DOCUMENT RESUME

ED 100 745 SO 008 025

AUTHOR Grannis, Joseph C.
TITLE On the Need for Criterion Referenced Research and Demonstration: A Reaction to the Model of National Assessment in Citizenship.
INSTITUTION National Council for the Social Studies, Washington, D.C.
SPONS AGENCY Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
PUB DATE [73]
NOTE 40p.; Related documents are SO 008 019-026

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS *Citizenship; *Educational Assessment; Educational Objectives; Educational Research; *Evaluation; Evaluation Criteria; Measurement Goals; *Measurement Instruments; Models; National Surveys; Participant Characteristics; School Environment; *Social Studies
IDENTIFIERS *National Assessment

ABSTRACT

This commissioned paper examines the National Assessment for Educational Progress model for citizenship and suggests an alternative model oriented more directly to educational and political goals. In the current model performances on the citizenship exercises can only be interpreted by the population characteristics of age, sex, parental education, race, region, size and type of community. The structural emphases of the model on population characteristics, rather than on the interactions of individuals with different environments, induces thinking in stereotypic terms. The basic question should be not who are good citizens, but what are the conditions that contribute to good citizenship. Data results in the age categories indicate a general tendency for the percentage of acceptable responses to increase with age, which denies Bruner's concept that anything could be taught in some honest way at any age. At the same time the model must recognize the identities and priorities of different social and ethnic groups which might contribute to differences in cognitive and affective development. The current model also ignores the differences in school environment. An alternative model would compare schools, programs, classrooms and take into account person-environment interactions and social and ethnic group goals. (DE)

ON THE NEED FOR CRITERION REFERENCED RESEARCH AND DEMONSTRATION: A REACTION TO THE MODEL OF NATIONAL ASSESSMENT IN CITIZENSHIP

Joseph C. Grannis
Teachers College, Columbia University

In this paper I want to underscore first of all two related meanings of the term "model," one emphasizing the structure of what is represented by a model, and the other stressing the phenomenon of influence. Next, I would like to note the potential interplay between these two aspects of the National Assessment model, with reference particularly to possible interpretations of the data in Citizenship Reports 2 and 9. Last I shall suggest an alternative to National Assessment, taking note of the gains National Assessment may have made, but at the same time orienting still more directly to educational and political goals, in this case to the fostering of citizenship.

SP 008 025

I. The National Assessment Model as Representation and as Influence.

This paper will dwell upon the obvious, and the distinction to be made here is no exception. Taylor's paper describing the National Assessment model and its uses[1] itself points to both the representation and influence as-

[1] Taylor, Bob L., "Potential Uses of the National Assessment Model at the State and Local Levels," University of Colorado, September, 1973.

pects of modeling. Thus my starting point is one of clarifying what he has written.

When we speak of the "model" of National Assessment we refer partly to some representation of its basic components and their articulation. Taylor's Diagram I[2] describes a succession of steps in the development, administration, and dissemination of the National Assessment instruments and findings. The multiple connections drawn in the diagram suggest complications in this process, as the first-cycle sampling plan was modified in anticipation of scoring and analysis, exercise development looked forward to the utilization of information, etc., and as these steps might be repeated in successive cycles.

Diagram I spells out, or differentiates, only steps of special concern to the assessor. A different diagram might elaborate the process between dissemination, utilization, and the redevelopment of objectives, suggesting different channels of national, state, and local dissemination, different uses of the assessment information as a function of different resources at these levels, and so on. Taylor's diagram reflects the putative neutrality of National Assessment vis-a-vis the uses of the assessment information. In various passages Taylor does suggest that this information calls for research, that it will be interpreted differently at state and local levels, and that it will lead to interventions in different spheres, economic, curricular, and

---

[2] Ibid., p. 3.

so forth. Still, it is the basic stance of National Assessment that these things be left to others to work out. From this standpoint Diagram I represents the National Assessment process accurately.

But a model is never neutral. Whether it has been drawn in physics or in politics and education, a model models a process of thought and action. In this simplistic way one can argue that what National Assessment models is assessment, and that complementary processes directed to changing what is assessed will be more an indirect, than a direct, result of National Assessment. It is not a trivial observation that the uses of National Assessment have in fact been, as described by Taylor, further assessment at the state and local levels. National Assessment's powerful coordination of money and technical expertise with academic and political judgment partly accounts for this imitation. Equally, however, it is the specificity of the model with respect to certain processes, and its lack of specificity with regard to others, that accounts for the imitation. Indeed, National Assessment itself can be said to be an imitation of a process that has been most influential in American education in recent decades, the testing of individuals for purposes that are a great deal more diffuse than the testing operations per se.[3] The criterion referencing of the National Assessment exercises might be seen

---

[3] White, Mary Alice and Jan Duker, "Models of Schooling and Models of Evaluation," Teachers College Record, 74, February, 1973, 293-307, makes the connection between testing and selection, a relatively diffuse, if still consequential process.

as a move beyond testing for its own sake, or beyond the parallel purposes of grading and selection associated with norm referenced tests. Whether this move is realized, however, and the testing thus does not again become an end in itself, depends on other processes that the National Assessment model does not explicate at present. If the assessment is supposed to be linked more closely to teachers' accountability than it has been in the past, there remains the risk that accountability will be displaced upon the population tested.

Let me now repeat this argument at another level of what is modeled by National Assessment, its representation of what citizenship is and how one goes about observing and explaining its occurrence or nonoccurrence. Reasoning simplistically again, let us first suppose that acceptable performances on the National Assessment citizenship exercises are the criteria of citizenship, i. e., that we would say someone is a good citizen if s/he performs in acceptable ways on the test, analogously to our saying that someone is a good driver if s/he passes a driving test.[4] How does National Assessment allow us to explain the occurrence or nonoccurrence of these performances? Age, sex, parental education, color, and location by region or size and type of community are the only variables that are systematically brought to bear on this question.

---

[4]This analogy is drawn, with much the same intent, in White, Naomi, "Statewide Testing Legislation for Minimal Skills: An Exploration of Some Issues," Educational Policy Research Center, Syracuse University, May, 1973.

One might conjecture that age represents different degrees of exposure to citizenship education, but whether it is explicit or implicit, or whether it stems from school, television, or community, defies conjecture except on a most ad hoc, exercise-by-exercise basis. Sex might relate to role opportunity, parental education to kinds of discourse in the home, color to aniindividual's treatment by others, and location to community needs and resources; but then we could scramble all of these conjectures too!, given the lack of specification of what the variables link with. What we are led to, then, if only by default, is an emphasis on the population characteristics themselves as explaining acceptable and unacceptable citizenship performances.

Caplan and Nelson[5] reported recently that, of 69 data-based, psychological research studies of blacks that were abstracted in the first six months' of the 1970 Psychological Abstracts, 82% were person-centered and only 18% situation- or environment-centered in their causal attributions. By far the largest category of variables Caplan and Nelson included under the heading "person centered" was "group membership (e. g., black or white)", which included 48% of all the 69 studies. Person-centered categories, the authors observe, are more available and more easily exploited than situational or environmental

[5]Caplan, Nathan and Stephen D. Nelson, "On Being Useful," American Psychologist, March 1973, 199-211.

categories. Discussing a variety of questions such as achievement motivation vs. the structure of the economy as causes of unemployment, Caplan and Nelson concluded:

> Whether the social problem to be attacked is delinquency, mental health, drug abuse, unemployment, ghetto riots, or whatever, the significance of the defining process is the same: the action (or inaction) taken will depend largely on whether causes are seen as residing within individuals or in the environment.[6]

Are we not dealing with a similar phenomenon? The effect of not specifying what age, sex, parental education, color, and location mean experientially is our attribution of good and bad citizenship to these qualities *per se*.

The National Assessment citizenship reports do speculate at certain points about the meaning of the population variables. The following excerpts from Report 9 illustrate the tentativeness with which this is done:

> Thirteen-year-olds in the Grade School group showed the greatest deficit of any group--10%--on the racial attitude exercises. Their responses to several questions describe where some of the deficits occurred, but don't explain why.... Perhaps as higher and higher proportions of the population get a high school education, the smaller proportion who do not become more distinctive in certain ways (*e. g.*, less accepting of other races).[7]

> Are Black youngsters less willing to tell a non-Black interviewer what they believe about racial discrimination? Or are they really less aware of racial discrimination than other 13-year-olds? The balanced results on these questions show smaller deficits (by at least half) than do the observed results, as we discussed earlier in this chapter (see page 51). At least one of the other characteristics on which the results are balanced is thus associated in some way with the results reported for Blacks.

---

[6] *Ibid.*, p. 201.

[7] Norris, Eleanor L., Vincent N. Campbell, Manford J. Ferris, and Carmen J. Finley, *National Assessment Report 9, 1969-1970 Citizenship: Group Results for Parental Education, Color, Size and Type of Community*, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, May, 1972. p. 25.

For example, perhaps the disproportionate number of Blacks whose parents have little education hear less discussion about acts of racial discrimination in the world.[8]

These and a dozen similarly limited speculations aside, the main tendency in the reports is to leave the findings to explain themselves. Presumably research to be designed and conducted by other agencies will lead to the development of interventions that will in turn result in changes in the assessment results the next time around. What may defeat this, however, is the psycho-logic of the model. The structural emphasis of the model on population characteristics, rather than on the interactions of individuals with different environments, induces thinking in these same stereotyping terms by those who receive the assessment reports.

The language of the citizenship reports consistently reinforces stereotyping:

> The two upper levels of parental education, particularly the Beyond High School group, excelled, as usual.[9]

> The typical performance of Blacks at all four ages shows deficits of about 9% on all Citizenship results combined.[10]

> The general picture is for performance on this goal to follow that on all Citizenship results fairly closely. Thus the Extreme Rural and Extreme Inner City respondents showed the greatest deficit in relation to the nation as a whole, and the Extreme Affluent Suburb respondents showed the greatest advantage....[11]

---

[8] Ibid., p. 57.

[9] Ibid., p. 37.

[10] Ibid., p. 46.

[11] Ibid., p. 92.

Of course, this is all very ironic. The reports feature as many exceptions to the general trends as they can find, and the avoidance of discussing the significance of the general trends is presumably calculated not to risk offense. But the patterns in the test results that are associated with the population variables are clues to individuals' interactions with their environments. That our thinking cannot afford to rest with these clues is the thrust of the next section of this paper.

The last point to be made in this opening argument returns to the analogy between a citizenship test and a driving test. Simplistically again, we have to notice that National Assessment presents what indivdduals say in response to various paper and pencil, interview, and discussion tasks as the basic model of what constitutes citizenship. Who can doubt that many moreindividuals know or will say that they should vote than actually vote in any election in this country, national, state, or local? Or that more believe they should oppose discrimination in a park, and can say how to do so, than are likely to put this belief and knowledge into practice? Kohlberg[12] has emphasized Hartshorn and May's classic failure to find differences between delinquents' and nondelinquents' knowledge of "right" and "wrong" actions, and has stressed instead the developmental level of the individual's justi-

[12]Kohlberg, Lawrence. "The Development of Children's Orientation toward a Moral Order," Vita Humana, 1963, 6, pp. 11-33.

fication of right and wrong. Kohlberg thus improves his capability of associating moral discourse with moral behavior, and, analogously, we might come closer in this way to associating citizenship discourse with citizenship behavior. The developmental approach, however, still leaves questions unanswered. Is it possible that those with "higher" levels of moral discourse are more capable of rationalizing their morality and immorality, and thereby more readily escape being branded delinquent or criminal? Our increasing awareness of "white collar" or copporation and political crime certainly gives credance to this question. Again, is it possible that different styles of language and discourse associated with socio-economic or racial/cultural differences affect the response of an individual to a verbal representation of a moral, or let us say a civic, problem, and affect equally an examiner's interpretation of the response? Here the questions we are raising with relation to what is known or shspected about moral judgment lead into the more general area of performance in a culturally standard verbal situation, in this case the formats that link testing and schooling so closely with each other. Labov's[13] finding that the task of talking to keep a rabbit from getting nervous (!) elicited much more talk from children who spoke Black English than questions addressed to the children by a sympathetic Black interviewer illustrates this concern.

---

[13]Labov, William. "Academic Intelligence and Black Intelligence?" Atlantic Monthly, June 1972, Vol. 229, No. 6, pp. 59-67.

Could all this have any bearing on who tends to "exhibit" less "knowledge and behavior considered desirable for citizens in our society"? Suppose the test was one of integrating a neighborhood, a bus, or a lunch counter, or that it was challenging a school bureaucracy, or that it involved hiring or working or playing or generally living with someone who had been in jail, as opposed to saying how one would do any of these things?

Part of what is at stake here is the relevance of the knowledge required by the assessment to the particular contexts in which different individuals enact their citizenship. This question applies not only to, say, who is more likely to be oriented to the courts as an institution for settling disputes about money, but who is more likely to have to accept doctors and dentists of a different color from one's self (just to indicate that the question can cut both ways!) Presumably the task force which is looking into the validity of the assessment exercises is considering this issue. What I would emphasize here is that the assessment is more likely to have content validity to the extent that it includes the enactment of citizenship in goal related contexts. This then connects with the other part of what is at issue, the format or structure of the setting in which citizenship is to be observed, which on linguistic grounds alone, as I have tried to argue above, is more likely to be valid as it is oriented to a citizenship goal, i. e., to a citizenship goal other than accept-

able performance on an exercise.

Not to mention how the assessment's having acceptable and unacceptable responses for each and every exercise contributes to the cultural standardization of "advantages" and "deficits"!

All of the above might seem to show that I have no use for information about what individuals say, or that I think differences would vanish if enactment, instead of saying, was assessed. Actually, I would like to know about both enactment and saying, or theknowledge and feeling saying represents. The National Assessment exercises do in fact include self reports of enactments, but, even accepting the reliability of self reports, these are still no more revealing in and of themselves than are sheer statements of knowledge and feeling. How knowledge and feeling facilitate or inhibit enactment is a question that especially concerns educators, along with how situational conditions facilitate or inhibit enactment. Does the civics text knowledge represented by many of the exercises contribute to effective action in some political situations and ineffective action in others? Or, to rephrase a different question asked in Report 9, what rules make most sense to different individuals participating in a common task?

Let us state the basic question still more generally. It is not who are the good citizens, but what are the conditions that contribute to good citizenship. Because National

Assessment is not designed to answer this question, however much it may speculate about it ex post facto. National Assessment does not model asking the question. We will now proceed further into the logic and psychologic of the problem this presents.

II. The Meaning of Patterns in the National Assessment findings.

The National Assessment reports tend to discuss the explanations of specific anomolies associated with one or another exercise, but they avoid discussing the meaning of the larger patterns that hit the reader full in the face. Here we shall take note of the some of the difficulties these larger patterns present. Our approach to this will be naive. We shall first orient to the data itself, asking what the patterns might be likely to suggest to the reader. Then we shall just begin to sort out some of the factors that are confused in these patterns, raising questions for an alternative program of research and demonstration.

A. Age as a Variable in Citizenship Performance.

In Report 2, straight percentages of correct or acceptable responses at different age levels are shown for each exercise dispositon. Thus the following results are exhibited:

	9	13	17	Adult
Report that the police do not have the right to come inside one's house at any time they want and can give as a reason legal guarantees, or reasons concerning privacy and permission of occupant to enter (in own words).[14]	20%	68%	90%	83%

[14]See page following for footnote.

	9	13	17	Adult
State that our legal system (courts, laws) is the means provided by government for settling an argument over money.[15]		50%	70%	87%
Last names of the persons now holding these offices....:	9	13	17	Adult
President (Nixon)	91%	94%	97%	98%
Vice President (Agnew)[16]		60%	75%	87%
Could give at least 1 explanation of what fighting was about in country named.		13	17	Adult
		53%	66%	77%
2 explanations		27	44	55
3 explanations		11	24	31
4 explanations		3	12	16
5 explanations[17]		1	4	7
Opportunity to read a greater variety of viewpoints and information was stated as a reason why it might be good to have newspapers in a city written and printed by more than one company.[18]	9	13	17	Adult
	37%	64%	88%	92%

As these examples illustrate, the general tendency in the data is for percentage of acceptable response to increase with age, though with some reversals between age 17 and adult. This trend is alluded to in various specific contexts in Report 2, for example in the following comment:

As seen above, adults consistently showed more knowledge about current conflicts than did the other ages. When asked in Exercise F4 to name some ways to avoid war, however, fewer adults than 13's and 17's named at least one way (77% of 13's; 88% of 17's; and 65% of adults). Even 9-year olds approached the adult achievement level, 60% giving at least one way to avoid war (Exercise F3). One explanation is that the adult view of the possibility

---

[14] Campbell, Vincent N., et al, Report 2, Citizenship: National Results, Education Commission of the States, National Assessment of Educational Progress, Denver, Colorado, November, 1970, p. 30.

[15] Ibid., p. 41.

[16] Ibid., p. 54.

[17] Ibid., p. 87

[18] Ibid., p. 104.

of avoiding conflict is both more sophisticated and more "jaundiced"; a larger number of adults indicated that they felt war to be unavoidable and a larger number at the younger ages gave simplistic, but acceptable answers ("stop fighting").

The numbers who gave at least three ways to avoid war demonstrates the more usual age trend of an increase in achievement up to age 17 with minor differences between 17's and adults (5% of 9's; 12% of 13's; 32% of 17's; and 25% of adults). A similar 9 to 13 age trend is shown in Exercise F5 which supposes competition between the U. S. and Russia concerning territorial rights on Mars. Almost twice as many 13's as 9's (75% vs. 45%) stated that the U. S. and Russia should discuss and settle these matters before men land on Mars.[19]

Notice that in the discussion above, as is the case elsewhere in the report, it is the departure from the "more usual age trend" that seems to call for explanation. How does one account for the main trend itself? Leaving aside for the present a consideration of how the exercises were constructed in relation to age expectations, and assuming the content or goal referenced validity of the exercises, the conclusion that best fits the pattern in the results is that they reflect maturation and/or the general accumulation of knowledge and experience with age. This is obvious, is it not? (And I did warn that this paper would dwell upon the obvious.)

What is less obvious is the significance of the conclusion for citizenship education. Many, but probably not all readers of this paper will be aware that social studies educators in the 1950's (I choose this date for convenience of discussion) had occasion to assess the significance

---

[19] Ibid., p. 84.

of similar findings arising out of a number of surveys of children's social concepts conducted during the Progressive Education era: Meltzer, 1925; Lacey, 1932; Pressey, 1934; Eskridge, 1939; Ordan, 1945; and Bates, 1947, to name those I am most familiar with.[20] All of these studies included elementary school pupils at different grades, and several of them included secondary school pupils as well. All tested children's attainment of the correct or conventional meanings of social concepts at different ages, and all found progressions in correct attainment with increasing age. Interestingly, only elementary social studies educators, with the exception of Wesley and Wronski[21] (Wesley himself was author of a "Test of Social Terms" in 1932[22]), paid attention to these findings, directly or indirectly, to judge from a comparison of elementary and secondary social studies methods texts of the 1950's. At least in the decade before the impact of sputnik, Bruner, and the structure movement in curriculum, it was elementary educators who were more concerned with the development of knowledge <u>over time</u> or age, while secondary educators focused on its acquisition <u>at a given time</u> or age.

---

[20] Meltzer, H., "Children's Social Concepts: A Study of Their Nature and Development," <u>Teach. Coll. Contr. Educ.</u>, No. 192, 1925; Lacey, JoynM., "Social Studies Concepts of Children in the First Three Grades," <u>Teach. Coll. Contr. Educ.</u>, No. 548, 1932; Pressey, L. C., "A Study in the Learning of the Fundamental Special Vocabulary of History from the Fourth Through the Twelfth Grades," Truman L. Kelley and A. C. Krey, eds., <u>Tests and Measurements in the Social Sciences</u>, New York, Scriberner's, 1934, 155-218; (continued on page following)

But which emphasis was correct? The elementary educators stressed a progression of learning from concrete to abstract, the necessity of building up manifold experiences with the referents of a concept, and the importance of teachers' avoiding empty verbalisms and premature formalization. What precipitated out of this, however, was such an emphasis on accomodating to the presumed "natural" pace of concept attainment reflected in the progressions of attainment with age, that Bruner's declaration that anything could be taught in some honest way at any age, hedged though it was in Piagetian conditions, came as a shock, particularly to elementary educators.

None of the social concept studies cited above, nor any others conducted in the years before 1950, tried to assess children's capability of learning a concept at a given age by systematically attempting to teach the concept to children. Lacey did interpret an observed acceleration

---

Eskridge, T. J., Jr., "Growth in Understanding of Geographic Terms in Grades IV to VII," Duke Univ. Res. Stud. Educ., No. 4, Durham, North Carolina, Duke Univ. Press, 1939; Ordan, H., "Social Concepts and the Child Mind," New York, King's Crown, 1945; Bates, F. L., "Factors Related to Children's Understanding of Social Concepts," unpublished doctor's thesis, University of California, 1947.

[21] Wesley, Edgar B. and Stanley P. Wronski, Teaching Social Studies in High School, Boston, D. C. Heath & Co, 1958.

[22] Wesley, Edgar B., "The Wesley Tests in Social Terms," Truman L. Kelley and A. C. Krey, eds., Tests and Measurements in the Social Sciences, New York, Scribner's, 1934, 219-226.

of concept learning at grade 3 as being the result of a more deliberately planned social curriculum than in grades 1 and 2. Though I do not know them, there may well have been many studies carried out during these years that weighed the relative merits of one method of teaching social concepts vs. another, as a variety of more recent studies have done.[23] Who, however, has set out to teach social concepts to a criterion at a given age, i. e., to teach until the pupils reached the criterion, and in this way studied the methods that resulted in different children's learning at that age? We may often teach with reference to a criterion, but in the regular or the experimental social studies or civics classroom we typically abandon a unit of instruction, to move on to another unit or to discontinue an experiment, while some pupils, at least, still have not met what we might hold as even a minimum criterion. Of course, this raises all kinds of other questions, about the desirability of convergent vs. divergent learnings, etc., but that is beside the point here. What I sense is that there are a variety of factors, perhaps especially in social studies and civics, that contribute to a criterion referenced system devolving into a norm referenced system. One of these factors is the very knowledge that individuals' differences with respect to attainment of a criterion held to be "reasonable" for a given age will tend to diminish as

[23] For references to these, see the annual reviews of research published in Social Education since 1968.

more of these individuals attain the criterion past that age.

Without understanding the issue completely, I think there is an ambiguity in National Assessment's own position on this matter. On the one hand, we have been told that "Outstanding local teachers familiar with each target age group (ages 9, 13, 17, adult) worked for weeks with our staff to break down each general objective in the most germane behaviors appropriate as goals for a given age group."[24] On the other hand, we learn that one criterion for exercise development was that "some exercises cover important attainments which nearly everyone is successfully achieving, some which very few people are achieving, and some which a middling number achieve."[25] Does the latter aim, together with the finding that National Assessment performances did in fact break down this way for each major citizenship goal,[26] imply that such a distribution should continue to be obtained in future assessments? Perhaps the aim of National Assessment is to provide information that will encourage state and local striving toward full attainment of the goals for all individuals. Taylor, however, has observed a tendency for the National Assessment data to be treated as a national norm

---

[24] Campbell, Vincent N. and Daryl G. Nichols, "National Assessment of Citizenship Education," Social Education, 32, March, 1968, p. 280.

[25] Campbell, Report 2, Citizenship: National Results, op. cit., p. 5.

[26] Ibid., p. 6.

against which both state and local test performances are being compared.[27] Given the statistical patterns within and between age groups that National Assessment models, I think this is inevitable.[28]

The National Assessment findings are likely to be taken as developmental norms unless several steps to counter this are taken. (1) The exercises would have to be broken down into their more truly developmental and nondevelopmental components. (2) The assignment of exercises to age levels would need to be done from the standpoint of what it is theoretically possible to expect any educable individual of a given level to achieve. And (3) experiments would have to be conducted to demonstrate that 100% of the educable individuals at an age level could achieve acceptable performances for the exercises at that level.

This is a rather drastic prescription, with many problems inhering in it. Its function in this argument is to underscore the fact, recognized by National Assessment but likely to be overshadowed by the psuedodevelopmental patterns in its findings, that National Assessment so far has not tried to take cognizance of what has in fact been deliberately taught, specifically in the schools; thus its findings should not be construed as suggesting limits to

---

[27] Taylor, op. cit., p. 20, p. 25.

[28] Ebel has made the interesting observation that the percentage system of grading that prevailed 40 years ago and before was a criterion referenced system. He discusses some of the reasons for its yielding to a norm referenced system in Ebel, Robert L., "Criterion Referenced Measurements: Limitations," School Review, 79, February, 1971, 282-288.

what individuals could be taught, or could learn, at the different age levels of the assessment.

But we did admit that this has its problems. One of these is distinguishing what we called above more "truly" developmental from nondevelopmental components of learning. It seems to me, for example, that any 9 year old could learn that policemen do not have the right of unrestricted entry to a home. At the same time, 9 year olds' reasons for such a belief would be less sophisticated than 13 year olds' reasons, focusing at the first level on, perhaps, the more direct consequences to the persons involved, and at the second on more general properties of the social system like the rights of individuals to privacy and the effects of not regulating entry into homes on other spheres of the social system. Instruction might then concentrate on all 9 year olds attaining at least the first developmental level of justification, and on all 13 year olds attaining at least

criterion in citizenship is more validly held for individuals or for the society as a whole. Consider the case for the latter standard first. Almond and Verba[30] have suggested that a political system might get overheated if too many of its citizens participated actively in the process of governance, beyond voting, and thus that there might be optimal levels of less than full participation for the stability of even a democratic system of government. (Almond and Verba were, indeed, rationalizing the levels of participation they found in American society.) Apart from whether or not one agrees with the value of stability in the system, this does lead one to distinguish between citizenship goals for the nation or for a region as a whole and goals for all the citizens of the nation or a region. Maybe only a few are needed to, for example, come up with many ways that war could be avoided or that discrimination in a park could be stopped. When Report 2 referred to "goals that only a select few were achieving,"[31] I wondered if a part of what this implied was that only a few are needed to achieve these goals.

The trouble with this reasoning, however, is that the many who do not achieve the goals in question may include precisely those whose interests are most at stake, viz., those most likely to be frontline soldiers in a war or most likely to be excluded from a park. Once we allow that less than all need to achieve one goal or another for the health

---

[30]Almond, Gabriel A. and Sidney Verba, The Civic Culture, Boston, Little, Brown and Company, 1965.

[31]Campbell, Report 2, op. cit., p. 5.

of the body politic as a whole, we run the risk of playing into just those features of the system that presumably account for some groups in the population being consistently "disadvantaged" no matter what the goal in question is. Furthermore, many of the goals set by National Assessment pertain to the rights and obligations of individuals qua individuals, in their everyday relationships to one another and to the law. Anyone of a relevant age who does not know that the courts are available for the resolution of a dispute over money might be disadvantaged in the event of such a dispute. Similarly, anyone who discriminates against another on the basis of his or her group identification contributes to the denial of the rights of the other to be treated as an individual.

A third problem arises from the fact that, especially if, in line with the position advocated earlier, instruction with respect to a particular goal at a given age was pursued until all individuals so instructed reached the goal criterion, there might not be time enough to reach All the goals held to be desirable for individuals at that age.

Still another problem is that it cannot be assumed that school is the most appropriate setting in which to intervene to try to bring individuals to a criterion. But then there is no way of knowing from the present design of National Assessment what part the schools have played in the performances on almost any of the assessment exercises. Even specifying which criteria the schools could effect would be progress from this standpoint. Especially for those criteria that we

have characterized as "enactments," it might be that school would have to be coordinated with other agencies or institutions, family, local government, citizens groups, television, and so on. National Assessment may assume that efforts of this sort will be an outgrowth of the publication of assessment results. It would be a more likely outcome if the contribution of such efforts to the attainment of goals was specified.

What all of this reasoning drives me to, then, is a different kind of neutrality from that which is modeled by National Assessment. Instead of publishing results that encourage normative thinking in the face of virtually no knowledge of the conditions of individual citizenship learning and development, an alternative strategy should concentrate on specifying the conditions that make full attainment possible for one or another criterion at a given age level. The freedom of state and local agencies or institutions to emphasize those goals that mattered most to them would thus be an enabling one, rather than the spurious freedom that sheer ignorance affords us.

B. Sex, Parental Education, Color, and Location by Size and Type of Community as Variables in Citizenship Performance.

The logic with respect to these variables is essentially the same as it has been above, though I will be repeating it in somewhat different terms. First, let us attend again to the psycho-logic of the matter. The very familiarity of the National Assessment findings, especially with respect to

parental education, color, and location, may tend to give them a normative significance, i. e., to imply that the obtained patterns are what we should continue to expect. For me, at least, this stereotyping is reinforced by the technique of comparing group performance levels with national levels, rather than with the criterion of 100%, so that there will regularly be groups with "advantages" and groups with "deficits", relative to each other (shades of norm referencing?), even when, say, it is 60% of one group and 50% of the other that has reached a criterion. Be that as it may, the familiarity of the results that I am referring to is their comparability to the findings of any number of assessments, from the Army Alpha intelligence testing program that might be said to have started it all, to the Coleman Report and still more recent studies. The sheer accumulation of these findings has contributed to their reification in recent years, so that in the Jencks analysis socio-economic status and color appear to be almost intractable constraints on education, or at least on education by the schools.

Let me focus this thinking in terms of the Coleman and Jencks analyses, raising a few questions at a very superficial level. First, these studies base their (somewhat differing) conclusions about the limited capacities of schooling to affect achievement on achievement defined in very close relation to general cognitive skill, with all of the cultural loadings that the phrase "general cognitive skill"

implies. As Jencks himself observes, the Coleman tests of Verbal Ability, Nonverbal Ability, Reading Comprehension, Mathematical Achievement, and General Information intercorrelated so highly that "The student who did well on one test and poorly on another was quite exceptional."[32] Assuming for the purposes of this argument that socio-economic and cultural factors reflected in the school will largely determine the school's effect on general cognitive skill, it remains possible that schooling can have a more independent effect in areas of achievement that are more specific to school instruction and, conversely, less generally diffused in the society at large. The International Education Assessment's current research on factors affecting achievement suggest, for example, that the effectiveness of science instruction is substantially more independent of non-school factors than is the effectiveness of instruction in reading or arithmetic.[33] Pursuing the principle that might explain this, we first have to account for IEA's finding that their Social Studies-Civics Education test results behave more like reading and arithmetic than like science!

The IEA researchers themselves attribute their finding to the permeation of citizenship knowledge, like the 3 R's, throughout life outside the schools. It makes sense from this standpoint that the citizenship results would reflect

---

[32] Jencks, Christopher et al, Inequality, New York, Basic Books, 1972, p. 55.

[33] "Scholars Weigh Education Data from Twenty Countries," Report on Education Research, 5, December 5, 1973, pp. 3-4.

varying access to, and occasion or power to use, the knowledge that the tests incorporate. But is not science also diffused throughout life in a modern society? Of course this is the case. One must reason then that science education and the science test exercises have been couched at a higher or more specialized level of knowledge than is reflected in every day life, and that citizenship education and/or the citizenship test exercises have not been defined at the same relatively high or specialized level. Studies of the learning of higher mathematics, as opposed to the common mathematics the Coleman (re-analyzed by Jencks) and IEA studies were concerned with, do show that differences in schooling make a difference.[34] What is needed, then, is to demonstrate a similar effect in citizenship.

What might be meant by "higher" or "more specialized" citizenship knowledge? One need not think of it simply as, say, secondary school or college level information and concepts, but rather as knowledge that is couched at a higher or more complex level than which seems to be reflected in citizenship achievement at any given age under conditions other than systematic instruction. Thus one would aim to teach at a given age what National Assessment or some other survey shows is not "normally" achieved at that age, or perhaps not even at a subsequent age.

Another, complementary way of construing what might be meant by higher or more specialized citizenship knowledge

---

[34]Dyer, Henry,

stresses, instead of official or ideological doctrine, a more skeptical, probing interpretation of citizenship. Neither National Assessment nor IEA has included knowledge exercises of this sort, although the IEA instruments do include attitude or opinion items that question the economic and political systems of the society. Can we say that the schools should also aim to teach to individuals of a given social background that which is not "normally" known to persons of that background in our society, be they rich or poor, and should cultivate questioning in that social sector of a sort not "normally" realized in that sector?

Arnoff's[35] investigation of factors related to the ability of children in grades two, three, and four to comprehend concepts of government points in the first direction suggested above. Arnoff designed a 5 week government curriculum to include (though not exclusively) concepts not ordinarily included in instruction for these grades, i. e., not included in social studies textbooks at these grade levels. Arnoff's results clearly showed the effect of instruction. 75% or more of second graders, for example, learned 23 new concepts of local, state, national, and general government: property tax, split ticket, subpoena, judge, etc., most of which were not included in second grade social studies textbooks. Furthermore, social class tended not to correlate with more or less learning of new concepts

---

[35] Arnoff, Melvin, "An Investigation of Factors Related to the Ability of Children in Grades Two, Three, and Four to Comprehend Concepts of Government," doctoral dissertation, Ohio University, 1963.

in this experiment, though mental age as defined by an intelligence test did correlate with new learning.

A field trial of the American Political Behavior course developed by the High School Curriculum Center in Government at Indiana University[36] points partly in the second direction indicated above. The APB course aimed to teach "facts and ideas about politics that have not been part of typical social studies curricula," [37] for example, facts about the proportions of different income groups that vote in American elections. Differences between experimental and control groups' pre-test/post-test knowledge gains were clearly demonstrated in all 9 communities involved in the field trial. On the other hand, comparable differences in political science skills achievement were demonstrated in only 4 of the 9 communities, and effects on students' attitudes were demonstrated in none of these communities. Furthermore, while there was some variation of student background characteristics, the communities involved did not include the rural and inner-city extremes identified in National Assessment, and almost all of the students were white.

One should not suppose that systematic instruction could eliminate the effects of population or student background characteristics, or indeed that this adequately represents what is desirable. The effects that National Assessment makes us principally aware of are those that stem

---

[36] Patrick, John J., "The Impact of an Experimental Course, 'American Political Behavior,' on the Knowledge, Skills, and Attitudes of Secondary School Students," Social Education, 36, February, 1972, 168-179.

[37] Ibid., p. 168.

from restrictions on different groups' access to and control of information; if citizenship education should aim to minimize these effects, still this would surely entail interventions outside the schools, and research and demonstration would need to be directed this way as well. At the same time, the schools must recognize the identities and priorities of different social groups, and, paradoxically, whatever equalization the schools accomplished with respect to the distribution of information and skills might contribute in some ways to heightened affective differences between groups.

A second problem in the Coleman study, one that the Jencks analysis only partially rectifies and that National Assessment simply ignores, is the question of comparability of school environments, in other than SES terms, between and within different locations. It is well known that Coleman compared whole systems as to library size, perpupil expenditure, level of teachers' formal education, and so forth, whereas these conditions might well vary between schools within a system. Jencks used Project Talent data to make these comparisons between individual high schools, and still obtained no effect independent of non-school factors. The IEA science findings do show such an effect, not for all the variables that might be thought to be relevant, but for "the opportunity to learn, the student time in hours per week and cumulative years, the curriculum emphasis, and the additional years of post-secondary preparation

of teachers."[38] In none of these analyses has classroom climate or methodology seemed to account for differences in students' achievement.

My own position is that one must compare not just programs, or schools, or even classrooms, in order to specify the environmental conditions that affect schooling's contribution to achievement. Rather, it is necessary to examine different settings within classrooms, or within school-related locations outside the classroom, in order to make headway with the question of educational environments.[39] In my approach to this problem over the past few years, I have found it useful to distinguish between high teacher control, joint teacher and learner control, and high learner control of each of various conditions within one or another classroom setting (or subsetting), and thus to be able to ask how much time the students in a given classroom spend under different conditions of control. Task options, pacing, teacher-learner and learner-learner interaction, teacher adaptiveness, and task performance criteria, are some of the conditions of settings that can be observed to vary with respect to control. So called "open classrooms", for example, can be seen to vary among themselves in the proportions of time that students spend under these different conditions of control, whether one is comparing time

---

[38] "Scholars Must Weight Education Data from Twenty Countries," op. cit., p. 3.

[39] Grannis, Joseph C., Columbia Classroom Environments Project Final Report, Contract No. OEC-0-71-0593, Teachers College, Columbia University, February, 1973.

for aggregates of students between two or more classrooms, or is comparing time for individual pupils within a single classroom.

This approach further distinguishes between different types of educational purpose: the transmission of knowledge, tradition, and experience to all learners in a setting alike; the operationalization of partly common, but also partly individuated, competencies or skills and concepts; and the cultivation of individual and collaborative venture, or exploration, problem solving, and expressive composition or construction. Intellectual, social, physical, and emotional aspects of learning and development are equally involved in each of these categories of purposes.

I have hypothesized both that the effectiveness of an educational setting will be partially dependent on the internal consistency or congruency of its controls, and that each broad type of educational puppose will be realized best in a different type of setting, viz., transmission in a high teacher control setting, operationalization in a joint teacher and learner control setting, and venture in a high learner control setting. It is apparent, then, that I believe that what has variously been called "classroom climate" or "classroom method" will someday be understood to relate to achievement, or to type of achievement, even though the results of research on this to date have been very confusing.

This paper has emphasized education in the schools,

but it has recognized at morethan one juncture above that going beyond learner or citizen population variables to the interactions of indivdduals with their environments would entail research and demonstration outside of the school as well. What are the effects on individuals' acknowledgement of racial discrimination, of this acknowledgement's actually contributing or not to ameliorating the discrimination? How is one's orientation to a legal system for settling an argument over money affected by one's having, or anticipating having, money enough to invoke the legal system in a dispute? Does involvement in the affairs of a responsive public agency or institution lead to greater participation in other public or community affairs?

My paradigm for these questions, and many more, is Kohn's[40] research on the conditions that determine fathers' values for their children. Social class, defined as occupational position and education, does relate in Kohn's data to whether parents tend to value self-direction or conformity to external authority, higher social class being associated more with the first, and lower social class more with the second of these value clusters. When the analysis controls for the degree of self-direction or autonomy that the fathers experience at work, however, social class differences tend to vanish, lower class fathers who experience autonomy in their jobs valuing autonomy for their children

[40]Kohn, Melvin L., Class and Conformity, Homewood, Dorsey Press, 1969.

as much as middle class fathers who experience this autonomy do. In other words, while social class is an approximation to the conditions that determine fathers' values for their children, the experience of fathers in their work describes these conditions more exactly. Is it not possible that research could similarly track down conditions associated with, but still independent of sex, parental education, color, and location, that would more exactly explain the associations of citizenship achievement with these variables? Further research could then be directed toward changing these conditions experimentally, so that, just as we might demonstrate in what way achievement other than that which is "normally" associated with different age levels is possible, likewise we could show how achievement other than that "normally" associated with different sex, parental education, color, and location statuses is possible.

III. An Alternative to the National Assessment model.

From various directions, our argument converges on the desirability of emphasizing person-environment interactions, and demonstrating the achievement that is possible under varying person-environment conditions. Consider one further vantagepoint that differs somewhat from ours so far. Etzioni,[41] in a discussion of organizational analysis, criticizes that assessment of organizations which focuses on goal attainment. "One of the major shortcomings of the goal model is that it frequently makes the studies' findings stereotyped as well as dependent on the model's assumptions."[42]

---

[41] Etzioni, Amatai, "Two Approaches to Organizational Analysis: A Critique and a Suggestion," Administration Science Quarterly, 5, 1960, 257-278. [42] Ibid., p. 258.

An organization frequently does not reach its goals effectively, and often has goals other than the ones it claims to have. It would be more useful, Etzioni concludes, to treat goals as cultural entities in themselves, and to ask how various internal and external conditions contribute to the realization of observed goals.

I do think National Assessment's attempt to define citizenship goals and to describe performances that represent the achievement of these goals is valuable. The fact that National Assessment anticipates modifying the goals and exercises to reflect changes in priorities between assessment administrations indicates at least some awareness of the goals as "cultural entities." But what is served by publishing only those statements of goals that represent the "consensus" of those consulted? Would not the publication of disagreements over the definition of goals, particularly disagreements stemming from different subcultural interpretations of citizenship, add further to the de-mythification and de-stereotyping of citizenship goals?

What stands in the way of this process is, once again, the National Assessment model, its structure and assumptions. Let me claim intuitively that a national test of citizenship simply cannot be consistent with cultural and political pluralism. National Assessment <u>is</u> first and foremost a test, and as such depends upon conformity to its goals for its authority.

An alternative to National Assessment would conduct

research and demonstration in specific relation to citizenship goals, but would recognize that these goals have different meanings and priorities for different individuals and groups. The alternative should, indeed, explore these differences explicitly, as an understanding of them would be essential to any application of the research and demonstration.

What I have in mind first of all, then, is a pluralism of inquiries to replace the monolithic National Assessment, inquiries that could be drawn upon differently by different agencies or groups, or rather, since anything can be drawn upon in this way up to a point, that encouraged different agencies or groups to order their efforts according to what is most meaningful and important to them.

An alternative to National Assessment should be truly criterion oriented. Its aim should be to use analysis and intervention to bring the observation of performances as close as possible to criterion, allowing that different agencies and groups will draw differently on this research and demonstration in subsequent applications of it.

Many separate studies would be involved, but their intent would be the same. The basic paradigm would be that which shows convergence on the achievement of goals as the result of reducing differences between groups, or aggregates, of different statuses, by specifying the conditions associated with these statuses that originally account for the differences in goal achievement. Is the seemingly lesser awareness of racial discrimination on the part of

Blacks a result of unwillingness to disclose this awareness in a test or interview? Then perhaps the difference would diminish as observations were conducted in settings in which the respondents felt more in control. Are parents of lesser educational attainment less involved in the politics of their children's schools because of a feeling that school had not been responsive to them as children? Comparing groups on the basis of this feeling might reduce the difference attributable to status; and intervention through the schools to respond to alienated parents might increase the parents' involvement in their children's schools.

Probably hundreds of studies that throw light on the specific conditions contributing to citizenship have already been reported. These could be indexed by goals, and again by conditions, so as to make this knowledge available to policy makers at different levels. Somewhat in example of this is a document prepared by the Social Research Group at The George Washington University, Research Problems and Issues in the Area of Socialization, 1972, part of which analyzes what is known about the development, the determinants, and the changing of intergroup and intragroup attitudes and behaviors. Of course, such an analysis points as well to what is not known. The following quotation from the Social Research Group's report represents a juncture that is frequently arrived at in their analysis:

> The significance of some of these findings on cooperation is far from clear. We still do not know how cooperation determines intergroup attitudes and behaviors or its

role as a factor in achieving a successful ethnic and social class mix. What does seem clear from the research is that cooperation sometimes leads to better intergroup relations, although much work is needed to determine the conditions under which cooperation produces an enhancing effect.[43]

Much work does need to be done. Some of it consists of formal studies in which the investigator attempts to control the principal variables and to predict outcomes precisely in advance. Often, however, these studies are too rigid to be able to deal with the unintended factors that enter into every complex action--which may partly account for formal studies' failing to confirm their hypotheses. Action research, in which the action is typically guided by more evolutionary goals, and in which "real time feedback" continuously regulates the participants' activity in relation to goals, compensates in flexibility for what it may lose in control. A formidable agenda of both formal and action studies, then, must be undertaken, all of it referenced with respect to goals and conditions alike.

Perhaps the case could be made that developing a knowledge base for citizenship education in this way would still require a determination of national performance levels as in National Assessment. In my opinion, however, this would be distracting at the least, and possibly destructive. An alternative way of construing the "national" significance of a study would be that it deals with problems that are best

---

[43]Sowder, Barbara J. and Joyce B. Lazar, Research Problems and Issues in the Area of Socialization, Social Research Group, The George Washington University, Washington D.C., 1972, p. 82.

approached with national resources, while a more local study deals with problems for which local resources are adequate, or, to put it differently, for which only local resources are likely to be appropriate. Thus a study could have national significance even if it did not have a national sample (which is not the same as saying it would not be carried out in a variety of communities). The effects of income maintenance on citizenship performance might be construed as a national question because only the federal government has the money to maintain incomes. The effects on citizenship performance of collaboration between schools and other community institutions or agencies might be construed as a local question because only local individuals or groups have the power to bring about this collaboration.

Clearly a citizenship research and demonstration program could go on forever, in so far as our awareness of what we do not know multiplies at the same rate, at least, as does our knowledge. But then National Assessment equally could continue forever, since the goals it sets will probably change in like manner. Substantial human and material resources have been invested in National Assessment, but this is not the only, or even the chief reason, to think that National Assessment will be continued. The technology of National Assessment, as of testing more generally, peculiarly lends itself to the collusion of academic ingenuity and political decision making. Were the questions this paper raises

foreseen? Probably some of them were, and others were not. If the present form of National Assessment is itself the result of compromises meant to render the project less dangerous politically, this does not bode too well for the alternative that I propose. On the other hand, what are the political consequences of what National Assessment hath wrought? These too will have to enter into the equation.