avoiding set. Objective responsibility.
Stage 2: *Instrumental Relativist Orientation.* Right action is that of instrumentally satisfying the self's needs and occasionally the needs of others. Awareness of relativism of value to each actor's needs and perspective. Naive egalitarianism and orientation to exchange and reciprocity.

CONVENTIONAL LEVEL

Moral value resides in performing good or right roles, in maintaining the conventional order and the expectancies of others.
Stage 3: *‘Good Boy—Nice Girl’ 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: *Law and Order 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.

POST-CONVENTIONAL LEVEL

Moral value resides in conformity by the self to shared for shareable standards, rights, or duties.
Stage 5: *Social-contract 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: *Universal Ethical Principle Orientation.* Orientation not only to actually ordained social rules but also to principles of choice involving appeal to logical universality and consistency. Orientation to conscience as a directing agent and to mutual respect and trust.

*Adapted from Turiel (1969, pp. 96-97)
moral dilemmas. Two procedures have been used by Kohlberg and his associates to stage score moral thought statements. The first procedure was a sentence coding process. The second procedure was a global rating process. In both, a rater guide was used to determine a profile for a subject consisting of a stage or a combination of stages. Besides a stage profile, each subject was given a Moral Maturity Score. The Moral Maturity Score was the sum of the product of the percentage of responses at each stage in a given profile multiplied by the ordinal value of the stage times 100. The range of the Moral Maturity Score was from 100 (all stage 1 responses) to 600 (all stage 6 responses). The second procedure for stage scoring appeared the easiest to use and did not result in subject profiles which differed markedly from those profiles produced when the sentence coding process was used (Blatt & Kohlberg, 1973).

The use of rating guides aided in the classification of qualitative data, and the skill used in classifying qualitative data (like the data produced by the responses to hypothetical moral dilemmas) was content analysis. Although not much has been reported about variables influencing the ability to use content analysis, inferences from the abilities required and variables found important in similar research techniques yielded several variables of probable importance (Hyman, 1970; Richards, Dohrenwend & Klein, 1965). First, intelligence seemed important, especially the ability to handle verbal symbols. Second, the demographic variables of sex, age, social class, educational background, and experience could be important. Finally, the ability to relate to others, or in other words the ability to be receptive of others, also seemed important.

The research of Rest, Kohlberg and Turiel (1969) suggested one other variable of probable importance. Their research showed that the moral stage of an individual influenced comprehension of moral thoughts of others. Therefore, the moral stage might correlate to stage scoring the moral thought statements of others.

In summary, the Porter and Taylor self training global rater manual was examined to determine if it aided teachers in stage scoring moral thought statements prior to instruction. This manual was the only published guide available to aid teachers, and it represented the easiest form of stage scoring moral thought statements. Second, the variables of sex, age, social class, educational background, experience, verbal ability, receptive ability, and moral stage were suggested as eight possible correlates to stage scoring ability.

METHODS

Subjects

The subjects of the study were sixty teachers involved in an experimental
Master of Education program at a middle Atlantic state university. Although the teachers were volunteers for the experimental program, they did not volunteer for this particular study. The majority (83%) of the teachers taught in elementary schools while the remainder taught in junior high schools. Most of the teachers (63%) taught all subjects with the others equally distributed among various disciplines. All the teachers were engaged in an elementary social studies education program during the time of this study. Other demographic and personality data on the teachers are given in Table 1 below.

Variables under Study

**Demographic variables.** The demographic variables investigated were sex, age, experience, and verbal ability. Experience was defined as teaching experience. The variables of social class and educational background were of less importance because the teachers were all college graduates and all were considered to be at similar socioeconomic levels. These latter two variables were eliminated from investigation since differentiation between teachers was difficult. Verbal ability was defined as the verbal score from the Graduate Records Examination.

**Personality variables.** The personality variables studied were the ability to relate to others (receptive ability) and moral stage (moral maturity). The ability to relate to others was defined as being cognitively open. Milton Rokeach's definition was accepted as an adequate description of the meaning of verbal receptive ability (Rokeach, 1960, p. 57):

"...the extent to which the person can receive, evaluate, and act on relevant information received from the outside on its own intrinsic merits, unencumbered by irrelevant factors in the situation arising from within the person or from the outside."

Moral stage was converted to moral maturity. Moral maturity was determined by obtaining the Moral Maturity Score produced through the global rating process of stage scoring moral thought statements (see earlier discussion of Moral Maturity Score).

**Description of moral stages.** The description of moral stages referred to the stage and sequence chart developed by Kohlberg and his associates (Kohlberg & Turiel, 1971, pp. 415-416), and was similar to that presented in Figure 1.

**Self training global rater manual.** The rater manual referred to an adapted version of one of the stories of the guide written by Porter and Taylor (1972, pp. 11-22).
Measurement Instruments

**Verbal ability measure.** The Graduate Records Examination test of verbal ability had both reliability and validity as inferred from the information from the total testing program. The program developers found that all tests in the program had reliability coefficients greater than .90 on the Kuder-Richardson Formula 20. The basis for validity of the various tests was their predictive validity for groups (Buros, 1972, pp. 1031-1032).

**Moral maturity measure.** The Kohlberg measure for determining the moral stages of individuals was constructed from four of the nine hypothetical dilemmas and interview questions found in his dissertation (Kohlberg, 1958, pp. 361-375). Stories II, V, VI, and IX were rewritten into "adult" form, updated, and made into a paper and pencil open-ended questionnaire. These particular stories were chosen because the situations they presented were unfamiliar to the teachers. The validity question of this measure was tentatively answered by Kohlberg on the basis of construct validity. The reliability of the nine dilemmas had been reported ranging from .76 to .31 among the situations with the median value being .51 (Kohlberg, 1969, pp. 369-397). The Spearman-Brown prophecy formula for split half correlations (odd-even) was used to estimate the reliability of the Moral Maturity Scores resulting from the four dilemmas used in this study. The resulting coefficient was .42.

The responses to the measure were stage scored with the aid of the global rater guide found in the Appendix of Kohlberg’s dissertation (Kohlberg, 1958, pp. 376-428). Each stage profile was converted into a Moral Maturity Score (see earlier discussion). Interrater reliability coefficients using Pearson product moment correlations were obtained on a random sample (30%) of the responses of the teachers. Three different forms of reliability were obtained. First, a reliability of .91 was found on assigning major stages to the teachers. Second, a reliability of .76 was found on assigning major and minor stages to the teachers. Third, a reliability of .70 was found on assigning Moral Maturity Scores to the teachers.

**Receptive ability measure.** The Rokeach dogmatism test form E was used. The measure had construct validity and a test-retest reliability ranging from .93 to .68. The original scoring used by Rokeach had a scale of +3, +2, +1, -1, -2, -3 in which +3 represented a totally close-minded response and -3 represented a totally open-minded response. In this study the scale was modified to +1 through +6 in which +1 represented a totally close-minded response and +6 represented a totally open-minded response. This change reversed the scale so that the higher scores indicated open-mindedness and the lower scores indicated close-mindedness (Rokeach, 1960, pp. 71-97).

**Stage scoring ability measure.** The testing instrument used to measure the ability to stage score moral thought statements (termed Moral Knowledge
Test) was constructed from typical stage responses to story III of Kohlberg's hypothetical moral dilemmas (Kohlberg, 1958, pp. 363-364). There were 24 items with each of the six stages represented by 4 items. In each stage, 2 of the 4 items represented "Orientations to Intentions and Consequences," and 2 of the 4 items represented "Motives for Engaging in Moral Action." "Orientation to Intentions and Consequences" and "Motives for Engaging in Moral Action" were the concepts of moralization that were used in the test. For each concept at each stage, 1 of the 2 items had a do choice and the other had a don't choice. These 24 items had an equal distribution of stage statements, an equal distribution of do and don't choices, and a sample of the concepts used in moralization (see Note). These items were adapted from the work of James Rest (Kohlberg, 1969, pp. 379-382). The test was scored by given +1 for each correct and 0 for each incorrect stage scored item.

The test had content validity based on the sampling of the types of stages, choices, and concepts used in moral thought statements. A form of "face" validity was, also, assumed because both Rest and Kohlberg agreed that these statements represented one of the six stages, one of the choices, and one of two concepts of moral thought described above (Kohlberg, 1969, pp. 370-382). The reliability of the measure was estimated by using the Cronbach alpha projected to a standard test (100 items). A reliability coefficient of .62 was found for the scores on the pretest, and a coefficient of .48 was found for the scores on the posttest. Although these reliability coefficients were low, they were adequately reliable for a group size of sixty teachers (Thorndike & Hagen, 1969, pp. 194-195).

Research Design and Procedures

Data collection. The data of age, sex, teaching experience, and verbal score on the Graduate Records Examination were obtained from the information folders for the experimental Master of Education program. The data on Moral Maturity Score, Rokeach dogmatism score, and correct scores for Moral Knowledge Test I (pretest) and Moral Knowledge Test II (posttest) were obtained over a period of three weekend sessions of the program. At the first session, the Rokeach measure was given, and the Kohlberg measure for determining moral maturity was handed out to be returned at the second session. At the second session, the teachers were given Moral Knowledge Test I (pretest) using only the stage description materials provided. The description was given to the teachers in order to preclude their having to memorize the descriptions of Kohlberg's six stages. Prior to administering Moral Knowledge Test I, the teachers were given information about Kohlberg's approach to moral development and moral education. Specifically, the teachers were shown a filmstrip (Kohlberg & Selman, 1972), given two articles (Kohlberg & Mayer, 1972; Kohlberg &
Whitten, 1972), and directed in discussions during two weekend sessions. Finally, most of the teachers (88%) had tried intuitively to stage score the moral thought statements of their students prior to the pretest, but had expressed difficulties about their attempts. After the pretest was returned, the teachers were instructed to take the same test home and use both the stage descriptions and the adapted version of the Porter and Taylor global rater manual provided and to stage score the same 24 moral thought statements (Moral Knowledge Test II). Two weeks later, the teachers returned Moral Knowledge Test II at the third class session.2

Two observations about the data collection method were made. First according to Kohlberg, the responses to hypothetical moral dilemmas and questions uncovered the underlying structure of moral thought and unless an individual understood this way of thinking, it could not be used properly (Kohlberg, 1971). So, taking the Kohlberg measure under an unsupervised condition was assumed to have no effect on the resulting Moral Maturity Scores for the teachers. Second, the testing influence from the Moral Knowledge Test I to the Moral Knowledge Test II was not deemed a contaminating factor. If the teachers had been given only the Moral Knowledge Test II, the teachers would have been able to do the exercise as many times as they wished before returning it because they were given two weeks to complete the exercise. Thus a prior exercise (Moral Knowledge Test I) was assumed to have no real influence on the results of Moral Knowledge Test II.

As with any social science research, the problem of missing data occurred. Despite all efforts, a complete data bank for all teachers was not obtained. Independent t tests were used to determine if there were significant differences between the correct scores on Moral Knowledge Tests I and II for the 50 teachers on whom all data was known and the 0 teachers on whom data was missing. No significant differences were found. Therefore, it was concluded that the missing data represented random loss.

**Statistical analyses.** The first question studied was:
1. Is there a significant linear correlation between the selected demographic and personality variables of the teachers and their ability to stage score moral thought statements and different stages of moral thought statements?

To test this question, Pearson product moment correlations on age, teaching experience, verbal ability, Rokeach score, and Moral Maturity Score with overall and stage correct scores for Moral Knowledge Test I and II were made. Since sex was a true dichotomy, point biserial correlations were used to test the relationship with correct scores for Moral Knowledge Test I and II. Procedures used were found in Bruning and Kintz (1968).

The second and third questions studied were:
1. Is there a significant difference between Moral Knowledge Test I and II in the ability of teachers to stage score moral thought statements and
different stages of moral thought statements?

2. Is there a significant difference within Moral Knowledge Test I and II in the ability of the teachers to stage score different stages of moral thought statements?

To test these questions, a three factorial analysis of variance test with \( n \) observations per cell, analysis of simple main effects tests and Newman-Kuels Sequential Range tests were used. The factors of Test and Stage were assumed fixed effects while the factor of Subject was assumed a random effect. The \( n \) observations per cell were the 4 items at each stage found in the Moral Knowledge Test. Procedures used were found in Winer (1971).

All tests of the null hypotheses associated with each question used a two tail test at the \( p = .05 \) level of significance.

RESULTS

Descriptive Statistics

The descriptive statistics for the dependent and independent variables are presented in Table 1. The demographic variables indicated that the teachers

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Standard Deviation</th>
<th>Range</th>
<th>Number Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>32.9</td>
<td>31.5</td>
<td>26.0</td>
<td>7.5</td>
<td>22.0</td>
<td>0</td>
</tr>
<tr>
<td>Teaching Experience</td>
<td>8.2</td>
<td>5.3</td>
<td>2.0</td>
<td>7.4</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>GRE Verbal Score</td>
<td>394.3</td>
<td>378.3</td>
<td>400.0</td>
<td>91.0</td>
<td>to</td>
<td>6</td>
</tr>
<tr>
<td>Rokeach Score</td>
<td>153.4</td>
<td>155.5</td>
<td>156.0</td>
<td>22.9</td>
<td>to</td>
<td>2</td>
</tr>
<tr>
<td>Moral Maturity Score</td>
<td>424.5</td>
<td>433.3</td>
<td>483.0</td>
<td>55.6</td>
<td>213.0</td>
<td>3</td>
</tr>
<tr>
<td>Moral Knowledge I Score</td>
<td>7.67</td>
<td>7.75</td>
<td>8.00</td>
<td>2.54</td>
<td>500.0</td>
<td>0</td>
</tr>
<tr>
<td>Moral Knowledge II Score</td>
<td>7.85</td>
<td>7.50</td>
<td>7.00</td>
<td>2.25</td>
<td>14.00</td>
<td>0</td>
</tr>
</tbody>
</table>

*There were 9 Male and 51 Female teachers in the study.*
were, on the average, female, around 33 years of age, and had 8 years of
teaching experience. Further, the Graduate Records Examination scores
indicated that the teachers were more than one standard deviation below the
norm mean of 500 on the measure of verbal ability. Although no norms
were available for the Rokeach dogmatism test, the design of the measure
provided for an open-minded score to fall within the range of 140 to 240.
Therefore, as a group these teachers were open-minded. The Moral
Maturity Score also had no norm reference. Since a Moral Maturity Score
of 400 was equal to a pure stage 4 (i.e., 4 x 100) and the mean Moral
Maturity score for these teachers was 424.5, it was noted that as a group
these teachers were thinking on the principle level of moral thought at
times. The score for Moral Knowledge Test I showed that the teachers
correctly stage scored only one-third of the 24 moral thought statements.
The results for scores of Moral Knowledge Test II were almost exactly the
same as for Moral Knowledge Test I.

The frequency distribution of stage scores assigned by the sixty teachers
to each moral thought statement in Moral Knowledge Test I and II are given
in Table 2. The table was constructed to reorder the 24 test items in order to
group by and identify according to stage, choice, and concept. The
frequency distribution for Moral Knowledge Test I demonstrated that these
teachers did not correctly assign stage scores to each of the 4 items within
each stage equally. Also, there was a difference in correct stage score
assignment between stages. In addition, when teachers incorrectly stage
scored different stage level moral thought statements, they assigned stage
scores at any of the other stage levels. The same general observations were
made by inspecting the frequency distribution for Moral Knowledge Test II.
Further discussion of these results is made in the next section.

Statistical Tests

First question. Pearson product moment correlations and point biserial
correlations of demographic and personality variables to the overall scores
of Moral Knowledge Test I and II produced only one significant result.
Verbal ability (r = .43) significantly correlated with the overall score for
Moral Knowledge Test I. No significant correlations were found with
overall scores for Moral Knowledge Test II.

Pearson product moment correlations were also done between the
demographic and personality variables and scores for each stage within each
Moral Knowledge Test. The results tended to conform to the ones found in
the correlations with overall scores for the Moral Knowledge tests. for
Moral Knowledge Test I, verbal ability significantly correlated with Stage 2
(r = .29), Stage 3 (r = .31), and Stage 6 (r = .35). Again, there were no
significant correlations with verbal ability and scores for each stage within
Moral Knowledge Test II. Four other significant correlations were
### Table 2

**Frequency Distribution of Stage Scores Assigned**

<table>
<thead>
<tr>
<th>STAGE</th>
<th>MORAL KNOWLEDGE I</th>
<th>STAGE Assigned</th>
<th>MORAL KNOWLEDGE II</th>
<th>STAGE Assigned</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DO INTENTION</td>
<td>STAGE CHOICE</td>
<td>CONCEPT 1 2 3 4 5 6</td>
<td>TOTAL 1 2 3 4 5 6</td>
</tr>
<tr>
<td>1</td>
<td>3 25 12 13 4 60 1 31 17 0 9 2 60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DON'T INTENTION 16</td>
<td>6 9 25 1 3 60 30 0 0 28 2 0 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DO MOTIVE 23</td>
<td>5 12 16 2 2 60 24 12 9 5 8 2 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DON'T MOTIVE 44</td>
<td>9 1 5 0 1 60 57 1 0 2 0 0 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 1 TOTALS</td>
<td>86 45 34 49 16 10 240 112 44 26 35 19 4 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6 6 9 25 3 60 30 0 0 28 2 0 60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DON'T INTENTION 23</td>
<td>4 6 2 6 2 60 24 12 9 5 8 2 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DON'T MOTIVE 44</td>
<td>9 1 5 0 1 60 57 1 0 2 0 0 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 2 TOTALS</td>
<td>31 86 39 32 39 13 240 34 77 54 34 10 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DO INTENTION 3</td>
<td>19 24 4 8 2 60 1 23 29 2 3 2 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DON'T INTENTION 13</td>
<td>24 8 9 5 1 60 16 14 1 27 1 1 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DO MOTIVE 23</td>
<td>5 11 12 22 8 3 60 18 9 6 24 2 1 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DON'T MOTIVE 44</td>
<td>9 27 13 6 1 60 14 7 9 22 5 3 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 3 TOTALS</td>
<td>31 86 39 32 39 13 240 34 77 54 34 10 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DO INTENTION 3</td>
<td>7 10 10 22 8 60 0 20 24 3 10 3 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DON'T INTENTION 13</td>
<td>24 8 9 5 1 60 16 14 1 27 1 1 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DO MOTIVE 23</td>
<td>5 11 12 22 8 3 60 18 9 6 24 2 1 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DON'T MOTIVE 44</td>
<td>9 27 13 6 1 60 14 7 9 22 5 3 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 4 TOTALS</td>
<td>31 86 39 32 39 13 240 34 77 54 34 10 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DO INTENTION 3</td>
<td>5 5 6 29 12 6 0 3 7 3 30 17 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DON'T INTENTION 13</td>
<td>2 3 51 10 13 6 0 0 2 42 10 6 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DO MOTIVE 23</td>
<td>7 7 29 7 8 2 60 0 7 28 4 15 6 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DON'T MOTIVE 44</td>
<td>9 27 13 6 1 60 14 7 9 22 5 3 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 5 TOTALS</td>
<td>31 86 39 32 39 13 240 34 77 54 34 10 240</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DO INTENTION 3</td>
<td>1 2 6 19 32 6 0 0 0 1 22 37 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DON'T INTENTION 13</td>
<td>5 9 9 17 19 6 0 1 4 2 8 23 22 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DO MOTIVE 23</td>
<td>5 5 9 20 16 6 0 6 7 6 19 16 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DON'T MOTIVE 44</td>
<td>9 27 13 6 1 60 14 7 9 22 5 3 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAGE 6 TOTALS</td>
<td>31 86 39 32 39 13 240 34 77 54 34 10 240</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The concept of "Intention" referred to "Orientation to Intentions and Consequences" and the concept of "Motive" referred to "Motives for Engaging in Moral Action."*
uncovered. For Moral Knowledge Test I, the Moral Maturity Score significantly correlated with Stage 2 \((r = .34)\), and the Rokeach dogmatism score significantly correlated with Stage 1 \((r = .26)\) and Stage 4 \((r = .28)\). Point biserial correlations between sex and scores for each stage were not done because of the insignificant correlations with overall scores for Moral Knowledge Test I and II.

In summary, the correlation tests demonstrated that verbal ability was the only significant correlation with overall performance on Moral Knowledge Test I, and no significant performance on Moral Knowledge Test I, and no significant correlations were found for overall performance on Moral Knowledge Test II. This finding for overall performance was supported by the findings for correlations with stage performance in Moral Knowledge I and II. However, it must be noted that none of the significant correlating variables had more than a moderate level of correlation with the overall and stage scores for Moral Knowledge Test I and II.

**Second and third questions.** The results of the analysis of variance test is presented in Table 3. The analysis between the main effects of Test showed no significant difference between scores for Moral Knowledge Test I and II.

Table 3: Analysis of Variance of Moral Knowledge Test I and II

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>17.604</td>
<td>59</td>
<td>.2984</td>
<td></td>
</tr>
<tr>
<td>A (Test)</td>
<td>.017</td>
<td>1</td>
<td>.0170</td>
<td>.094</td>
</tr>
<tr>
<td>A x Subject</td>
<td>10.629</td>
<td>59</td>
<td>.1802</td>
<td></td>
</tr>
<tr>
<td>B (Stage)</td>
<td>11.022</td>
<td>5</td>
<td>2.2044</td>
<td>10.438*</td>
</tr>
<tr>
<td>B x Subject</td>
<td>62.290</td>
<td>295</td>
<td>.2112</td>
<td></td>
</tr>
<tr>
<td>AB</td>
<td>2.665</td>
<td>5</td>
<td>.5330</td>
<td>3.125*</td>
</tr>
<tr>
<td>AB x Subject</td>
<td>50.314</td>
<td>295</td>
<td>.1706</td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>481.750</td>
<td>2160</td>
<td>.2230</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>636.291</td>
<td>2879</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at \( p < .05 \)

Because the second questions required a test for differences between Tests at each Stage level and since the interaction of Text x Stage was significant (see Table 3), a test of simple main effects was performed. Table 4 presents the results of these tests. The analysis of Test at each Stage level showed that
Stages 1 and 3 were significantly different between Moral Knowledge Test I and II. Referring to Table 2, the teachers improved on Stage 1 items but declined in correctly assigning stage scores for Stage 3 items.

Since the third question asked for the determination of differences between Stages at each Test level and the interaction between Test x Stage was significant (see Table 3), a test of simple main effects for Stages at Test levels was performed (Table 4). For both Test levels there was a significant
difference between Stages. A Newman-Keuls test was performed on the total correct stage scores for both Moral Knowledge Test I and II (Table 5). The results for Moral Knowledge Test I showed (a) Stage 3 was significantly different from Stage 5 and 4; (b) Stage 1, 2, and 6 were significantly different from Stage 5. No other significant differences were noted. The results for Moral Knowledge Test II showed (a) Stage 1 was significantly different from Stage 2, 3, 4, 5, and 6; (b) Stage 6 and 3 were significantly different from Stage 4.

In summary, the overall test for differences between Moral Knowledge Test I and II showed that the overall performance on the two tests were not
Table 5: Newman-Keuls Sequential Range Test for Stages at Test Levels

Moral Knowledge I

<table>
<thead>
<tr>
<th>Tj</th>
<th>57</th>
<th>67</th>
<th>81</th>
<th>86</th>
<th>86</th>
<th>101</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_1</td>
<td>B_4</td>
<td>B_6</td>
<td>B_2</td>
<td>B_1</td>
<td>B_3</td>
<td>γ</td>
</tr>
<tr>
<td>B_3</td>
<td>—</td>
<td>24*</td>
<td>29*</td>
<td>29</td>
<td>44* — 6</td>
<td>29.12</td>
</tr>
<tr>
<td>B_4</td>
<td>14</td>
<td>19</td>
<td>19</td>
<td>34* — 5</td>
<td>27.83</td>
<td></td>
</tr>
<tr>
<td>B_6</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>20 — 4</td>
<td>26.20</td>
<td></td>
</tr>
<tr>
<td>B_2</td>
<td>0</td>
<td>15</td>
<td>3</td>
<td>15 — 3</td>
<td>23.92</td>
<td></td>
</tr>
<tr>
<td>B_1</td>
<td>15 — 2</td>
<td>19.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Moral Knowledge II

<table>
<thead>
<tr>
<th>Tj</th>
<th>55</th>
<th>60</th>
<th>77</th>
<th>82</th>
<th>85</th>
<th>112</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_4</td>
<td>B_3</td>
<td>B_2</td>
<td>B_1</td>
<td>B_6</td>
<td>B_1</td>
<td>γ</td>
</tr>
<tr>
<td>B_4</td>
<td>—</td>
<td>5</td>
<td>22</td>
<td>27*</td>
<td>30*</td>
<td>57* — 6</td>
</tr>
<tr>
<td>B_3</td>
<td>17</td>
<td>22</td>
<td>25</td>
<td>52* — 5</td>
<td>27.83</td>
<td></td>
</tr>
<tr>
<td>B_2</td>
<td>5</td>
<td>8</td>
<td>35* — 4</td>
<td>26.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B_3</td>
<td>3</td>
<td>30* — 3</td>
<td>23.92</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B_6</td>
<td>27* — 2</td>
<td>19.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Significant at p < .05

Different. Inspection of performance on stages between and within each Moral Knowledge Test indicated that there were some differences. These latter findings support the conclusion that the Porter and Taylor global rater guide did have some effect on the performances between Moral Knowledge I and II, but did not aid the teachers in changing overall performance.

DISCUSSION

Limitations

The results and conclusions of this study are limited to the sample, measurement instruments and procedures used. The subjects may not have
been typical of all elementary school teachers, and the conclusions made apply only to those teachers like the ones in this study. Furthermore, the measurement instrument used to determine moral maturity and the Moral Knowledge Tests had low estimates of reliability. These reliability problems qualify the interpretations of results and the conclusions. Because these measures may have produced inaccurate information, results associated with these measures may have been susceptible to errors in measurement. Finally, generalizations are restricted to the unsupervised self training global rater manual treatment used. Other treatments may produce different results and conclusions.

Conclusions

This study supported the generalization that teachers cannot stage score moral thought statements with an adequate degree of correctness by using the Porter and Taylor global rater manual and self training. However, information concerning factors related to the ability to stage score moral thought statements was meager. None of the demographic and personality variables examined were strongly related to stage scoring ability in either Moral Knowledge Test I or II. These variables were studied as being related linearly to the ability to assess moral thought statements. It could be that they do not relate in a linear fashion, or there are other more strongly correlated variables that were not investigated.

A rater manual is supposed to equalize raters so that differences in demographic and personality variables are controlled. Although the variable of verbal ability, which significantly correlated with Moral Knowledge Test I, was removed with Moral Knowledge Test II, the teachers did not improve in their overall stage scoring abilities. There was something influencing the teachers in such a way as to cause them to do poorly in stage scoring moral thought statements. Tests on differences in stage performance between and within tests indicated that the Porter and Taylor global rater manual and self training had an effect. However, the effect was merely to help stage score Stage 1 moral statements better and Stage 3 statements worse between the two tests.

At present, the exact reasons for these results are not evident. However, reviewing the distribution of stage scores assigned the 24 moral statements by the teachers raises some interesting points. One of the major tenets of stage scoring was that a rater look at the “thinking” used to justify a choice made and not what choice was made or what aspects were used in justifying the choice made (see: Kohlberg, 1971, pp. 163-180). The differences between correct stage scores of items at each stage may indicate that the teachers were stage scoring on the basis of the content of a moral statement (i.e., the choice and/or aspects) rather than the “thinking” used in a moral statement (i.e., how the aspects were used to defend the choice made).
is, the teachers may have had a certain choice and/or aspect expectation for each stage. The changes in correct stage score for certain Stage 1 and 3 moral thought statements between Moral Knowledge Test I and II may have been the result of the Porter and Taylor global rater manual "equalizing" the teachers' expectations of the appropriate content at each stage rather than aiding the teachers in stage scoring the "thinking" used at each stage. This speculation that the teachers may have been influenced by the content of moral thought statements should be examined in future research as one of the possible factors related to stage scoring ability.

Whatever the reasons why these teachers assigned stage score, it is evident that the Porter and Taylor manual and the self training were inadequate in helping these teachers successfully use the global rating process of stage scoring moral thought statements. Therefore, research needs to be done on ways to assist teachers to stage score moral thought statements (i.e., a better rater manual and training). Until research indicates that teachers can properly stage score moral thought statements, the use of the advocated instructional procedures requiring stage scoring of moral thought statements of students should not be used. It is imperative that social studies educators refrain from indicating that teachers can correctly stage score the moral thought statements of their students. For if teachers tried to stage score the moral thought statements of their students, they could incorrectly stage score Stage 1 moral thought as Stage 6, or Stage 6 as Stage 1.

Therefore, until better materials and procedures are developed for teachers, some logical alternative is needed to replace the instructional procedure of stage scoring moral thought statements in order to provide models of "next higher stage." This alternative should be one which would enhance the probability of exposing students to the next higher stage without requiring teachers to stage score moral thought statements.

REFERENCES


FOOTNOTES

1 In order to demonstrate the items on the Moral Knowledge Test, the following 4 examples are given. The examples are similar to the items used on the Moral Knowledge Test and represent the four basic contents and four of the six stages used on the Test:

Stage 1, Choice do, Concept of “Orientation to Intentions and Consequences”:

Heinz should steal the drug for his dying wife. It isn’t actually bad to take the drug. It isn’t like Heinz didn’t ask the druggist if he could pay later. The drug that Heinz would take is only worth $200. So, he isn’t really taking a $2000 drug.

Stage 2, Choice don’t, Concept of “Orientation to Intentions and Consequences”:

Heinz should not take the drug for his dying wife. The druggist isn’t wrong or evil, he just wants to sell the drug for a profit. That is what a person is in business to do, to make a profit.

Stage 3, Choice do, Concept of “Motives for Engaging in Moral Action”:

No one will think Heinz is bad if he steals the drug for his dying wife, but his family will think he is an inhuman husband if he doesn’t steal the drug. If Heinz lets his wife die, he will never be able to look anybody in the face again.

Stage 4, Choice don’t, Concept of “Motives for Engaging in Moral Action”:

Heinz is desperate and he may not know he is wrong when he steals the drug for his dying wife. But he will know he did wrong after he is punished and sent to prison. He will always feel guilty for his dishonesty and lawbreaking.

For a more detailed description of the 24 items, the reader is referred to the discussion by Rest and Kohlberg (Kohlberg, 1969, pp. 370-382).

2 No formal investigation was conducted as to whether the teachers used the Porter and Taylor self training global rater manual for Moral Knowledge Test II. Because of the simplicity of the adapted version of the Porter and Taylor manual, it was believed logical to assume that the
teachers would use the manual. The adapted version of the Porter and Taylor manual was returned to the researcher by the teachers after the experiment. It was observed that all the manuals returned had been handled, and therefore, most likely used. In addition, informal discussions between course instructors and teachers after the experiment further convinced this researcher that the teachers had used the manual during the study.
Decisions to develop innovative curriculum materials in a particular subject area represent the beginning stages of formative curriculum evaluation (Scriven, 1967). They imply judgments about the inadequacies of existing curricula, the ability of curriculum developers to diagnose pupil needs effectively and to develop appropriate teaching strategies and course content. This conception of curriculum, specifically in terms of answering Tyler's (1949) basic questions about the purposes of education, the identification and organization of specific educational experiences, and measurement to determine whether or not the purposes are in fact being attained, characterizes much of the curriculum development work in social studies.

Whereas earlier social studies development projects focused on a single discipline (e.g., political science, economics, anthropology), the Family Life Curriculum Project is interdisciplinary; it draws from history, sociology, and anthropology for its content. More importantly, this project converges on a topic of vital societal concern: the condition of the family as an institution. With one exception (American Sociological Association, 1968) most other projects did not consider the study of the family as pertinent to social studies. This omission occurred despite the growing body of social science research on the family, child development and child rearing, male and female roles, courting, mate selection and marriage which appears in the historical, psychological, sociological and anthropological literature.

The apparent neglect of these topics in the high schools may have stemmed partly from a conviction that family life is too private a matter for educators to teach, and partly because home economics education rather than social studies has traditionally served as the single source for family life information in secondary schools. Since such courses exist without links to the disciplines of sociology, psychology (Allen and King, 1970), history and anthropology, and lack a theoretical framework drawn from the social sciences, they do not systematically examine the condition of the family in modern societies. Hence, a need was perceived for a family curriculum which is interdisciplinary and cross-cultural.

*The work reported in this paper was made possible by a grant from the Buhl Foundation.
This need was reinforced by evidence drawn from contemporary life. Frequently cited statistics on the American family reveal the trauma afflicting this institution and provide evidence of significant changes which have occurred in the American family in the recent past (Bronfenbrenner, 1974). In addition, the literature on adolescent development reveals that youth are generally ignorant about such vital family-related issues as sex, courtship, marriage and parenthood (Bronfenbrenner, 1974; Kerchkoff, 1973). When queried about these issues, the information they possess is largely incorrect (Thornburg, 1974). Yet, adolescents, their parents and teachers agree that the schools should provide up-to-date, honest, non-judgmental curriculum materials about sex mores, life styles, working mothers, day care, and parent-adolescent relationships (Kerchkoff, 1973; Liffy, Acock, Payne, 1974; Somerville, 1972) and help students clarify their own attitudes toward sexuality, sex roles and the nuclear family (Bayer, 1975; Dunn, 1960).

This paper describes the development, trial and experimental evaluation of four family life courses for secondary school students. It presents a formative evaluation strategy for assessing the impact of these courses on the knowledge, skills, attitudes and preferences of students, using feedback from the developers, teachers and students. The strategy amplifies the typical formative evaluation process used for curriculum development. This is the development-trial-evaluation-revision cycle (Scriven, 1967). The cycle of course development and evaluation reported here was based on a sequential approach which involves developing a first set of two courses, testing, evaluating and revising these, and then using the results from these courses to aid in designing the next set of courses. This strategy is illustrated in Figure 1.

Four general goals were identified by the project staff in developing new course materials. The first, namely cognitive development, is to present students with accurate information about adolescent, courtship, marriage, and parenthood, and to develop their comprehension, application and analysis skills. The second goal, namely affective development, is to assist adolescents with the task of shaping their own believes and values about the roles of family members and the functions of the family unit within a societal perspective. Third, students should find the material interesting, attractive and challenging. Finally, teachers should find the materials clear and easy to use. These multiple goals reflect concern for the “differentiated outcomes” expected from traditional educational inputs and special outcomes from innovative curricula (Tuckman, 1975; Walker and Shaffarzick, 1974).

In order to accomplish these goals several inputs were used. The staff of the Family Life Curriculum Project argued that the closeness of the courses to the developmental concerns of adolescents would stimulate them to learn. Each course should contain a social science framework on the role
relationships of adolescents to siblings and adults in the family. In addition, each course should include a substantial emphasis on contemporary life, while including a historical component. The staff believed that if students liked the courses, they would be motivated to learn from them. Further, the courses should stimulate discussion among peers and with parents thereby fostering the development of well-founded beliefs about family life. Richard and varied visual components were used to reduce the routine of daily reading assignments.

Four six-week courses served as the means for combining these inputs in varying proportions as shown in Table 1. Each course contains a text,
worksheets and tests for students, and a teacher’s guide. All four courses emphasize the family as a universal social institution which takes on variable forms according to historical and cultural context. “Family Life in Two Societies: Japan and the Kibbutz” by Stuart Lazarus, hereafter referred to as Course 1, and “American Youth: Past and Present” by Mark Tierno, hereafter referred to as Course 2, were developed, field-tested, and evaluated during the project’s first year. Both courses employ a conceptual structure referred to by developmental sociologists as the family life cycle, and examine the changing norms, roles, role clusters, and socialization processes of family members in historical and contemporary contexts. The materials stress these concepts and require the use of inquiry skills to classify and analyze evidence about the topic. Vivid primary sources in the form of documents, letters, diaries, statistics, sketches, drawings and photographs are used as curriculum materials to link abstract concepts to information.

After these courses were evaluated, two more courses: “Courtship and Marriage: The American Experience” by James Paces, called Course 3, and “Perspectives on the American Family” by Gerald Clarke, Course 4, were developed. Course 3 presents students with information about dating and marriage patterns during two historical periods and the present. Students examine the roles and behavior of dating adolescents and young adults preparing for marriage, comparing dating and marriage patterns prevalent in the past with contemporary phenomena to detect continuity and change. Course 4 emphasizes family forms and functions, kin relationships and changing roles for family members in a modern and historical context.
We used a modified four-group experimental design (Campbell and Stanley, 1966) to assess course impact on cognitive and attitudinal test scores. Since we could not randomly assign students to treatment and control groups, we refer to the research design as a modified four-group design. Group 1 in this design receives pre-tests, then the curriculum and then is retested at the conclusion of the field trial. Group 2 takes the pre-tests and post-tests, but no curriculum materials. Group 3 takes the post-test only. Students in Group 4 receive the curriculum materials and the post-tests.

For each post-test cognitive and attitudinal measure we wished to posit a model which can be used to explain the values taken by the measure and which simultaneously could be used to estimate the effect of the experimental course on that measure. The standard linear analysis of variance model is \( y = Xb + e \), where \( y \), the dependent variable, is the post-test measure, \( X \) is the design matrix, \( b \) is the unknown coefficients, and \( e \) is random error. For our experiment the design matrix based on the modified four-group experimental design includes a mean, a curriculum effect, a pre-test effect, and an interaction between the pre-test and curriculum. An interaction may be reasonably expected because identical pre- and post-tests were administered, and thus the experimental Group 1 results may demonstrate "retest effects."

The Solomon four-group design is such that the complete model may be readily interpreted as a problem in ordinary least squares regression. In particular, each type of effect is reflected in the design as a single variable assuming binary values. Furthermore, there are no side constraints imposed on the coefficients—the design matrix is of full rank.

The mean score of all students in a particular classroom, rather than the scores of the individual students is used as the independent variable in order to lessen the influence of individual differences on the analysis. However, classes within a school demonstrate sufficient variability so that correction for inter-class differences is necessary. In particular, classes differ in years of schooling (grade) of their students, a variable which may also be considered a surrogate for age. The general level of scholastic achievement varies from class to class, especially where "tracking" is operative, but also because of possible self-selection by students motivated by teacher or course reputation, course title or schedule conflicts. Sex composition of classes may vary, often due to the same factors as above with a tendency for women to select family life courses (Mason, 1974). Thus, we employ mean year in school, grade point average and sex as control variables.

The schools from which our sample classes are drawn differ dramatically in the demographic characteristics of the populations they serve, and these
differences appear as inter-school and hence inter-class differences. There is thus strong *a priori* justification for correcting for school differences, which we do with dummy (0 or 1) variables for schools included in the analysis of each course. Including the control variables as covariates, \( Z \), our model is thus \( y = Xb + Zc + e \), the standard model for a linear analysis of covariance.

A "no effect" hypothesis is a statement that an estimated coefficient is zero. The test of this hypothesis is, of course, what is done in an analysis of covariance, but is more readily obtained, probably more familiar, and certainly more easily presented in the equivalent regression context.

Our remarks concerning the general structure of the model pertain equally well to the analysis of either cognitive or attitudinal measures. We recognize the existence of more parsimonious sub-models, but report the results for the complete model because of its strong *a priori* justification. In particular, deletion of the interaction term causes no qualitative change in our conclusions.

Each course was tested in at least three schools. Six schools participated in one or more courses, and of these, one school was exposed to all four courses. In all, the experiment for Course 1 involved 25 classes in three schools, Course 2 involved 26 classes in three schools, Courses 3 and 4 involved 18 classes each in four schools. The number of students enrolled in a course ranged from a low of 394 in Course 3 to a high of 496 in Course 2. Average class size varied from 17 in Course 1 to 23 in Course 4.

The six schools in this study cover the spectrum from a low-income inner city school to an elite suburban school. School A, one of the largest public high schools in Pennsylvania, is located in a working class satellite city of Pittsburgh. School B serves a white working class ethnic industrial suburb of Pittsburgh. Schools C and D are in an elite residential municipality adjoining Pittsburgh. School E serves an inner city population in the Pittsburgh public school system. School F is in a lower middle class suburb of Cleveland. All are high schools except for School D which is a junior high school.

All the schools were predominantly white; School E, with 29 percent black students, was the most racially mixed. The percent of students who eventually attend college ranged from School E with 15 percent to School C with 80 percent. We selected classes from the various grades 7 through 12 in order to ascertain empirically the appropriate level of each course, but the majority of students were tenth graders. In most of the classes about half the students were women, but Course 1 in School A had two-thirds women, while Course 3 in Schools C and F had more than three-fourths women.

The courses were fitted into the social studies course offerings. All the teachers in the experiment were volunteers and all were experienced social studies teachers. The developers of all four courses were placed in School B as teachers of their own materials, as in-service directors instructing six of
the school's social studies teachers in proper use of these materials, and as supervisors of these teachers during the actual field trial. Thus, School B teachers received intensive in-service training. A few classroom observations and site visitations before and during the field trial represented the only interventions in the other schools. Except for the teachers in School B, none of the remaining 50 teachers received special preparation.

The cognitive tests for each of the four experimental courses were developed in a similar manner. Each developer constructed a pool of multiple four-choice questions closely keyed to the content of his course. A pilot test using this initial form was administered to approximately one hundred high school students and college freshmen, and the results analyzed critically both subjectively and according to standard item analysis techniques (Thorndike, 1971). Poorly worded items, items answered correctly by too few or too many students, and items containing under or over-utilized incorrect response categories were rewritten or discarded. The resulting form was re-evaluated after its administration to over two hundred students; a few additional items were eliminated from the scoring as a result of this analysis, but the test form was not altered.5

A goal of each course was to develop students' inquiry skills, in addition to simply expanding their knowledge. Items on each test were classified accordingly using Bloom's Taxonomy (1956). Test questions demanding knowledge only correspond to Bloom’s Level 1; those we consider as requiring comprehension, application and analysis fall in Level 2 and above. We employ this partition in analysis of the cognitive tests, referring to cognitive Level 1 and cognitive Level 2 and above, separately. The cognitive tests vary in length, ranging from 8 to 28 items. Scores on each of tests were normalized to lie between 0 and 1; however, no comparability between scores is intended.

Attitudinal measures differ between the 1974 courses (1 and 2) and the 1975 courses (3 and 4). For the first two courses three measures were employed. A 21-item Familism scale was adapted from the Family Scale of Sletto and Rundquish (1936). Identification with one's parents was tapped through two items developed by Simmons, Rosenberg and Rosenberg (1973)—"If someone said something bad about your mother, would you feel as if they had said something bad about you?", and a second question referring to the respondent's father. Responses were normalized for each measure to be between 0 and 1; a high school indicates a pro-family orientation for each measure.

For Courses 3 and 4 the developers collaborated in constructing their own attitude test. Pilot testing eliminated items which evoked extreme responses on five-point Likert scales. The resulting pre-test was administered and the results subjected to factor analysis. Two primary orthogonal factors emerged; the first indicated questions of sex role attitudes and the second included questions concerning the relationship of the adolescent with his or
her family, especially involving parental authority and family "togetherness." Items contributing significantly to either one of these two factors were used, with the appropriate sign, to form an unweighted sum for each of the two dimensions. The Sex Role Ideology measure is a composite of thirteen items, the Family Orientation measure uses ten; higher scores on the normalized measures indicate more modern sex role conceptions and a more pro-family orientation.

Student reactions to the curriculum units were obtained through an open-ended questionnaire and replies were categorized through content analysis. Responses to the "least-liked" and "most-liked" aspects of each of the four courses were among the data collected. Comments, criticisms, and suggestions for improvement were incorporated in refining the first pair of courses, and the second pair of courses benefitted from the earlier experience. This questionnaire was used again at the conclusion of the later courses to provide input for their improvement. In light of the useful comments received the first year, an additional evaluation questionnaire was given to the second year's groups. This questionnaire, developed by the Alpha Phi Omega fraternity, employs Likert-type scales and has been extensively used in evaluating university level courses. We sought additional evidence of course impact through questions about the extent to which students felt influenced by the material, specifically if they felt they had learned something about their own families, and if they had discussed the course with family members or peers.

Teachers were asked to complete a form each day about the effectiveness of each lesson, pupil achievement, reading level, use of various media, links to previous lessons, classroom activities and interaction. These daily evaluations were used each week to advise teachers on instructional strategies for subsequent lessons. The course developers reported detailed impressions based on their own classroom experience and the feedback they received from other teachers.

RESULTS OF THE TRIAL AND EVALUATION

Cognitive Performance

The results presented in Table 2 show that all four courses were effective in improving students' knowledge of concepts, facts and generalizations (Level 1) about the family. The curriculum effects are strong in each course but especially in Courses 1 and 2. In Courses 2 and 3, grades also influence Level 1 cognitive skills, so that the better students achieve higher test scores. While the explanatory model for all courses fits well as evidenced in the high F-ratios, the Course 4 curriculum effect is noteworthy because no other independent variables contaminate it.

When we consider the effects of the courses on Level 2 (inquiry skills) the
Table 2: Regression Results for Four Courses on Level One Cognitive Tests

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Course 1</th>
<th>Course 2</th>
<th>Course 3</th>
<th>Course 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum</td>
<td>0.27***</td>
<td>0.12***</td>
<td>0.10**</td>
<td>0.08**</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.03)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Pre-test</td>
<td>-0.05</td>
<td>-0.09**</td>
<td>-0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.03)</td>
<td>(.03)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.05)</td>
<td>(.05)</td>
<td>(.05)</td>
</tr>
<tr>
<td>Year in school</td>
<td>-0.06</td>
<td>0.00</td>
<td>-0.05</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.06)</td>
<td>(.04)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Grades</td>
<td>0.11</td>
<td>0.16**</td>
<td>0.25***</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(.10)</td>
<td>(.06)</td>
<td>(.06)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.08</td>
<td>0.19</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(.08)</td>
<td>(.10)</td>
<td>(.10)</td>
<td>(.12)</td>
</tr>
<tr>
<td>School A</td>
<td>-0.14***</td>
<td>-0.11***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School C</td>
<td>0.01</td>
<td></td>
<td>0.10**</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td></td>
<td>(.04)</td>
<td>(.07)</td>
</tr>
<tr>
<td>School D</td>
<td></td>
<td>-0.09</td>
<td></td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.22)</td>
<td></td>
<td>(.25)</td>
</tr>
<tr>
<td>School E</td>
<td></td>
<td></td>
<td>-0.23**</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.09)</td>
<td>(.06)</td>
</tr>
<tr>
<td>School F</td>
<td></td>
<td></td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.04)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.91</td>
<td>0.14</td>
<td>0.35</td>
<td>0.52</td>
</tr>
<tr>
<td>F-ratio</td>
<td>15.89***</td>
<td>9.56***</td>
<td>11.44***</td>
<td>11.46***</td>
</tr>
<tr>
<td>df</td>
<td>8/16</td>
<td>8/17</td>
<td>9/8</td>
<td>9/8</td>
</tr>
<tr>
<td>R² adj</td>
<td>.81</td>
<td>.69</td>
<td>.89</td>
<td>.89</td>
</tr>
</tbody>
</table>

**p < .05

For Tables 2-6, the regression coefficient or beta (B) for each independent variable is presented with the standard error of each B in parenthesis below it. Only those variables are considered acceptable predictors whose B is at least twice the standard error of B.
When we consider the effects of the courses on Level 2 (inquiry skills) the results are similar to Level 1 results but less strong (see Table 3). Course 4 curriculum effects on Level 2 test scores are clearcut and uninfluenced by the other independent variables. While Course 1 and 3 show curriculum effects, Course 2 shows none at all. Good students perform better on inquiry skills in Courses 2 and 3. School effects are less important on Level 2 scores except that School A students score significantly lower than the mean.

The cognitive test results point to the beneficial effects of all four courses on Level 1 performance. Three of the four courses affect Level 2 cognitive skills. We conclude, then, that the cognitive effects of these courses are clear and positive.

Although students' knowledge about the family and family-related topics was high before they enrolled in these courses (Angrist, Mickelsen, Penna, 1976), course impact on student learning in the cognitive domain is significant. Courses 1 and 2 show the greatest impact with Courses 3 and 4 following closely behind. These differences may be explained by the relatively new material in Course 1, namely the study of family life in Japan and the Kibbutz and the historical emphasis in both Courses. Because students know little of this material to begin with, it is not surprising that the curriculum could strongly influence test scores. The curriculum effect is weaker for Courses 3 and 4 which contain more contemporary material, although it is still significant.

The gains in the knowledge area probably reflect the teaching mode in subject matter courses. Traditionally, classroom teaching in social studies focuses on the transmission of a body of knowledge from teachers and texts to students. The acquisition of knowledge dominates teacher-student transactions; thus the mode of teaching for knowledge is familiar. Conversely, teaching for non-conventional goals, namely inquiry skill development, requires classroom methodology which facilitates the formulation of hypotheses and their verification, a mode with which teachers may have less experience. Although detailed and repeated prescriptive suggestions were included in the teacher's guides accompanying each of the four courses, teacher comments indicated that they seldom used the guides as models for classroom behavior. As a result, the scores in the area of inquiry skill development (Level 2) are more modest than the scores yield by knowledge acquisition (Level 1). While Course 2 showed no Level 2 effects, Courses 1 and 3 yielded modest curriculum effects on inquiry skills with somewhat stronger effects in Course 4. These stronger effects may be explained by the conscious decision of the developer of Course 4 to include in the curriculum materials inquiry terminology, questions and activities designed to develop students' ability to reason without a specific family-related topic.
Table 3: Regressions Results for Four Courses on Level Two Inquiry Tests

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Course 1</th>
<th>Course 2</th>
<th>Course 3</th>
<th>Course 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum</td>
<td>0.13*</td>
<td>0.03</td>
<td>0.11*</td>
<td>0.09**</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.05)</td>
<td>(.05)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Pre-test</td>
<td>-0.10</td>
<td>-0.10**</td>
<td>-0.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.11</td>
<td>0.04</td>
<td>0.02</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.06)</td>
<td>(.07)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Year in school</td>
<td>-0.10</td>
<td>-0.00</td>
<td>-0.00</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.08)</td>
<td>(.05)</td>
<td>(.07)</td>
</tr>
<tr>
<td>Grades</td>
<td>0.04</td>
<td>0.24***</td>
<td>0.30***</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(.14)</td>
<td>(.08)</td>
<td>(.09)</td>
<td>(.07)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.04</td>
<td>0.09</td>
<td>0.02</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>(.11)</td>
<td>(.14)</td>
<td>(.15)</td>
<td>(.13)</td>
</tr>
<tr>
<td>School A</td>
<td>-0.14**</td>
<td>-0.27***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School C</td>
<td>0.00</td>
<td>0.04</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.08)</td>
<td>(.06)</td>
<td>(.07)</td>
<td></td>
</tr>
<tr>
<td>School D</td>
<td></td>
<td>-.27</td>
<td></td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.29)</td>
<td></td>
<td>(.27)</td>
</tr>
<tr>
<td>School E</td>
<td></td>
<td></td>
<td>-0.18</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.13)</td>
<td>(.07)</td>
</tr>
<tr>
<td>School F</td>
<td></td>
<td></td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.57</td>
<td>0.12</td>
<td>-0.27</td>
<td>-0.88</td>
</tr>
<tr>
<td>F-ratio</td>
<td>4.47***</td>
<td>5.01***</td>
<td>8.90***</td>
<td>12.69***</td>
</tr>
<tr>
<td>df</td>
<td>8/16</td>
<td>8/17</td>
<td>9/8</td>
<td>9/8</td>
</tr>
<tr>
<td>R^2 adj</td>
<td>.48</td>
<td>.50</td>
<td>.87</td>
<td>.90</td>
</tr>
</tbody>
</table>

***p < .01  **p < .05  *p < .10
School Differences in Cognitive Performance

We expected to find school differences in cognitive performance for two reasons: (1) the disparity in school quality evident both in the academic reputations of the schools and in the proportions of their graduates who go to college; and (2) the special differences between School B where the developers taught and supervised the other treatment teachers, and the remaining schools, where regular teachers presented the experimental courses.

Although we cannot untangle the various elements contained in the "school" variable, our expectations are rather clearly confirmed on the first point, and only partially so on the second. Based on the percent of graduates going to college as a measure of academic emphasis, School C should lead in cognitive scores with Schools D, F, B, A and E ranked in that order. This pattern is confirmed for each set of schools taking a course. When the regression coefficients for schools in Tables 3 and 4 are ranked in relation to the means on Level 1 and 2 scores, we find Schools B, C, D and F are generally at or above the mean, while Schools A and E are below. On both Level 1 and 2 scores, for Courses 1 and 2, School B ranks at the top and higher than expected for its modest academic standing. This does not occur for Courses 3 and 4. Based on feedback from the developers of the later courses, it appears that they spent less time and effort in helping teachers use the materials than had their predecessors.

Attitudes

Compared with the cognitive results, attitudinal effects due to the courses are weaker and less frequent. For the several sets of measures used with students in Courses 1 and 2, the results are presented in Tables 4 and 5. Course 1 demonstrates a curriculum effect on each of the three attitudinal measures. In contrast, Course 2 has no impact on any of these attitudes.

As Table 4 indicates, Course 1 influences students to be less familistic; while Course 2 shows some pre-test and grades effects, the curriculum effect on attitudes is nil.

Students' Identification with their parents is negatively affected by Course 1 but not by Course 2 (see Table 5). Scores on Identification with one's mother are influenced also by the pre-test, and interactions between pre-test and curriculum. Course 2 shows a strong effect of sex; boys score higher on identification with both fathers and mothers, perhaps reflecting the male tendency to protect the family's honor.
Table 4: Regression Results for Courses 1 and 2 on Familism Scale

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Course 1</th>
<th>Course 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum</td>
<td>-0.11*</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Pretest</td>
<td>0.01</td>
<td>0.07*</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td>(.03)</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.07</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.04)</td>
</tr>
<tr>
<td>Year in school</td>
<td>-0.01</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Grades</td>
<td>-0.12</td>
<td>-0.13**</td>
</tr>
<tr>
<td></td>
<td>(.19)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.03</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(10)</td>
</tr>
<tr>
<td>School A</td>
<td>-0.07</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td>(.04)</td>
</tr>
<tr>
<td>School C</td>
<td>-0.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td></td>
</tr>
<tr>
<td>School D</td>
<td></td>
<td>-0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.21)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.35</td>
<td>2.05</td>
</tr>
<tr>
<td>F-ratio</td>
<td>2.20*</td>
<td>3.61**</td>
</tr>
<tr>
<td>df</td>
<td>8/16</td>
<td>8/17</td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td>.21</td>
<td>.38</td>
</tr>
</tbody>
</table>

**p < .05
p < .10
### Table 5: Regression Results for Courses 1 and 2 on Identification with Mother and Father

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>1-Mother</th>
<th>Course 2-Mother</th>
<th>1-Father</th>
<th>2-Father</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1-Mother</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curriculum</td>
<td>-0.16***</td>
<td>0.03</td>
<td>0.14**</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.05)</td>
<td>(.06)</td>
</tr>
<tr>
<td>Pretest</td>
<td>-0.08*</td>
<td>0.09**</td>
<td>-0.03</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.05)</td>
<td>(.05)</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.12**</td>
<td>0.01</td>
<td>0.14*</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.05)</td>
<td>(.07)</td>
<td>(.07)</td>
</tr>
<tr>
<td>Year in school</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>(.06)</td>
<td>(.07)</td>
<td>(.07)</td>
<td>(.10)</td>
</tr>
<tr>
<td>Grades</td>
<td>0.14</td>
<td>0.03</td>
<td>-0.08</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(.09)</td>
<td>(.07)</td>
<td>(.11)</td>
<td>(.10)</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.12</td>
<td>-0.36***</td>
<td>-0.09</td>
<td>-0.38**</td>
</tr>
<tr>
<td></td>
<td>(.07)</td>
<td>(.12)</td>
<td>(.09)</td>
<td>(.17)</td>
</tr>
<tr>
<td>School A</td>
<td>-0.04</td>
<td>0.09**</td>
<td>-0.07</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.06)</td>
</tr>
<tr>
<td>School C</td>
<td>-0.05</td>
<td></td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.05)</td>
<td></td>
<td>(.06)</td>
<td></td>
</tr>
<tr>
<td>School D</td>
<td></td>
<td>-0.08</td>
<td></td>
<td>-0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.25)</td>
<td></td>
<td>(.36)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.16</td>
<td>0.52</td>
<td>1.01</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(.04)</td>
<td>(.04)</td>
<td>(.04)</td>
</tr>
<tr>
<td>F-ratio</td>
<td>2.84**</td>
<td>4.19***</td>
<td>2.76**</td>
<td>2.34*</td>
</tr>
<tr>
<td>df</td>
<td>8/16</td>
<td>8/17</td>
<td>8/16</td>
<td>8/17</td>
</tr>
<tr>
<td>$R^2_{adj}$</td>
<td>.31</td>
<td>.44</td>
<td>.30</td>
<td>.20</td>
</tr>
</tbody>
</table>

***$p < .01$  **$p < .05$  *$p < .10$
Courses 3 and 4 used different attitude measures than the first two courses. Despite the fact that these two measures, the sex Role Ideology and Family Orientation scales, derive more directly from course content, they yield no effects due to taking either course. Grades influence Sex Role Ideology so that better students in Course 3 are more modern in sex role conception (see Table 6). In Course 4, girls are more pro-family in orientation than boys, that is, they are more accepting of parental authority and communication with parents.

Across all the attitude models, it is interesting that the School variable is unimportant. But it is also true that none of the independent variables either consistently or strongly affects the attitudinal measures.

Student Preferences, Teacher and Developer Reactions

We obtained systematic feedback on all four courses from students through their comments but student ratings were only available for Courses 3 and 4. The ratings on each question are favorable, particularly on the amount of work required, the rate and depth of course presentation, and the relative emphasis on facts and broad concepts. For the remaining concepts about assignments, the tests and intellectual stimulation, the ratings are middle range, none very low or very high. Students looked forward to Course 3 more than to Course 4, and in general, they rated the former course more favorably than the latter.

For all four courses, students responded to questions asking what features they most liked and least liked. Their replies were categorized through content analysis. Some themes cut across courses, while others pertain only to a given course. Class participation by students as individuals and in group discussion is a positive feature in all the courses. The cross-cultural focus of Course 1 made a strongly favorable impression on 47 percent of the students who took that course. The emphasis on courtship in Courses 3 and 4 yielded sizeable positive reactions.

Critical comments stress dullness or boredom with repetition of ideas or concepts, especially in Courses 1 and 4. A negative reaction to historical material occurred mainly for Course 3. Course 2 is criticized for the large number of worksheets, assignments and homework.
Table 6: Regression Results for Courses 3 and 4 on Sex Role Ideology and Family Orientation Scales

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Course</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-Sex Role</td>
<td>4-Sex Role</td>
<td>3-Family Orientation</td>
<td>4-Family Orientation</td>
</tr>
<tr>
<td>Curriculum</td>
<td>0.05 (.04)</td>
<td>0.03 (.05)</td>
<td>0.04 (.03)</td>
<td>-0.00 (.01)</td>
</tr>
<tr>
<td>Pretest</td>
<td>0.06 (.03)</td>
<td>-0.00 (.04)</td>
<td>0.01 (.03)</td>
<td>-0.00 (.01)</td>
</tr>
<tr>
<td>Interaction</td>
<td>-0.05 (.05)</td>
<td>0.03 (.07)</td>
<td>0.00 (.04)</td>
<td>0.01 (.02)</td>
</tr>
<tr>
<td>Year in school</td>
<td>-0.04 (.04)</td>
<td>0.00 (.08)</td>
<td>0.03 (.03)</td>
<td>0.02 (.02)</td>
</tr>
<tr>
<td>Grades</td>
<td>0.18** (.06)</td>
<td>0.07 (.08)</td>
<td>-0.01 (.05)</td>
<td>0.04 (.03)</td>
</tr>
<tr>
<td>Sex</td>
<td>0.11 (.11)</td>
<td>-0.08 (.15)</td>
<td>-0.13 (.10)</td>
<td>0.12** (.05)</td>
</tr>
<tr>
<td>School C</td>
<td>0.00 (.04)</td>
<td>0.08 (.09)</td>
<td>0.05 (.04)</td>
<td>-0.02 (.03)</td>
</tr>
<tr>
<td>School D</td>
<td>0.04 (.32)</td>
<td></td>
<td></td>
<td>0.06 (.10)</td>
</tr>
<tr>
<td>School E</td>
<td>-0.14 (.09)</td>
<td>0.00 (.08)</td>
<td>0.04 (.08)</td>
<td>-0.00 (.02)</td>
</tr>
<tr>
<td>School F</td>
<td>-0.07* (.04)</td>
<td></td>
<td>0.03 (.04)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.33 (.33)</td>
<td>0.25 (.40)</td>
<td>0.29 (.12)</td>
<td>0.15 (.74)</td>
</tr>
<tr>
<td>F-ratio</td>
<td>3.51** 9/8</td>
<td>1.29 9/8</td>
<td>0.60 9/8</td>
<td>4.12** 9/8</td>
</tr>
<tr>
<td>df</td>
<td>9/8</td>
<td>9/8</td>
<td>9/8</td>
<td>9/8</td>
</tr>
<tr>
<td>R² adj</td>
<td>.70</td>
<td>.40</td>
<td>.12</td>
<td>.74</td>
</tr>
</tbody>
</table>

**p < .05  *p < .10
We wondered whether any "spillover" occurred leading students to discuss the courses with significant others. Such reported influence did occur especially in discussions with friends or classmates. In all courses, students believed they learned about their own families, least of all in Course 2 (23 percent) and most of all in Course 4 (51 percent). Course 1 generated the least discussion either with parents (18 percent) or with peers (35 percent), while Course 3 was most discussed of all the courses (46 percent with parents and 61 percent with peers), perhaps because the topic is more controversial. Additionally, based on the evaluation of Courses 1 and 2, Courses 3 and 4 incorporated student activities to promote a higher degree of interaction between students and their parents and peers; evidently, these were successful.

Most of the teachers communicated reactions informally to project staff; only about half completed the daily forms and the final questionnaire. The course developers reported the most detailed reactions. In all cases, the developers discovered that they had prepared too much material for a six-week course. Virtually all the teachers skipped some parts of the curriculum using their own discretion in what to omit; but some dwelt overly long on short lessons. The teachers who completed questionnaires offered various comments, both positive and negative. The students' anti-historical bias and the redundancy noted by the developers were mirrored in some teachers' criticism: "... constant repetition of Colonial and Victorian periods became boring." "Our students were concerned about what's happening now and wanted to rush through ... the history." We cannot be sure whether this anti-historical view was initiated by students or teachers!

Some Input-Output Differences

In the curriculum development process, the project staff intended to create specific features for each course (see Table 1). The courses on adolescence (Courses 2 and 3) emphasize the students' experience and compare this with youth in earlier periods. This element of "closeness" to the students is lacking in Courses 1 and 4. The use of social science concepts and historical material was very explicit in the first two courses and more subtle in the last two. Readings for the later courses were increased in complexity and length. Courses 1, 2 and 3 are strongest in use of audio-visual materials while Course 4 relies mainly on lengthy text material, essay writing and discussion. But all courses, especially 1 and 4, were disliked for their heavy use of concepts. Thus even the subtler conceptual framework of Courses 3 and 4 was criticized. The "closeness" to adolescence intended in Courses 2 and 3 did not lead to the expected cognitive or attitudinal results. Course 2 which deals directly with adolescence did not elicit any special student reaction to its content.
Instead, Courses 3 and 4 were liked for their courtship and family emphasis. All the courses were valued for their use of group discussion and classroom participation, yet none was noted by students for its visual aids—an explicit feature of Courses 1, 2 and 3. Perhaps students are saturated with visual materials in other courses or through heavy TV exposure and become inured to them.

Balancing Student Cognitive Performance and Preferences

Generally students liked the courses and learned from them. But this conclusion led us to ask whether their learning due to a course had anything to do with their enjoyment of it. Students rated Courses 3 and 4 at the end of the six-week field trial. We analyzed the relationship of the students’ rating of three items to their cognitive test scores. The three items were: (1) How eager were you to take this course at the beginning of the semester? (2) Do you enjoy this course? (3) In general, how would you rate this course? We used partial correlation to ascertain if there was any connection between how students performed in the course and their course ratings for the three items, while controlling for grades, year in school, sex and school.

The results revealed no such connection. All their rp’s were close to zero, suggesting that the two matters are not related: students whose knowledge is significantly affected by a course may have either favorable or unfavorable judgments about the course.8

Influencing Attitudes through Cognitive Material

Educators and curriculum developers tend to assume that improvement in cognitive skills and attitude change go hand-in-hand in the educational process. Curricula which offer both kinds of learning, cognitive and affective, are now fashionable. But it is not known whether these two elements influence each other. Does increased knowledge of subject matter alter attitudes about it? Or does a change in attitude affect a student’s academic achievement? Studies of college students suggest that these two elements interact so that with each additional year of college, students become more liberal politically, more expressive personally and more tolerant of diversity (Feldman and Newcomb, 1969). For high school students what little evidence exists does not support this pattern (Angrist, Mickelsen, Penna, 1976).

While we could not unravel the causal process, we analyzed the relationship for students in Courses 3 and 4 through partial correlation. The associations considered were between an individual’s change on Level 1 and Level 2 cognitive test scores, and change in the two attitude scales, Sex Role Ideology and Family Orientation. Again, this analysis controlled for grades, year in school, sex and school. The rp coefficients were low, ranging
from -.02 to -.24. Only two of the eight coefficients were statistically significant, both for Course 3, and with a 2-tailed test; these were \( r_p = -.22 \) \( (p < .05) \) between changes in Level 1 cognitive score and changes in Family Orientation, and \( r_p = -.24 \) \( (p < .02) \) between Level 2 score changes and changes in Family Orientation. These results indicate that students whose cognitive skills improved in the course became less pro-family in orientation. While this result is not surprising, it contradicts our hope that students who know more about the family as an institution would be more favorably disposed toward it. We found no association between cognitive test scores and Sex Role Ideology. Because of the few and rather weak associations, we conclude that there is insufficient evidence to claim a link for these students between their mastery of cognitive skills and changes in attitudes about the family.

**CONSEQUENCES OF THE EVALUATION**

Results from the trial and evaluation of the four courses led to useful design and revision decisions in two ways: (1) we designed Courses 3 and 4 largely on the basis of the results from the first two courses; and (2) all the courses were revised using cognitive performance, attitudinal data, student preferences and teacher reactions as guides. In particular, we reviewed course inputs purported to accomplish the curriculum goals.

The design of Courses 3 and 4, when compared with Courses 1 and 2, included the following: (1) subtler use of social science concepts; (2) higher reading level; (3) more challenging material emphasizing problem-solving skills and based on the recognition that students enter these courses with more knowledge about the family than project staff had expected; (4) proportionately more contemporary material; (5) more emphasis on parent-teen relationships and family roles; (6) more focus on sex roles; (7) closer links between course content and the measures used to tap attitudes. In general students were more critical of the first two courses and less critical of the second two. This suggests that most of the intended improvements in the later courses were successful. This conclusion is also confirmed by the stronger "spillover" effects of the later courses. Courses 3 and 4 more often led to students' perception that they learned something about their families and to more discussion with both parents and peers.

Revising the courses to affect family attitudes remains quite another matter. We have concluded along with Patrick (1972) that attitudes are hard to affect (Cf. Deutscher, 1973). Clearly, influencing attitudes by means of cognitively oriented materials did not work in these short-term courses. Perhaps cumulative exposure to family courses over time might lead to the desired attitude change (Cf. Angrist, Mickelsen, Penna, 1976).
CONCLUSIONS AND RECOMMENDATIONS

Our experiences in designing, testing and rigorously evaluating an interdisciplinary family life curriculum produced rich information regarding the development process.

Construction of curriculum materials to teach new knowledge of concepts, facts, and generalizations seems to be a straightforward task. Since adolescents have entered the cognitive developmental stage of formal operational thought, curriculum emphasis should be on reasoning skills. Indeed, if the ability to pursue meaning is sought (Chall, 1967), then content focusing on cognitive levels above knowledge is needed, and materials should be designed to develop skills in comprehension, application and analysis.

We believe that students' liking of a course is important, but developers should not assume a relationship between learning and a simplistic notion of liking. Emphasis on topics which were seemingly close to adolescent concerns apparently generated only transient enjoyment; rather our findings suggest that unfamiliarity and remoteness of the topics sustained student interest. Additionally, attempts to increase course attractiveness by using extensive audio-visual materials seem unwarranted for this type of curriculum. The most important variable in course design remains the level of cognitive challenge of the materials.

Feedback collected by developers teaching in the schools and site visits suggest that as many teachers ignored our extensive teacher's guides as implemented them. Those lessons in each course which required the use of alternative pedagogical strategies in the classroom, attempting to link formal classroom learning to home and community, faltered or were ignored. Many teachers held on to familiar practices, reshaped the unfamiliar into conventional modes of instruction and avoided seemingly controversial aspects of the curriculum content. These propensities suggest that power over pedagogical matters should not be solely the teacher's (Lortie, 1975; Waller, 1961). Instructional strategies, their intent and content, can easily be conveyed to students via their texts, lessening the possibility that such strategies will be ignored.

We were unable to detect attitude changes induced by the curriculum, even when the measures closely reflected the curriculum content. In retrospect this is not surprising. The courses were short, the materials stressed the cognitive aspects of learning, the students had prior beliefs about the family, and teachers probably ignored aspects of the curriculum which raised questions about traditional family patterns and sex role orientations. Whether these, or the insensitivity of our instruments, explain our negative results, developers should be alerted to the fact that detecting attitude change from course materials may not be achievable in the short run.
As have other curriculum developers, we recognized the need for early assessment of curriculum materials before developing materials for the entire project. The initial development experience provided input to the subsequent work, involving the development of new courses as well as revision of the earlier ones. Valuable and timely feedback is afforded by a sequential approach to the development cycle; a curriculum structured as a series of mini-courses or modules may readily exploit such an approach.

REFERENCES


FOOTNOTES

1 Although the model generated by Tyler's basic curriculum questions has been challenged recently (Schwab, 1971; Walker, 1971), it continues to be implemented by developers and evaluators engaged in curriculum work, including the authors of this paper.

2 Although we use means for each class of students, these are not weighted by class size. Thus equal importance is attributed to each class. But inter-group class size comparisons suggest no consistent relationships between class size and experimental group.

3 Variables reflecting the percent of graduates attending college from each school, and father's and mother's education were omitted from the model because they become statistically non-significant when the School variables are included. For each dependent variable, the post-test class mean was normalized from 0.0 to 1.0. The independent variables were scored as follows: curriculum 1 = Group 1 or 4, 0 otherwise; pre-test 1 = Group 1 or 2, 0 otherwise; interaction 1 = Group 1, 0 otherwise; year in school 12 = 12th grade ... 7 = 7th grade; grade point average 4 = A ... 1 = D; sex 1 = female, 0 = male.

4 Discussion of the equivalence between regression and analysis of variance is presented by Burke and Schuessler (1974), among others.

5 Standards applied for inclusion of each item were: difficulty index between 0.3 and 0.7, index of discrimination above 0.27, point biserial correlation (adjusted for self correlation) above 0.3. The Standard Kuder-Richardson coefficients were 0.76, 0.67, 0.85 and 0.84 for the cognitive tests used in Courses 1 through 4, respectively.

6 Each item retained for a measure based on the factor analysis had a factor loading of at least 0.33 on one of the two dominant dimensions and negligible loading on the other.

7 Students' opinions were obtained only from those taking the courses (Groups 1 and 4); there are no controls here. Detailed tables showing student responses and course ratings are available from the authors.

8 Other analyses with change in score (Group 1 students), and scores and changes in excess of those predicted from Group 2 (no curriculum) also yielded insignificant results.
This study was designed to examine the changing nature of children's concepts of their own and other countries as they progress through the elementary school years. A sample of 1,456 children drawn from grades one, three and six in two centers in Canada, Australia and the United States commented on their own and seven other countries. Their responses were analysed for egocentrism, evaluative comment and stereotypes, and dominant associations were described. Comparisons were made between the responses of children in different places and at different grade levels.

Educators concerned with the development of international understanding have shown increasing interest in the early development of concepts of people in different countries. These concepts are important because they filter subsequent data (Child, 1968); and because they are likely to be learned uncritically without knowledge of alternatives or standards for judging information, or even unconsciously, through imitation and identification (Greenstein, 1965).

Children's concepts of other nations and peoples are shaped by many agencies: parents, peers and the media, as well as the schools. Most early school programs deal with the child's immediate environment and only sketchily with distant places. By the time other countries are systematically studied in upper grades, early concepts are already established and shape perceptions of subsequently presented data.

Several recent studies have explored the nature of children's concepts of their own and other nations. Piaget and Weil (1951) studied the development of the concept of homeland and other countries through a series of interviews with 200 children aged from 4 to 15. They postulated a natural tendency toward egocentrism which changed in form through the integration of new data and as the child progressed through stages of reciprocity.

Initially egocentrism takes the form of an unconscious assumption by children that their own perspective is the only one possible. With increasing awareness of their own country and of other countries and points of view different from their own, egocentricity takes on new forms which are farther removed from the children's initial center of interest. Piaget and Weil refer to these new forms as sociocentricity. From this perspective the children's concept of homeland develops gradually as they broaden their center of interest to take in more than their immediate environment. Central to this process is the gradual recognition that their immediate surroundings
are in and part of a more inclusive concept of their homeland. At a first stage, about age 6, children studied by Piaget and Weil recognized that they were in Geneva, and in Switzerland but could not understand the relation of these to each other. They drew Geneva and Switzerland beside each other, not one within the other, and identified themselves as either Genovese or Swiss, but not both. At a second stage, children could relate Geneva within Switzerland spatially, but were still unable to identify themselves as Genovese and Swiss at the same time. At a third stage, about age 10, children were able to synthesize the data correctly and had a realistic and inclusive notion of their homeland. These same stages were apparent in discussions of other countries and in student’s ability to think from the point of view of others. Only at stage three could children correctly apply the concept of foreigner to Swiss and non-Swiss according to the perspective of the observer.

An alternative way of viewing the development of children’s concept of their own and other countries is to consider changing sets of attributes of the concepts as their labels gradually acquire meaning. In the early stages of development concepts of particular countries may be almost devoid of meaning. One of the first attributes we might expect to find is that other countries are somewhere else, and different, but with specific differences not known. Thus we might expect differences rather than similarities to be noted, which in turn would color subsequent perceptions. This tendency might be increased by two other factors. First, teachers and parents attempt to make the other countries seem (different and therefore) interesting. Second, the process of recognizing similarities is thought to be a more complex cognitive task than awareness of differences. The first few pieces of information defining the concept may be specific or general, and either high or low in emotional content. But since they are the only information available to the child, they are likely to be over-generalized. If the only data a child has on Brazil is a picture of an Indian village in the jungle, mention of a person as Brazilian will immediately conjure up that image.

It would be reasonable to expect the initial set towards differences to influence subsequent perceptions and result in a continuation of stress on differences. On the other hand an increase in knowledge about other nations might gradually reduce stereotyping, though many original stereotypes would tend to persist and others which were part of the adult world might be added.

Jahoda (1962) interviewed 144 children aged 6 to 11 in four Glasgow schools to explore their attitudes towards other countries. Younger children noted strange, unusual or exotic features of other countries and were attracted by them. Older children were more aware of people and politics, and less of physical characteristics. They reacted negatively to people who were perceived as strange or who were on the “wrong” side of the cold war.

Lambert and Klineberg (1967) studied the perception of foreign peoples
held by 3,300 6, 10 and 14 year-old-children in 11 countries. They noted a strong tendency to refer to certain national groups as "not like us" and concluded that these groups were used as contrasts in developing concepts of homeland and own group. Six-year-olds in particular stressed the differences of foreign peoples much more than the similarities. Lambert and Klineberg found that emotional loadings tended to be positive though there were variations between nations and age levels. This lack of prejudice against people of other nations who are different is in marked contrast with Jahoda's findings for older children and with studies of attitudes towards minority groups within nations (Allport, 1954).

PROCEDURE

This study of children's associations with their own and other countries was conducted in 1973 and 1974 in Canada, Australia, and the United States, three countries with many similar traditions but important differences in educational and political systems. Two centers were identified in each country and in each center tests were administered to children in grades one, three and six. Two centers were used for each country to allow within-country as well as between-country differences to emerge, and an attempt was made to collect data from middle and lower class children.

An attempt was made through local contacts in schools, school districts or departments of education to identify three schools within each center which drew students from different socio-economic levels within the city. In each school one or two classes at each level (depending on the size of the school) completed test instruments. This pattern was followed fairly closely with the exceptions of Ontario, where all data were obtained from one school, and California, where the sample of grade one children was drawn from only one school. At grades three and six all students in the selected classes who were willing to participate completed the test, and at grade one a random sample of 10 students in each class was selected. The instruments were administered by the teachers following detailed instructions prepared by the researcher. Initially it was hoped to gather data on socio-economic variables but this was not possible in several centers and socio-economic data based on school location was considered too unreliable to serve as a basis for analysis. The schools were all in suburban areas close to major cities and what socio-economic data were obtained suggested that the sample population in each area was predominantly lower middle class. There was no attempt to make the sample representative of the countries as a whole and this must be borne in mind in considering the results.

The test instrument was a very simple one. Students were asked to write whatever they thought of in responses to the names of eight countries, England, China, Australia, Japan, Brazil, Canada, Nigeria and the United States of America. The order of countries was altered for Canadian and
Australian children by exchanging the name of their country with that of the United States so that their own country appeared last. Students were given a booklet with the name of each country printed at the top of a blank page. Grade 3 and 6 children were given one and one half minutes to write their response for each country. The questions were read to the grade 1 children and their responses noted by a recorder. This open-ended procedure was adopted to minimize the possible distorting effect of more detailed written or interview questions. Its major disadvantage is that certain ambiguous responses cannot be probed, but it does have the advantage of eliciting any dimensions of concepts of nations which are significant to the children.

Stereotypes were expressed as a percentage of general statements to total responses. A general statement was one which carried the implication that it was true of all or a class of people. If a statement was qualified it was considered as not general. Thus "they wear black clothes" was considered stereotypic, while "some people wear black clothes" was not.

Egocentrism was a percentage score for the proportion of a child’s responses which implied his reference point was himself or his own country. The comments "big" or "close to Japan" with reference to another country were not considered egocentric while "bigger than us" or "a long way off" were considered egocentric.

In both these cases percentages rather than raw scores were used to provide single scores which could serve as indices of children's tendency to think in stereotypes and to be egocentric in their thinking about their own and other countries.

Evaluative comments were subdivided into five categories and a score was
calculated by adding the number of responses falling into each category. The categories used were negative evaluation, positive evaluation, non-evaluative-similar, non-evaluative-different and non-evaluative-other. Comments such as funny looking, slant-eyed, powerful or small were scored as non-evaluative. To be rated as evaluative a response had to include a direct evaluation such as “nice,” “I would like to live there“ or “a bad place for kids.”

Analysis was done by four research assistants who worked in pairs after a period of initial training. Five test instruments from each level in each center, a total of 90 instruments, were coded twice and the test-retest reliability was calculated. The following reliability figures were obtained: stereotypes .87, egocentrism .83, evaluative comment—own country .81 and evaluative comment—other countries .90.

Statistical comparisons were made using one-way analysis of variance with differences accepted as significant at the .05 confidence level. When overall significant differences were found, additional comparisons were made among the groups using the LSD procedure for multiple range tests (Kirk, 1968) to determine which groups were responsible for the differences.

RESULTS

Mean number of responses by center and grade level are shown in Table 2. As might be expected the number of responses to other countries increased substantially in the higher grades and the same tendency can be seen for responses to own country. The most likely explanation for the number of responses to own country in grade 3 being about the same as for grade 1 is the method of collecting data. In grade 1 the teacher wrote down what the children said whereas in grade 3 the children completed the test instrument themselves.

Mean scores on stereotypes for own and other countries by center and grade level are shown in Table 3.

For other countries there were significant differences between grade levels for each center and for the total score, and in each case the score for children in grade 1 was significantly lower than for the other grades. In all centers except Victoria the scores for grades 3 and 6 children were very close, but for Victoria the grade 3 score was the highest of any center and the grade 6 score the lowest. The difference between the two scores was significant. In only one case, at grade 1 level, was there a significant difference between centers with the score for Georgia being higher than the others.

For own country there were significant differences between grade levels in five centers and for the total score. In these comparisons there was a tendency for scores to be higher at grade 3 than at grade 1 and lower again at grade 6. The Georgia sample did not follow this pattern and scores for
that center were about the same at each level but for Ontario and British Columbia the difference between grade 1 and grade 3 scores was not significant.

In comparisons between scores for own and other countries the grade 1 scores for own country were significantly lower in every center, the grade 3 scores for own country were significantly higher for California, Ontario and British Columbia, and the grade 6 scores for own country were
significantly lower for California, Victoria and British Columbia.

The relatively low scores for grade 1 children for other countries may be somewhat misleading. Probably the best interpretation for the low figure is that it resulted from general lack of knowledge. Many did not respond at all to some of the countries and a number of those who did simply listed animals, places or events without any explanation or comment. The coding was done in such a way that some statement was needed before a response could be coded as a stereotype. A much higher proportion of responses to own country by these children were in the form of statements.

The index of egocentrism (Table 4) shows a consistent drop at each grade level for both own and other countries indicating a shift towards a less personal view of the world.

<table>
<thead>
<tr>
<th>Center</th>
<th>Other Countries</th>
<th></th>
<th></th>
<th>Own Country</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 1</td>
<td>Grade 3</td>
<td>Grade 6</td>
<td>Grade 1</td>
<td>Grade 3</td>
<td>Grade 6</td>
</tr>
<tr>
<td>Georgia</td>
<td>26.71</td>
<td>17.15</td>
<td>6.14</td>
<td>24.54</td>
<td>24.09</td>
<td>5.60</td>
</tr>
<tr>
<td>California</td>
<td>20.43</td>
<td>4.74</td>
<td>2.14</td>
<td>9.32</td>
<td>4.92</td>
<td>3.40</td>
</tr>
<tr>
<td>Tasmania</td>
<td>15.00</td>
<td>10.29</td>
<td>4.37</td>
<td>15.89</td>
<td>4.89</td>
<td>.90</td>
</tr>
<tr>
<td>Victoria</td>
<td>14.71</td>
<td>12.57</td>
<td>7.14</td>
<td>31.19</td>
<td>13.95</td>
<td>8.36</td>
</tr>
<tr>
<td>Ontario</td>
<td>7.83</td>
<td>7.69</td>
<td>4.14</td>
<td>8.93</td>
<td>14.53</td>
<td>5.06</td>
</tr>
<tr>
<td>British Columbia</td>
<td>9.17</td>
<td>5.13</td>
<td>4.80</td>
<td>20.43</td>
<td>13.44</td>
<td>7.53</td>
</tr>
<tr>
<td>Total</td>
<td>15.50</td>
<td>9.53</td>
<td>4.67</td>
<td>20.57</td>
<td>12.67</td>
<td>5.07</td>
</tr>
</tbody>
</table>

This pattern was reflected in significant differences for all centers for both own and other countries with the exception of Ontario where the difference between grade 1 and grade 3 was not significant for other countries and reversed for own country, Georgia where the difference between grade 1 and grade 3 for own country was not significant, and Victoria, where the difference between grade 1 and grade 3 for other countries was not significant.

In comparisons between own and other countries the following significant differences were found: grade 1 children in California were more egocentric in references to other countries than to their own, grade 1
children from Victoria and British Columbia were more egocentric in references to their own country than to others, grade 3 and grade 6 children from Tasmania were more egocentric in references to other countries than to their own, and grade three children from both Ontario and British Columbia were more egocentric in references to own country than to other countries. In other comparisons of egocentrism between own and other countries by center and grade level scores were very similar and no other significant differences were found. The general pattern of results in comparisons between own and other countries must be considered inconclusive. In a majority of cases there were no significant differences, and the differences that were found were in opposing directions. Insufficient data were available to probe in depth into reasons for the differences that were found.

A summary of evaluative comments is set out in Table 5. Probably the most outstanding feature of the responses is the very small proportion of evaluative and comparative comments. By far the greatest proportion at all levels fell into a category for other (non-evaluative, non-egocentric) responses. The number of responses per student was much lower for other countries at the grade 1 level, and increased for grade 3 and grade 6 students. However, the proportion of responses falling into various categories was remarkably stable.

While the number of evaluative responses was small there were far more positive than negative comments at all levels for both own and other countries. Some isolated comments reflected very negative attitudes towards particular countries or peoples but these were rare and were not concentrated in any one of the centers.

As expected there were many more comments relating to differences in references to other countries than there were to similarities, and this pattern extended to each grade level for all centers except grade 6 in Ontario. One possible reason for this variation from the general pattern is proximity to predominately French speaking Quebec. At the grade 1 level, there seemed to be a set towards differences and a generalized impression of particular countries that had a partial knowledge base. Australians spoke differently, had different hair, clothes, toys and houses and many wild animals. However, the animals listed were simply wild, not necessarily Australian. Kangaroos and koala bears made the list but no more often than giraffes, ostriches or lions. As might be expected characteristics which differed were directly related to student’s experience—houses, shoes, hair, toys, schools and hair ribbons being common examples. China brought out images of different hair, different houses, different schools, different shoes and different toys, and chinamen with chopsticks and slanted or pointed eyes.
Table 5: Evaluative Comments Per Pupil for Own and Other Countries

<table>
<thead>
<tr>
<th>Center</th>
<th>Negative</th>
<th>Positive</th>
<th>Similar</th>
<th>Different</th>
<th>Other</th>
<th>Number of Responses</th>
<th>Negative</th>
<th>Positive</th>
<th>Similar</th>
<th>Different</th>
<th>Other</th>
<th>Number of Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>Georgia</td>
<td>.02</td>
<td>-0.</td>
<td>.08</td>
<td>.56</td>
<td>.96</td>
<td>1.62</td>
<td>.03</td>
<td>.11</td>
<td>.19</td>
<td>.08</td>
<td>4.56</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>.01</td>
<td>.01</td>
<td>.08</td>
<td>.12</td>
<td>1.02</td>
<td>1.44</td>
<td>-.01</td>
<td>.18</td>
<td>-.01</td>
<td>.40</td>
<td>3.21</td>
</tr>
<tr>
<td></td>
<td>Tasmania</td>
<td>-0.</td>
<td>.02</td>
<td>.07</td>
<td>.31</td>
<td>1.12</td>
<td>1.52</td>
<td>-.01</td>
<td>.19</td>
<td>-.01</td>
<td>.11</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>Victoria</td>
<td>-0.</td>
<td>.01</td>
<td>.14</td>
<td>.34</td>
<td>1.72</td>
<td>2.11</td>
<td>-.01</td>
<td>.19</td>
<td>-.01</td>
<td>-.01</td>
<td>8.42</td>
</tr>
<tr>
<td></td>
<td>Ontario</td>
<td>.05</td>
<td>.03</td>
<td>.03</td>
<td>.10</td>
<td>1.35</td>
<td>1.56</td>
<td>.10</td>
<td>.13</td>
<td>-.01</td>
<td>-.01</td>
<td>7.07</td>
</tr>
<tr>
<td></td>
<td>B.C.</td>
<td>-0.</td>
<td>.07</td>
<td>.03</td>
<td>.06</td>
<td>1.47</td>
<td>1.63</td>
<td>-.01</td>
<td>.10</td>
<td>-.01</td>
<td>-.01</td>
<td>2.83</td>
</tr>
<tr>
<td>Mean</td>
<td>.01</td>
<td>.02</td>
<td>.07</td>
<td>.25</td>
<td>1.27</td>
<td>1.65</td>
<td>.02</td>
<td>.13</td>
<td>.06</td>
<td>.06</td>
<td>.06</td>
<td>5.18</td>
</tr>
<tr>
<td>Grade 3</td>
<td>Georgia</td>
<td>.04</td>
<td>.25</td>
<td>.13</td>
<td>.26</td>
<td>1.80</td>
<td>2.48</td>
<td>.03</td>
<td>.43</td>
<td>.22</td>
<td>.03</td>
<td>2.55</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>.01</td>
<td>.08</td>
<td>.05</td>
<td>.29</td>
<td>3.37</td>
<td>3.79</td>
<td>.02</td>
<td>.09</td>
<td>.06</td>
<td>-.01</td>
<td>6.35</td>
</tr>
<tr>
<td></td>
<td>Tasmania</td>
<td>.02</td>
<td>.16</td>
<td>.10</td>
<td>.22</td>
<td>4.50</td>
<td>5.00</td>
<td>-.01</td>
<td>.32</td>
<td>-.01</td>
<td>.04</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td>Victoria</td>
<td>.01</td>
<td>.13</td>
<td>.17</td>
<td>.18</td>
<td>2.50</td>
<td>2.99</td>
<td>-.01</td>
<td>-.01</td>
<td>-.10</td>
<td>-.01</td>
<td>5.10</td>
</tr>
<tr>
<td></td>
<td>Ontario</td>
<td>.02</td>
<td>.21</td>
<td>.01</td>
<td>.10</td>
<td>2.48</td>
<td>2.72</td>
<td>-.01</td>
<td>-.67</td>
<td>-.01</td>
<td>-.01</td>
<td>3.36</td>
</tr>
<tr>
<td></td>
<td>B.C.</td>
<td>.09</td>
<td>.15</td>
<td>.04</td>
<td>.21</td>
<td>2.69</td>
<td>3.18</td>
<td>.12</td>
<td>.33</td>
<td>-.01</td>
<td>-.01</td>
<td>3.24</td>
</tr>
<tr>
<td>Mean</td>
<td>.03</td>
<td>.16</td>
<td>.08</td>
<td>.21</td>
<td>2.89</td>
<td>3.36</td>
<td>.03</td>
<td>.31</td>
<td>.07</td>
<td>.01</td>
<td>.01</td>
<td>4.54</td>
</tr>
<tr>
<td>Grade 6</td>
<td>Georgia</td>
<td>.01</td>
<td>.10</td>
<td>.10</td>
<td>.05</td>
<td>7.64</td>
<td>7.90</td>
<td>.03</td>
<td>.12</td>
<td>.04</td>
<td>.10</td>
<td>9.49</td>
</tr>
<tr>
<td></td>
<td>California</td>
<td>.12</td>
<td>.17</td>
<td>.13</td>
<td>.24</td>
<td>4.76</td>
<td>5.42</td>
<td>.03</td>
<td>.51</td>
<td>-.01</td>
<td>.08</td>
<td>7.81</td>
</tr>
<tr>
<td></td>
<td>Tasmania</td>
<td>.17</td>
<td>.25</td>
<td>.12</td>
<td>.92</td>
<td>5.70</td>
<td>7.16</td>
<td>.03</td>
<td>.21</td>
<td>-.01</td>
<td>-.01</td>
<td>8.03</td>
</tr>
<tr>
<td></td>
<td>Victoria</td>
<td>.03</td>
<td>.13</td>
<td>.04</td>
<td>1.24</td>
<td>5.39</td>
<td>6.83</td>
<td>.04</td>
<td>.18</td>
<td>.05</td>
<td>.03</td>
<td>10.26</td>
</tr>
<tr>
<td></td>
<td>Ontario</td>
<td>.39</td>
<td>.46</td>
<td>.46</td>
<td>3.26</td>
<td>6.53</td>
<td>11.10</td>
<td>.42</td>
<td>.82</td>
<td>.49</td>
<td>5.95</td>
<td>12.65</td>
</tr>
<tr>
<td></td>
<td>B.C.</td>
<td>.09</td>
<td>.17</td>
<td>.04</td>
<td>1.15</td>
<td>4.18</td>
<td>5.63</td>
<td>.08</td>
<td>.38</td>
<td>.01</td>
<td>.13</td>
<td>5.77</td>
</tr>
<tr>
<td>Mean</td>
<td>.14</td>
<td>.21</td>
<td>.15</td>
<td>1.14</td>
<td>5.70</td>
<td>7.32</td>
<td>.11</td>
<td>.37</td>
<td>.10</td>
<td>1.05</td>
<td>.01</td>
<td>9.00</td>
</tr>
</tbody>
</table>

**Center**
- Georgia
- California
- Tasmania
- Victoria
- Ontario
- B.C.

**Number of Responses**
- Own Country
- Other Countries
Comments from children at the grade one level demonstrated varying ability to include themselves in their concept of their own country and this is reflected in the proportions of similar and different references shown on Table 3. Some trace of this was still present at grade three level, but in grade six it had been replaced by references to differences within the country.

Responses of students to their own country suggested a wide range in knowledge and in the development of reciprocity and decentering as defined by Piaget. While a comment from a Southern child that "I can understand American" might be capable of several interpretations, there were some statements such as "I live near there" and "they speak different" which suggest inability to include themselves in their concept of the United States. On the other hand "they speak different from us, they think we speak funny too" from a six year old with reference to China, suggests a high degree of reciprocity.

At grade 3 and grade 6 levels the range in comments was narrower than at grade 1 with greater concentration on certain insights. Children no longer appeared to be listing as different any aspects of their own environment that occurred to them. Instead they had become more discriminating as their knowledge base improved. Clustering of concepts for children in the same school class suggests a strong school influence at this stage. Awareness of national and international events increased substantially, particularly at grade six but associations with people or places which were seen as different were usually positive, as can be seen in the following examples.

Australia—"thar are gowalo beears the galopen gormay comes from thar. not very warm. Lots of trees nice houses they have wered axents. I would like to live thar."

David—Canada, grade 6

Japan—"nice, pearl harbor, war, festival, superstitious, dragons, fish, Explosion '70, kites"

Mary—United States, grade 6

At each level, but particularly at grades three and six there appeared to be a national image reflected in commonalities among the children's responses in all centers. The image was less dominant when children responded to their own country because a number of specific and local associations were included, and perceptions of other countries were colored by their own country's relationship with them. However there was still a surprising amount of agreement between different centers. For Australia the image was one of exotic wild animals, hot deserts, sheep and the flying doctor with occasional references to Canberra and the Sydney Opera House. Canada, viewed from outside was large and cold with mountains, moose, skiing and hockey. The United States had a president and 50 states, cars, movie stars, the White House, Indians, big buildings and the Statue of Liberty. Major
associations with Japan were cars, ships and heavy industry, fish, rice and chopsticks, language differences and slanted eyes. Pretty Japanese ladies were frequently mentioned by the grade one students in all areas, and war, Hiroshima and Pearl Harbour were commonly mentioned by the Americans.

The American children were more conscious of their own national monuments, government institutions and personalities than either Canadian or Australian children. However, the fact that this perspective on the United States was shared by the non-Americans suggests that their apparent greater political awareness is a reflection of a different national image projected inside the country as well as outside, rather than greater political maturity or more effective political education in American schools.

CONCLUSIONS

Many early impressions which appear stereotypic disappear or are modified as students acquire additional data. Some, such as awareness of physical differences in the Chinese and Japanese are more resistent to change. Other associations relating to national and international events begin to appear between grades 3 and 6. The overall effect is for the amount of stereotyping in references to other countries to remain fairly constant after a sufficient knowledge base has been acquired.

The concept of homeland develops as one would expect from the children’s immediate environment to the nation as a whole. In the process greater attention is given to government and national leaders. This effect is particularly strong and comes early for American children. There is a strong indication of reductions in stereotyping in references to own country between grades 3 and 6. Whether this resulted from greater maturity or additional knowledge could not be determined from the data.

The amount of egocentrism diminishes as children grow older and acquire additional data. There appears to be wide variation in readiness for the kind of reciprocity described by Piaget and Weil. Some children appear to have the ability at a fairly advanced level in grade 1 at the age of six. A few more have it at grade 3 though not enough to affect the index of egocentrism. At grade 6 the ability is common though far from universal. This finding does not contradict Piaget and Weil’s notion of stages in the development of reciprocity but it does suggest greater variation in individual development than they imply through the use of age level references.

For the majority of students, associations with their own and other countries were non-evaluative. This finding may be partly a function of coding since the assumption was made that concepts that are emotionally loaded for adults are not necessarily loaded for children. This differs from the position taken by Lambert and Klineberg and may explain the
difference in results.

A related finding, this time supporting a conclusion of Lambert and Klineberg, is that differences are much more likely to be associated with positive than with negative evaluations. In fact large majority of evaluations across the whole sample were positive.

Emphasis on differences rather than similarities when considering other countries is of major concern only if it leads to distortions in the students' perceptions, if it inhibits subsequent acquisition of more complete information, or if differences are associated with negative attitudes. The evidence found in this study suggests that none of these undesirable consequences occur. Emphasis on differences seems to precede formal study of particular countries in school indicating that it is a general cultural phenomenon rather than a consequence of approaches taken in school. Certainly school materials do emphasize differences and in so doing support this trend. However since undesirable consequences do not appear to follow there seems no need to change this approach.

The associations children have with other countries are subject to school influence. In a number of cases certain associations came in batches from individual classes or schools and were not shared by other students from the same center. If changes in students' perceptions were considered necessary they could be at least partly brought about through curriculum changes.

The American, Canadian and Australian children had very similar knowledge and attitudes towards each of the countries listed in the study. While there were variations between centers in stereotyping, egocentrism and evaluative comment, there were no systematic differences which could be attributed to country of origin. Perhaps the most interesting aspect of this similarity was the fact that the image children had of their own country was shared by children of the same age from the other countries, and that this effect was apparent as early as grade 1. The mechanisms by which these images are projected and communicated to children at such an early age are apparently quite powerful and worthy of careful study.

The very wide range in individual ability to perform logical operations ascribed by Piaget to particular stages of cognitive development underscores the difficulty of matching the cognitive level of children in classes with curriculum materials of appropriate complexity. The background knowledge of children within classes varied as widely as their level of psychological development.

This suggests a need for considerable variation in curriculum for individual children, and either careful matching of children with topics of study, or extensive student choice.
REFERENCES

ARTHUR W. FOSHAY, ASSISTED BY WILLIAM W. BURTON, “Citizenship as the Aim of the Social Studies.”

This article seeks to present a single, coherent aim for the Social Studies. It argues that the Social Studies have never had a coherent programmatic intent, and that such an intent is both desirable and possible. The intent is called "citizenship," this being an ancient and honorable goal of American education. This term is defined according to John Rawls' principle in his *Theory of Justice*, as those activities that seek to promote justice between men and their institutions. From this definition, two criteria for Social Studies programs are derived: that programs should (a) deal directly with a conception of justice, and (b) provide for the direct practice of citizenship.

Four widely used texts are examined in the light of these criteria. All four fail to meet them, and are criticized for certain misleading portrayals of the American political situation. Two more recent "reform" programs are examined. They meet the criteria somewhat better, though not wholly.

The article closes with a broadly sketched indication of what a social studies program would be like if it met the two criteria. It would be a twelve-year program, with an emphasis on various forms of social and political behavior intended to deepen the concept of citizenship as indicated by Rawls.

SHIRLEY S. ANGRIST, RICHARD MICKELSEN AND ANTHONY N. PENNA, "Development and Evaluation of Family Life Courses."

The development, trial and evaluation of an interdisciplinary family life curriculum for high school students is described. Using a sequential formative evaluation strategy with an experimental design, the impact of these courses on the knowledge, inquiry skills, attitudes and preferences of students was assessed. Feedback from developers, teachers and students was used to revise the four courses. All courses improved students' knowledge about the family, three influenced inquiry skill development, only one course affected student attitudes toward the family. Students showed preferences for contemporary and cross-cultural rather than historical materials. The evaluation results suggest that attitudes are difficult to influence; curriculum designers should concentrate on intellectual challenge, trying to combine course attractiveness to students with high cognitive levels. Additionally, instructional strategies should be conveyed to students as well as to teachers.
JOHN D. NAPIER, "The Ability of Elementary School Teachers to Stage Score Moral Thought Statements."

The study examined whether sixty elementary school teachers could stage score moral thought statements using a published rater manual and what factors were related to their stage scoring ability. The subjects took a pretest using only descriptions of the six moral stages originally researched by Lawrence Kohlberg. Next the teachers were given the self training rater manual and a posttest. Correlations of the demographic and personality variables studied with scores on the pre and posttests showed no significant correlations with the posttest and only one significant (but moderate) correlation with the pretest scores. Analysis of variance tests on the pre and posttests showed Stage 1 and 3 scores significantly different but showed no overall difference. The major conclusion was that the rater manual and training used was ineffective.

A. GUY LARKINS AND SALLY E. OLDHAM, "Patterns of Racial Separation in a Desegregated High School."

Patterns of racial separation were studied in a small-town desegregated southern high school with a population of 825 students, two-thirds of whom were black. The primary study sample consisted of the approximately 200 students enrolled in a mandatory American History and Government course. Data were gathered over a three month period through participant and non-participant observation, interviews, standardized achievement tests, and examination of class rosters and other school records. The data indicated that ability grouping tended to separate students by race, as did student selection of social studies electives and of seat location in social studies classrooms. Verbal interaction between students in social studies classes was intraracial, and social studies teachers tended to interact verbally with members of one race or the other but not both. Data concerning adult authority roles and student roles in extracurricular activities also indicated patterns of racial separation within the school. Recommendations are given for decreasing racial separation in desegregated schools.
D. I. ALLEN, “Children’s Association with Their Own and Other Countries.”

1,456 children from three grade levels in two centers in Canada, Australia and the United States completed open ended test instruments designed to explore their concepts of their own and seven other countries. Their responses were coded and ratios calculated to provide scores on stereotyping, egocentrism and evaluative comments. Differences between centers, countries and grade levels were tested for significance by analysis of variance. Some early stereotypes remain in the upper grades while others disappear and are replaced. Overall, stereotyping regarding other countries remained constant after a minimum knowledge basis had been acquired. Stereotypes about the students’ own country were less frequent at grade 6. Egocentrism diminished sharply at the higher grade levels, though there was considerable individual variation. Most comments were non-evaluative with an emphasis on differences in references to other countries. The image of each country was shared by children from the other countries. Early consciousness among American children of their national leaders, systems of government and national monuments was interpreted as a reflection of a national image rather than of greater political maturity.