CSE's INTERNATIONAL MONOGRAPH SERIES IN EVALUATION

Impact of a Major National Evaluation Study: ISRAEL'S VAN LEER REPORT

by Arieh Lewy & Marvin C. Alkin

Center for the Study of Evaluation
University of California
Los Angeles

Impact of a Major National Evaluation Study:

ISRAEL'S VAN LEER REPORT

Arieh Lewy
and
Marvin C. Alkin

April 1983

CSE INTERNATIONAL MONOGRAPH SERIES IN EVALUATION

Center for the Study of Evaluation
UCLA Graduate School of Education
University of California, Los Angeles
Los Angeles, California 90024

This monograph series is supported all or in part by the National Institute of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

TABLE OF CONTENTS

The Center for the Study of Evaluation (CSE) at UCLA is a national research center devoted to improving the quality of schooling through systemic evaluation practices. Created in 1966 as a result of a national competition, CSE has worked continuously to develop more valid methodologies for testing and evaluation and has promoted vigorously the use of evaluation for more reasoned educational decision-making.

Eva L. Baker, Director

Joan L. Herman, Associate Director

UCLA Graduate School of Education
145 Moore Hall
Los Angeles, California 90024
(213) 825-4711

ACKNOWLEDGEMENTS

This report was completed under the auspices of the Evaluation Practices Program of the Center for the Study of Evaluation, University of California at Los Angeles. Our thanks to the Center and its director, Eva Baker, for the continuous support of this and other program efforts.

Special acknowledgement should go to Laura Kent, whose editing assistance added to the professionalism of the final manuscript. Phyllis Jacobson ably assisted in coordinating administrative details related to production.

INTRODUCTION

This monograph documents the impact of the so-called Van Leer Study (Minkovich et al., 1977), a major evaluation study of achievement in the primary schools of Israel, originally conceived as a counterpart to the Coleman Study in the United States and the Plowden Report in Great Britain. Using a case-study approach, we seek to examine in depth the effects of the Van Leer Study on the policy and actions of the Israeli Ministry of Education, the Knesset (Parliament) and on subsequent research actvities.

THE STUDY OF EVALUATION IMPACT

Evaluation is intended to be more than a theoretical exercise. As Weiss (1974) notes, unless evaluation "gains serious hearing when program decisions are made, it fails its major purpose" (p. 314). Because evaluation is supposed to contribute to program operations, it is not surprising that the utilization, or the impact, of evaluation studies has been a matter of concern during most of the history of this field.

One's conclusions about whether evaluation has an impact will depend largely upon one's definition of evaluation and one's expectations. For example, one view holds that evalu-

ation is, or should be, the sole or major influence on any program decision that is made consciously, rationally, and at an observable point in time. Since this expectation is rarely met, evaluation so defined can be said to have failed in its major purpose. According to an alternative view, elaborated by Alkin and his colleagues (1979) and by Patton (1978), evaluation and other kinds of information may work together, over time and in a variety of ways, to influence decisionmaking. Since this expectation is more frequently met, evaluation so defined does achieve its major purpose.

This alternative definition of evaluation and utilization, which grew out of empirical studies, refers to a phenomenon that is usually subtle in its impact, that can assume a variety of forms, that depends on the combined technical skills and personal attributes of both the evaluators and the key decisionmakers, and that is mediated by the political, social, and organizational context of the specific educational program. Given the complexity inherent in this view of evaluation and its utilization, it is critical that the phenomenon be carefully studied, in order to add to our understanding of how impact takes place and to suggest ways for improving evaluation's potential utility.

So far, relatively few empirical studies of evaluation utilization and impact have been conducted in accordance with this alternative view. Studies of evaluation use at the national level have consisted chiefly of questionnaires

and interviews, occasionally enhanced by reference to such products as evaluation reports and other program-related documents. For example, Alkin and his colleagues (1974) examined the impact of evaluation on decisionmaking about bilingual education programs. This study, based on questionnaire results and examination of program documents, not only noted instances of evaluation utilization but also contributed to the alternative definition of evaluation and utilization offered above. Patton and his associates (1975) conducted interview studies of national health programs which also yielded instances of utilization and supported the alternative definition. Similarly, Reisner and her associates (1982) interviewed compensatory education administrators at the state level and found instances of evaluation impact. At the national level, Leviton and Boruch (see Chapter 6 of Boruch and Cordray, 1980) documented instances where evaluation information of various types was used to modify laws and regulations or to alter management practices; their data sources were interviews and citations in congressional and executive documents. This study fails to provide a cohesive picture, however, because it gives attention to any and all evaluations.

By way of contrast, the present study focuses on a single national evaluation study and documents those events following its publication which can be regarded as demonstrating impact. The importance and visibility of the Van Leer

Study, and the greater manageability associated with observing impact in a small country, offered a perfect opportunity for furthering research knowledge on evaluation impact.

METHODOLOGY

In carrying out this study, we relied heavily on close examination of all relevant primary-source documents: not only the Van Leer Report itself, but also other documents that reflected the reactions of individuals, groups, agencies, and the general public to the Report. We have restricted ourselves to written documents. Though fully aware of the possibile partiality of such kinds of information, we wanted to avoid exegetics and rationalization. We did not want to listen to verbal explanations of why something had happened or to post-hoc justfications such as "What I really meant when I wrote this is such-and-such."

Further, the issue is sensitive, and we did not want to make people uncomfortable. For instance, the chief Scientist of the Ministry of Education--whose published critique noted some methodological flaws in the study and who expressed reservations about some of its findings--might be embarrassed were he asked to comment directly on these matters. Our failure to conduct interviews does not mean that we lacked the insights that might be provided by such interviews; all these kinds of feelings are expressed in the documents themselves.

Thus, we used all available public documents, as well as some documents that were not

publicly available: i.e., newspaper stories, conference proceedings, committee minutes and other accounts of committee meetings, the files of the Chief Scientist including correspondence, and various press releases and written statements. Each of these sources is described in more detail below.

1. **Newspaper stories:** To assure complete coverage of all news stories without wasting time and effort examining every issue of every newspaper over a period of years, we concentrated our attention on newspaper clippings from those periods when momentous events had occurred in relation to the Report: namely, from a month or so before publication to several months following publication, after which media reaction died down; the time of the Knesset debate; and those periods when committee conclusions or recommendations were announced.

2. **Conference proceedings:** A number of conferences dealt exclusively or primarily with the Van Leer Report. Most noteworthy were the prepublication meetings that were held to announce the upcoming report and to familiarize the research and education communities, as well as the general public, with some of its findings. Though no formal proceedings emerged from these meetings, they did produce documentation from which data could be gathered: e.g., published programs listing the speakers and the various topics covered. (Though one of the authors of this monograph was present at some of these prepublication meetings, we have based our

work solely on the printed documentation.) In addition, the annual seminar of the Pedagogical Council of the Teachers Union devoted six of its meetings to the Van Leer Study; the proceedings of these meetings are available in published form. Finally, in the case of the seminar sponsored by the School for Training Educational Senior Staff (STESS), we relied on notes taken by participants.

3. **Committee minutes and other accounts of committee meetings:** The Educational Committee of the Knesset gave brief consideration to the Van Leer Report; its minutes are available to the public but do not contain a great deal of information. This committee appointed a subcommittee to consider the Report more extensively, but the meetings of this subcommittee were not open to the public. Nonetheless, we got permission to examine these minutes, which proved to be lengthy and detailed, covering about six meetings and running approximately 70 pages. We also examined the minutes of the Ad Hoc Committee of the Ministry of Education, which are public documents. This committee provides policy advice to the head of education in government. The minutes of the meetings relevant to the Van Leer Report also ran about 60-70 pages each.

4. **Files of the Chief Scientist:** The Chief Scientist of the Ministry of Education and his staff have primary responsibility for providing the expertise and technical skills necessary to review funded studies, make recommendations, and commission new research. The

role is similar to, though more extensive than, that of the director of the National Institute of Education in the U.S. Documents from the Chief Scientist's files that proved to be especially valuable included copies of interim reports (which gave insight into how the Van Leer Study was shaped and changed over time), correspondence with the authors and others (some 40 or 50 letters to and from the Chief Scientist, which revealed points of view at various times), and press releases (which presented the official view).

5. **Other sources:** These included letters written to the Knesset committee and maintained in a file to which we were given access.

Finally, we examined all research publications in Israel that cited the Van Leer Report in order to get some idea of the Report's impact on subsequent research.

GENERAL FEATURES OF THE VAN LEER STUDY

The first step in planning what became known as the Van Leer Evaluation Study was taken in 1969, when a senior faculty member of the Hebrew University School of Education applied for a grant from the Van Leer Foundation to conduct a large-scale achievement survey of the Israeli primary schools. The requested grant was awarded in 1970, and the actual work of carrying out the study began. When in 1972 it became clear that the funds provided by the Van Leer Foundation were insufficient, the Ministry of Education allocated supplementary funds for the study. The final report (Minkovich et al., 1977), published simultaneously in Hebrew and in English, mentions the Coleman Report in the U.S. and the Plowden Report in Britain (both of which had attracted great interest in Israel) as the progenitors of this study of Israel's primary schools.

There was widespread national agreement about the need for a comprehensive survey of the primary school system. One of the most powerful arguments for the study was that in 1971 the annual SEKER examination had been abolished. This examination, an omnibus achievement test given in the last year of the primary school,

not only had served as a screening device for admission to secondary schools of various types but also had provided an annual summary of achievement in the school system. Secondary analysis of these data yielded valuable information about the achievement levels of various subgroups in the school population (Orthar, 1967; Smilansky and Yam, 1969). The abolition of this examination deprived the authorities and the public of a rich information source and created a demand for alternative information sources. Two achievement surveys were conducted in the mid-70s but were limited in scope: The IEA study (Lewy et al., 1978) dealt only with the subjects of reading, English as a foreign language, and civics; and a longitudinal study of educational achievement carried out in 1971-3 (Lewy and Chen, 1977) focused on only three grades and covered relatively few of the subjects taught in the schools. These surveys received little publicity, and their results were communicated mainly to the research community rather than to a broad audience.

These circumstances, then, set the stage for conducting a large-scale study, which from its inception aroused great interest in Israel. This chapter describes the general features of the Van Leer Study: its aims, audiences, target, framework, purpose and limitations.

AIMS OF THE STUDY

The principal aim of the Van Leer study was to provide a comprehensive picture of the Israeli primary school in three major areas:

1. resources (facilities, budget, and manpower) and their allocation to schools of various types;
2. student achievement in major subjects (reading, arithmetic, biblical studies, geography including basic concepts in social studies, and sciences); and
3. home and school variables associated with achievement level.

A minor aim was to provide separate statistics within each of these three areas for specified subgroups of the population, particularly ethnic and socioeconomic subgroups.

In the United States, a distinction is made between "assessment" (systematic information, in the form of a summary of achievement measures, about the output of an educational system) and "evaluation" (judgments about the merits or weaknesses of a particular entity such as a process or product). In Israel, no such distinction is made. Therefore, although the study was basically an "assessment," it is usually referred to as the Van Leer Evaluation Study. Moreover, despite its survey character, the study's results were meant to have evaluative overtones; that is, both the authors and the sponsors expected the descriptive summaries of achievement scores to allow inferential judgments about the Israeli educational system. Not only would these test scores yield comparative data about subgroups of the population but also they would provide the basis for absolute and definite judgments about the degree of congru-

ence between curricular objectives and actual attainments. The study was also regarded as an evaluation because it was intended to evaluate the success of continuous efforts to raise the achievement level of children of Asian-African origin (see Appendix A, Glossary of Terms) and to reduce the achievement gap between them and children of European-American origin.

In its evaluative aspects, the study specifically notes the success or failure of activities carried out in the past. But it was intended as more than a mere summary of past events. The study had a prospective as well as a retrospective slant: The authors believed that well-documented knowledge about past achievements and failures could be used to make changes that would eliminate flaws. Indeed, they themselves tried to recommend specific future actions that would lead to improvement.

Despite the authors' intentions, the recipients of the Report attached less significance to the recommendations than to the descriptive data summaries, thus lending support to an observation made frequently in the management literature: administrative authorities view reports produced by experts as only one of several legitimate input sources which should affect their decisionmaking. No matter what the scientific merit of some finding, they believe that it is their duty and privilege to formulate their own recommendations (Dror, 1979; Self, 1972).

AUDIENCE FOR THE STUDY

The final report was prepared in a form which fit the needs of the research community. Five hundred pages in length, it contains a detailed description of the study's methodology and is laden with statistical tables. Nevertheless, the authors viewed the report as having a multiple clientele and so made arrangements to meet the varying needs of each group of clients. The spokesperson of the Ministry and the Principal Investigator prepared a brief, and widely circulated, summary of the study. Other audiences were reached mainly by oral communication: after the Report was published, the authors gave a series of lectures to various audiences, including the Education Committee of the Knesset (Parliament), the staff of the Ministry of Education, the Pedagogical Council of the Teachers Union, the members of the Curriculum Center, the supervisory staff of the primary school system, and the Pedagogical Secretariat (the Ministry of Education's chief decisionmaking body for pedagogical matters whose members are the Chief Supervisors for school subjects and for regional districts, and whose framework includes the Committee for Primary Education and the Standing Committee for Secondary Education). The mass media--daily newspapers, popular weekly magazines, radio, television--were also briefed. The authors tailored both the content and the style of each presentation to the particular audience being addressed. It should also be noted that, while the study was being conducted,

brief reports on its progress were issued from time to time, thus creating an awareness of the importance of the study, arousing the expectations of a wide range of potential audiences, and setting the stage for the publication of the final report.

TARGET OF THE STUDY

By "target" is meant the particular phenomenon whose merits or shortcomings constitute the focus of an investigation. The target of an evaluation may be: a series of actions, the products used within the framework of an action, or the persons associated with a particular action. The results of an assessment imply praise or blame not for the learners whose achievement is being measured but rather for the system which is supposed to take care of their needs. The target may be the educational program operated by the system, those who operate the system, or both.

In program evaluation studies, evaluators frequently emphasize that the target of evaluation is the program itself and not the persons who create it or operate it. Though negative findings may disappoint those who believed in the merits of the program, they will not be blamed for the failure. In the world of scientific experimentation, failures are considered to be unavoidable co-occurrences of success.

This stance toward program evaluation reflects a principle commonly agreed upon by evaluators, but in practice, the taint of personal blame is often difficult to avoid. Program

developers are inclined to perceive any negative results as an accusation that they personally have failed, and this interpretation is particularly likely when assessment or evaluation studies are conducted within a hierarchically structured management system.

Attitudes toward a study and its results are largely a function of the level at which the study was initiated and at which the results are received. A study on one rung of the ladder is perceived by those on the rung directly below as an examination of competency. They feel they are to blame for any flaws detected in the system. Those who are several rungs below may be more neutral in their attitudes.[1] The initiator of the evaluation can scarcely be hurt by the results of the study. Signs of success will enhance his/her prestige, and signs of failure will reflect credit on him/her for being coura-

---

[1] Studies in organizational behavior devote great attention to overt coalitions formed within the framework of organizations of various types. Bacharach and Lawler (1981) discuss the tendencies of lower-level employees to form coalitions against upper-level management. Tacit agreements between working groups of different hierarchical ranks that do not result in implicitly formulated policy statements or demands are not discussed in basic textbooks of organizational theories. Nevertheless, one may assume that tacit agreements between nonadjacent hierarchical ranks of an organization may sometimes impose stress on those who occupy an intermediate position between them both on behalf of their superordinates and subordinates.

geous enough to call for an evaluation study with the aim of identifying flaws. Besides, the initiator is in a position to pass blame on to others.

The Van Leer Study was initiated at the highest executive level of the educational hierarchy-- the Ministry of Education-- and was carried out by an independent team from a university. As a result, many chief executive officers felt, not without reason, that they were the "target" of the evaluation--that the results of the study would constitute a kind of verdict upon their actions.

FRAMEWORK OF THE STUDY

The **target population** of the study comprised pupils in the first, second, fourth, and sixth grades of the Jewish State School System. Excluded populations were learners in institutes for the retarded, disturbed, and physically handicapped (about 3.5 percent of the age group) and children in the religious Independent School System (5 percent of the age group). A parallel evaluative study of Arab primary schools was conducted, and its major results were published later in a separate volume (Bashi, 1981).

A stratified random **sample** was drawn from the total population. In schools which had more than two classes at the same grade level, two classes were selected randomly. The total sample consisted of 98 schools, 614 classes, and 17,700 pupils.

Two types of **variables** were considered: (1) student characteristics and (2) school vari-

ables. The student characteristic variables included family background (e.g., parents' education, occupation, and country of origin; students' interaction with parents), personal characteristics (e.g., motivation and verbal abilities), and school achievement in five curricular areas (reading, arithmetic, biblical studies, geography, and science). The school variables included characteristics of teachers (training, experience), school programs (extracurricular activities, innovative programs, services available), and physical conditions (facilities and instructional materials).

Data were acquired from a variety of instruments. Information about the schools was obtained through a questionnaire answered by the principal. The teachers answered questions about their personal backgrounds and attitudes. The students completed a questionnaire about their family background. (In the first grade, a random sample of children from each class was interviewed.) They also took a test in reading and arithmetic. Fourth- and sixth-graders were administered an intelligence test and a personality, motivation, and attitude test. Half of the pupils in each class completed a test in Bible, geography, and science. The data collection was carried out in 1973.

STUDY PURPOSE AND LIMITATIONS

Studies of the input-product type have frequently been criticized for their inability to explain an observed success or failure. If a program does not achieve some anticipated out-

come, one can not be sure whether the problem lies in program design or in expectation. Researchers have expressed concern about the danger of evaluating the impact of "nonevents" in input-product studies, which are designed to examine the outcomes of a particular program. The difficulty increases when the study focuses on a complex cluster of programs, for in such instances, one program differs from another not only in method and strategy but also in goal priorities (Walker and Schaffarzik, 1976). These were basic problems faced by the research team for the Van Leer Study.

Since one purpose of the study was to evaluate the success of the Israeli primary schools in reducing achievement gaps over a long period of time, the problem of evaluating a "black box" (Fullen and Pomfret, 1977) also arises. Such a study may offer certain summative results, which provide information about the attainment of goals, but these cannot easily be translated into diagnostic hints and prognostic suggestions. Examining achievement profiles (to see in which domains achievement is relatively high and in which domains it is dangerously low) may result in some prognostic suggestions, but taking quick action based on such profiles could easily change the profile patterns without curing the illness. To give an example: achievement profiles from the Van Leer Study revealed that the learners had extremely poor map-reading skills. If more systematic map-reading instruction were instituted, the "cure" might reduce

achievement in domains where students had previously excelled. In the Van Leer Study, the "black box" phenonemon imposed limitations on the validity of suggestions based on data analysis, and supported the claim that the validity of such suggestions could only be established through further inquiries by means of further surveys, experimentation, or committee work. Indeed, in many cases the same findings were cited by different groups to support conflicting prognoses.

FINDINGS OF THE VAN LEER STUDY

This chapter not only summarizes the findings presented in the Van Leer Report itself but also discusses two other documents which pertain to those findings: a summary prepared by the principal investigator, and a critique prepared by the Chief Scientist of the Ministry of Education. Finally, the chapter concludes with an account of the events that preceded the release of the report.

THE STUDY FINDINGS

The Van Leer Report presents data based on summaries of the responses of students, parents, teachers, and school principals to hundreds of questions and test items, as well as data extracted from official school records filed at the Ministry of Education. Almost 500 pages long, it contains 11 chapter summaries (see Appendix C), along with hundreds of tables and graphs. Even a single chapter can be overwhelming: For instance, the chapter on mathematical achievement comprises 38 single-spaced pages, including five figures and six tables. Moreover, as is the case with most scientific reports, data summaries are presented sequentially, without regard to the relative salience of particu-

lar issues, and no boldface headlines are used. In short, the data overload is substantial, and it is easy to see how readers of the Report may have had difficulty in identifying the major issues.

Two navigational aids were available to help readers get through the dense forest of data summaries. The first was a brief document, prepared by the Principal Investigator, which summarized the major findings and provided concise interpretations of them[1]. This summary served as a guide to, and often as a substitute for, the huge research report itself. Indeed, in some circles, it came to be identified as the findings of the study. In addition, the Principal Investigator gave a more complete exposition of the data by means of oral presentations to various groups. The written digest and the oral presentations constituted one navigational aid.

Second, some readers were guided by their own conceptions of the most striking problems faced by the system and assumed that the Report contained answers to their questions. Those

---

[1] A second summary, prepared by an official spokesperson for the Ministry of Education, differed somewhat in its interpretations, complicating public reception of the Report. See Appendix B for an example of differences in interpretation between the spokesperson's and the Principal Investigator's summaries. We have chosen not to deal with the spokesperson's summary, since its optimistic view did not have a great impact on the events following the release of the study results.

familiar with the Israeli educational system could weigh appropriately the different parts of the Report, draw from it those conclusions most relevant to their own orientations, and raise questions publicly. Others with a more superficial knowledge of the educational system turned their attention to those issues which made newspaper headlines.

Thus was generated a list of major issues which subsequently constituted the focus of discussions and debates about the study. Neither of these navigational aids could, however, ensure an "objective" reading of the Report. From the outset, there was contention over just what the major findings were; consequently, there is ambiguity about the extent to which these findings had impact. The matter of which findings had what impact is addressed in Chapter 4. Here, it will suffice to describe the format of the Report itself, to present the main outlines of the Principal Investigator's summary, and to review the Chief Scientist's critique.

FORMAT OF THE REPORT

The Van Leer Report comprises 18 chapters. Chapters 1-4 are introductory, describing the background of the study, its objectives, the research design, and the sampling procedures. The next four chapters give general information. Chapter 5 analyzes family characteristics and background information (e.g., home conditions relevant to the intellectual development of the child). Chapter 6 deals with school facilities, programs, and resource allocations, as well as

with the correlation between school resource variables and achievement. Chapter 7 summarizes results on the backgrounds and attitudes of teachers and principals; it includes 19 tables providing univariate information on 130 variables related to teachers' views on a variety of school phenomena (e.g., teaching practices, effectiveness of contact with parents, characteristics of disadvantaged learners, expectations concerning the achievement of disadvantaged learners, satisfaction with work). No less loaded with data is Chapter 8, on the personality characteristics of learners; it discusses some 62 variables, summarizing the results of a locus of control test and of a general self-image and scholastic self-image scale and covering such topics as the relationship between effort and learning success; the perceived attitudes of parents, teachers, and peers toward the pupil; interest in studies; level of aspiration; importance attached to interest; attitudes toward teacher, school, and class; attitudes toward behavior in school; and conceptions of the ideal pupil.

The central section of the Report (Chapters 9-14) deals with student performance on various tests (six general ability tests, plus tests on achievement in reading, mathematics, biblical studies, geography, and science). Chapter 9 contains information about the development of the tests and about test results. Three scores were derived for each test: (1) a total test score, (2) a minimum knowledge score (the score

on a subset of ten items intuitively judged to represent a "minimal curricular demand"), and (3) a score on a set of anchor items which were common to two consecutive grade levels (i.e, grades 1 and 2; grades 2 and 4; and grades 4 and 6). Chapter 10 discusses differences between various student populations on these ability and achievement tests. The student groups are classified along two dimensions: (1) origin by generation (seven such categories were defined: e.g., both parents second-generation Israelis of Asian-African origin; Father of Asian-African origin and mother of European-American origin); and (2) parents' educational level (five such categories were defined).

Chapters 11-14 discuss the results obtained in each subject area (language, mathematics, biblical studies, and geography and science), by grade level and by population group. As an example, the data for the mathematics test are presented in Table 1. In addition to the information on means and standard deviations, information was provided on the proportion of students making scores of at least 60 percent correct responses (considered by the Principal Investigator to be a "pass" grade). The progress made by various subgroups on the set of anchor items from one grade level of the sample to the next was also detailed.

Chapters 15-17 use the achievement data as a basis for examining interrelations between achievement and the other criterion variables identified as relevant factors for purposes of

Table 1

Percentage of Students with at Least 60 Percent Correct Answers on the Mathematics Tests

	Total Pop.	Cultural Origin by Generation							Parents' Educational Level in Years				
		AA1	EA1	AA2	EA2	GEN3+	FaAA	FaEA	0	1-4	5-8	9-12	13+
Grade 1	42	30	56	34	76	57	55	58	15	24	28	49	68
Grade 2	56	44	74	59	83	73	69	72	26	39	43	65	83
Grade 4	59	47	75	64	88	81	57	63	32	45	48	70	82
Grade 6	42	36	66	53	80	65	56	61	21	27	30	49	74

AA1 - Asian-African, first generation
AA2 - Asian-African, second generation
EA1 - European-American, first generation
EA2 - European-American, second generation
GEN3+ - Third-Generation or greater
FaAA - Asian-African on paternal side; European-American on maternal side
FaEA - European-American on paternal side; Asian-African on maternal side

the study. Chapter 15 compares levels of achievement in various types of elementary schools (religious and nonreligious; schools with a majority of disadvantaged or advantaged pupils; integrated schools; the Arab and the Jewish school systems). Chapter 16 presents a hierarchical model for partitioning the variation in reported achievement test scores. Chapter 17 pulls together all the data bearing on the impact of integration.

The final chapter, Chapter 18, presents a summary and conclusions.

PRINCIPAL INVESTIGATOR'S SUMMARY

In his summary, which took the form of a "handout" for use at lectures and discussions, the Principal Investigator clustered the findings in three major groups: (1) equality of resource allocation, (2) achievement in various subjects, and (3) the achievement gap between the Asian-African and European-American groups.[2]

Equality of Resource Allocation.

The study revealed that both home and school variables had high zero-order correlations with student achievement. Among the school variables significantly correlated with achievement were the seniority of the teachers, the educational level of the principal, teachers' attitudes

[2] This discussion follows the structure of the handout, with details added from a published transcript of six lectures delivered by the Principal Investigator shortly after the release of the Report (Teachers' Union, 1978).

toward disadvantaged children, regularity of student attendance, and time scheduled for teaching a certain subject. When the home background variables were statistically controlled, however, the correlations between school variables and achievement dropped to non-significance; this was true for all school variables except time scheduled to teach a particular subject and teachers' attitudes. In other words, conditions are worse in schools for disadvantaged learners than in schools for advantaged ones. This finding came as a surprise, since the Ministry of Education--in an effort to reduce the previously demonstrated achievement gap between the Asian-African and European-American groups--had previously initiated various support programs and granted special privileges to schools with a large concentration of disadvantaged students. It was, therefore, expected that these schools would no longer be inferior to schools with a high concentration of advantaged learners. It became apparent that "advantaged" schools supplement their budget with parental donations and are more alert in taking advantage of available funds, whereas lack of adequate staff at "disadvantaged" schools prevents them from benefitting fully from the program grants allocated to serve their needs.

Achievement in Various Subjects.

The Principal Investigator interpreted the results of the achievement tests not only in terms of individual achievement but also in terms of the achievement of the whole system. In compar-

ing achievement at various grade levels and in various subjects, he viewed the achievement tests as being criterion-referenced and having equal content validity for each grade level in each subject field. Following the tradition of teachers in assigning school grades, he considered an individual achievement score of 70 percent correct responses as "good" and of 60 percent correct responses as "satisfactory". Correct response percentages below 60 percent were considered "unsatisfactory". The Principal Investigator decided that if at least 50 percent of a defined subpopulation got "good" grades and another 25 percent got "satisfactory" grades, then it could be said that the level attained by that particular population was satisfactory. While admitting that this criterion was arbitrary, the Principal Investigator defended it by saying that the standard was commonly accepted among teachers in schools and universities. He also indicated the possibility of utilizing other scores but claimed that the particular pattern described above constituted a moderate standard and a realistic challenge for an educational system.

The same criteria were used to evaluate the achievement of the whole system. In these terms, the system's "grades" were: reading, good; mathematics and biblical studies, satisfactory; geography and science, unsatisfactory. Such overall scores were computed for subgroups too. Thus, for example, it was concluded that, in grade 4, only the first- and second-genera-

tion European-American groups (EA1, EA2) and the group of third-generation or greater students from any cultural background (GEN3+) (constituting respectively 12 percent, 16 percent, and 8 percent of the total fourth grade population) reached the "normative" level in the mathematics test (see Table 1).

Most teachers accepted this approach. The content validity of the test items was demonstrated by relating test items to curricular content specified in the national syllabus. For the most part, teachers were unaware that only one aspect of content validity had been demonstrated: i.e., that the items dealt with prescribed curricular content. They tended to ignore another aspect of content validity: whether the items constituted an appropriate sample of the total curricular content. However, the simplicity of this approach to establishing content validity had great appeal to the mass audience, as well as to educational policymakers.[3]

The Principal Investigator defended his assumption by referring to the universal prac-

---

[3] The major criticism of this approach, coming from experts in testing and measurement, focused on the assumption that a fixed proportion of correct responses in a variety of tests represents the same level of success. Experts in psychometrics claimed that such an assumption could be accepted only if careful documentation were presented about all aspects of content validity, including the complexity levels of the items.

tice of assigning grades to students, and even of calculating grade-point averages from grades in various subjects.

Achievement gap between Asian-African and European-American groups.

The third focus of the Principal Investigator's summary was the question of whether the Israeli educational system had reduced the achievement gap between children of Asian-African and of European- American origin. The learner's origin was defined as the country of birth of his/her parents and grandparents. Thus, one can talk about children of Asian-African origin whose parents were born in Israel, and children of Asian-African origin whose parents immigrated to Israel. Within these larger groupings, separate statistics were computed for children both of whose parents were of Asian- African origin, one of whose parents was of Asian-African origin, and both of whose parents were of European-American origin. Within each group, further differentiations were made between children whose parents were born in Israel and those whose parents immigrated to Israel.

The Principal Investigator's summary highlights three study findings. First, there is still a considerable gap in achievement (approximately .80 of a standard deviation) between the Asian-African and the European- American groups. Second, the achievement of children with one parent of Asian-African origin falls between the achievement level of the two groups. Third, the achievement of third-generation Israelis (those

whose parents were born in Israel) is higher than the achievement of second-generation Israelis (those whose parents immigrated to Israel), regardless of country of origin. The Principal Investigator concludes that the educational system contributes to the achievement of all children, even though it does not reduce the gap between Asian-Africans and European-Americans; only intermarriage succeeds in reducing the gap.

THE CHIEF SCIENTIST'S CRITIQUE

The Ministry of Education requires that all the studies it finances be reviewed and evaluated by the Chief Scientist. Such a review serves two purposes: first, to determine whether the researcher fulfilled all obligations as specified in the research contract; and, second, to allow the Ministry to criticize those statements and conclusions of the researcher that do not seem to have a satisfactory level of scientific validity.

In most Israeli ministries, the Chief Scientist is a person from the academic community who is employed part time. The Chief Scientist's function is to create a link between policymakers and research by calling the attention of the policymakers to relevant research findings and by initiating research studies which may provide useful input for decisionmaking. The Chief Scientist plays a central role in distributing research grants and receives research reports for projects which the Ministry finances or supports (Kugelmass, 1981).

The Chief Scientist's critique of the Van Leer Study was released along with the full Report. The Chief Scientist made highly appreciative comments about the study, describing it as one of the most important to be carried out in Israel. He praised the systematic work invested in developing the instruments, the adequacy of the sampling procedure, and the meticulous data collection. He emphasized the implications of the study for policy decisions in the future. Nevertheless, he criticized several methodological procedures and expressed doubts about the validity of some findings. His criticism touched on two main issues: the conclusions derived from differences in achievement between generations of immigrants, and the validity of comparing achievement in various subjects on the basis of the percentage of correct responses to test items. On the first issue, the Chief Scientist claimed that the higher achievement level of the third-generation immigrant children, compared with second-generation immigrant children, does not necessarily prove the effectiveness of the Israeli educational system, since there could be alternative explanations for this difference. For instance, it may well be that immigrants from the same country arriving in Israel at different times differ from each other in their educational levels.

His most serious reservations concerned the legitimacy of making comparisons across subject areas (i.e., stating that Israeli students had

performed better or worse in one subject than in another). This issue evoked highly emotional responses from various groups of readers, creating controversy between professional evaluators (who disagreed with the Principal Investigator on methodological grounds) and educators (who tended to accept the Principal Investigator's approach because of its simplicity).

Even though the Chief Scientist's comments were formulated in terms of methodological requirements, the publication of his critique raised fears of political exploitation. In his reaction to the critique, the Principal Investigator emphasized that any comment which called into question the scientific validity of the study might be used by the Ministry of Education as an excuse for disregarding the study's recommendations. To defend his approach, the Principal Investigator marshalled both procedural and substantive arguments. He said, for instance, that at the planning stage of the data analysis, a group of experts--among them a representative of the Ministry of Education--had agreed to his working plan, as had his co-authors. He claimed that actual school grades are based on scoring procedures similar to those employed in the study. In addition, he published examples of items, and statistics about the proportions of students who had answered them incorrectly, in order to demonstrate the low level of knowledge in subjects where his summary implied unsatisfactory results. He claimed that the Chief Scientist's criticisms should have been handled

as an academic comment and not given such great publicity, which could turn out to be counter-productive by weakening the impact of the study.

The debate between the Principal Investigator and the Chief Scientist at times took on a personal tone. This aspect of the controversy received exaggerated attention from the mass media stirring emotions even in the Knesset debates.

ANTECEDENTS OF THE REPORT'S RELEASE

The Report was published in 1977, eight years after the proposal for funding was submitted and four years after the data collection was completed. Many researchers who had been involved as consultants or as participants in some preparatory activity eagerly awaited its publication. The heads of departments within the Ministry were anxious to know what it would say about issues related to their departments. Shortly before the Report was released, the Jerusalem Van Leer Foundation, in cooperation with the Hebrew University of Jerusalem, conducted two international seminars devoted to general problems in educating the disadvantaged learner. One of them was a two-day lecture program open to the public. Israeli and foreign experts spoke on the achievement of disadvantaged groups of Asian-African students. American lecturers included James Coleman, Ralph Tyler, Ernest Hilgard, and Lawrence Cremin. Several department heads from the Ministry of Education also delivered lectures. Though the program dealt

only with general problems, the findings of the Van Leer Study were hinted at in the discussions.

The second seminar, led by Coleman, was restricted to an invited group of researchers from Israeli universities. Since various parts of the Report were presented to the participants, some Israeli researchers were able to acquaint themselves with portions of the study findings and to comment upon them.

These seminars were designed primarily to allow Israeli scholars and educators to discuss the theoretical and practical issues involved in educating the disadvantaged and to share with other countries the Israeli experience in dealing with disadvantaged populations. Though not conceived as discussions of the Van Leer Study, they nonetheless had some direct bearing on it. First, the seminars were "time fillers," reminding the audience that the Van Leer Report would soon be released. Second, they raised curiosity and expectations about the findings of the study. Finally, since they entailed prepublication release of some of the findings to selected Israeli experts, they allowed the authors of the Report to gauge the extent of consensus about its conclusions and implications and to prepare themselves for reactions to its actual publication.

Once in print, the Report still had to leap two hurdles, one political and the other procedural. The first hurdle was the parliamentary elections which took place in March 1977. It

was decided that the Report should be published after the elections in order to prevent its findings from being exploited in the political campaign. Second, as has been pointed out, the Ministry of Education's regulations require that studies supported by the Ministry should first be submitted to the Ministry and examined by its Chief Scientist, who would then write a critique, to be published simultaneously with the Report. The Chief Scientist needed two months to review the voluminous Report. This last delay in publication created some tension, since the public knew the Report had been presented to the Ministry but did not understand why the results were not released immediately. The newspapers obtained some partial results of the study and started to publish those which they judged to be "sensational", accusing the Ministry of intending to "pigeonhole" the Report and to conceal its findings from the public. There were inquiries on behalf of members of the Knesset about the fate of the study, and when the Report was finally released for publication in July of 1977, its message was not entirely unknown to various audiences. Nevertheless--or, perhaps, exactly for that reason--it became a bestseller, and the 1,000 copies of the first release disappeared from the market immediately.[4] (As a matter of fact, it was not really

[4] This number of copies in a country with a total population of several million was quite large; to be comparable, 50,000-75,000 copies of a research report would have to be distributed in the United States.

put on sale; it was distributed to researchers, to educational, social, and political leaders, and to institutions upon request.)

THE IMPACT OF THE VAN LEER STUDY

From conception to follow-up studies, the Van Leer Study represents a long and complex process (see Table 2). In this chapter, we will focus on the impact of the study, tracing the chronology of events after the release of the Report.

EARLY REACTIONS

The results of the Van Leer Study elicited strong emotional reactions and considerable frustration, even though they were consistent wtih previously published achievement surveys (Lewy and Chen, 1977; Orthar, 1967). People had hoped that the study would hold "surprises," would show that the educational system was at least partially successful. Instead, it proved again what was already known, thus intensifying the frustration. Reactions came from numerous groups and institutions. For at least two weeks, most newspapers headlined various aspects of the study. The Knesset (Parliament) devoted a plenary session to discussing its implications; the Teachers Union decided to conduct a seminar on the findings. These first reactions focused only on those educational problems which had been of concern in earlier years. People seldom asked what they could learn from the

Table 2

Partial Chronology of the Van Leer Study and Its Impact

1969	Conception of the study
1970	Award of grant by the Van Leer Foundation
	Initiation of the study
1972	Award of supplementary funds by the Ministry of Education
1973	Data Collection
1977	Prerelease seminars
	Chief Scientist's critique
	Release of Van Leer Report
	Principal Investigator's Summary
1977	Initial Knesset debate
	Seminar of the Pedagogical Council of the Teachers Union
1977-79	Work on recommendations by the Educational Committee of the Knesset and the Ad Hoc Committee of the Ministry of Education (and their subcomittees)
1978-81	Replication studies and supplementary analyses
1980 (January)	Knesset recommendations
(May)	Ministry of Education's Ad Hoc Committee's recommendations
(July)	Minister of Education's report to the Knesset on steps taken to implement recommendations
1983 on	Continuing impact

study or questioned which findings were valid and which were not. Rather, each individual and each constituency posed questions related to longstanding issues of personal interest, hoping that the answers provided by this national large-scale educational evaluation would satisfy their particular needs.

Problems arose when people tried to use the findings to support their own particular views on educational issues, as was the case in the mass media's coverage, the Knesset debate, and even the Seminar of the Pedagogical Council of the Teachers Union. Over the protests of the Report's authors, some of the findings were brought to bear on issues to which they really had no relevance. At the same time, many of the study findings evoked no reaction whatsoever, either from the public or from leading educators.

Mass Media Coverage.

Because the mass media--newspapers, radio, television--treated the Report as national rather than educational news, it rapidly came to the attention of the entire nation. Typically, mass media coverage was sensational rather than analytic. For instance, despite the importance of the study, and the headlines devoted to it, no newspaper or professional journal saw fit to publish a scientific review of the Report, nor apparently did any scholar or researcher see fit to offer such a review for publication. Rather, the reaction was mainly political. Even the critique of the Chief Scientist was treated by

the press chiefly as an instance of disagreement within the academic community; its content and the validity of its arguments were ignored.

Two examples will serve to illustrate how press coverage distorted the study findings and thus aroused heated debate over sensitive issues. The first concerns the achievement gap between the religious and the nonreligious school systems. The study found that, in terms of raw scores, the religious schools performed worse in all subject fields, including biblical studies. Some newspapers presented this finding as evidence of the inferiority of the religious schools, even though anyone with a basic knowledge of research methodology realizes that such an inference is unjustified, since differences in raw scores on the output measures may reflect differences in the entry-level abilities of the children. Indeed, as is well known, the religious schools in Israel enroll a much larger proportion of disadvantaged children (75 percent) than do the nonreligious schools (33 percent). Moreover, the results of a covariate analysis reported in the study showed that, when socioeconomic background was taken into account, the output of the two school systems did not differ significantly. In fact, the religious schools did slightly better than the nonreligious schools with respect to biblical studies, prabably because they devoted more time to the subject. Nevertheless, the newspapers ignored the results of the covariate analysis and headlined instead the raw score differences, over

the strong protest of the authors. Commenting on the situation in an interview, the Principal Investigator claimed that the issue was intentionally falsified for the political benefit of a certain group campaigning against the religious schools.

Similarly, the media chose to interpret descriptive data on the achievement of disadvantaged children in integrated schools (where from 40 to 60 percent of the learners were disadvantaged) and in nonintegrated schools as evidence against the merits of integration. As the Report clearly states, and as the Principal Investigator later emphasized, only experimental studies can provide definitive evidence on the question of the merits or shortcomings of integration, but the headlines given to this issue by the newspapers gave the impression that the question had been settled and that integration had a negative effect on student achievement.

It would be unjust not to mention here that some newspapers made serious efforts to grasp the deeper meaning of the study findings and to offer constructive criticisms about the operation of the system. Our purpose in mentioning these two examples of distortion is to demonstrate that the authors' views do not always determine the interpretations imposed on research findings. Often the audience imposes its own preconceptions and thus arrives at an interpretation which is not only contrary to the interpretation of the authors but also unsupported by the weight of the evidence.

The Debate in the Knesset (Parliament).

On June 27 and 29, 1977, about a month after the publication of the Report, a plenary session of the Knesset was devoted to a debate on the implications of the study.[1] Three members of the Knesset insisted on this debate, two of them from the opposition party, and one from the party of the parliamentary majority (but not from the party of the Minister of Education; in a coalition government, a small political party may hold important ministerial portfolios). To understand the climate of the debate, one needs some background information on the parliamentary scene at that time. The election for the Knesset, held in March 1977, resulted in a victory for the National Unity (Likud) Party and the consequent installation of a new coalition government. A representative of the National Religious Party was named Minister of Education, a position which for the previous 30 years had been held by members of the Labour Party. The Van Leer Study dealt with the achievement of the educational system during the period when this system was headed by a member of the Labour Party; thus, the new Minister had to respond to questions related to the responsibility of his predecessor. It seems logical that the new minister should not be blamed for the failures of the previous Ministry of Education.

---

[1] The minutes of these sessions appeared in the Knesset Proceedings, 1977, Issue 3, pp. 135-43.

Complicating the situation even further, the state of Israel maintains both a religious and a nonreligious system of education. The religious school system, a subsystem of the state schools, is nondenominational and enjoys a high level of autonomy in determining the school curriculum, maintaining its own teacher training institutes, hiring faculty, and so forth. Its leadership has always been closely associated with the National Religious Party, the party of the new Minister, who had in fact been promoted from a leadership role in the religious school system. Thus, the alleged failure of the religious school system (no matter how unfair the allegation) could prove embarrassing to him.

At the beginning of the debate, the new Minister remarked that education is a major national problem of concern to all Knesset members and that it should not be turned into a political football. Despite his hope, political issues were not fully avoided. Two members of the left wing of the Labour Party, whose interest in education had been demonstrated by their previous parliamentary actions, expressed their fear that the new Minister would require that the schools increase the time devoted to religious studies by reducing the time allocated to secular studies, adducing purported evidence from the Van Leer Study that such a change was undesirable. One speaker took the opportunity to criticize Israeli educational research in general. Pointing to some aspects of the Van Leer Report, he claimed that such research,

though paid for by Israeli taxpayers, was strongly academic in orientation and was addressed to the community of researchers--especially researchers in the United States--rather than to educators or the general public.

The Knesset debate relied heavily on the comparative data about various grade levels and subject fields (even though these data had been severely criticized by the Chief Scientist), probably because these findings were easy for nonresearchers to comprehend. Even the findings on the achievement gap between the Asian-African and the European-American groups were discussed in terms of the percentage of learners passing the arbitrarily set "score" for satisfactory results, rather than in terms of standard deviations. Though this emphasis--in the Knesset debate and in later discussions--lends support to the contention that research findings are often presented in such a way as to be inaccessible to the lay reader, it also suggests that whenever readers do not fully understand the presentation of the findings, they will substitute their own simplified version. In other words, scientific jargon invites erroneous conceptions.

Seminar of the Pedagogical Council of the Teachers Union.

Early in 1977, prior to publication of the Report, the Pedagogical Council of the Teachers Union decided to focus its annual seminar (which usually comprises 10-12 half-day sessions) on

the Van Leer Study, inviting presentations by the research team and (for the sake of balance) by the Chief Scientist. One lecture was to be devoted to the Israeli results of another large-scale achievement survey conducted by the International Association for the Evaluation of Educational Achievement (IEA) (Lewy et al., 1978; Walker, 1976).

To understand the context of this seminar, one should realize that the Teachers Union in Israel concerns itself not just with professional matters but with all pedagogical aspects of school life, organizing inservice training courses for teachers and taking positions on various issues related to the functioning of the educational system. For four decades, it has played an influential role in education. For instance, prior to the formation of a State of Israel in 1948, it helped to create a network of Hebrew Schools whose bias reflected the Israeli and the International Labor Movement ideologies. Not until 1951, was this network fully absorbed into the State School System. The leaders of the Teachers Union have continued to maintain an interest in educational matters; and its Secretary has been a member of the Knesset's Educational Committee for more than a decade.

At six of the 11 sessions of the seminar (which was attended by 11 members of the Pedagogical Council and 33 invited participants), the research team talked about the major findings of the study. Indeed, the published proceedings of this seminar (Teachers Union, 1979)

came to be regarded as a valuable source of information on the research team's conception of the study. At each of these sessions, the Principal Investigator gave an introductory lecture and then opened the meeting to questions from the audience. These questions tended to center on those policy implications of the study which were of personal interest to the participants: e.g., grade repetition, ability grouping, extension of the school year. Even those few questions that dealt with issues directly raised in the study tended to concern themselves with policy issues rather than with substantive findings. For example, one participant claimed that any comparison of the achievement of different ethnic groups should be condemned as implying the inferiority of certain ethnic groups.

Although the seminar was academic in nature, it resulted in the creation of an ad hoc committee which formulated 17 recommendations covering such topics as the following: defining minimum competency requirements in each subject; equalizing resource allocation to schools of various types; strengthening the contact between school and community; and focusing teacher training on problems related to the education of disadvantaged children.

This set of recommendations was probably the first to be formulated by an external group reacting to the study findings. It was followed by sets of recommendations from other groups such as the Knesset Educational Committee and

the Standing Committee for Primary Education of the Ministry of Education.

The STESS Seminar.

The School for Training Educational Senior Staff (STESS) is a government-operated center for preparing school supervisors, principals, inspectors, etc., that is independent of the Ministry of Education and is housed in a resort hotel. At the instigation of one of the co-authors of the Van Leer Study, STESS decided to hold a seminar designed to create a greater sensitivity toward the findings of the study and to explore the data more fully. The seminar director (Report co-author) had not approved of the methodology of utilizing percentage scores to compare achievement in various subjects and at various grade levels, arguing that item analysis rather than whole-test scores might produce more valid results. Since the committees appointed later adopted a similar methodology, the substance of the STESS seminar merits closer attention.

The seminar participants examined the frequency distribution of item averages across classes for the advantaged and the disadvantaged populations. Other studies had used aggregated item data as a basis of analysis (Araisian and Madaus, 1978; Lewy, 1973), but ths approach was applied in a novel way to the Van Leer Study findings. The participants focused mainly on very easy items, hoping to discover why a large proportion of the disadvantaged population had answered them incorrectly. Tables were compiled to indicate the percentage of the classes in the

advantaged and disadvantaged groups in which all students had answered a given item correctly, (i.e., the facility level of the item was 1.00). To illustrate, the following data relate to three relatively easy items from the fourth-grade reading test:

Percentage of classes in which all students answered a particular item correctly

	Schools for Advantaged Students	Schools for Disadvantaged Students
Item No. 17	100	71
Item No. 27	100	54
Item No. 29	100	29

In all schools for advantaged learners, all the students marked the correct answer. In the schools for disadvantaged learners, the facility level per class on these items differed considerably. The seminar participants attempted to explain these differences in terms of the linguistic features of the items, the similarity of item distractors, the familiarity of learners with the situation described, and so forth, though to some extent the psychometric principle of "ceiling effect" may explain the differences.

Item analysis of this kind was performed also by the committees appointed by the Ministry of Education. To some extent, it is simply a technique for familiarizing people with the

results of the study. At the same time, it represents an attempt to translate the findings into recommendations for action.

THE UNAVOIDABLE SOLUTION: MORATORIUM

The words had all been said. The news media, the Knesset, and various professional organizations had all dealt with the study, and apparently there was nothing further to discuss. While the importance of the Van Leer Study was generally acknowledged, people did not know exactly what to do with the results. The summary of the study had included suggestions for action, but the executive branch of the educational system was not fully convinced of their validity. People were tired of speeches but still not ready to act. Among the choir of voices commenting on the study, the voice of the Ministry of Education was not heard very loudly; there was no clamoring for action. This may be understandable. The very complexity of the study, the wealth of data it contained, constituted a good argument for refraining from immediate action. The leaders of the Ministry claimed they needed more time to examine the findings and to derive from them implications for action. They paid tribute to the information contained in the Report, but they preferred to formulate their own action policy rather than merely to implement the recommendations included in the study.

Their solution was to form a committee. Committees are not only necessary for implementing change but also valuable for establishing a

moratorium. This moratorium served two purposes. First, some people suspected that the situation had changed since the data were collected in 1973, and the moratorium gave the committee an opportunity to check on the validity of the findings in 1978. Second, a system cannot implement the recommendations of an external agency without first consulting its own leaders. A moratorium gave the executive leadership the freedom to modify the study recommendations on the basis of their own accumulated experience in the field and their own judgments about the feasibility of a given action.

The Knesset, too, had reached a level of fatigue. After a single public session of debate, it agreed to transfer the issue to its Educational Committee for further treatment. After some three or four weeks of considering the Report, this Committee in turn experienced fatigue and so formed a subcommittee to continue consideration and to formulate recommendations for action.

Thus, after the first upheaval of public reaction, a period of quiet committee work followed. Both committees conducted follow-up studies to examine the changes in the system over the four or five years following data collection, hoping to show that the system is characterized by continuous improvement and by gradual elimination of the flaws pointed out in the Van Leer Report. While both committees worked intensively and continuously, they were not highly visible, so the Van Leer Study

dropped from public view. Not until two years later, when the Committees completed their work, did it again come to the attention of a broad audience.

Knesset Educational Committee.

The Educational Committee of the Knesset first convened on July 20, 1977 and over a period of three weeks, devoted four sessions to discussing the Van Leer Study results. The Chairman of the Committee, who had been the Minister of Education when the study was commissioned and carried out, opened the first session with the following words: "This is a prominent study, a milestone in the history of educational research in this country. It summarizes results of efforts invested in compensatory education during the past 20 years, and it contains guidelines for action in the future."

Experts from various academic fields, department heads of the Ministry of Education and the Municipalities, and the study's Principal Investigator and one of its co-authors were invited to testify before the Committee. Most Committee members participated in these sessions, which provided an opportunity to discuss various aspects of the educational system and to extend the parliamentary debate to more people. The proceedings of the Committee (unlike the proceedings of plenary sessions) were not open to the public. Only Committee decisions brought for approval to the plenary sessions shed light on the work done in Knesset committees. After four sessions, six Knesset members were appoint-

ed to a subcommittee that was created to formulate "suggestions for resolutions."

The subcommittee worked for 27 months, holding eight formal sesions and making numerous site visits to schools for disadvantaged learners. At its first session on August 10, 1977, it decided to invite external comments and recommendations on how to reduce the achievement gap between various groups within the school population. Accordingly, letters were mailed to faculty members at schools of education and teacher training institutes asking them to formulate recommendations for action based on their study of the Van Leer Report findings.

Response to this invitation was disappointing. Only a single team from one of the schools of education submitted a written reaction, though numerous reminders were sent out. There are several possible explanations for this lack of response: the subcommittee's failure to set a definite deadline for submitting reactions; a conviction on the part of many leading researchers that since they had already testified personally before the subcommittee, they need not submit written statements. Indeed, representatives of the major schools of education had appeared before the subcommittee, but records of their presentations reveal that, for the most part, they simply repeated views and recommendations already familiar from their previous writings, occasionally citing study findings that supported their recommendations.

The general public was notified through the newspapers about the work of the subcommittee, but only about eight or ten people approached the subcommittee and indicated a willingness to testify. These respondents can be divided into three groups: representatives of private institutes, producers of instructional materials, and teachers. Several representatives from private institutes in the field of non-formal education offered suggestions for reducing achievement gaps through, for example, private initiatives in sports, tutoring, and social interaction workshops to increase productivity among school principals. Authors of textbooks and instructional materials of other types pointed to the success of their own approaches in presenting recommendations. Teachers who had introduced innovations in their schools approached the subcommittee seeking wider diffusion for their innovation or offering personal services for disseminating their innovative approach. These teachers can in turn be divided into two groups: those who got positive feedback from their schools and were thus encouraged to seek greater publicity, and those who were unable to implement their ideas for lack of resources and who hoped that the subcommittee would help them to get the needed resources.

One of the most conspicuous activities of the subcomittee was the organization of a one-day workshop on "Achievement in the Primary Schools in Israel," conducted at the Parliament building on July 2, 1979. About 120 persons

attended the workshop, which was organized on behalf of the Parliamentary Committee for Education. Among the participants of the workshop were the Minister of Education and other top administrators, Knesset members, researchers and university professors, and invited teachers and principals.[2]

On December 19, 1979, the subcommittee brought before the plenary session of the Education Committee 12 "suggestions for resolution," clustered into five groups. Examples of suggestions from each group are as follows:

1. **Resources:** (a) Improve the housing conditions of the disadvantaged group of the population. (This suggesion was based on the assumption that improved housing conditions would indirectly affect the educational achievement of the children.) (b) Introduce compulsory kindergarten for children age 3-4 in depressed areas.
2. **Curriculum:** Ensure that, by at least the second grade, all children acquire basic skills in reading, writing, and arithmetic; schools should maintain continuous follow-up on the educational progress of each child.

---

[2] The workshop was chaired by its newly elected chairperson, Ora Namir, M.P. The proceedings were published by the Parliamentary Committee (1979).

3. **The learner and his family:** Expand the scope of adult education (since 13 percent of school children have parents who never completed elementary school).
4. **Teachers:** Create a salary system which grants special benefits to teachers in schools for the disadvantaged.
5. **Principals:** Establish an institute for training principals for schools for the disadvantaged.

Ad Hoc Committee of the Ministry.

The Director General of the Ministry of Education nominated a committee (for convenience, referred to here as the "Ad Hoc Committee of the Ministry"), which was charged with responsibility for reviewing the implications of the Van Leer Study and formulating recommendations of ways to eliminate flaws from the educational system. The committee consisted of four members: the Chief Scientist (who was in charge of translating into operational terms the result of all studies financed by the Ministry), the Director of the Curriculum Center, the Chairman of the Standing Committee for Primary Education, and the Head of the Religious School Network (all three of whom were directly responsible for monitoring the primary school system and thus for implementing the recommendations). Since this same committee had also served as a monitoring committee for the Van Leer Study during its last two years, the committee members were already familiar with the sampling design, measurement instruments, and so forth.

The Ad Hoc Committee, in turn, established subcommittees in six subject areas covered by the study: reading, mathematics, geography, science, biblical studies in nonreligious schools, and biblical studies in religious schools. Additionally, a Resource Allocation Committee was nominated, chaired by the person in the Ministry in charge of those activities. Each of these subcommittees worked for approximately a year and prepared a written report.

The Subject Matter Subcommittees.

Each of the six subject-matter subcommittees consisted of 5-6 persons (teachers and curricular experts). In most cases, they were chaired by a Chief Superintendent of the Ministry responsible for the teaching of that particular subject. The subcommittees were assigned to examine test-item data from the study. In several subject areas, data had been collected at four grade levels (first, second, forth, and sixth grades), and each of these tests contained approximately 200 items. After examining the items to check their adequacy for testing achievement at a particular grade level and their content validity vis-a-vis the curriculum, the subcommittees were supposed to suggest ways of improving achievement related to these items. They were asked to devote special attention to that subset of items which had been defined by the researchers as minimum achievement requirements.

The work patterns of these subcommittees varied, as did their reports. Typically, a

subcommittee held from four to eight meetings throughout the year. In addition, several subcommittees did small-scale empirical work to support their contentions. For example, the subcommittee for biblical studies in nonreligious schools initially claimed that many test items were unduly difficult owing to linguistic factors unrelated to knowledge of the Bible. To test their claim, the subcommittee simplified the linguistic structure of the items without changing their content and then compared the difficulty level of the two versions of the items. The conclusion was that the initial claim was erroneous: Linguistic variation did not affect the difficulty level of the items.

After examining the content validity of the test items, only the science subcommittee was severely critical: It claimed that some items were statements lifted from textbooks and that others had multiple correct responses or no correct response at all. All the other subcommittees found the tests to have appropriate content validity.

The subcommittees suggested several actions for program improvement: e.g., starting with the fourth grade, certain subjects should be taught by specialists; greater use should be made of audiovisual aids in the higher grades.

The Van Leer Study, which was conducted at a time when curricular reform was being implemented in Israel, had concluded that the quantity of materials prescribed by the old curriculum was excessively high and that, since the re-

quirements of the new curriculum were even more complex and voluminous, students would have even greater difficulties coping with them. Most of the subcommittees took issue with the claim that the curriculum was overloaded, attributing low student achievement to the modular approach dominant in the schools, an approach which enables teachers to select certain parts of the curriculum for emphasis. Several subcommittees opposed any reduction in curriculum content, suggesting instead that a set of core requirements (content and skills) be specified for each subject. Indeed, all the subcommittees recommended (though not very enthusiastically) that such requirements be established. It should be added that, when they made their recommendations, the subcommittees were aware that the Standing Committee for Primary Education favored establishing a Core Requirement for each subject at each grade level. Thus, their recommendations may be seen as reflecting acquiescence with this decision.

The Resource Allocation Subcommittee.

The work of the Resource Allocation Subcommittee was extremely complex and difficult. It was supposed to provide up-to-date information about resource allocation by documenting changes in the system since 1973. Discovering that there was a high level of departmentalization in allocating resources to schools and that some departments could not provide data about how they distributed resources among disadvantaged and advantaged schools, the subcommittee recommended that in the future each department should

report its activities in such a way as to indicate their contribution to work improving the situation of schools for the disadvantaged.

FURTHER STUDIES AND ANALYSES

Several replication studies and a number of supplementary analyses were conducted subsequent to the Van Leer Report. They are mentioned here because some of them affected later administrative decisions or actions and because the very fact that they were conducted may be viewed as one impact of the Van Leer Study on research and on the research community.

Replication Studies.

In a formal communique that accompanied the release of the Van Leer Report, the Public Relations Officer of the Ministry of Education expressed doubts that the results reported in the study were up-to-date and promised that a follow-up would be conducted to examine changes in the system since 1973. Two such studies were performed, one in 1978 (a year after publication of the Report) and the other in 1980 (when both the Knesset and the Ministry of Education had already approved measures for action). Thus, the second replication study clearly could not affect the full range of decisionmaking related to the Van Leer Study. Nevertheless, we will briefly report below on both replication studies to illustrate that the actual impact of a study lasts longer than its formal bureaucratic impact.

Resource Allocation Studies.

In 1978, the Chief Scientist's research assistants conducted a study which used the Van Leer instrumentation to examine resource allocation practices at a subsample of the schools surveyed in 1973. Thus, the data collected in 1978 provided information on changes whch had occurred since 1973. This study (Raziel, 1978), which became known as the Raziel Bulletin, greatly increased the prestige of the Ministry of Education, especially in the eyes of the Knesset's Educational Committee, but also in the eyes of the general public. It was impressive for its clear message and its timing. While others were busy debating the significance of the Van Leer Study findings, the Raziel Bulletin represented positive action and shed light on issues which until then had been topics for speculation only.

The findings revealed increases in the resources allocated both to schools for the advantaged and to schools for the disadvantaged. On several variables, the gap between the two groups of schools had been reduced: indeed, on some variables, parity had been achieved. On most variables, however, the magnitude of the gap had remained constant since 1973. Thus, the Raziel Bulletin could point to some favorable changes in the system.

This evidence that a gap in resource allocation still existed between "advantaged" and "disadvantaged" schools, despite continuing administrative efforts to increase the resources of schools for disadvantaged learners, prompted

the Ministry of Education to initiate yet another study of resource allocation. In 1981, eight years after the data for the Van Leer Study had been collected, a team revisited the schools surveyed in the Van Leer Study. Besides using the same instruments that had been used in the original study and in the Raziel (1978) study, the team collected additional information about resource allocation from a sample of approximately 2,500 fifth-grade children and their teachers. Although the results of this study (Davis and Sprinzak, 1983) have not yet been established, interim reports indicate that further progress has been made in equalizing resources. Indeed, "disadvantaged" schools now lead "advantaged" schools with respect to some resource components. Thus, some of the shortcomings in the educational system revealed by the Van Leer Study have apparently been rectified. Though no evidence on impact is directly available, one may assume that the Van Leer Study was at least partially responsible for effecting these changes.

The Jerusalem District Achievement Study.

Initial reactions to the Van Leer Study's findings with respect to achievement paralleled initial reactions to its findings on resource allocation; that is, school supervisors claimed that the situation was better in 1977 than it had been in 1973. It is not surprising, then, that immediately after publication of the Report, some people were calling for a replication

of the achievement survey. Collecting achievement data from thousands of students is a more demanding task, however, than collecting resource information from 150 school principals. Therefore, the new survey of achievement data had to wait until 1979, when some members of an ad hoc committee exploring alternative modes for a Ministry of Education feedback system collected data on the reading and mathematics achievement of fourth grade children in the Jerusalem school district (Egozi, 1981). Because of budget limitations, this replication study was more restricted than the Van Leer Study had been with respect to sample size, number of grade levels tested, and number of subject areas covered. Nevertheless, comparative data became available. The findings were disappointing. Mathematics achievement had not changed, and reading achievement had improved only slightly. Although experts at the Curriculum Center attributed this slight improvement to the dissemination of a new language curriculum, which reached the schools in 1973-74, the possibility exists that the improvement is simply a reflection of a greater familiarity with the tests on the part of teachers or of better preparation for taking multiple choice type reading tests on the part of students.

Supplementary Analysis Studies, 1977-1980.

Using the wealth of information available in the Van Leer data base, several researchers (all of whom had participated in the original study) re-

analyzed the data to investigate issues not directly examined in the Van Leer Study. Four such reanalysis studies will be mentioned here.

In a secondary analysis of the intelligence test data from the Van Leer Study, Peled (1980) used Guttman's Facet Analysis Type definition to identify specific differences between children of Asian-African origin and those of European-American origin with respect to intellectual ability patterns. Moreover, in a separate multiple regression analysis, she found that home background variables explained 19 percent of the variance in ability for European-American children and 13 percent for Asian-African children.

Bashi (1977), one of the co-authors of the Van Leer Report, used the data base to examine the impact of class composition on student self-concept. He found that, after the student's relative class standing is taken into account, there is no relation between class composition and academic self-concept; this holds true for both ethnic groups in the fourth grade and for the European-American group in the sixth grade. Moreover, once achievement in relation to that of the total population is taken into account, there is no relation between class composition and future-success-oriented self-concept.

The question of whether teachers discriminate against pupils of Asian-African origin in assigning grades was examined by Cahan (1977), who used the Van Leer Study data base to compare the grades received by children of different

ethnic origins, matched with respect to scores on objective tests. He found no evidence of teacher discrimination.

Finally, Egozi (1978) used the Van Leer data base to check the validity of a "disadvantaged" index (based on students' background characteristics) that is used by the Ministry of Education in allocating compensatory resources to schools. His reanalysis led to suggestions for minor changes in the index.

A common feature of these four studies is that they examine issues not treated in the original study. So far, no one has attempted to conduct secondary analyses of the Van Leer Study data for the purpose of examining the validity of the original findings or of checking their robustness across various methods of data analysis.

FURTHER ADMINISTRATIVE/LEGISLATIVE ACTIONS

The studies and analyses described in the preceding section were carried out after the release of the Van Leer Report and, for the most part, after the formation of the Knesset subcommittee and the Ad Hoc Committee of the Ministry of Education. Meanwhile, both of these bodies were studying the findings, exploring their implications, and formulating recommendations. The next steps in the chronology involve the release of those recommendations and the administrative and legislative actions then taken on the basis of those recommendations.

Epilogue at the Ministry of Education.

Early in 1979, the Ad Hoc Committee of the Ministry of Education pulled together the recommendations of its subcommittees and produced a document containing a series of recommendations which served as a basis for discussion at various forums within the Ministry: the Pedagogical Secretariat,[3] the Directory of the Ministry, and finally the Standing Committee for Primary Education. After devoting three sessions to discussion of the Ad Hoc Committee's recommendations, the Standing Committee approved the suggested recommendations, with only slight modifications, in May 1980.

The Ad Hoc Committee made 36 recommendations, clustered in eight groups:

1. **Curriculum:** In each subject matter and for each grade level, a list of basic concepts and basic skills should be defined. The Curriculum Center should focus on providing programs for the disadvantaged. The development of programs for heterogeneous classes also deserves special attention.
2. **Planning and Reporting:** All units within the Ministry dealing with resource allocation should maintain

[3] The sessions in the Pedagogical Secretariat shed additional light on the process of accepting and implementing recommendations. Appendix D presents details about one of these sessions, held on March 4, 1979, immediately after presentation of the Ad Hoc Committee's report.

records about the distribution of resources between schools for the advantaged and the disadvantaged.

3. **Integration:** Appropriate teaching methods should be developed for integrated classes. A committee should be established to examine the implications -- in terms of educational policy, school organization, and economic resources -- of integrating the primary schools.
4. **Number of Hours Per Class:** A system should be developed to monitor and control the utilization of supplementary teaching time allocated to the schools for the disadvantaged.
5. **Preservice Teacher Training:** The teacher training institutes should include in their program studies of the new curricula; should ensure that their students have an appropriate knowledge of mathematics; and should increase the study of problems specifically related to teaching the disadvantaged. Special financial benefits (for housing arrangements) should be expanded to all teachers who teach in schools for the disadvantaged.
6. **Inservice Training:** Intramural inservice training should be increased at the teacher training institutes, and follow-up studies of inservice programs should be designed.

7. **School Services:** Existing recommendations concerning services (medical, social worker, psychological, etc.) should be examined with the goal of creating cooperation with corresponding community services.
8. **Subject Matter:** In reading, exercises in which learners are required to "translate" pictures into words (and vice versa) should be increased. In mathematics, more audiovisual aids should be used. In geography, more classroom teaching time should be devoted to map-reading skills.

The recommendations are policy statements: They tend to be stated in very general terms, and to have an exhortative tone. The task of working out operational details is imposed on the heads of various departments in the Ministry. The recommendations are addressed to those issues which had constituted the focus of discussions in the Ministry and among Israeli educators during the previous few years, regardless of whether they had been dealt with directly in the Van Leer Report. For example, several recommendations concerned the provision of better educational programs for heterogeneous classes, though the Van Leer Report had not touched upon this issue. Further, though the Report had pointed to low achievement levels in various subject areas, it did not suggest what type of organizational setting might best raise achievement levels. Nevertheless, the Ad Hoc Committee

recommended strengthening mixed ability class arrangements and asked the Curriculum Center and the Pedagogical Secretariat to devote attention to the production of appropriate learning materials for such classes.

One of the interesting features of the Ad Hoc Committee's summary document is that it specifies the department within the Ministry which should assume responsibility for carrying out each of the recommendations. The heads of the departments were also asked to prepare reports about the progress made in implementing the recommendations, and the Standing Committee was asked to appoint a working committee to follow up the implementation of the recommendations and to monitor the progress reports prepared by the department heads responsible for carrying out the recommended activities.

Epilogue in the Knesset.

The Knesset Educational Committee approved the recommendations of its subcommittee and presented them to the full Parliament on January 1, 1980. At that time the Minister of Education was asked to report to the Knesset within six months about steps taken to implement the recommendations. On July 28, 1980, the Minister presented his response, reporting on arrangements made by the Pedagogical Secretariat to increase the resource allocations to schools for the disadvantaged and on the Ministry's decision to define minimum achievement requirements and to instruct the supervisory staff to emphasize the teaching core of objectives. Other points

covered by the Minister were as follows: Adult education programs aimed at teaching parents to help their children with homework had been launched; teacher training institutes were giving more attention to the topic of dealing with disadvantaged learners; the salaries of teachers working in villages with overwhelmingly disadvantaged populations had been increased; and additional resources had been allocated to training school principals for "disadvantaged" schools. The Minister's report concluded the Knesset's dealing with the Van Leer Study.[4]

THE SECOND DECADE OF THE STUDY

The Van Leer Study was conceived in 1969; data collection was carried out in 1973; and the report published in 1977. In 1982-83, when this monograph was being published, one could say that the study has not yet been fully completed. Even though its findings were consistent with those of previous surveys, many people were reluctant to accept them. The authors of the Report found signs of both success and failure in the educational system. That second-generation learners had a higher level of achievement than immigrant children was interpreted as evidence that the situation was gradually improving. On the other hand, the achievement gap between the two ethnic groups, the relatively

---

[4] See Knesset Proceedings (in Hebrew), Fourth Session (January 1, 1980), Issue II, pp. 1262-1301. For the Minister's response, see Issue 36, p. 3478.

low overall achievement levels, and inequalities in resource allocation were judged to be indications of the school system's failures. The first reactions to the Report focused on the unfavorable findings. Both the mass media accounts and the Knesset discussions had a gloomy tone.

In an interview with a journalist from Maariv (July 4, 1977), the director general of the Ministry of Education, Mr. E. Shmueli, reacted to the study by saying that the results reflected the situation in 1973. Since that time, he said, the situation could have improved because of the Ministry's efforts to improve the resource allocation balance. He cited subsequent replication studies to document this balance. As noted, a second replication study is still in progress, and its results are not yet fully known.

Social scientists frequently measure the impact of a study by counting citations in scholarly journals and books. The Citation Index is a widely recognized, though not universally accepted, measure of impact. It may well be that the Van Leer Report will gradually come to occupy longer columns in the Citation Index. Its real impact, however, might be better measured by counting the number of times the Report is mentioned by educational administrators and teachers as they plan programs for the entire educational system or for a particular class. In 1982, five years after publication of the Report, one still hears frequent references to

the study. One may even detect the influence of its findings on certain decisions made within the school system. While researchers mine the data base for evidence in support of new hypotheses, the administrative leaders of the system try to prove that its findings are no longer valid--examples of two different kinds of utilization.

It appears that the findings of the Van Leer Study will continue to be mentioned in years to come, until it becomes clear that they are of historical interest only and no longer reflect the actual situation prevailing in the educational system.

ACHIEVING IMPACT

In many cases only a thorough, long-term study can detect all the impact of a particular evaluation. Impact may be gradual and cumulative; it may be manifested in policy actions, in attitude changes, in general conceptual understanding, or in all of these. The research literature suggests several modes of evaluation use. For instance, Caplan and his colleagues (1975) distinguish between "instrumental" use (when evaluative information is used directly in making programmatic changes) and "conceptual" use (when evaluative information influences policymakers' thinking without having any documentable effect on action). The distinction between these two types of use is perhaps clearer in theory than in practice, as is obvious in the case of the Van Leer Study. A third type of use mentioned in the literature is "symbolic" use (Pelz, 1978), when an evaluation is conducted for symbolic purposes; e.g., merely to satisfy the requirements of external agencies. Alkin (1975) provided a number of examples of symbolic evaluation use.

Even within these categories, further distinctions may be made: for instance, in the levels at which the various types of use take

place (Kennedy et al., 1980) instrumental use may occur with respect to policy actions at the national government level or in the Ministry of Education. Likewise, and perhaps even more indicative of the breadth of level, conceptual use may occur at the ministerial level, at the classroom or local school level, and at various administrative levels in between. Conceptual use of evaluative information is also exemplified when researchers and other scholars discuss the findings of a particular study or are prompted to conduct follow-up studies. In short, it is nearly impossible to document completely all of the effects of a particular evaluation report.

The Van Leer Study offers an unusual opportunity to document the manifest and direct impact of a particular evaluation study not only at the local level (in Israel) but also at the international level. Few studies have ever elicited such varied and dramatic reactions, as is evident from the impact chronology presented in the previous chapter. The uses made of the Van Leer Report findings were not instrumental, in the strict definition of the term; rather, they may be viewed primarily as conceptual or, at best, as hybrids somewhere between the instrumental and the conceptual. While it is true that the Ministry of Education ultimately produced a set of recommendations, as did the Knesset, these policy actions were filtered through the medium of various committees and other intermediate or peripheral agencies. The primary

impact of the Van Leer Study was conceptual: the study data provided the basis and the stimulus for various deliberations that eventually led to a variety of policy actions. The conceptual use of the Van Leer Study findings is undoubtedly consistent with the nature of the study, whose purpose was to provide information that would have lasting significance for the Israeli educational system and that could be used in various ways for various decision purposes. Thus, the Van Leer Study constitutes a classic case of a large-scale evaluation study which had significant and documentable conceptual uses.

FACTORS ENHANCING IMPACT

An analysis of the events chronicled in the previous chapter suggests several factors that enhanced the intensity of the reactions to the Van Leer Study and thus heightened its subsequent impact. Many of these factors have been noted in the literature on utilization and impact, and we will comment briefly on some of that literature within the context of the discussion that follows.

Technical Quality of the Report.

The Van Leer Report received generally high marks for its technical quality and methodological rigor. It was judged to be an excellent study by fellow researchers in Israel, and even its critics praised the study as a whole, directing their criticisms only at specific aspects of the study. For example, as was noted earlier, some questions were raised over certain psy-

chometric issues, and even the two co-authors of the Report disagreed with some of the methodological procedures imposed by the Principal Investigator. Nonetheless, the report en toto was generally perceived as being of superior technical quality, and this factor served as a basis for the attention given to the study recommendations in the various policy deliberations.

The bulk of the research literature on evaluation utilization (and especially that part of the literature which bemoans the lack of evaluation utilization) indicates that methodological rigor is a necessary but not sufficient condition for utilization (see Alkin, in press). That is to say, if an evaluation study is methodologically weak, that deficiency will serve as a basis for rejecting its findings; but high technical quality does not guarantee that the findings of a study will be accepted and used.

It may be that technical quality was a more important factor in the case of the Van Leer Study than in many of the studies cited in the research literature that focus on state and local government levels. The Van Leer Study was commissioned to increase conceptual understanding about the Israeli primary school system rather than to produce information for specific decisions. Thus, the implication is that program sponsors had higher expectations about the report's technical quality and methodological rigor. All of which is to say that high technical quality was an even more necessary condition of use than is usually the case.

Involvement of Prominent Scholars.

The Van Leer Study was also accorded credibility because a large group of scholars--experts from both the Israeli and international research communities--participated as ad hoc consultants in the preparation of the study plans and at various stages in the execution of the study. This large-scale involvement, which is rare and probably could only happen in a small country like Israel, not only contributed directly to the quality of the study but also helped to assure its acceptance and to minimize sharp criticism against it.

While the importance of report credibility is frequently mentioned in the research literature (see Alkin and Law, 1980; King, Thompson, and Pechman, 1982; O'Reilly, 1981), most observers see it as an extension of the personal credibility of the evaluator. Here, in addition to the credibility that emanated from the scholarly reputation of the evaluators, credibility was assured through the involvement of prominent scholars who, in essence, placed their imprimatur upon the Van Leer Report.

Characteristics of the Evaluators.

The characteristics of the evaluators, and especially of Professor Minkovich as Principal Investigator, also enhanced its impact. During several decades of scientific work in Israel, Minkovich had established a reputation as a dedicated researcher, a charismatic speaker, and an ardent fighter for the implementation of his

pedagogical ideas. His participation in various educational activities and on government committees had, in the past, generated public controversy, which usually ended with the acceptance of his ideas. Thus, his prestige played an important role in compelling people to react to the Van Leer Report.

Alkin, Daillak and White (1979) have documented the importance of the characteristics of the evaluator as a factor influencing the utilization of evaluative information. Most of the literature focuses on such characteristics as credibility, orientation toward providing user information, and rapport with audiences. In the case of the Van Leer Study, the most important evaluator characteristic was probably the personality and dynamism of the Principal Investigator.

A further comment is in order here: during the period when the committees were functioning (1977-79), Professor Minkovich passed away. Thus, the reduced public interest in the Van Leer Study during this period may in part be attributable to his absence from the scene. Had the Principal Investigator been alive when the committees finished their work, his personal involvement in the course of events would almost certainly have created a higher level of interest in the consequences, thus increasing the study's impact.

Interested Users and Crucial Issues.

The Van Leer Study was requested by the Director General of the Ministry of Education, and the

Ministry was the actual recipient of the Report. As was pointed out earlier, the fact that the study was initiated at the top level of a hierarchical system further assured that attention would be paid to the findings. The initiators of an evaluation are not often personally jeopardized by its results, since they are in a position to make program judgments based on the findings, but those at lower levels may be threatened. Thus, there was the highest possibility that the Ministry of Education and the Knesset would pay attention to the study since it focused chiefly on actions below the level of Director General. Further, the recent change in government, with the consequent installation of a new Minister of Education, relieved the administrative hierarchy of responsibility for any negative findings.

Not only were there interested users of the Report, but also the study addressed critical issues, and was the topic of numerous debates, discussions, and conferences--thus assuring its visibility to the public at large as well as to the research community and government sector. As noted earlier, the abolition of the annual SEKER examination in 1971 had deprived Ministry officials and the public of a valuable source of information about the Israeli school system and had created a demand for the Van Leer Study.

One of the most critical issues addressed by the Van Leer Study was the comparative achievement of children from different ethnic backgrounds. In the years immediately prior to

the study, the Ministry of Education had initiated administrative changes within the school system designed to reduce the achievement gap between children of Asian-African origin and those of European-American origin, and both Ministry officials and the public were anxious to know if those efforts had been successful. Thus, the potential for impact was high, in terms of both the issues involved and the presence of interested potential users of the findings.

Patton and his associates (1975) maintain that the presence of interested users is the most important determinant of evaluation utilization. In the case of the Van Leer Study, no single individual stands out as the prespecified user of the evaluation results (as Patton would have it); nonetheless, a mood of anticipation, interest, and intent to act on the results was evident within the administrative and legislative hierarchy.

The relevance of the issues considered in the study heightened the anticipation and increased the likelihood that the findings would be utilized. A variety of researchers have dealt with this factor. For instance, Alkin, Daillak and White (1979) note that the more the findings or recommendations of an evaluation reflect an important program concern, the more likely it is that they will be used in decision-making. Similar observations have been made by Braskamp and Brown (1980) and by King, Thompson, and Pechman (1982).

COMPLEMENTARY DATA

Yet another factor associated with the high impact of the Van Leer Report was the availability of complementary data. Policy actions do not take place in a vacuum; information beyond that supplied by the evaluator is always available. An evaluation tends to be seen as valuable when it adds new dimensions to, or substantiates, data already available. In essence, this "triangulation" of data, which had been discussed by Popham (1975), legitimates the information source.

Such was the case with the Van Leer Study. In making a set of recommendations based on study data, the Ad Hoc Committee of the Ministry of Education noted that the Van Leer data were consistent with trends noted by the highly prestigious Etzioni Committee (a government-appointed body, headed by a high court judge, that was charged with formulating recommendations to improve the status of teachers in Israel and thus avert the threat of a general teachers' strike). The Ad Hoc Committee's recommendations gained additional credibility from their congruence with the recommendations emanating from the Etzioni Committee.

Antecedents of Report Publication.

Several events that occurred prior to the release of the Van Leer Report contributed to the interest it aroused and intensified its ultimate impact. For instance, while the study was being conducted, progress reports and other reminders

of the study's existence were periodically furnished to the Ministry of Education and the Israeli research community. Although the actual findings were not divulged, information was provided on the scope of the study, the topics covered, and the research methodology. Emphasis on the attention by the researchers to maintaining high technical quality of the study helped to assure a more favorable response.

Further, the sample for the study comprised 15 percent of Israel's primary schools, from all parts of the country. Thus, hundreds of teachers and thousands of parents were involved in, or at least aware of, the study. Clearly, then, the activities surrounding the study generated high anticipation and increased the possibility of impact.

Institutionalized Mechanism for Impact.

The final factor in explaining the impact of the Van Leer Study is the existence, within the Israeli Ministry of Education, of the Office of the Chief Scientist, which is responsible for creating links between the Ministry and the research community, analyzing and criticizing research findings, and translating these research findings into recommendations for action. Thus, the utilization of research findings is not left to chance but is systematically monitored by the Chief Scientist. Talking about his role as a facilitator of evaluation utilization, the Chief Scientist has said: "The Chief Scientist may have to create particular organizational mechanisms to help maximize the research

utilization. In terms of the resources available to the Chief Scientist, this may suggest that even more emphasis should be given to the problem of research utilization" (Kugelmass, 1981). Indeed, quite apart from his official function in creating mechanisms to review the recommendations and consider the implications of evaluation studies, the Chief Scientist may--through his perception of the importance of a study and of its potential as a basis for making improvements and implementing change--influence evaluation utilization.

A number of writers (e.g., King, 1982) have commented that the evaluator can create a demand for evaluation information by developing mechanisms that facilitate use of evaluation information by decisionmakers. In the case of the Van Leer Study, the mechanism was organizational: that is, arrangements within the Ministry of Education itself were conducive to evaluation use. Very little has been written about this kind of mechanism. According to Alkin et al. (1982), some states in the U.S. require funded projects to report the extent to which evaluation information from the previous year's report was utilized, and this requirement may facilitate impact. Similarly, Raizen and Rossi (1981) and Alkin (1980) have mentioned evaluation use reports as an institutionalized mechanism that may be a factor in assuring impact.

OTHER ISSUES

This investigation of the impact of the Van Leer Study was requested by the Ministry of Educa-

tion, but the Report itself made difficult reading for any decisionmaker concerned with policy. Indeed, a Knesset member used the Report to attack Israeli educational researchers in general, claiming that they were more interested in communicating with their peers than with educators, teachers, and parents.

Would the study have an even greater impact if it had been written in a style more accessible to the lay reader? The question is difficult to answer. After all, the findings of most large-scale surveys are presented as scientific reports but may nevertheless have widespread effects on the public. A large and detailed report on the findings of an international achievement survey carried out by the International Association for the Evaluation of Educational Achievement gained a large readership, whereas a one-volume summary report (Walker, 1976) appended to the findings had no great impact.

The charge that researchers are interested in publishing for their peers only seems to have some justification. Some observers have suggested that a study should produce several reports, aimed at different audiences and written in different forms and styles (Alkin and Fink, 1974). The issue of single versus multiple reports deserves further examination.

Amount of Information.

The amount of information presented in the Van Leer Report is overwhelming, and no reader can grasp it all in a single reading. The authors

themselves were well aware of this fact. In his debate with the Chief Scientist, the Principal Investigator charged that the Chief Scientist had overlooked certain facts stated in the Report. Further, he noted that some topics had been broken down into subtopics and were dispersed in various sections throughout the book. This could create the impression that the authors had omitted important facts related to their findings, when in reality, the fact was stated somewhere else in the Report. In essence, the report contains a compendium of data, and the reader may have to search for specific information.

The findings that attracted the greatest attention (and thus may be said to have had the greatest impact) were related to topics that were already of considerable interest (e.g., the achievement gap, integration, the relative effectiveness of the religious school system) but represented only a small proportion of the data contained in the Report. Given this fact, one may ask whether it might not have been preferable, in the planning stages, to focus the study on a more narrow range of issues, thus reducing the amount of data collection and producing a more limited set of data summaries. It is difficult to answer this question. Certainly, greater selectivity might have been desirable.

On the other hand, the very scope of the study, and the abundance of information it produced, carry certain advantages. First, the

study itself was in part a kind of search for unknown phenomena. And, indeed, it yielded some unexpected findings which became sensations. Second, the Report has a "handbook" function (which would not have been possible without the plethora of data), enabling the reader to extract information about emerging issues and to obtain baseline information on the details of broad topics. Third, the peripheral data summaries serve to cushion and support the major findings, adding to the scientific prestige of the Report.

The larger question is: Just how useful are large-scale studies that may require a decade to complete? The Van Leer Study had an impact that could not have been achieved by a series of shorter and more focused studies, even if the latter had produced just as much, or more, information. Large-scale studies constitute road signs in the history of research. Functional social science research demands a rhythm of large-scale and small-scale studies; the large-scale studies establish prestige for the scientific endeavor, whereas the small-scale studies are necessary to solve the many problems that arise in the process of social actions. The Kinsey Report probably would not have impressed the world if it had first been published as a small book containing a precise summary of what was in the actual report, even though most readers did not really want to know more than could be condensed in a summary.

Use and Misuse of Data.

The prodigious size of the Van Leer Report had several consequences not anticipated by the authors. First, it laid the basis for the moratorium on action. Nobody could, or wanted to, act immediately, without careful and thorough consideration of the data. The Knesset needed two additional years before it was ready to formulate recommendations; the Ministry approved its last recommendation three years after the Report's release. This does not mean that nothing happened in the system during the moratorium period. It is reasonable to believe (although impossible to document) that, at both the Ministry and the local school levels, the study findings directly affected actions. The committee work itself was worthwhile, insofar as it involved leaders in serious consideration of Israel's educational problems and attempts to correct those problems.

The moratorium certainly did not please the authors of the Report, since it meant that they had to wait a long time before they knew whether the leaders of the system accepted or rejected their findings. Had the Report focused on only a few problems, structured the recommendations to relate directly to those problems, and invested further effort in proving that the data were relevant to the recommendations, then a quicker decision might have been reached. On the other hand, the authors may have understood that immediate implementation of the recommendations was undesirable and that policy deci-

sions should be based on an integration of the evaluation data and recommendations with experience accumulated in the system.

A second adverse consequence of the abundance of data contained in the Report was that these data were exploited to support conflicting views and suggestions for action. In some cases, this use of the findings was justified; that is, the data clearly provided evidence that favored one point of view over another. In other cases, the data were inconsistent, allowing decisionmakers to give greater weight to whatever bit of information supported their particular opinions. In still other cases, the conflicting interpretations were merely exegetic exercises performed on imprecisely formulated sentences. Again, more concise data summaries might have prevented such misuse of the study findings.

A third unanticipated consequence was that the recipients of the Report (chiefly the mass media, but sometimes the politicians) interpreted the results of the study in ways that directly contradicted the authors' intentions. Indeed, on one occasion, the Principal Investigator accused journalists of deliberately falsifying the results, for political purposes. His accusation may have been justified, but it is equally possible that the journalists, confused by the mass of findings, made honest mistakes of interpretation. (This would explain why they rejected the Principal Investigator's request that they publish corrections to some of their stories.)

The lesson to be learned from this experience is that, when it comes to controversial issues, authors should be careful to state their own views clearly. Indeed, it may not be enough to state what conclusions can be drawn from the findings; perhaps one should also make explicit mention of what conclusions **cannot** be drawn on the basis of the evidence.

REFERENCES

Airasian, P.W. & Madaus, G.F. A post hoc technique for identifying between program differences in achievement. Studies in Educational Evaluation, 1978, 4, 1-8.

Alkin, M.C. Evaluation: Who needs it? Who cares? Studies in Educational Evaluation, 1975, 1(3), 201-212.

Alkin, M.C. Improving Title I Technical Assistance: An Evaluation Utilization. Paper presented at the American Educational Research Association Annual Meeting, Boston, 1980.

Alkin M.C. Utilization of evaluation findings. INNOTECH Journal, 7(1), Regional Center for Educational Innovation and Technology, Manila, Philippines, in press.

Alkin, M.C., Daillak, R.H., & White, P. Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage Publications, 1979.

Alkin, M.C., & Fink, A. Educational product report: A user orientation. In G. Borich (Ed.), Educational product evaluation. Englewood Cliffs, NJ: Educational Technology Press, 1974.

Alkin, M.C., Kosecoff, J., Fitz-Gibbon, C., & Seligman, R. Evaluation and decision making: The Title VII experience. Los Angeles: CSE Monograph No.4, UCLA Center for the Study of Evaluation, 1974.

Alkin, M.C., & Law, A. A conversation on evaluation utilization. Educational Evaluation and Policy Analysis, 1980, 2(3), 73-79.

Bashi, J. Effects of ethnic class composition on self concept. Megamoth, 1977, 23, 124-133. (Hebrew with English abstract)

Bashi, J., Cahan, S., & Davis D. Educational achievement in Arab primary schools in Israel. Jerusalem: Hebrew University, School of Education, 1981. (In Hebrew)

Boruch, R.F., & Cordray, D.S. An appraisal of educational program evaluations: Federal, state, and local agencies. Evanston, IL: Northwestern University, 1980.

Braskamp, L.A., & Brown R.D. New directions for program evaluation: Utilization of evaluation information. San Francisco: Jossey-Bass, 1980.

Braskamp, L.A., Brown, R.D., & Newman, D.L. The credibility of a local educational program evaluation report: Author source and client characteristics. American Educational Research Journal, 1978, 15(3), 441-450.

Brown, R.D., Newman, D.L. & Rivers, L. Perceived need for evaluation as an influence on evaluator impact on decision making. Educational Evaluation and Policy Analysis, 1980, 2, 67-73.

Cahan, S. Ethnic discrimination in teachers' evaluation of pupils' achievements. Megamoth, 1977, 23, 238-247. (Hebrew with English abstract)

Caplan, N., Morrison, A., & Stambaugh, R.J. The use of social science knowledge in policy decisions at the national level: A report to repondents. Ann Arbor, MI: University of Michigan, Institute for Social Research, 1975.

Carr, E.H. What is history. London: Penguin Books, 1961.

Davis, D., & Sprinzack, D. An eight year follow up of resource allocation in Israeli schools. (Temporary title). Jerusalem: Hebrew University, School of Education, 1983.

Dror, Y. Think tank: A new idea in government. In C.H. Weiss & A.H. Barton (Eds.), Making bureaucracies work. Beverly Hills, CA: Sage Publications, 1979.

Egozi, M. Recommendations for changes in the disadvantaged index. Jerusalem: Ministry of Education and Culture, Division of Planning, 1978. (In Hebrew).

Egozi, M. Achievements in the fourth grade of the Jerusalem district. Jerusalem: Szold Institute, 1981.

Fullen, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, 1977, 47, 335-397.

Kennedy, M.M., Apling, R., & Neuman, W. The role of evaluation and testing information in public schools. Cambridge, MA: The Huron Institute, 1980(a).

Kennedy, M.M., Neuman, W., & Apling, R. The role of evaluation and testing information in Title I programs. Cambridge, MA: The Huron Institute, 1980(b).

King, J.A., Thompson, B., & Pechman, E.M. Improving evaluation use in local schools. Washington, D.C.: National Institute of Education, 1982.

Kugelmass, S. Considerations toward a policy of evaluation research: The case of the Chief Scientist in the Israeli Ministry of Education. In A. Lewy et al., Decision oriented evaluation in education: The case of Israel. Philadelphia, PA (USA)/ Rehovoth, Israel: International Science Services, 1981.

Lewy, A. Discrimination among individuals versus discrimination among groups. Journal of Educational Measurement, 1973, 10, 13-24.

Lewy, A. Timeliness and efficiency in serial evaluation activities. Educational Evaluation and Policy Analysis, 1981, 3, 55-66.

Lewy, A., & Chen, M. Differences in achievements. Evaluation in education (Vol.1 No.1). Oxford: Pergamon Press, 1977.

Lewy, A., Rapaport, C., & Rimor, R. Educational achievements in the Israeli school system. Tel-Aviv: Ramoth, 1978. (In Hebrew)

Maslow, A.H. Toward a psychology of being. New York: D. Van Nostrand, 1962.

Minkovich, A., Davis, D., & Bashi, J. An evaluation study of Israeli elementary schools. Jerusalem: The Hebrew University, School of Education, 1977.

O'Reilly, C.O.,III. Evaluation information and decision making in organizations: Some constraints on the utilization of evaluation information. In A. Bank, R.C. Williams, & J. Burry (Eds.), Evaluation in schools districts: Organizational constraints. Los Angeles: CSE Monograph No. 10, UCLA Center for the Study of Evaluation, 1981.

Orthar, G. Educational achievements as related to socio-cultural background of primary school graduates in Israel. Megamoth, 1967, 15, 220-230. (Hebrew with English summary)

Parsons, T., & Shils, E. Toward a general theory of action (3rd ed., 1954). New York: Harper & Row. P. 167.

Patton, M.Q. Utilization focused evaluation. Beverly Hills, CA: Sage Publications, 1978.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., Grench, D.B., & Blyth, D.A. In search of impact: An analysis of utilization on federal health evaluation research. Minneapolis: University of Minnesota, Department of Sociology, 1975.

Pedagogical Secretariat. Minutes of Sessions, 1979. Jerusalem: Ministry of Education, 1979. (Mimeogr.)

Pedagogical Secretariat, Plenary Session, 1976/6, Minutes. Jerusalem: Ministry of Education and Culture, 1976.

Peled, S. Patterns of mental ability of Israeli children of Oriental and Western origin. Doctoral dissertation, New York, Columbia University, 1980.

Pelz, D.C Some expanded perspectives on use of social science in public policy. In J.M. Yinger & S.J. Cutler (Eds.), Major social issues. London: The Free Press (Division of MacMillan Publishing Co., Inc.), 1978.

Popham, W.J. Educational evaluation. Englewood Cliffs, NJ: Prentice Hall, Inc., 1975.

Raizen, S., & Rossi, P. (Eds.) Program evaluation in education: When? How? To what ends? Washington, D.C.: National Academy Press, 1981.

Reisner, E.R., Alkin, M.C., Boruch, R.F., & Millman, J. Assessment of the Title I evaluation and reporting system. Washington, D.C.: U.S. Department of Education, 1982.

Raziel, O. Compensation for whom - a follow up investigation of resource allocation. Jerusalem: Ministry of Education, The Resource Allocation Committee, 1978. (In Hebrew)

Self, P. Administrative theories and politics. London: George Allan and Unwinn, 1972.

Smilansky, M., & Yam, Y. The relationship between family life, ethnic origin, father's education and students' achievement. Megamoth, 1969, 16, 248-273. (In Hebrew)

Teachers Union. Kuntres Proceedings. Tel-Aviv: Teachers Union, 1979.

Walker, D. The IEA six subjet survey: An empirical study of education in twenty-one countries. Stockholm: Almqrist and Wiksell, 1976.

Walker, D., & Schaffarzik, J. Comparing curricula. Review of Educational Research, 1976, 44, 83-111.

Wimsatt, W.K., & Beardsley, M.C. The intentional fallacy. In W.K. Wimsatt: The verbal icon. Lexington, KY: University of Kentucky Press, 1954.

APPENDICES

APPENDIX A

GLOSSARY OF TERMS

Asian-African origin

Refers to individuals who immigrated to Israel, or whose parents or grandparents immigrated to Israel, from developing Asian or African countries. About three in five Israeli schoolchildren are of Asian-African origin. The term "Asian-African" excludes Israel and South Africa and thus does not fully correspond to the geographical meaning of the term.

Biblical studies

The study of the Bible constitutes a compulsory school subject in all Jewish schools in Israel, including the nonreligious state schools. Biblical studies are coordinated with language arts and social studies.

Chief Scientist

Most Israeli ministries employ, on a part-time basis, an expert from the academic community who is in charge of coordinating research activities supported by the ministry. The Chief Scientist examines the quality of the studies and formulates operational recommendations for the minister based on the results of the studies.

Cultural origin

The term refers to various cultural traditions of Jews in the Diaspora. In some cases a single country is characterized by a strongly unique cultural tradition (for example, the Yemenites,

or Jews of Kurdistan). But most Jews share the common cultural heritage of a larger region (e.g., Sefardim Jews whose ancestors are of Spanish extract, or Jews of East European countries). The present study distinguishes between the cultures of Oriental Jews (Asian-African origin), Western Jews (European-American origin), and those whose ancestors were born in Eretz Israel.

Curriculum Center of the Ministry of Education and Culture

The Center operates as a department of the Ministry of Education and Culture. A research team of about 80 persons is in charge of producing curriculum materials in all subjects and for all grade levels. The Center was established in 1966. It has produced innovative instructional materials for most school subjects and coordinated curriculum development activities carried out by various academic units within the country.

Disadvantaged schools

Schools in which the majority of students are classified as disadvantaged learners. The status of disadvantaged learner is determined by an index in which a predetermined weight is given to the cultural origin of the parents, the parents' socioeconomic status, and the size of the family. Schools are entitled to receive some financial support from the Ministry of Education and Culture to provide compensatory education to disadvantaged children.

European-American origin

Refers to individuals who immigrated to Israel, or whose parents or grandparents immigrated to Israel, from Europe, America,or Australia.

Generation

In the present study, the term applies to the immigration of the parents. Differentiation is made between first-generation immigrants (children whose parents were born in Israel), second, and third generation immigrants whose grandparents or previous ancestors were born in the land of Israel.

Independent schools

This system of Orthodox religious schools, which accounts for about 5 percent of the Israeli school population, has retained its independence; its program is supervised by a committee which is not controlled by the Ministry of Education, and it maintains an independent system of teacher training institutes.

Integration

The enrollment in a single school of children from both disadvantaged and advantaged groups. Schools in which the disadvantaged learners constitute 40-60 percent of the enrollment are considered integrated schools.

Kibbutz schools

The schools in Kibbutim (collective agricultural settlement) are regular state schools and may belong either to the Religious or to the non-religious State School Subsystems. The Kibbutz

enjoys a relatively high level of independence in shaping its school program. Kibbutz schools are characterized by openness to innovations, by a relatively non-formalized relationship between teachers and learners, and by a relatively high level of autonomy granted to the individual learner to structure his school program.

Knesset

The Israeli Parliament.

Labor Party

One of the large Israeli political parties. From the establishment of Israel in 1947 until 1977, the Labor Party was the leading political party of the country, and the prime minister was a member of that party. Since 1977 the Labor Party has been an opposition party.

National Religious Party

A political party which draws its support from the religious population of Israel. It is a minority party (in various elections, it has accounted for about 10 percent of the vote). The Religious Party participates as a partner of the coalition in most Israeli governments.

National Unity Party

The Likkud, the party which has been in control of the government since 1977.

Nonreligious schools

The largest educational subsystem in Israel, absorbing about 70 percent of the whole student population in Israel.

Pedagogical Secretariat

A committee of the Ministry of Education consisting of its top executive officers, which is in charge of determining and supervising the pedagogical content of the school program and the instructional procedures in classes. The Pedagogical Secretariat operates through two standing Committees, one dealing with primary education and the other with secondary education.

Religious schools

A subsystem of the state school system. In the areas of humanities, traditional Jewish studies, and social studies, the curriculum of the religious schools is parallel to but different from that of the nonreligious schools. In science the program is almost identical with that of the nonreligious schools except that less time is generally devoted to studying science.

SEKER

A multiple-choice omnibus test, which was administered for 16 years (until 1971) in Israeli schools at the terminal grade of the primary school (grade 8). The SEKER examinations served as a selection device for the academic-oriented secondary schools. Pupils who had not passed the SEKER examination were admitted to vocational secondary schools. In 1966 the Israeli parliament approved an educational reform which resulted in the establishment of a junior high school for children in grades 7-9. In this setting, continuous observation of the learner's

progress replaced the one-shot SEKER examination, and in 1971 the SEKER was abolished.

School for Training Educational Senior Staff (STESS)

Maintained by the Ministry of Education, this institute provides continuing education for educational administrators, school principals, and outstanding teachers. Its courses run from a few days to several months.

Szold Institute or Henrieta Szold National Institute for Behavioral Sciences

An independent research institute. It publishes the journal Megamoth (in Hebrew with English abstract) which is Israel's most prestigious journal in educational thought and educational research.

Teachers Union

Two Teachers Unions exist in the country. The majority of the General Teachers Union members are primary school teachers. A group of secondary school teachers also belong to this union. There exists a separate Secondary Schools Teachers Union which admits secondary school teachers only. The General Teachers Union has a long tradition of involvement in shaping the cultural climate of schools in Israel. Until 1952, the year of the establishment of the state school system, it was active in the leadership of the Labor Movement School System.

Van Leer Foundation

An independent institute located in Jerusalem. It supports research related to various social

and cultural issues and sponsors international and national symposia of distinguished scholars and lectures for the public.

APPENDIX B

A COMPARISON OF TWO REPORT SUMMARIES

Summaries of the Van Leer Report were prepared by both the Principal Investigator of the study and an official spokesperson for the Ministry of Education. These summaries differ somewhat in their interpretations of some findings, as the following excerpts show.

Excerpt from the Principal Investigator's Summary

Given the fact that at the primary school there is a high level of ethnic segregation, one may group schools according to their ethnicity characteristics (i.e., which ethnic group constitutes the majority of the student population) and compare the schools on variables of interest. On one hand the Ministry implemented a positive discrimination in favor of learners from Afro-Asian origin, insofar as it allocated special funds for increasing the number of teaching hours in schools according to the proportion of the disadvantaged children in a school. Since 95% of the disadvantaged (as defined by the Ministry) are of Afro-Asian origin, this means that they are the beneficiaries of this arrangement. These special funds for financing additional teaching hours enables the operation of various fostering programs, such as ability grouping, afternoon classes, tutoring groups, etc. But in contrast to these advantages in favor of the Afro-Asian learners, we detected a clear direction of inferiority of the schools of Afro-Asian learners with regard to most other variables, which are crucial in determining the pedagogical standard

of the school and its changes to attain high achievement level. They included the following variables: the educational and professional level of the teaching staff and their work experience, innovative programs, special enrichment programs and pedagogical initiatives of teachers and principals within the framework of regular and extracurricular activities, the educational equipment and its utilization, services, etc. In this respect one should note two issues:

1) Numerous innovations and services, which were originated with the aim of promoting disadvantaged groups of children, are implemented with a higher level of frequency in schools for the advantaged than in schools for the disadvantaged learner.

2) The superiority of the disadvantaged school from the point of view of budget is in practice highly diminishing or even fully disappearing, due to the fact that in the schools of the advantaged parents frequently support innovative programs through voluntary contributions, and also due to the fact that the salaries of the teachers and principals in schools of the advantaged are higher since the basis of the salary is the educational level and the years of working experience of the teacher or of the principal.

Excerpt from the spokesperson's summary

The researchers emphasized that the state of Israel implements a policy of 'positive discrimination' in favor of schools with population of overwhelmingly

Afro-Asian origin, which is realized through increasing the number of hours per week in such schools. Nevertheless despite such allocations, it turns out that in some domains--and mainly in locations far way from cities--some schools of the disadvantaged population are characterized by inferiority in several school variables which appear to be important for improving achievement such as the educational level of teachers, their work experience, the utilization of innovations, special educational programs.

Two issues should be noted here:

1) Many innovations which were initiated by the Ministry of Education for improving the conditions in the schools of the disadvantaged had a better diffusion in regular schools than in the schools of the disadvantaged.

2) These summaries refer to 1973, i.e., before the REVAHA (in Hebrew, "affluence") program started, which had the aim of increasing educational resources in the areas described in this study as inferior.

The reasons for the differences between these two interpretations of the same facts are not difficult to discern. The reader may attribute a higher degree of validity to the Principal Investigator's interpretation on the grounds that he is interpreting his own Report. Similarly, the reader may discount the interpretation of the Ministry spokesperson on the grounds that he did not fairly describe the intention of the authors. Such a claim is possible

only if one compares the two excerpts. But the fairness of the spokesperson's statement should be examined not in light of the Principal Investigator's interpretation but rather, in relation to the vast amount of data contained in the Report.The Principal Investigator does not have an exclusive right to summarize the findings and interpret their meanings. The right of the spokesperson to prepare a digest of the findings is no less legitimate. After all, Plato writes in his Apology that most children in the street understand the poems better than the poets who wrote them. And, in a more modern work, Wimsatt and Beardsley (1954) claim that the author's intent is not relevant in the process of interpreting a work of literature.

There is probably nothing new in the claim that the researchers' views and opinions constitute a subjective component in a scientific study. As the present comparison shows, the subjective component reappears at various points in the conduct of a study. First, it appears in selecting variables, then in operationalizing them. Again a subjective component appears in structuring the data analysis and in determining which data summaries to include. The Van Leer Study produced thousands of pages of data, from which the authors had to select a small subset for the Report. In this respect, the writing of a research report resembles the work of the historian. As Carr (1961) points out in his popular book, the historian decides what is important enough to be recorded, and what is devoid of historical significance (p. 11). The historian is necessarily selective. The belief in a hard core of historical facts that exists objectively and independently of the interpretation of the historian is a fallacy (p.

12). The publication of a research report does not necessarily end the authors' subjective involvement in the study. The formulation of a digest or summary, as a last step of the authors' involvement, completes the series of subjective steps.

APPENDIX C

The following information summarizes ten chapters in A. Minkovich, D. Davis and J. Bashi's **An Evaluation Study of Israeli Elementary Schools**, Published by The Hebrew University of Jerusalem School of Education.

Summary of Chapter 5
Family Characteristics and Background Information

The chapter covered the following areas: the principal dimensions used in presenting results of the study, family and home conditions, family interactions and learning opportunities outside the home and school.

The dominant cultural origin group, AA1, comprises over 50 percent of the sample at all grade levels. Next in size are EA1 and EA2 with about 15 percent each, and smallest are GEN 3+,FaAA, and FaEA (3 percent to 8 percent). AA1 is also the dominant group in older, more settled towns and villages. This is an indication of their integration in all types of settlements. These groups differ greatly in terms of parents' education: the parents of EA children have more formal schooling than do those of AA children; the parents of second generation children have more formal schooling than do the parent of first generation children; and the GEN 3+ and mixed cultural origin groups are about midway between the AA and EA groups on parents' schooling.

The lowest parent education group (MAX 0 and Max 1-4) are almost entirely composed of AA1 children. On the other hand, the highest education group (MAX

13+) contains children from all cultural origin-generation groups.

Differences in conditions and activities in the home are found among the various educational level groups. The lower educational level groups (MAX 0 and MAX 1-4) differ from the other groups in the following respects: more unemployed parents, larger families, greater housing density, poorer learning conditions (e.g., own desk or quiet room), fewer learning aids (e.g., dictionary, encyclopedia, childrens' books), lack of intellectual opportunities for verbal interactions, fewer enrichment activities both in and out of the home, and less encouragement of reading and questioning. In addition, the most advantaged children (MAX 13+) on the average enter compulsory kindergarten with a one year advantage in nursery school attendance over the most disadvantaged children (MAX 0).

The above results present a gloomy picture.The syndrome of educational disadvantage is characterized by a wide variety of problems. Fortunately, some of these are amenable to change (e.g., nursery school attendance, availability of books, and family interest and help in learning). Furthermore, there are two promising aspects of the problem. First, in cases where both parents lack formal schooling a possible agent for home intervention programs is the older sibling, who has surpassed his parents' level of education. Second, the percent of parents in the lowest educational level categories will decrease as more students who have completed at least eight years of compulsory education become parents.

Summary of Chapter 6
School Facilities, Programs, and Staff Allocations

Our major aim in this chapter was to compare advantaged and disadvantaged schools on a large number of school characteristics. The data indicate that of all areas studied, the disadvantaged schools are superior in only two--number of children per class and number of hours devoted to remedial education. The large number of school hours in disadvantaged schools reflects the "positive discrimination" policy of the Ministry of Eduction. However, the following should be noted:

1. One-third of the disadvantaged schools do not take advantage of the extended school day (YULA), a remedial program offered by the Ministry of Education, apparently because appropriate staff and conditions are lacking.
2. Only small groups of disadvantaged pupils, not the entire population of a disadvantaged school, benefit from much of the extra time allotted by the Ministry for such special activities as remedial classes.
3. While disadvantaged schools get far more extra hours from the Ministry for remedial activities, advantaged schools have access to other resources for use in extra-curricular activities and for educational materials.

In certain areas there are little or no differences between disadvantaged and advantaged schools. But in light of the greater and more demanding needs of the disadvantaged schools, they are actually in an

inferior position. The availability of a social worker can be cited as an example.

In most other areas, the disadvantaged schools are inferior to the advantaged. They lack rooms for special services and equipment for various educational activities. Their equipment is less adequate, extracurricular activities are less frequent and less varied, and innovative activities are practically non-existent.

In our discussion, we emphasized the disturbing fact that certain programs intended for advancing the disadvantaged (such as educational guidance and psychological services) are actually concentrated in the advantaged schools and, to a certain extent in the integrated schools, but are rarely to be found in disadvantaged schools.

Summary of Chapter 10
Differences Between Pupil Populations on Ability and Achievement Tests

The relative positions of the cultural groups are similar on all tests of the present study. The lowest position is consistently occupied by first generation Asian-African children (AA1) on all tests at each grade level. Similarly, second generation European-American children (EA2) consistently occupy the top position on all tests and at all grade levels. In the five cultural groups between the two extremes, first generation European-Americans (EA1) and third generation children, who culturally constitute a heterogeneous group (GEN 3+), achieve somewhat higher than second generation Asian-African (AA2) and the two cultural intermarriage groups (FaAA, FaEA).

Except for the nonverbal ability tests on which between group differences are the smallest, very little variation between tests in respect to mean group differences is obtained in the present study. The average mean difference between Asian-African and European-American children is .80 SD. This difference is somewhat smaller than the differences found in Israel in previous studies.

The mean differences between the seven cultural groups reflect the relationship of school achievement with cultural origin and generation in Israel. EA children are superior in their achievement to AA children and in each of these cultural groups second generation children perform better than first generation children. The relationship between cultural origin-generation and achievement prevail even when differences in parents' education are taken into account.

As in former Israeli evaluation studies, the largest differences between student groups are obtained on the geography test and the smallest differences on the Bible test. Differences in mathematics are somewhat smaller than in language. This result is in accord with the results of an IEA data analysis indicating that the student's home characteristics, as opposed to school characteristics, have a greater impact on achievement in language than in mathematics.

Variations in between group differences at the various grade levels are small and in many cases rather inconsistent. A detailed analysis of these variations led us to the conclusion that between group mean differences of the present study do not support the hypothesis of a cumulative deficit of culturally disadvantaged student groups.

Summary of Chapter 11
Results of the Language Tests

In the results of the achievement on the language tests, as reflected by the means (expressed as percentages of correct answers) and by the percentage of pupils who succeeded according to the criteria applied in this study, a rise appears from the first to second grade in the groups classified according to cultural background, or the educational level of their parents (except MAX 0). However, in the fourth grade, most of these groups show a slight drop in achievement, which is then recompensated by the rise in the sixth grade. These fluctuations in achievements are, apparently, in part the result of differences in the test structure and of group compositions at the various grade levels. However, there is reason to assume that these fluctuations also reflect differences in the demands and objectives of the curriculum of each of the grade levels: the rise in the achievement level in the second grade reflects the progress made by most of the children in reading ability (this progress is relatively small in MAX 0 and consequently the achievement level of this group does not rise in the second grade); the decline in achievement in the fourth grade comes from a rather ungradual rise in curriculum demands, as reflected in the learning materials for this grade level. However, the two additional school years enable most of the children who showed a backslide in achievement level to return to their earlier achievement level.

The purpose of the criteria for success used in this study was not only to enable a comparison of achievement between groups of pupils but, also, to

determine the actual achievement level of each group in accordance with the expectations of the curriculum. On the basis of these criteria, we found that a certain percentage of pupils, from all groups and all grade levels, do not attain satisfactory mastery in language skills. This percentage is largest in AA1 where 40 percent of the pupils in all four grade levels did not obtain scores that would be considered satisfactory according to accepted school standards. This percentage is particularly large in MAX 0 (60 percent) which is, in fact, a subgroup of AA1.

As expected, the situation is more satisfactory regarding the minimum demands of the curriculum. Except for AA1, all cultural background groups achieve almost complete mastery at every grade level. In AA1, however, the percentage of pupils who do not attain adequate achievement in even the minimal demands of the program, is relatively high (15%). The percentage is even higher in the MAX 0 subgroup (30% in first and second grades and 25% in grades four and six). It is reasonable to assume a certain bias of results due to errors of measurement. Nonetheless, in our opinion, these results do demonstrate an adequate approximation to reality and, therefore, merit attention by the leadership in the educational system.

The results for the anchor items demonstrated significant progress on the part of all the groups in the material of prior grade levels. In the groups of intermediate and low achievements, this relative progress was the highest in second grade and lowest in the sixth grade.

The problems of gap size between pupil groups in language and of the stability of the gaps across grade levels was examined by several modes of

comparison. On the comparison of means expressed as standard scores we found a difference of about .80 SD between AA and EA. When the means were expressed as percentages of correct answers on the tests we found that AA pupils lag behind EA pupils on 20% of the subject matter studied. With the comparisons based on percentages of pupils succeeding according to the lenient criterion, the difference between AA and EA is 30% on the total tests and 15% on the minimum items.

For most types of comparison, we did not obtain consistent and convincing evidence of increases in gap size between AA and EA. On the other hand, by comparing the final achievement levels on the anchor series of low achievement groups in higher grade levels with the initial achievement levels of high achievement groups in lower grade levels, we found indications of gap increases during elementary school.

Summary of Chapter 12
Results of the Mathematics Tests

Results of the mathematics tests for all groups at the four grade levels were presented in this chapter. Results included means, expressed as percents of correct answers in tests, and percent of pupils who passed according to each of the two criteria (at least either 60% or 70% of test items), for total tests, and for minimum and anchor items. Analyses and comparisons were done in the following topics:

A. Achievement at the four grade levels: We found a gradual increase in achievement level from first to fourth grade, and a

sharp drop in sixth grade. A more detailed analysis of sixth grade results revealed that on the section of the test covering problems and operations on whole numbers the progress found in earlier years continued. The decrease was evident with regard to simple fractions, and decimals, an area apparently more difficult to understand and master than whole numbers.

B. Mathematics achievement versus language achievement: Mathematics achievement was lower than language achievement in all groups at all grade levels (on total tests and on subtests). Achievement differences between these two curricular areas was most evident in low achievement groups (AA1, in the cultural origin groups and MAX 0, and MAX (1-4) in the parents' educational categories).

C. Achievement of the different groups on total tests: AA1 achieved the lowest, especially in sixth grade, where only 36% of this group passed according to the lenient criterion. The achievement level was high in the EA2 group in which 80% of the pupils attained this required mastery. In the other groups only 53% to 66% of the pupils passed. Poor achievement was especially marked in the MAX 0 group in which only 20% on the average passed according to the lenient criterion, and in the sixth grade this reached only 12%.

D. Achievement on minimum items: Low and medium achievement groups on the total tests showed a high level of achievement on the minimum items, but this was somewhat lower than their achievement on the minimum language items.AA1 groups, especially, scored lower on mathematics than on language minimum items (84% passed in language vs. 72% in mathematics). The

difference was somewhat larger in MAX 0 which is a subgroup of AA1 (71% passed in language and only 62% in mathematics).

E. Achievement on anchor items: At every grade level, progress on items from earlier grade levels was evident. Progress, defined as learning growth estimate, was relatively greater in the lower then the higher grades, and in the lower achieving than the higher achieving groups.

These results reflect the differences in the initial achievement levels: the higher a group's (defined by cultural background or parents' education) initial achievement level in a lower grade, the smaller the increment of pupils succeeding in the next grade, even if all pupils succeed in the anchor items. When we calculated the growth index of each group in relation to the percentages of failures on the initial achievement level (the percentage of increment at the final achievement level calculated from the portion of those who did not meet the criterion on the initial level), we found that the growth rates were larger for the high achievement groups than for the low achievement groups.

F. Size and stability of gaps between pupil groups: Examining differences between groups using means and percentages of success according to defined criteria led us to conclude that there are no consistent and significant variations in the gaps between student groups at different grade levels. Comparing means expressed as standard scores we found that the difference between AA and EA was .75 SD. With means expressed as percentages of correct answers we found a gap of 15% between these groups. Using percentages of success according to our lenient criterion we found a gap of 30% on the total tests and 20% on the minimum items.

Evidence of widening gaps between student groups appeared on the anchor items. Comparing the final achievement level of the AA groups (mainly AA1) in higher grades to the final achievement levels of the EA groups (mainly EA2) in lower grades, we found a lag of one school year between them in second grade that increased to two or more years in the fourth and sixth grades.

Summary of Chapter 13
Results of the Bible Tests

Only the fourth and sixth grades of the research sample were tested in Bible. In the fourth grade sample a single test was given, whereas in the sixth grade sample, pupils of the State schools and of the State Religious schools were examined in separate tests. The achievement level in Bible in the total grade samples and in the various groups of pupils was found to be relatively low in comparison with that of the language tests. This was reflected in the means and success percentages according to the criteria of mastery, on total tests and minimum items, as well as in the growth rate in the anchor items. This finding may be explained by the fact that the demands on the pupil were higher in the Bible than in the language tests. Success in the Bible tests required skills and abilities necessary for succeeding in the language tests, but also put additional demands on the student such as the recall of specific contents, the comprehension of words, expressions, and grammatical forms not used in modern spoken or written Hebrew, as well as the understanding of more complex concepts than those which appeared in the language tests.

However, the achievement level in Bible was also lower than achievement in mathematics. This result seems extremely puzzling, as more instructional time is usually devoted to Bible in Israeli elementary grades than to any other subject. The results on our tests seem to reveal two weak points in the Bible curriculum for the elementary grades:

1. The overloading of study materials for which the number of teaching hours, although greater than for other subjects, is apparently insufficient for mastery learning.
2. The level of difficulty and complexity of some parts of this material seems to cause considerable learning difficulties for elementary school children and especially, for disadvantaged students.

Between group mean differences in the Bible tests were smaller than on all the other tests of our study. A similar finding was reported some years ago in the analysis of the SEKER results (Ortar, 1967): mean differences between AA and EA in this analysis were smaller in Bible than in the other subjects. The interpretation offered for this was that the relative achievement level in Bible is higher for AA than in the other subjects due to the extensive time devoted to Bible studies in the elementary grades. In addition, the traditional atmosphere prevailing in many of the AA homes helps to instill a positive attitude in the children towards the subject.

In our research, which undertook to examine not only mean differences, but also differences in achievement level according to criteria of mastery, we found that this explanation, with regard to the

State schools, needs an important modification. The achievement level in Bible, relative to that in language and mathematics, is lower in the secular schools in all groups of pupils. However, this relative inferiority is smaller for AA pupils than for EA pupils. (For example, in the fourth grade, 46% of the pupils in AA1 succeeded in mathematics by the lenient criterion, as compared with 41% in Bible, whereas in EA1, the success percentages were 75% and 64% respectively.) It seems then, that the reduction of differences in Bible achievement between AA and EA, as compared with differences in language or mathematics, reflects more a low achievement of EA1 in Bible than a high achievement level of AA1. In other words, the additional time devoted to the Bible studies in the State schools appears to have a slight positive effect on the achievement level of AA, while no effect is observed in the achievement of EA.

The picture is different regarding the State Religious schools. Here, the mastery level in Bible is only slightly lower than that of language and higher than the mastery level in all the other subjects. The relatively high achievement level of the religious schools in the Bible tests is also reflected in the differences between the two school systems. The differences between the two types of schools in Bible, as compared with the differences in the other subjects, decrease in the fourth grade, and completely disappear in the sixth grade when the particular difficulty encountered by State Religious school pupils in the study of Deuteronomy is taken into account. Considering the differences in the composition of the student population of both types of schools, the relative achievement level in Bible can be viewed as being higher in the State Religious

schools than in the State schools. This is, apparently, a combined result of the increased time devoted to the subject and the highly positive attitudes of teachers and parents of religious school children towards the Bible.

In contrast to the state schools and, due apparently to the influence of the two factors mentioned above, the relative achievement level in Bible is higher for all of the groups of religious school pupils. The narrowing of the differences between AA and EA in the State Religious schools occurs, therefore, not because of the relative drop in achievement level of EA but, rather, because of the relative rise in the achievement level of AA. By way of summary, the following picture emerged from the results of the Bible test:

A. Comparison with the language and mathematics tests: In the fourth and sixth grades of both secular and religious school streams, the achievement level in the Bible test is lower than in the language tests, both in terms of the total test and the minimum items. In the secular State schools, the Bible achievement level of the fourth grade is lower than the achievement level in the mathematics test as well, whereas, in grade six this is true regarding whole numbers only, and does not pertain to the results in fractions (in which the achievement level is particularly low at this grade level). In the State Religious schools, on the other hand, the successes in the Bible test are higher than the achievement level in the mathematics tests at both grade levels.

B. Comparison of the schools in the two educational streams: In the State religious schools the achievements in the Bible test are lower at both

grade levels than those in the secular State schools; however these differences are smaller than the differences, in the language and mathematics tests, and disappear completely in the minimum items. Moreover, when the comparison in the sixth grade is made on the contents common to both Bible tests, no difference is evident between the two types of schools in the average achievement level. In fact, we find a higher achievement level in these contents in the State Religious schools, when the comparison is made between similar groups in terms of cultural background or educational level of parents (for example, in the State Religious schools, 64% of the AA1 pupils attained success according to the lenient criterion, as compared with 50% in the secular State schools).

C. Comparison between the cultural background groups: The relative position of each cultural background group in the Bible tests is generally not different than in the language and mathematics tests. However, the differences in average achievement between the groups are smaller than in the latter tests. In the State Religious schools a reduction of differences is caused mainly by the relatively high rise in the Bible achievement of the AA groups, whereas, in the secular State schools the differences are reduced as a combined result of a certain rise in the achievement of AA and the drop in the achievement level (relative to other tests) of EA. The comparisons based on data from the Bible tests indicated gaps of the following sizes between the AA and EA groups:

1. The average differences between means presented as standard scores reached .60 SD.

2. Differences between means expressed as percentages of correct responses on tests reached more than 10%. The interpretation given to this finding was that AA students lag behind EA students on an average of 10% on the subject matter studied in class.
3. The differences between the AA and EA groups in respect to the percentages which attained mastery learning in Bible according to the criteria employed in our study, reached an average of 25%.

The problem of a growing gap between the groups of pupils was investigated in the results of the Bible tests. The results here were similar to those of the language test, insofar as no consistent evidence was found for the widening of the gap between the different groups in the comparisons of means of percent of pupils who succeeded according to the defined criteria. The comparison of achievement levels in contents learned in previous years revealed that the lower the SES level of the group, the slower its rate of mastery learning in these contents. However, since the first and second grades were not tested in Bible, it is impossible to establish whether there exists a trend towards an increase in this lag, as has been noted in the language and math tests.

Summary of Chapter 14
Results of the Geography and Science Tests

A. Achievement in the Geography Tests

The tests in geography examined in grades four and six the pupils' mastery in three areas: ecologi-

cal, social and economic geography; descriptive, physical geography; and map reading skills. The achievement levels in the total test and in the minimum items were slightly higher in the sixth than in the fourth grade. In general, the achievement level for geography in both grades was considerably lower than the respective achievement levels in the language and Bible tests. In the fourth grade this situation was also true regarding achievement in mathematics, while in the sixth grade, no substantial difference was found between results in these subjects. This equivalence mainly reflects the relatively low achievement of sixth grade pupils in the mathematics test, especially in decimals and fractions.

According to the grading system used in the schools, the average achievement in the geography test at both grade levels is between "pass" and "barely pass." Only in EA2 does the average achievement come close to a grade of "fair." This is moreover, the only group that could be considered to have learned the subject to the point of mastery, according to the criteria applied by this study (the success of 50% of pupils according to the stringent, and of another 25% according to the more lenient criterion).

Achievement in geography, as compared with achievement in the other subjects, is also lower on the basic curriculum requirements as represented by the minimum items: 15% in EA1 and 40% in AA1 did not arrive at a satisfactory achievement level in the basic contents of the subject, even according to the lenient criterion.

As expected, progress was observed in the sixth grade on the anchor items common to the tests for

both grade levels. However, in contrast to the fourth-sixth grade anchor series on the other subject matter tests, the growth rates on the geography anchors (differences between initial and final achievement levels) are smaller for the low achievement group than for the high achievement groups.

The breakdown of the tests into the three topics mentioned above showed that the weak point in the fourth grade is map reading skills.In the two other areas, achievement at this grade level is not significantly lower than in the other subjects. In the sixth grade, as stated, a notable improvement occurs in map reading in comparison to the fourth grade. However, even at this grade level the mastery of this skill is still relatively low (60% of the pupils do not attain a level that could be described as "passing").

Surprisingly, we found that in comparison with the fourth grade, the sixth grade mastery level was lower in the area of ecological, social and economic geography, as well as in the area of descriptive physical geography. One of the reasons for this, apparently, is the broad scope of the curriculum requirements in the sixth grade relative to the number of instructional hours devoted to the subject in the fifth and sixth grades.

When means are presented as standard scores the differences between pupil groups on the geography tests are larger than on the other tests.The average difference between AA and EA reaches about .85 SD. When means are expressed as percentages of correct answers, the differences between EA and AA is over 15%, and 30% between the two extremes of parents' education categories (MAX 0 and MAX 13+). Comparisons based on percentages of success reveal a

difference of 35% between EA and AA on the total test and 25% on the minimum items. The differences between MAX 9 and MAX 13+ on these comparisons reach 50% and 45% respectively.

The fluctuations of these differences between grades are not sufficiently large or consistent to indicate a clear trend toward the widening or narrowing of the gap between different groups.

On the comparisons made with the geography anchors between the final achievement levels (grade six) of low SES groups and the initial achievement levels (grade four) of high SES groups, the results are similar to those found on the other tests: a lag of several school years between the types of pupil groups. The size of the time is seen to be a function of the size of the socio-economic distance between the groups compared. In the absence of data for the lower grades, we cannot determine if this gap widens over the years in geography, as it does in language and mathematics.

B. Results in Science Tests

Three topics were included in the fourth grade science test: zoology, botany and general organic and inorganic phenomena.

The achievement level in the fourth grade was relatively higher than in the sixth grade. This was reflected both in the means and in the percentage of pupils who passed the tests according to the two criteria of success. Achievement in the sixth grade science test was the lowest, in comparison with results of this grade level in other tests. In the fourth grade science tests, on the other hand, most of the pupils attained an achievement level which was

similar to that found in the Bible tests and higher than the level attained in geography.

A breakdown of the tests according to topics revealed that the fourth grade achievement was highest in zoology and lowest in botany. In the sixth grade, achievement was particularly low in anatomy.

Based on the grading system used in the schools, the average achievement in science for the fourth grade was "pass," and in the sixth grade, "unsatisfactory." The explanation offered for the differences between the two grades in the achievement levels for science was that the knowledge acquired at the end of the fourth grade is the product of learning accumulated over three-four years. In the sixth grade, however, the content taught in science is new and more specific, so that the material learned in previous grades can be of little assistance. In addition, it is possible that the learning program in anatomy and in inorganic nature is more overloaded than other subjects and not sufficiently based on observations and experimentation.

In preparing the science tests, we had difficulty defining which content and skills should be viewed as representing the minimal demands of the program. Accordingly, it was difficult to interpret the results in the minimum items of the science tests. However, since these minimum items were rather easy, in the opinion of experts, the results obtained from these items strengthen conclusions based on the results found in the total tests. These findings seem to verify that achievement in science is lower than achievement in all other subjects at both grade levels. (In the fourth grade, as mentioned, the only area in which the achievement level was even lower was map reading in geography.)

We found in the anchor items that the growth rate was lowest in comparison with anchor series for grades four-six in the other subjects. These results reflect, even more clearly than is observed in the total tests, the fact that in the sixth grade science program there exists less overlap or connection with material learned in the past and, therefore, less mutual reinforcement between the old and new study contents. In the areas where achievement levels of the various groups were compared the following gaps appear between EA and AA: a dfference of over .80 SD between means presented as standard scores; a difference of 15% between means expressed as percents of correct answers; and a difference of 25% on the total test and 15% on the minimum items in respect to the percentages of pupils succeeding according to the lenient criterion. Like the fourth-sixth grade anchor series in the other subjects, we found a two year lag regarding mastery in content from previous years of study between the groups of varying SES level (essentially between AA1 and the EA groups).

We certainly cannot determine from the results in the science tests, or from the tests in Bible and geography, which were also not given in the lower grades of the sample, whether this lag is greater in the sixth grade than in the lower grades. However, the size of the lag in terms of years of study for the fourth-sixth grade anchor series, is similar to that found in all tests. Since in the anchor series of the language and arithmetic tests, which were given to the first and second grades as well, a gradual growth of the lag can be observed from the second to the sixth grade, it is possible to conclude by way of extrapolation that this tendency may be characteristic of all subjects taught in the elementary schools.

Summary of Chapter 15
Levels of Achievement in the Various Types of Elementary Schools in Israel

In this chapter the achievement levels of different types of schools were compared. In the first part, Jewish schools were compared according to the percentage of disadvantaged pupils and according to whether they belonged to the State or State Religious systems.

15.5.1 Advantaged, Integrated and Disadvantaged Schools

The purpose of comparing schools according to the percent of disadvantaged pupils was to examine the effectiveness and validity of the definition of the disadvantaged pupil adopted by the Ministry of Education in 1974 in order to rank schools according to their eligibility for special assistance. Thus, on the basis of data provided by the Ministry, the schools in our sample were classified as advantaged (24% disadvantaged pupils or less), integrated (25%-75%), or disadvantaged (76% or more).

This classification was found to have substantial predictive power: the correlations with average achievement levels on the mathematics and reading tests reached an average correlation of .70. The close connection between percentages of disadvantaged pupils and achievement levels was also presented by mean differences between the three schools categories in terms of standard deviations. Achievement in disadvantaged schools was lower than that in advantaged schools by one standard deviation. The correlations and mean differences pointed, then, to the usefulness of the definition of the disadvantaged pupil based on cultural background, parents'

educational level, and number of children in the family.

The Ministry's definition of certain groups of pupils as advantaged is based on an analysis of the SEKER of 1972 that divided the eighth grade population into various socio-cultural groups. A group was defined as disadvantaged if less than 41% of the pupils achieved a score of 70 on the SEKER, and the severity of the designation was determined in inverse proportion to the number that failed to reach this level. An analysis of the results of the achievement tests of this study according to similar criteria, and for all three school types, produced similar results.

When comparison between the SEKER analysis and results on our achievement tests was based on a score of 70 on our tests (according to the accepted standards in the school that view 70 as a grade of "fair"), we found that on the average in the advantaged schools 58% of the pupils examined achieved at least 70 while only 18% did so in the disadvantaged schools. But even when the comparison was based on a score of 60 on our achievement tests (close to the median score, as was 70 for the SEKER), we found that 31% of the pupils in disadvantaged schools scored 60 or more while 73% did so in the advantaged schools.

However, when each subject was analyzed separately, we found that more than 40% of the pupils in disadvantaged schools achieved a score of 60 or more on the reading tests (on the average, 45% of the pupils in disadvantaged schools scored 60 or higher on the reading tests, 64% in the integrated schools, and 75% in the advantaged schools.)

In comparing our results to those of the SEKER for various student groups cutting across school

subjects, we found that the achievement level of AA1 was higher on our tests with 60 (the median) as the criterion, but lower with 70 ("fair") as the criterion. The EA1 achievement level was the same on both studies using the median score in our tests but lower using the criterion of "fair" as the normative grade (a score of 70). The present study has produced convincing empirical evidence of the usefulness of the Ministry's 1974 definition of groups of disadvantaged pupils. However, this empirical confirmation seems to us to be a source of great concern, since our data indicate that in the nine years between the time the eighth graders tested on the 1972 SEKER started school (in 1966), and the time (1973) when our tests were administered, no significant improvements have occurred. This is despite changes which were designed to diversify and improve compensatory programs and extend their benefits to younger ages.

Our results also point to two lines of potential revision in the definition of the disadvantaged pupil:

1. Differential diagnosis according to different subjects.
2. Distinction between two levels of severity when dealing with the disadvantaged based on differences in achievement levels according to optimal and minimal requirements of the curriculum.

15.5.2 The State Religious Schools

Disadvantaged pupils comprise 73% of all pupils in the State Religious schools and only 33% in the State schools, which leads us to expect a lower level of achievement in the State Religious schools.

In our study, however, we hypothesized that the size of the differences on the various tests will be a function of the differences in the amount of time (measured by the average number of weekly hours of instruction) devoted to teaching the different subjects. As a frame of reference for examining this hypothesis, we used the size of the differences on the language tests, since both school systems devote the same amount of time to teaching this subject.

We found that the means on the language tests in State Religious schools are lower, on the average, by .40 SD than those in State schools. The only exception was in the first grade where achievement in the religious schools was relatively high in reading due, no doubt, to the special efforts made in teaching the pupils to read the prayer book. In comparing the percentages of pupils who succeeded on at least 60% of the items on the reading tests, we found that the ratio between the school systems was 100:75 (that is, for every 100 pupils scoring 60% or higher in the State schools only 75 did so in the State Religious schools).

State Religious schools devote on the average 1.5 more weekly hours to Bible studies than the State schools. Parallel to this we found only a small difference on the Bible tests in achievement levels for grades four (about .20 SD) and no difference for grade six. When the two school systems were compared according to groups of pupils from similar cultural background or SES, we found that in the religious schools achievement levels on the Bible tests were higher for both AA groups and for the lower parents' educational level categories. The most reasonable explanation for the findings on the Bible tests is that the extra hours of instruction in Bible offset

the expected differences due to socio-cultural factors. It is also likely that the particularly positive attitude toward the Bible fostered in the State Religious schools and in these students' homes plays a role in compensating for socio-cultural differences.

Due to the extra hours devoted to teaching Bible (and other religious subjects not examined in the study), less time is devoted in the State Religious schools to mathematics, science and geography. The difference in instruction time between the school systems in regard to mathematics is small (0.2 weekly hours) but some impact is, nevertheless, apparent on achievement. Mean differences are somewhat larger than for the language tests, and the ratio of successes according to the lenient criterion is 100:65, compared to 100:75 for reading.

The difference between the two systems in time devoted to science and geography is larger (1.6 weekly hours), and, as a result, the greatest differences in achievement levels appear for these two subjects. Mean differences reach an average size of .50 SD and the ratio of successes according to the lenient criterion is 100:50.

The importance of this analysis lies in the fact that it confirms the positive correlation between time devoted to teaching a subject and achievement level in it. (Similar evidence for this relationship was found for absences from school and is presented in Chapter Sixteen.)

15.5.3 Elementary Education in the Arab and Jewish School Systems

An evaluation study of elementary education in the Arab school network was carried out parallel to

the present study in the Jewish schools.However, in the Arab sample were included only grades four, six and eight. Due to difference in language of instruction and curriculum, the achievement test of the Arab students differed in language and content, which in turn made it difficult to compare the results for the two school networks.

Nevertheless, the comparisons carried out of achievement levels in grades four and six for reading, mathematics, science, and geography did yield several meaningful results.

The reading comprehension tests given in the language spoken and used for instruction (Hebrew in the Jewish schools, Arabic in the Arab schools) were based on the curriculum requirements. Therefore, we felt that the tests were comparable in this case despite the differences in language and content. The results indicated that the average reading achievement level for grades four and six in the Arab schools is similar to that in the Jewish schools for these grades. In comparing the Arab schools to the two AA groups it was found that the average reading achievement level of the Arab sector was somewhat higher than that of AA1, but lower than that of AA2.

Two-thirds of the items on both mathematics tests given in the Arab schools were taken from the tests for Jewish pupils. A comparison of these anchor items shows that Arab achievement level in grades four and six is lower than that for Jewish students. The average Arab achievement was even somewhat lower than that for AA1 pupils.

Comparing achievement in geography and science was more problematic. Due to different emphases in the curriculum (mainly in geography) in each network and because many items on the tests for the Jewish

schools were shown in a pilot study to be too difficult for many Arab pupils, it was possible to include only a small number of anchor items on the tests for the Arab schools. The results indicate that these anchor items were the most difficult ones for the Arab pupils. Since achievement on these items was considerably lower in Arab as compared to Jewish schools, one may conclude that the overall achievement level in the Arab school network is lower than in the Jewish schools.

In summarizing these comparisons we may say that the level of reading comprehension in the Arab schools is certainly satisfactory. The achievement level is lower for mathematics and lowest for science and geography. Given that the achievement levels in the Jewish schools for science and geography were quite low, it is clear that the situation in these subjects in the Arab schools is highly unsatisfactory.

Summary of Chapter 16
Variations in Students' Achievements and Their Concomitants

In correspondence with other studies, we found that most of the variation in achievement is within classes. However, we demonstrate that the obtained between school variation is indicative of large and important inequalities in school achievement outcomes. These inequalities are the same at all grade levels, as well as when calculated separately for the Asian-African and for the European-American subsamples. However, the inequality is slightly smaller for reading comprehension in comparison to mathematics, geography and science.

Most of the variations in achievements among classes and schools can be explained by variations in a small group of student body, class and school characteristics. However, only a small part of the explained variation can be attributed uniquely to student body characteristics. This is a result of the fact that the distribution of school resources is related to variations in student body characteristics.

If we look at variations in individual achievement on reading and mathematics within classes we see that only a small part is explained by variations in the 11 basic home background variables. However, reading comprehension is influenced more than is mathematics, a reflection of the verbal nature of most interactions in the home. These interactions are more likely to influence language ability than knowledge in mathematics. There is no difference between Asian-African and European-American children in their sensitivity to home and school characteristics.

We also examined the effect on achievement of other individual variables after controlling for home background characteristics. The first of these, number of absences from school, is a potent predictor of achievement at grade six but not at grade four. This may be due to greater complexity of content at the upper grades or to the fact that absences at the upper grades reflect motivation. However, the fact that number of absences has a stronger relationship with reading than with mathematics is so unexpected that we suggest further investigation before drawing strong conclusions.

As found in other studies, the family verbal interactions variable is strongly related to achieve-

ment after controlling for static home background variables. In spite of the problem of attributing causality on the basis of this result, we feel that home intervention programs which endeavor to manipulate this variable should be considered.

If we accept the view that locus of control, anxiety, self image and satisfaction with the class are, at least partially, inputs to the process of learning, then our results indicate that it would be worthwhile to attempt to manipulate these variables to improve achievement.

Many findings of survey studies in education do not allow clear cut conclusions for policy decisions. This, in large, is due to the complex interrelationships among many variables examined in these studies. Regression analysis was intended to cope with this problem. However, as we have emphasized in this chapter, regression analysis cannot overcome the difficulties in survey studies due to their post hoc nature, the lack of theory to reduce the mass of interrelated variables to a mangeable, meaningful set, and the global nature of the variables examined.

After his extensive reanalysis of the EEOS data (Coleman, 1966), Smith has come to the following conclusions in regard to survey studies: "Until we adopt the experimental model, we will continue to flounder in the swamp of uncontrolled plausible hypotheses." (Smith, 1972, p. 316). Although many problems exist in experimental studies as well, we concur with Smith's view that experimental studies enable us to arrive at more clear cut conclusions than do survey studies. However, we do not accept the extreme position of rejecting the importance of survey studies. An extensive survey can reflect various aspects of reality in education which cannot

be reflected by experimental studies, and such reflections can be very useful to policy makers. In addition, the descriptive components of a survey study and the "uncontrolled plausible hypotheses" derived from the analysis of these components can be very important as baseline data and guidelines in the development and interpretation of careful experimental studies.

A promising approach to evaluation research in education is the integration of the extensive survey study with a series of intensive and well controlled observational and experimental studies.

Summary of Chapter 17
Cultural Integration in Schools

The school integration of the various cultural origin-generation groups has been progressing naturally as a result of the social and economic mobility of AA families. To a certain degree, the educational policy of many municipalities has accelerated the process. However, the Ministry of Education has not espoused any active, systematic policy on integration in the elementary schools. An analysis of the seven cultural origin-generation groups in our study reveals an inverse relationship between the extent of integration and SES among AA (versus EA). We find that greater extent of integration in classrooms is correlated with higher SES of AA pupils and lower SES of EA pupils.

Second generation AA children study in culturally mixed classrooms more often than first generation AA children while the opposite is true for EA children. Due to differences in demographic stability,

school integration is more extensive in the new towns and large cities than in the small farming communities and the older, established city neighborhoods.

An examination of the relationship between integration and achievement demonstrates that the percentage of EA students in the classroom has virtually no predictive power when other classroom composition variables, such as parents' educational level, family size, and father's occupation, are controlled. These results corroborate the findings of similar studies conducted in Israel and abroad.

In the light of these findings, decisions regarding integration should be based on social criteria rather than on the expectation that integration will affect achievement. Nonetheless, it is possible that planned school integration, accompanied by special educational activity, can facilitate the progress of disadvantaged students.

The analysis of teachers' views on integration indicates that the majority of teachers favor integrated classes at the elementary school level and believe that integration will raise the level of achievement of the disadvantaged students while improving social relations among the various cultural groups. At the same time, however, many teachers express concern that the achievement level of the more gifted students will be lowered and that teachers will be unable to adapt their teaching methods in order to effectively accommodate the needs of children in a very heterogeneous classroom.

In our comparison of disadvantaged, integrated, and advantaged schools teachers and principals in the integrated schools are more likely to report problems related to student behavior and maintenance of

facilities. However, judging by the input invested in integrated schools, it appears that neither the Ministry of Education, the municipalities, nor the parents consider integration to be a special educational enterprise warranting support and special aid.

APPENDIX D

THE DEBATE IN THE PEDAGOGICAL SECRETARIAT

The Pedagogical Secretariat makes decisions on pedagogical issues for the Ministry of Education. Its members are the District Directors of the Ministry and the heads of the major pedagogical departments. Within its framework two standing committees operate, one for Primary and one for Secondary Education. The meeting which took place on March 4, 1979, was attended by the members of the Ad Hoc Committee in charge of formulating the recommendations.[1] Fourteen persons actively participated in the discussion, which lasted more than three hours. Their comments supported or opposed one or another point(s) on the list of recommendations. If one were to be given a matching test with a job description of the participants in one column and the speeches of the participants in the other column, one could easily guess which speech was delivered by the head of the Curriculum Center, which by the Chief Supervisor of Science, and so forth. Although many participants were well versed in the findings of the Van Leer Report, and ready to quote from it to support their arguments, in most cases these arguments were not generated by the findings of the study. A comparison of the proceedings of this session with the proceedings of 1978 meetings reveals that, over the interval, participants had familiarized themselves with the details of the study and could therefore

---

[1] The Pedagogical Secretariat, Plenary session, 1976/6, Minutes, Ministry of Education and Culture, 31 pp. mimeogr.

support their point in more sophisticated ways (e.g., people spoke about partial correlations rather than about zero-order correlations).

A broad range of basic pedagogical questions were raised in the discussion. The major topics were: integration, the relationship between time and achievement, and the desirability of implementing a "minimum requirement" program.A recurrent theme was the problems of educators responsible for running the schools.

Some of the comments made at the discussion merit mention because they shed light upon the interaction of decisionmakers with researchers. One of the participants expressed his reluctance to accept the finding that the only school characteristic significantly related to student achievement is the teacher's belief in the educability of the disadvantaged child. (Despite the sophisticated level of the discussion, the question of the direction of the effect was not raised. People assumed that this finding meant that attitude affects achievement.) According to this participant, such a finding may be interpreted to mean that less emphasis should be given to teacher training, a recommendation which cannot be accepted. It may be that a different kind of training is needed, but what kind? The system has experimented for years to improve teacher training, and the major message of empirical studies has been that the methods employed have not produced the desired changes. But no alternative ways have been suggested. Thus, such a finding is of little value to the planner (Pedagogical Secretariat, 1979). Another speaker supported his suggestion by quoting a recommendation from the Report. When he was told by his opponents that the Report's recommendations were

not substantiated by the reported findings, he replied: "Suppose that it is not anchored in findings; the recommendation itself still exists" (Pedagogical Secretariat, 1979, p. 30).

These two examples demonstrate that policymakers frequently seek support for their own ideas, rather than advice. The more complex the findings, the more likely it is that they will be used to support an already existing opinion rather than to change the mind of the decisionmaker. This is so because the suggestion contained in a complex set of findings does not clearly point to a direction for action. Indeed, in the discussion of the Pedagogical Secretariat, the same statement from the Van Leer Report was sometimes used to support contradictory suggestions for action.