FOUCAULT, POWER/KNOWLEDGE,
AND
THE RECENT LITERATURE ON SCHOOL IMPROVEMENT

By
MURRAY ROSS

B.A., The University of British Columbia, 1973

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in
THE FACULTY OF GRADUATE STUDIES
(Deptartment of Social and Educational Studies)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
November 1987
© Murray D. Ross, 1987
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Social and Educational Studies

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date October 7, 1987
This thesis examines some of the recent literature on school improvement: namely, teacher effectiveness research, school effectiveness research, and four of the commissioned reports on education that were published in the United States during 1983. The analysis relies on a number of central concepts from the social theory of Michel Foucault, in particular his notions of power/knowledge and discipline. It is argued that these bodies of educational research are in themselves either inadequate or inappropriately employed in policy discussions, and that as a result the manipulation of students and teachers seems reasonable and necessary. It is further argued that the teaching practices and educational policies called for in this research are likely to produce unintended, and undesirable consequences which are completely at odds with the stated goals of school improvement.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>A presentation of the research problem</td>
<td>1</td>
</tr>
<tr>
<td>Two</td>
<td>A discussion of Foucault's theories on power, knowledge and discipline</td>
<td>7</td>
</tr>
<tr>
<td>Three</td>
<td>A discussion of teacher effectiveness research with a preliminary sketch of the utility of applying Foucauldian theory to it</td>
<td>32</td>
</tr>
<tr>
<td>Four</td>
<td>A discussion of school effectiveness research with a preliminary sketch of the utility of applying Foucauldian theory to it</td>
<td>87</td>
</tr>
<tr>
<td>Five</td>
<td>A discussion of the recent commissioned reports on American education</td>
<td>128</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

One of the most puzzling features of much of the educational research over the last 25 years is the manner in which students have been characterized in the research literature. Typically, students are depicted as being the passive recipients of one or another of various educational "treatments" or "inputs", or as being subject to the influence of a host of stimuli that will, it is hoped, produce the desired educational "outcomes".

Various critics of this conception of the student have not failed to notice its peculiarity. John Meyer, for instance, has observed that in the prevailing research traditions "the student is seen as a mildly intelligent monkey (or occupant of a monkey-like role), constrained by the immediate distribution of rewards." (Meyer, 1980, p. 25) In Meyer's view, "too much research looks at the technology of teaching as if it were the mechanical action of a person on an object." (Meyer, 1980, p. 53) In a similar fashion, Alan Tom (1984) has seen a connection between just this sort of mechanistic conception of the student and the failure of process-product research to yield significant findings.

Essentially, in Tom's view, process-product research is flawed because it conceives of learning solely in terms of what teachers do. In effect, teaching "behaviours" or processes are said to lead directly to student "outcomes" or products, hence the category, process-product research. Teachers, according to
this view, cause learning. What is so striking about this conception of teaching and learning, aside from the fact it has failed to produce truly effective teaching methods, is that in such a conception, the student remains passively subject to whatever the teacher does to induce learning. Tom calls this conception of teaching the "billiard ball hypothesis", where the teaching/learning relationship is understood in terms of an analogy between teaching and shooting a ball in billiards. The pool player (teacher) aims the cue ball (his behaviour) so that it will strike the target billiard ball (the student) at exactly the right angle to cause the billiard ball to go into the pocket (the achievement of what the student is supposed to learn). (Tom, 1984, p.82)

Similar to Tom's objection, is Richard Murnane's. In his review of school effectiveness research, Murnane (1981) points to the peculiar methodological basis of this research tradition. An underlying assumption of school effectiveness research is that teaching and learning can be viewed as a stable, well-defined production process, similar to that of growing hybrid corn. In fact, according to Murnane, the production of hybrid corn is the substantive area from which this research tradition stems. As he notes, the chief difference between corn production and education is that "in corn production, the key inputs, seed, water, and fertilizer, are inanimate and their productivity depends only on the resource mix and the weather...[while] in education, the key resources are students and teachers." (Murnane, 1981, p.13) Ironically enough Murnane
himself is not free of the language of "people as resources".

Aside from investigations with a narrowly pedagogical focus there is a significant body of educational research devoted to broad social issues like the need for equity and equal educational opportunity, or the matter of education and the national interest. Educational philosopher, Kenneth Strike, has observed that in America, schools are seen as "the basic social institution in which opportunities to develop marketable talents are distributed." Quite rightly Strike has seen that what is assumed in this is that schools can teach "the kinds of talents which are important in economic competition." (Strike, 1983, p. 185)

Notions such as these have generated a large and expanding literature devoted on the one hand to the distributive effects of schooling, and more recently on the other hand to the relationship between schooling and the material well-being of the entire social body. Some of this earlier literature greatly influenced the courts, and led to legal decisions and legislation which altered the lives of millions of people. The findings contained in this literature quite literally generated the power needed to move large numbers of people from one school to another in a distant neighbourhood. They similarly provided the grounds for funding compensatory education programs and altering admission standards to colleges and universities so that minorities underrepresented in these institutions might have better opportunities for further education. The more recent research, on the other hand, has made it seem imperative that policy makers tight-
en up course requirements for graduation from high school, and insist upon higher standards of achievement within the schools. Whether concerned with the apparent lack of equity in American education and American life or simply concerned with an apparent decline in educational standards and the implications such a decline might hold for the American way of life, one of the things the studies in this literature have in common is a conception of the student as "an educational resource" that should be either more equitably distributed or more fully developed in the interests of national security or prosperity or both.

What, then, are we to make of these conceptions of the student and of effective teaching, especially when they so obviously fail to square with our daily experience of human actions and seem to subordinate the interests and welfare of the individual to those of one group or another? Before we answer this question it may help to summarize what is at work in this educational research. Firstly and most obviously students and teachers have become the objects of study. Secondly the findings that emerge from this study have provided grounds for action and the basis for policy. The relationship between power and knowledge seems evident enough. The knowledge generated in this research generates at the same time the power to shape human action. Teachers learn new methods of teaching and are expected to use them or have good reasons for not doing so. School administrators learn what are the features of an effective school and are expected to introduce or maintain them. Legislators and
their constituents learn of the effects of racial composition of classrooms on achievement and feel the pressure to legislate a particular racial mix in the schools. Or legislators find confirmation in the research of what they supposed all along and feel compelled and empowered to act on the basis of this knowledge. Seen in this way the aforementioned research takes on a somewhat different character. No longer does our question seem to be simply concerned with the characterizations of students as billiard balls and monkeys or with research programs modelled on systems for the production of hybrid corn.

STATEMENT OF THE RESEARCH PROBLEM

It seems reasonable to ask, in light of the foregoing discussion of the relationship between power and knowledge: What is the significance of the fact that these conceptions of students, teachers, and learning function and circulate in educational discourse as if they were true?

It may often be assumed that the history of attempts to improve the schools will reveal little more than the reasons one group or another had for wanting to reform schools, along with the arguments and evidence needed to document the superiority of their particular program. What appears to be lacking is an analysis of the relationship between the knowledge that is gathered and produced in educational research, (along with the social practices that emerge from this knowledge), and the workings of power. For if it can be shown that "power produces
knowledge (and not simply by encouraging it because it serves power, or by applying it because it is useful); that power and knowledge directly imply one another; that there is no power relation without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time power relations", then there may be grounds for our being apprehensive about the operation of power and knowledge in the educational world, an apprehension that may have implications for the way in which participants in policy debates about education view themselves and the debate itself. (Foucault, 1977, p.27) As to why we might be apprehensive about the workings of power/knowledge in the educational world we must turn to Michel Foucault.
French philosopher-historian Michel Foucault has been concerned with the idea of political power in virtually all of his writing, especially in the books and articles written since 1970. In these Foucault rejects both what he calls the traditional, juridical conception of political power as well as the Marxist conception of power. In their place he argues for a radically new conception of power. He suggests both the traditional and Marxist conceptions are incomplete and anachronistic. "The traditional conception of power", Foucault says, "[is that of] an essentially judicial mechanism...which lays down the law." (1980, p. 183) In this view the law derives its power from the transfer of individual powers to the sovereign or the state. Thus power is taken to be a right, which one is able to possess like a commodity, and which one can, in consequence, transfer or alienate, either wholly or partially, through a legal act or through some act that establishes a right, such as takes place through cession or contract. Power is that concrete power which every individual holds, and whose partial or total cession enables political power or sovereignty to be established. This theoretical construction is essentially based on the idea that the constitution of political power obeys the model of a legal transaction involving a contractual type of exchange." (1980, p. 88)

The Marxist conception, on the other hand, is similarly economistic and rejected as incomplete for that reason. In it "power is conceived primarily in terms of the role it plays in the maintenance simultaneously of the relations of production and
of a class domination which the development and specific forms of the forces of production have rendered possible." (1980, p. 88)

Foucault doubts that power can be modelled on a commodity that is held, accumulated, ceded and recovered, although he readily admits that power relations are asymmetrical in modern society. He doubts as well that power is always in a subordinate position relative to the economy or that it is destined to maintain and reproduce the relations essential to the functioning of the economy. In its place or rather alongside this conception there operates another form of power - bio-power. Bio-power, according to Foucault, "brought life and its mechanisms into the realm of explicit calculation and made knowledge-power an agent of the transformation of human life." (1979, p. 143) It exists at two poles: one pole is the human body, the other is the human species.

One of these poles - the first to be formed, it seems - centered on the body as a machine; its disciplining, the optimization of its capabilities, the extortion of its forces, the parallel increase of its usefulness and its docility, its integration into systems of efficient and economic controls, all this was ensured by the procedures of power that characterized the disciplines: an anatomo-politics of the human body. The second, formed somewhat later, focused on the species body, the body imbued with the mechanics of life and serving as the basis of the biological processes: propagation, births and mortality, the level of health, life expectancy and longevity, with all the conditions that can cause these to vary. Their supervision was effected through an entire series of interventions and regulatory controls: a bio-politics of the population. The disciplines of the body and the regulations of the population constituted the two poles around which the organization of power over life was deployed. The setting up, in the course of the classical age, of this great bipolar technology... characterized a power whose highest function was perhaps no longer to kill, but to invest life through and through. (1979, p. 143)
The first pole, therefore, represented a power over bodies, disciplinary power. And while there is, in a sense, nothing new in having political power exercised over bodies, there is a fundamental difference between disciplinary power as Foucault conceives it and the age-old power of life and death exercised by sovereigns. Disciplinary power applies itself more continuously and to more people directly than sovereign power in itself could hope to. Moreover, the threats that lie behind it are at the same time less severe and more numerous than the floggings, hangings and mutilation that made up the catalogue of judicial punishments. It would be behind the "gentle mask of Man" expressed in the human sciences that these threats would operate. These threats, alongside a range of rewards, would serve in the disciplinary regime to transform individuals into more productive, more obedient people. Disciplinary power would begin and for most of its history be exercised in institutions designed to improve or rehabilitate the people in their care. In these institutions we have not only the exercise of disciplinary power, but the production of knowledge as well. Power brought individuals to these institutions and would for varying lengths of time keep them there, subject to a range of penalties and gratifications. It would also extract from individuals, by means of supervision, a knowledge of human beings that would form the foundation of many of the social sciences and would ultimately allow disciplinary power to move beyond an institutional setting and spread to society itself. One of the conditions which made
this expansion possible lay in the development of the second pole of bio-power.

The second pole - power over the species in the form of regulatory control - would be exercised over the social body and would be dependent upon a concise mapping of it. Beginning with demography in the 18th century and a general interest in all matters relating to the growth and care of populations we have the development of the knowledge of the human species, knowledge that could be used to organize life and maximize its usefulness. Administrators in 18th century Europe approached the population as something to be known, taken care of and made to flourish. Foucault (1979) observes that "it was necessary to analyze the birthrate, the age of marriage, the legitimate and illegitimate births, the precocity and frequency of sexual relations, the ways of making them sterile or fertile, the effects of unmarried life or of the prohibitions, the impact of contraceptive practices." (p. 25-26) In this indirect manner 18th century bureaucrats concerned themselves with human sexuality and its effects on society as a whole. By the mid 19th century, according to Foucault, the focus had shifted to medical analyses of sexuality. By the end of the century the understanding and regulation of sexuality appeared to give access to the future well-being of society and to provide the means of understanding the individual human psyche.

Like so many of the ideas he has introduced, Foucault developed this notion of power through his historical studies. In
Discipline and Punish Foucault became concerned with the rituals of sovereign power connected with the public torture and execution of criminals in the 17th and 18th centuries. In these rituals he noticed what he took to be a characteristic feature of sovereign power - the right to decide life and death. It was not that the sovereign could exercise this power over his subjects in an absolute and unconditional way; rather the right to take life depended upon external enemies who might overthrow him or contest his rights. In the international sphere such threats justified armed retaliation; that is, the sovereign could require his subjects to take part in the defense of the state. By this means the sovereign had the power to expose his subjects to death and thus he exercised an indirect power over their lives. In the case of a subject who dared to attack him directly or violate his laws, the sovereign could exercise a more direct power over the subject's life: the offender could be put to death.

It is with just such circumstances - the execution of a regicide - that Foucault begins Discipline and Punish. The opening pages recount the gruesome public torture and execution of Damiens the regicide in 1757. The sentence instructed his executioners to tear Damiens' flesh with red hot pincers, burn the hand that held the murder weapon, pour a mixture of boiling oil, molten lead and burning resin into the wounds inflicted by the pincers, draw and quarter his body with four horses, then burn the remains until nothing but ashes were left. Eyewitness accounts reveal the manner in which the executioners bungled the
job, how the horses, untrained in executions, failed to tear the limbs apart quickly, necessitating the executioners hacking away at the conscious Damiens while the horses strained at their traces. All the while confessors exorted Damiens to confess publicly to his crime.

In this horrible event Foucault sees manifested the power of the sovereign and the limits to that power. Public torture and execution were clearly political rituals designed to demonstrate the superior power of the sovereign and thereby serve as a deterrent. Within the framework of juridical theory, violations of the king's law amounted to attacks against the king himself. As such, the king, with justification, could respond in kind, but with an excessive force that showed the power of the law. Not only would such rituals demonstrate the power of the king in all its "murderous splendor", the addition of confession on the scaffold would serve to validate the whole proceedings. The ritual was not then simply one of power but one of power and knowledge. It became possible in these rituals to re-establish the truth that crime didn't pay, as well as the truth of the charges themselves. Torture in this way brought together a complex of power, truth and bodies. For Foucault it provided a simple and clear model for the relationship between power and knowledge, a relationship that would preoccupy Foucault from the early 1970s until his death a decade later.

Power produces knowledge (and not simply by encouraging it because it serves power or by applying it because it is useful); that power and knowledge directly imply one
another; that there is no power relation without the correlative constitution of a field of knowledge, nor any knowledge that does not presuppose and constitute at the same time power relations." (1977, p. 27)

One of the chief effects of power in Paris on March 2, 1757 was a knowledge of pain and its connection to the will of the sovereign and his laws. While this lesson was no doubt most keenly understood by Damiens, (even though its effect ended with his death), it was generally intended that it would also be understood by those present, the king's other subjects. The sovereign's power led to knowledge effects ("the constitution of a field of knowledge"), which in turn, at least among the prudent, constituted a strengthening of the power relation.

The king's power over his subjects in criminal matters was essentially a power to decide whether subjects were to continue to live or to die. This particular feature of sovereign power is, for Foucault at least, closely related to its other features. For Foucault, sovereign power was exercised as a means of deduction; it consisted of a right "to appropriate a portion of the wealth, a tax of products, goods, and services, labour and blood, levied on the subjects. Power in this instance was essentially a right of seizure: of things, time, bodies, and ultimately life itself." (1979, p. 259) However it could not be applied uniformly and continuously. It was too easy for subjects to escape this network of power.

Eighty years after the execution of Damiens, bodies were
being controlled in a different, less spectacular way. Instead of transgressors being subjected to physical torture, they were being imprisoned, and while in prison subject to a regime of subtle and relatively painless coercions. Time was first of all tightly organized, and within the spaces of time in the day prisoners were uniformly put to their tasks; each pair of hands, arms and legs were required to carry out the same functions in identical ways. According to the regulations for the conduct of prisoners in Paris' House for Young Prisoners, prisoners were required to rise and retire at the same time each morning and night and work at the same time throughout the day. "At quarter to six in the summer, a quarter to six in the winter ... [prisoners] must wash their hands and faces ... In the evening at the first drum roll they must undress, and at the second get into bed." (Foucault, 1977, p. 6-7) These seemingly trivial details are for Foucault the indications that a new type of power was in the process of being invented. "In the seventeenth and eighteenth centuries, we have the production of an important phenomenon ... the invention of a new mechanism of power ... [one] possessed of highly specific procedural techniques, completely novel instruments, quite different apparatuses..." (Foucault, 1980, p. 104) Power still possessed its circular relationship with knowledge; its chief difference, however, lay in that it was a form of power unrelated to sovereign power. As we have seen Foucault has termed it "disciplinary power". It would no longer be dealing with "legal subjects over whom the ultimate dominion was death, but with living beings, and the mastery it would be able to exercise over them would have to be applied at the level
of life itself; it was the taking charge of life, more than the threat of death, that gave power its access even to the body." (1979, p. 148)

This new mechanism of power is more dependent upon bodies and what they do than upon the Earth and its products [seen by Foucault as the traditional targets of sovereign power]. It is a mechanism of power which permits time and labour, rather than wealth and commodities, to be extracted from bodies. It is a type of power which is constantly exercised by means of surveillance rather than in a discontinuous manner by means of a system of levies or obligations over time. It presupposes a tightly knit grid of material coercions rather than the physical existence of a sovereign. It is ultimately dependent upon the principle, which introduces a genuinely new economy of power, that one must be able simultaneously both to increase the subjected forces and to improve the force and efficacy of that which subjects them. (1980, p. 104)

The point of this new form of power is to produce "subjected and practiced bodies", "docile bodies". In Foucault's view, "discipline increases the forces of the body (in economic terms of utility) and diminishes these same forces (in political terms of obedience)." (1977, p. 138) Unlike sovereign power which manifested itself through the right of death, this new power - bio-power - manifested itself as the power over life. Whereas in preceding periods the soul was the object and target of power, the body and the manner in which it could be "manipulated, shaped, trained" was the focus of this new power. As we have seen, Foucault noticed that by the late 18th century bodies were brought increasingly under control not only to a greater degree than ever before, but in a different manner. While in the past bodies had been controlled by the threat of physical punishment,
a threat which was periodically made manifest in the form of spectacular public executions, by the late 1700s bodies came to be controlled by various forms of discipline. In these forms of discipline Foucault saw "an uninterrupted, constant coercion" supervising the actions of the body in order to make it both docile and useful by the ever finer partitioning of time, space and movement.

Among the forms of discipline most closely examined by Foucault are those associated with penal discipline. It would be from the prisons that technologies of discipline would emerge and spread to other institutions and sites: monasteries, armies, workshops and schools. In contrast to the prisons of the late Middle Ages where prisoners were locked away from sight and more or less forgotten for long periods of time, Foucault provides us with the panopticon, Jeremy Bentham's design for a modern prison. The panoptican was a design for a prison which consisted of a large courtyard containing a tower at its center, and on the perimeter, tiers of cells containing inmates. In each cell were two windows: one facing the tower and the overseer, the other admitting light, thus permitting the continuous supervision of the silhouetted inmate, who, significantly, could not see his overseer. Because the inmate could not tell when he was under scrutiny, he was forced to assume he was always under scrutiny. In Foucault's words, "the panopticon provides an inspecting gaze which each individual under its weight will end by interiorization to the point he is his own overseer." (1977, p.
156) Foucault sees in this design the guiding principle of modern disciplinary power: supervision. In a manner similar to that of public torture and execution, the panopticon, for Foucault demonstrates the workings of power-knowledge, but with an important difference. Power in the case of the panopticon is applied continuously and to everyone, including the overseer, but with little or no violence. Moreover it can operate anywhere so long as the following conditions obtain. First individuals need to be placed within a "field of visibility", a panoptic field. Once visible, they need to know they are visible and need to be aware of a range of penalties and gratifications associated with behaving in reference to some standard or norm. "He who is subjected to a field of visibility, and who knows it, assumes responsibility for the constraints of power; he makes them play spontaneously upon himself, he inscribes himself in a power relation in which he simultaneously plays both roles; he becomes the principle of his own subjection." (1977, p. 202-203) What is significant for Foucault is the fact that this form of power spread to so many other localities. That Bentham's panopticon was never fully realized is unimportant to Foucault. What is important is that the technologies of discipline spread to become a "general formula of domination". This sort of disciplinary power, he says,

may be identified neither with an institution nor with an apparatus; it is a type of power, a modality for its exercise, comprising a whole set of instruments, techniques, procedures, levels of application, targets; it is a "physics" or an "anatomy" of power, a technology. And it may be taken over either by "specialized" institutions (the penitentiaries or
"houses of correction" of the nineteenth century), or by institutions that use it as an essential instrument for a particular end (schools, hospitals), or by preexisting authorities that find in it a means of reinforcing or reorganizing their internal mechanisms of power..." (1977, p. 215)

Discipline was to operate throughout society by means of three instruments -- hierarchical observation, normalizing judgement and the examination -- and was to serve as a means of making bodies more productive at the same time it made them more obedient. Hierarchical observation refers to the connection between visibility and power in situations where groups of individuals are arranged in such a way they can be easily observed. The result of such spatial arrangement is a "technology" for the production of knowledge of human actions. This knowledge produces in turn effects of power. Once space is organized and arranged to facilitate observation of those within it, and once those within it are brought into view, it becomes possible to know them and thereby alter them. Where direct observation is not possible it then becomes necessary to develop indirect supports or "relays" that would over time connect the information of accumulated periods of time. Power in this Foucauldian scheme is not seen as the possession of any particular individual or group, but rather as a machine or apparatus through which power and individuals are distributed in a continuous field. With normalizing judgement norms are established and depending on the degree of one's variation from the norm one is either punished or rewarded. A chief effect of this system of gratification and punishment is the distribution
of individuals along a continuum. The third instrument of discipline is the examination, an instrument which combines the techniques of hierarchical observation and normalizing judgement to produce a normalizing gaze through which individuals could be judged and classified. The examination as an instrument of discipline reveals the manner in which it is possible to bring about the "subjection of those who are perceived as objects and the objectification of those who are subjected..." (1977, p. 185) The examination mechanism is capable of producing three effects which serve to link power and knowledge. The first effect is the transformation of the field of visibility into the domain of power. The second effect is the collation of files, documents, and records. The third effect of the examination is the constitution of individual cases. In this manner it becomes possible, that as a result of the examination, individuals are situated in a field of visibility, subjected to scrutiny and objectification, and thereby made subject to the exercise of power. In Discipline and Punish Foucault devotes considerable space to a discussion of these instruments or "technologies of discipline" as they operate in prisons. In addition to discussions of this sort, however, there is a discussion of discipline in schools. Foucault notes the spatial arrangements of desks and chairs that permit those in authority to survey and supervise pupils, so that in the event of some transgression the overseer could not help but notice, a fact not lost on pupils.

Secondly, he notes the presence of all sorts of norms, some that will be used to justify punishment, others whose purpose is
to permit differentiation of individuals insofar as they deviate from the norm.

In short, the art of punishing, in the regime of disciplinary power, is aimed neither at expiation, nor even precisely at repression. It brings five quite distinct operations into play: it refers individual actions to a whole that is at once a field of comparison, a space of differentiation, and the principle of a rule to be followed. It differentiates individuals from one another, in terms of the following rule: that the rule be made to function as a minimal threshold, as an average to be respected, or as an optimum toward which one must move. It measures in quantitative terms and hierarchizes in terms of value the abilities, the level, the "nature" of individuals. It introduces, through this "value-giving" measure, the constraint of a conformity that must be achieved. Lastly, it traces the limit that will define difference in relation to all other differences, the external frontier of the abnormal....The perpetual penalty that traverses all points and supervises every instant in the disciplinary institutions compares, differentiates, hierarchizes, homogenizes, excludes. In short, it normalizes." (1977, p. 195)

Even though it so often appears to be in the general interest that discipline is demanded, there is a cost for productivity and efficiency; for methods of discipline rely on norms, whose effects and whose power lie as much in the production of failures, deviants and delinquents as they do in successes and paragons. In the case of schools, achievement and excellence are relative terms and can be understood only in terms of the relations of rank among students, where the standard by which students are judged is most frequently the norm.

The Normal is established as a principle of coercion in teaching with the introduction of a standardized education and the establishment of the ecoles normales
(teachers' training colleges) ... Like surveillance and with it, normalization becomes one of the great instruments of power at the end of the classical age. For the marks that once indicated status, privilege, and affiliation were increasingly replaced - or at least supplemented - by a whole range of degrees of normality indicating membership of a homogeneous social body, but also playing a part in classification, hierarchization, and the distribution of rank. In a sense, the power imposes homogeneity; but it individualizes by making it possible to measure gaps, to determine levels, to fix specialities, and to render the differences useful by fitting them one to another. It is easy to understand how the power of the norm functions within a system of formal equality, since within a homogeneity that is the rule, the norm introduces, as a useful imperative and as a result of measurement, all the shading of individual differences. (1977, p. 184)

Students therefore and by extension their teachers are pushed towards normalcy. Yet the examination will unmask those characteristics of the individual that fail to meet the norm or are abnormal in some other way. It will after a time allow each individual to be made into a case, "a case which at one and the same time constitutes an object for a branch of knowledge and a hold for a branch of power." (1977, p. 191) Each case represents an individual who may be "described, judged, measured, compared with others..., who has to be trained or corrected, classified, normalized, excluded." (1977, p. 191)

In order for these ranks to be efficiently determined, time must be fully utilized. "Time measured and paid must be time without impurities or defects; a time of good quality, throughout which the body is constantly applied to its exercise." Thus wasting time has become one of the cardinal sins in a disciplined environment. In school, punctuality is stressed, units of work
marked off in minutes and days, efficacy of instruction measured by pupil time on task.

In order for such ranking to be accurate, examination, in the broadest sense, is necessary. By this Foucault means not only formal examinations, but close supervision as well. The more examination comes to characterize a school, the more disciplined a place it becomes, and, the more likely it is to be efficient in producing knowledge. The school becomes "a sort of apparatus of uninterrupted examination that duplicates along its entire length the operation of teaching." (1977, p. 136) By means of supervision schools develop a host of micro-penalties:

micro-penalties of time (lateness, absences, interruptions of tasks), of activity (inattention, negligence, lack of zeal), of behaviour (impoliteness, disobedience), of speech (idle chatter, insolence), of the body (incorrect attitudes, irregular gestures, lack of cleanliness), of sexuality (impurity, indecency).

(1977, p. 178)

With each examination there comes a ranking, where each person is marked off from the others. In this way he is individualized, objectified in numbers, language and reports. By such means individuals become known to themselves and others as "slow average" or "bright", dyslexic or gifted, although for the very young student these categories may operate under the innocent sounding labels of "robins" or "bluejays". But examination does more than rank individuals, as important a function as this is. "The examination [enables] the teacher, while transmitting his knowledge, to transform his pupils into a
whole field of knowledge...The examination in the school [is] a
constant exchanger of knowledge; it [guarantees] the movement of
knowledge from the teacher to the pupil, but it [extracts] from
the pupil a knowledge destined and reserved for the teacher. The
school [becomes] the place of elaboration for pedagogy...The age
of the "examining" school marked the beginnings of a pedagogy
that functions as a science." (1977, p. 198) This is so much more
the case today in the instance of laboratory schools, or the
schools upon which teams of educational researchers descend with
diagnostic "instruments" and a desire to study teacher-pupil
interactions or the relative merits of one instructional method
over another.

A notion that is central to Foucault's thought,
especially in his historical studies of prisons, asylums, and
hospitals, is that when human beings began to be treated as
subjects of scientific study, they began to become subjects in
another sense. In the Middle and Classical Ages human beings in
Europe were subject to the power of a sovereign; they were his
subjects. By the time of Freud human beings were well on their
way to becoming subjects in this same sense, but to another, more
diffuse power; one that was both within and outside the person.
According to Foucault the object of his historical studies "has
been to create a history of the different modes by which, in our
culture, human beings are made subjects." (Rabinow, 1985, p. 10)
The first mode he calls "dividing practices". By this he means
the practices by which individuals have been physically isolated
and confined. The most famous examples of this isolation and confinement are the isolation of lepers in the Middle Ages, the confinement of the poor and the insane in the Hospital General in the Paris of 1656, the new classifications of disease and the associated practices of clinical medicine in early 19th century France, the rise of modern psychiatry and its entry into hospitals, prisons and clinics, and lastly the medicalization, stigmatization and normalization of sexual deviance. At their most essential, dividing practices are "modes of manipulation that combine the mediation of science or pseudo-science and the practice of exclusion." (Rabinow, 1985, p. 8) Dividing practices as they are defined here formed the basis for a number of Foucault's books: *Madness and Civilization*, *The Birth of the Clinic*, as well as *Discipline and Punish*. In these, Foucault focussed on four main themes. Firstly, he discussed the objectification of individuals who were first drawn from the undifferentiated masses such as the vagabonds in 17th century Paris, or the juvenile delinquents from working class districts in 19th century France. Secondly, he considered the interconnections of dividing practices with the formation and development of the social sciences. Thirdly, Foucault investigated the historical relationship between these modes of classification, control, and containment and an emerging tradition of humanitarian rhetoric on reform and progress. Lastly, Foucault examined the application of these procedures of power and knowledge to dominated groups or to groups that were formed and given an identity through the dividing practices (the mad, the delinquent, the criminal, etc.)
The second mode can be termed "scientific classification". It grows out of modes of inquiry which concern themselves with building conceptions of human beings or types of human beings. These conceptions objectify in human understanding features of humanity that go to make up "human nature". In economic theories such as those of Ricardo or of Marx, the human subject is defined in terms of his labour or the relations that make up the world of productive labour; in criminology the delinquent is a "type" of human being whose distinctiveness is defined not so much by what he does as by what he is. "The delinquent is to be distinguished from the offender by the fact that it is not so much his act as his life that is relevant in characterizing him." (1977, p. 251) In a related fashion one who commits acts defined by psychology as homosexual acts is one who is to be understood in terms of the type of person he is, the type of imperatives that precipitate his actions. His actions are intelligible in terms of the type of person he is, the scientific category to which he belongs. As a measure of the influence of social science, since the 19th century, the law, which had previously only concerned itself with the criminal act, had been forced more and more to consider the nature of the criminal in its deliberations. To this list one might add the modes of inquiry which likewise try to give themselves the status of sciences and which objectify the subject that teaches and the subject that learns.

The third mode may be termed "subjectification". It
concerns "the way a human being turns him- or herself into a subject." In addition to "operations on [people's] own bodies, on their own souls, on their own thoughts, on their own conduct", subjectification involves as well the mediation of an external authority figure -- a physician, a priest, a psychoanalyst. For example, Foucault demonstrates that during the nineteenth century there developed an enormous literature on sex, partly because people had become convinced that an understanding of one's sexuality led to an understanding of oneself. To use Freud's interpretation of dreams as an example, the sexual imagery of dreams, combined with the psychoanalyst's hermeneutic powers could be used to open the windows on one's inner being. Out of this collaboration of speaking subject and erudite interpreter grew a discourse that pose sex "as a drive so powerful and so irrational that dramatic forms of individual self-examination and collective control were imperative in order to keep these forces leashed." (Dreyfus & Rabinow, 1982, p. 169) One could come to understand oneself by making oneself a subject of one's own investigation.

With respect to Foucault and the state, it is plain that Foucault noticed a number of key historical changes in the state's relation to the individual. It seems to Foucault that by the 16th century in Europe the concern of government began to extend from the sovereign and his manner of ruling down to how the sovereign's subjects lived. By this time a great many treatises on government began to emerge where the subject of
these works was society itself. Unlike earlier discourse, these treatises focussed on a wide range of topics: the "governing of a household, souls, children, a province, a convent, a religious order, or a family". (Foucault, 1979, p. 8-10) Rabinow (1984) notes that political reflection was extended to include almost all forms of human activity. "Society was becoming a political target." (p.15) According to Foucault these treatises on government became linked to the rise and growth of centralized state bureaucracies along with the demand for detailed knowledge about the country and its inhabitants. "The art of government and empirical knowledge of the state's resources and condition - its statistics - together formed the major components of a new political rationality which is still with us." (Foucault, 1980, p. 14) Among the most important of the state's "resources" are its people and their ability to reproduce themselves. The concern for the vitality of populations and the utility of individuals came by the 19th century to be united in the concept of sexuality. Attention turned to "the body, to life, to what causes it to proliferate, to what reinforces the species, its stamina, its ability to dominate, or its capacity for being used." (Foucault, 1979, p. 269) Foucault (1979) recognizes in this concern for the strength of the species the seeds of modern racism and its concern with "protecting the purity of the blood and ensuring the triumph of the race." (p. 271) This concern makes a problem out of the body; that is, the body of individuals and the social body in general. Bodies are a problem from this perspective because they can be both "scarce and numerous, submissive and restive, rich and poor, healthy and sick, strong
and weak, ... more or less utilizable, more or less amenable to profitable investment, ... with greater or lesser prospects of survival, ... with more or less capacity for being usefully trained." (1979, p. 279) Thus it became imperative that any characteristic of human sexuality dangerous to health be known and controlled.

In his attempt to understand the Victorian sensitivity regarding sex, Foucault was struck by how horrified Victorians appeared to be of sexual matters, especially if those matters had little to do with conjugal, procreative sex. More puzzling given this prudishness was the manner in which sex was such a large part of their lives, a point well documented by Peter Gay in *The Bourgeois Experience*. Ultimately, claims Foucault, this interest in sex, confined in many cases to sexual abnormalities and perversions, produced a conception of sexuality that was central to the way human nature would be defined. Sexuality "came to be seen as the very essence of the individual human being and the core of personal identity. It was possible to know the secrets of one's body and mind through the mediation of doctors, psychiatrists and others to whom one confessed one's private thoughts and practices." (1979, p. 68) This social construction of "sexuality" became, according to Foucault, another means of social control. As individuals sought access to the core of their personal identities through sexual discourse with "experts", the knowledge contained in this discourse created power effects touching children (whose irrepressible onanism was legendary),
women (seen as possessing limitless sexual appetites) and perverts (who stood as testimony to the fragile limits of human self-control). Foucault notes "the sexualization of children was accomplished in the form of a campaign for the health of the race (precocious sexuality was presented from the eighteenth century to the end of the nineteenth as an epidemic menace that risked compromising not only the future health of adults but the future of the entire society and species); the hysterization of women...was carried out in the name of the responsibility they owed to the health of their children, the solidity of the family institution, and the safeguarding of society...In the case of [the] psychiatrization of perversions...the intervention was regulatory in nature...and had to rely on individual disciplines and constraints." (1979, p. 268) The discourse on abnormality, for Victorians the overwhelming focus of investigation in sexual matters, helped to reinforce the notion of the powerful, irrational nature of sexuality, a notion that served to legitimate the confinement of deviants, the close supervision of children and the narrowly defined social role of women. It was these people, in the Foucauldian scheme of things, who represented the high cost of "normality" and a healthy population.

Foucault has noted in his study of discourses on sexuality that social research often does not develop along methodologically rigorous lines. With respect to the early discourses on sexuality Foucault observed:
When we compare these discourses on human sexuality with what was known at the time about the physiology of animal and plant reproduction, we are struck by the incongruity. Their feeble content from the standpoint of rationality, not to mention scientificity, earns them a place apart in the history of knowledge. (1978, p. 69)

It seems to Foucault that the aim of so much of the discourse on sexuality "was not to state the truth, but to prevent its very emergence." The discourse on sex was imbued with "a systematic blindness: a refusal to see and to understand... evading the truth, barring access to it, masking it: these were so many local tactics which as if by superimposition and through a last minute detour, gave a paradoxical form to a fundamental petition to know." (Foucault, 1979, p. 69)

One might ask the question of educational research, particularly as it regards students and their relation to teachers through learning, whether there is in the research "a systematic blindness... a refusal to see and to understand", and if there is, what significance does this hold?

It seems clear that Foucault has much that is relevant to say in a discussion of research that stresses norms and testing. This seems equally true for the testing of individuals or the social body as it is represented by students at any one time. It seems also that his work has a bearing on understanding some of the likely effects of the hierarchical structure of grades and certification in schooling. As well it may be that his work can
throw some light on the relationship between competing bodies of research and power, a theme that invites examination in the matter of whether or not schools and teachers "make a difference."

One final remark is needed on the applicability of Foucault in a discussion of educational reform literature. Foucault isolates as a form of power - disciplinary power. He illustrates the manner in which this power operates institutionally. He makes it possible therefore to consider somewhat separately from the aims of educational reform the practices that are likely to emerge from the implementation of recommended reforms. Is it the case that the ends (increases in valued skills and abilities) justify the means (among them the disciplinary mechanisms discussed thus far)? Or rather is it the case that there is in this literature a relative disregard for these ends; indeed that these ends serve mainly to justify discipline and mask its operation? Perhaps there is nothing more sinister in this literature than an innocent failure to apply the same amount of care and attention to the goals of reform as to the means of effecting them. Yet the well-documented persistence of failure in other institutions that aim at transforming human beings should give us pause for thought.
CHAPTER THREE: TEACHER EFFECTIVENESS RESEARCH

The section on teacher effectiveness research which immediately follows will make clear the extent to which the literature is characterized by a systematic blindness with respect to the nature of students, how they learn, the adequacy of the constructs by which their efforts and those of their teachers are measured, as well as the intelligibility of the concepts upon which teacher effectiveness research rests. This blindness stems from our reliance on a research paradigm that promises a degree of certainty and predictability which it cannot deliver. Our quest for certainty has resulted in the development and refinement of classroom practices which may be viewed as examples of what Foucault means by disciplinary mechanisms. While these practices appear to promote generally desirable levels of scholastic achievement, it will be argued they do just the opposite. Because these classroom practices appear to promote achievement, their use is widely supported and is, in some cases, demanded. Doing one's job as a teacher or an administrator comes to mean acting according to the prescriptions in this literature, at least to some degree. What is significant, of course, is that these practices tend not to promote any such thing as scholastic achievement, except where achievement is understood in trivial terms. In fact, the use of these practices tends to necessitate their further use, since the goals which justified their use in the first place inevitably fail to be reached. That researchers persist in recommending their use is another of the puzzles which will be considered in the following pages.
To understand how such a state of affairs could come about, it helps to consider the context of crisis from which it emerged. There has grown, since the mid-1970s, a conviction that the American public school system is a failure. At various times in the history of American schools, of course, similar convictions have emerged. But there is reason now to believe the current crisis is more menacing. Newspaper editors, business leaders and university professors in particular have complained about declining standards and rising rates of illiteracy among high school graduates. Data from the early 1970s which has re-emerged in recently commissioned reports suggest that "23 million American adults are functionally illiterate by the simplest tests of everyday reading", while SAT scores over the last 20 years are said to reveal consistent declines in achievement in science, and even greater declines in English. The dominant view among those who decry the lowering of standards is that the current crisis is the result of the educational reforms of the 1960s. Chester Finn (1982) sums up this position nicely.

The sad fact is that for close to two decades now we have neglected educational quality in the name of equality. Trying to ensure that every child would have access to as much education as every other child, we have failed to attend to the content of that education. Seeking to mediate conflict and forestall controversy over the substance of education, we begin to find ourselves with very little substance needed. Striving to avoid invidious comparisons among youngsters we have stopped gauging individual progress by testing ... Hesitant to pass judgement on lifestyles, cultures and forms of behaviour we have invited relativism into the curriculum and pedagogy. (p. 32)

It is interesting to note in Finn's remarks the assumptions
that educational quality is in some way naturally at odds with equality, and that objective standards exist for judging students' work. Although these assumptions do not figure significantly in the literature on teacher effectiveness, they do represent a common theme in much of the literature on school reform.

In addition to an atmosphere of alarm over declining standards, there are several other aspects of the social context of reform in the 1970s and 1980s which are noteworthy. The first is the public concern over the rising costs of education. As early as 1970, Patrick Moynihan suggested that in terms of costs and benefits, educational spending had very likely passed the point of diminishing returns. (Hodgson, 1979, p. 56) In addition, there has been a shift in the attitude of policy-makers and educators from equating quality with inputs such as instructional materials, qualified staff, etc. to equating quality with outcomes or test results. A third factor is the growing belief that evaluation methods can be perfected to the point where precise objectives for learning can be set with confidence while the failure to reach them means failure, plain and simple. (Ornstein, 1983, p. 42)

Understandably then, much of the focus of educational research since the 1970s has been on improving instructional effectiveness in ways that can be measured, and which are compatible with a mood of fiscal restraint. The sort of considerations which were common in the 1960s, such as a concern
for the way in which schools could be instruments of social change, or the appreciation that individual differences and interests influence school success, tend to be much less evident in the current literature on effectiveness. Discussions of what teachers should do tend to be couched in technical rather than therapeutic language. Rather than insisting that schools do more than they have to help the disadvantaged, teacher effectiveness advocates have argued that teachers can do a better job of teaching by relying on proven methods.

Unfortunately, attempts to improve instructional methods, which have most often taken the form of attempts to discover the relationship between teacher "behaviours" and student achievement, have not been successful. Their lack of success, according to Garrison and Macmillan (1984), can be attributed to four assumptions which undergird the research. Teacher effectiveness researchers have tended to assume that anything which exists does so in some measurable amount, with the result that statistical analysis of quantifiable variables has become the main investigative technique. This assumption in turn depends upon a number of subsidiary assumptions. One, that nature is uniform over time. Two, that it is possible in every case to give a causal explanation for phenomena. Three, that our knowledge depends ultimately upon our experience of the world. Four, that the only phenomena suitable for scientific inquiry are those which can be quantified and measured.

The second major assumption in teacher effectiveness research is that teaching brings about student achievement in a
causal manner. One discovers these causes by finding the statistical correlation between measures of teacher behaviour and measures of student outcomes. Within this research tradition it is not legitimate to attend to the intentions of teachers or students, since intentions are too difficult to measure. Ideally researchers are to make as few inferences as possible in their observations.

It is further assumed that investigations into the causal relationship between teaching and learning will yield rules which can be used to guide instructional planning and classroom management. Lastly, it is assumed that the content of teaching is not relevant in assessing the relationship between teaching and learning.

The chief objection Garrison and Macmillan have with teacher effectiveness research is its failure to see intentions as central to both teaching and learning.

The problem of intention is central. The process-product tradition explicitly ignores the intentions of teachers and learners in its investigations. In part this is for reasons having to do with basic methodological assumptions. Intentions cannot be investigated by the usual low-inference methods embraced by this tradition. But teaching is a human activity, and like all human activities it is intentional, a matter of moods and tenses, aspirations, beliefs, and goals... Process-product research is conducted in the language of analysis of variables, but its application is in the intentional language of action and belief.(p.18)

Garrison and Macmillan point out that if we look at human "behaviour" as if takes place without a social context of
beliefs, dispositions, goals, desire and the like, we will probably construct a false view of why we act as we do. By means of an analogy between a chess game and teaching Garrison and Macmillan illustrate how process-product researchers are likely to get it wrong.

Imagine a researcher who knows nothing of chess but who wishes to understand how chess players play the game, with a view, eventually, to training novices and being able to assess players' ability... Our researcher might begin by forming two groups of players, one group consisting of strong players and the other group containing weak players... If our researcher should then proceed -- by analogy with many an educational researcher -- by determining the frequencies with which pawns, knights, bishops, rooks, and the other pieces on a chess board are moved by representatives of the two groups in the course of their games, then his understanding of the players and of the game will be slight indeed, no matter how extensive or how objective may be his research. For although differences between the groups will be found to exist, little will be gleaned of players' strategies and tactics, which are what distinguish the more able from the less able players. (p.19)

While it strikes Garrison and Macmillan that such an enterprise as this is bound to fail, they insist they would be willing to accept process-product research if it produced promising results. This it has not done. By their estimates, empirical research on teaching shows correlations between teacher behaviour and student outcomes that range from .10 to .20. While correlations of this magnitude may be viewed as significant in medical research where it safer to assume that nature is uniform and that strictly causal explanations are acceptable -- a point made elsewhere by Gage (1983) -- they cannot be viewed similarly where there is reason to believe intentions are intervening variables. Gage's
(1983) protests notwithstanding, correlations are measures of association, not cause. The values quoted above would be considered low for any research endeavour; too low certainly to betoken causal relationships.

Beginning with what appears to be an obvious truth -- teachers and what they do cause learning -- teacher effectiveness researchers have focussed their attention on the identification of a repertoire of successful teaching methods or "behaviours". Duncan and Biddle (1974) state this assumption in the following way:

There seems to be no more obvious truth than that a teacher is effective to the extent that he causes pupils to learn what they are supposed to learn ... If teachers do vary in their effectiveness, then it must be because they vary in the behaviours they exhibit in the classroom. (p.13-14)

It fell to researchers, therefore, to identify effective teaching "behaviours" and see if these behaviours could consistently produce gains in learning over other methods or behaviours. In this way a research agenda has developed.

It is worth noting at the outset a number of the features of this statement by Duncan and Biddle. Few people, I believe, would strongly object to the first sentence. They might object to what seems the peculiar use of cause, but would likely be inclined to agree that success as a teacher must in some way be connected with how much is learned by students. The second
sentence, however, is more problematic. It seems self-evident that there are variations in the extent to which students learn which cannot be attributed to the teacher. So effectiveness cannot refer to outcomes alone, but must be judged in relation to circumstances over which teachers have little or no control. These may include student characteristics such as intelligence, interest, health, energy, emotional well-being or any number of others. Or, they may include factors other than teacher influence which have an impact on learning: the adequacy of instructional materials or the state of the school's physical plant. There is little reason to believe we know enough about the number and strength of possible intervening variables to suggest that with the currently available statistical methods we are able to isolate the learner from the many factors that influence achievement.

Throughout most of its history, the research into teacher effectiveness has consisted in attempts to develop a method of isolating teacher effects on achievement from these other mediating influences. In some cases this has meant adopting behaviourist language, while in others it has meant redefining important concepts in ways that make it possible to measure outcomes. In this way the research findings have come to appear useful and compelling. When researchers conclude that teaching method x produces higher levels of reading achievement than teaching method y, and can substantiate this claim with data, it is difficult to ignore such a claim. However, when the most fundamental of educational goals, literacy and numeracy, come to
be operationally defined for the purposes of conducting the research, a great deal gets lost.

Our goals, it would seem, are being defined by the limits of the means we have to measure our success in reaching them. "Reading achievement" and "mathematics achievement" are given operational definitions in the research which allow investigators to measure teaching effectiveness on pencil and paper tests. As it turns out, such constructs as these have caused considerable confusion; in part because they deviate so sharply in meaning from those we are accustomed to in ordinary language and in part because researchers have not taken the necessary pains to make the difference between the two as clear as they might. If more people realized that reading ability as it is defined in this research means little more than the ability to decode vocabulary in short, unrelated passages, they might feel less inclined to believe the teaching methods which bring about gains on such tests are so valuable after all.

Process-product research is objectionable for a number of reasons. Two shall be emphasized here. The recommendations for teaching that emerge from this research can be objected to on the grounds that they are morally hazardous, and insofar as teacher effectiveness research influences teaching practice, potentially disastrous to our efforts to educate the young.

It strikes many of us as intuitively false to suggest students are objects which "learn" when teachers "behave" in
certain ways. Anyone who has ever tried to teach a child anything knows that a child's learning is dependent on a number of factors, his wanting to learn primary among them. As well, there is, in teacher effectiveness research, little or no mention of the role "understanding" is to play in any of this. Any of the familiar features of coming to understand something such as paying attention or noticing -- those features associated with the conscious efforts of the learner -- are missing from accounts of learning in this literature. Nor is there much discussion of the point any of this teaching might or should have; as if it were possible to view teaching merely as a collection of techniques which needn't be linked to a defensible conception of education or the sort of goals likely to inher in such a conception.

It seems clear that process-product research ignores the agency of a certain class of persons, in this case students in school. At the same time it provides teachers with reasons for acting in a particular way, yet, in effect, denies the significance of "reason for acting" explanations of human behaviour in school settings. Instead it relies on a simple causal theory of learning and fails thereby, in the teaching practices it recommends, to respect the rationality of the person. Why does it matter that students be treated as persons? How do the teaching practices which have been derived from this research militate against the likelihood that teachers will treat their students with the respect they deserve as persons?
One of the most notable features of talk about persons or respect for persons is the lack of agreement among philosophers about what is meant by "person". For most people it seems the concept is an unproblematic one. Invariably they insist that a person is a human being, that a human being and a person are one and the same. Philosophers, on the other hand, have tended to disagree that the concept of a person is quite so simple. In fact The Encyclopedia of Philosophy admits that "it is somewhat misleading to speak of Locke's, or any other philosopher's theory of persons, for there is no antecedently agreed upon definition of what these should be theories of. Rather we must at best speak of this or that philosopher's use of "person" as a semitechnical term in his system." (p.111) Generally, philosophers are agreed that persons, as opposed to things, are valuable in themselves. "That a person is distinct from a (mere) thing, and that any human being, insofar as he is a person, is in consequence of this status to be treated in a special manner, are two of the logical features of this concept." (p.111) Where much of the disagreement arises is over the question of what it is that makes a person valuable and distinct from that which has value as a means to someone's end. Things, most of us would agree, may be used for our own purposes. They are valuable precisely because to some degree they may be of use to us in the pursuit of our goals. Persons, on the other hand, must not be used merely as means to our ends; they are as Kant has said "ends in themselves and sources of value in their own right." (p.111)

But in what does the value of persons consist? A common way
of going about answering this question has been to identify or attempt to identify the attributes of a person. The most commonly cited attributes relate in one way or another to self-consciousness. For Kant, "that which is conscious of the numerical identity of itself at different times is insofar a person."

(p.111) Leibniz characterized a person as that which is capable of retaining "consciousness, or the reflective inward feeling of what it is: thus it is rendered liable to reward and punishment."

(p.111) Similarly, Wolff held that persons are distinct from animals simply because persons have "a consciousness of having been the same thing previously in this or that state." It is this feature of persons that helps make it possible for us to claim responsibility for our actions; for without an identity over time we could not be or become moral agents. A related attribute, and one dependent on self-consciousness, is rationality, defined in part by the condition of self-consciousness over time. Locke put the matter in this way: A person is a thinking intelligent being that has reason and reflection and can consider itself as itself, the same thinking thing, in different times and places; which is inseparable from thinking, and seems to me essential for it."

(p.111)

From these, something of the character of the concept of persons begins to emerge. We can see from the above that we cannot, in every case, treat people as things, or as means to our ends, since our concept of persons involves the recognition that persons are moral agents who have the capacity to stand in moral
relationships with others. This however tells us very little about how we should treat persons. In response to this problem moral philosophers have developed the principle of respect for persons as the basis of our moral relationships with persons. To respect a person in this context is to value or hold them in esteem. But we are soon driven to an examination of the attributes of a person in order to see just what it is we value and why.

Downie and Telfer (1970) suggest that "human beings ... ought to be respected for what is valuable in them." (p.30) In their view persons are to be seen as ends in themselves and valuable in their own right, much as Kant saw them. "Hence, to respect a person as an end is to respect him for those features which make him what he is as a person and which when developed, constitute his flourishing." (p.15) For Downie and Telfer the feature of a person that is valuable and which should be esteemed is the capacity for the exercise of rational will. This capacity is characterized by two other features: self-determination and the ability to follow rules.

The exercise of rational will is to be seen, then, as something at once (in old fashioned terminology) cognitive, conative and affective. It is the ability to exercise such a will in self-determination and rule-following which gives human personality its intrinsic value. (p.22)

Respect for these capacities has two necessary components: an attitude of "active sympathy" and the inclination to consider the applicability of other individuals' rules to ourselves.
Active sympathy and the readiness to consider others' rules as being applicable to us are, in this view, "independently necessary and jointly sufficient to constitute the attitude of respect which is fitting to direct at persons, conceived as rational wills." (p.29) The problem with Telford and Downie's account of respect for persons is, however, that it leaves open the question of why we should consider rational will so valuable that it should form the basis of our respect for persons. Why not value some other attribute of persons as most important? Moreover, what are we to do about the fact that there are wide variations among human beings in the extent to which they are rational. Do these variations permit differential treatment of individuals according to the extent they appear to act rationally? Put another way, how can respect be something which is due all persons; are there circumstances under which respect is not due a person? One might also ask these additional questions regarding respect: is respect a single kind of attitude, or is it even primarily a moral attitude which can only be directed toward other persons? Darwall (1977) in Two Kinds of Respect attempts to answer these questions by distinguishing between the two different ways in which persons may be the object of respect, or more precisely the two different kinds of attitude which are both referred to by the concept of respect. These two attitudes of respect he terms recognition respect and appraisal respect. Recognition respect, he suggests, is that form of respect which must be given to some feature of its object when one deliberates on what to do:
The most general characterization which I have
given recognition respect is that it is a
disposition to weigh appropriately some feature
or fact in one's deliberations. Strictly speaking,
the object of recognition respect is a fact. And
recognition respect for that fact consists in
giving it the proper weight in deliberation. Thus
to have recognition respect for persons is to
give proper weight to the fact that they are
persons.

(Darwall, 1977, p.39)

Recognition respect, on this account, is commonly extended to
such things as persons, or persons in specific roles such as
judges or ministers, or anything that could be involved in any
sensible deliberation about what to do. A partial list might
include the law, another's feelings, or nature. This sort of
respect recognizes or takes into account the fact of something in
deliberations about what ought to be done. The appropriate object
of this sort of respect is any fact that should be taken into
account in deciding what to do. Related to this sense of respect
is the notion of moral recognition respect. This notion suggests
that if some fact or feature of an appropriate object of respect
were treated inappropriately, that is, if an essential feature
were not respected, it would be legitimate to say such a lack of
respect is morally wrong. This form of respect regards its object
as "requiring restrictions on the moral acceptability of actions
connected with it." (p. 39) This is the sort of respect we most
often mean when we talk about respect for persons. It is, as we
shall see, quite different from appraisal respect.

Appraisal respect is an attitude of respect whose objects,
In Darwall's view, are persons or some feature of persons that are individually distinctive and praiseworthy. Darwall gives as examples of this attitude a respect for someone's integrity, or for some other quality. "Appraisal respect, then, consists in an attitude of positive appraisal of that person either as a person or as engaged in some particular pursuit." (Darwall, 1977, p.38) Unlike recognition respect, appraisal respect does not call for an appropriate behaviour with respect to its object. We may or may not have this sort of respect for someone or we may have it to varying degrees. It has two aspects: respect for persons as persons, and respect for their accomplishments. Appraisal respect for persons as such refers to an appraisal of them in light of the moral attitudes they hold, while respect for a person's accomplishments within a particular pursuit refers to the developed features of their character that have a bearing on their being successful in their pursuits. Such features are not talents or capacities for which a person can take no credit, but instead character traits such as resoluteness which are a product of the will or can be taken as constitutive of the will. Taken together these two aspects represent respect for those long term dispositions which identify a person as a moral agent. "Those features of persons which are appropriate grounds for appraisal respect are their features as agents -- as being capable of acting on maxims, and hence for reasons." (p.43) Character then is inextricably bound up with a conception of persons as moral agents who deserve respect. The distinction between recognition respect and appraisal respect helps us get at the ambiguity
surrounding the concept of respect. While appraisal respect may not be owed to everyone, and may be owed to varying degrees, recognition respect is due to all persons by virtue of the fact they are persons. These two kinds of respect are connected with one another in a number of important ways. Firstly, there is a connection between the basis for appraisal respect and the facts which are taken as appropriate objects of recognition respect. For example, if a person is considered unworthy of respect because he is lazy (an appraisal), then determination or drive are taken to be appropriate objects of recognition respect. Those who would appraise a lazy man in this way are committed to recognition respect for determination or drive. Secondly, as argued by Chelsom (1982),

our appraisal of persons depends on whether they show appropriate recognition respect for considerations which merit it, those features of persons as moral agents capable of acting and acting for reasons. The only beings who are appropriate objects of appraisal respect are those who themselves are capable of recognition respect, that is capable of acting deliberately. (p.32)

As helpful as all this seems, there is a certain emptiness to it. The suggestion that there is a difference between respecting persons for their accomplishments and respecting them because of the "fact" they are persons, or the "fact" they are moral agents possessing a rational will tells us nothing about why we should respect them. It only provides us with the circular notion that we should value that which is valuable and esteem the estimable. Alongside the lists of attributes that appear to constitute a person is talk of "merit", "appropriate objects of
respect" and of giving "proper weight to the fact they are persons." The lists of attributes provide us with nothing more than a list of what we commonly view as most valuable in human beings. We are left to conclude that we should value that which we believe to be valuable in human beings.

Daniel Dennett gives one of the better, recent accounts of the concept of a person, one that demonstrates its normative nature and emphasizes the regulative function this concept performs in the language. Dennett identifies six necessary conditions of personhood, of which the first three are mutually interdependent:

1) persons are rational beings.

2) persons have as a central attribute a state of consciousness to which psychological or mental or intentional predicates are ascribed.

3) for something to count as a person we must have a certain attitude toward it based on its intentionality. Dennett calls this attitude the Intentional stance and says "nothing to which we could not adopt the Intentional stance with its presupposition of rationality, could count as a person." (p.180) It is not so much that once we recognize another "being" as being intentional we feel we should treat him as a person. In Dennett's view "it is not the case that once we have established the object fact that something is a person we treat him or her or it in a certain way, but that our treating him or her or it in this certain way is somehow and to some extent constitutive of its being a person." (p.178)

4) the object toward which this personal stance is taken must be capable of reciprocating in some way.

5) persons must be capable of verbal communication.

6) persons can be distinguished from other entities by their being conscious in a special way, by their having a sort of self consciousness characterized by second order volitions. It seems clear from Dennett's account that he sees the concept of a person as being a construct that both identifies the attributes of a person
(rationality, verbal ability) that are the attributes of a human being and that are at the same time criteria for being a moral human being. In his account of the conditions of personhood, however, he does not give the same amount of weight to each condition. He clearly emphasizes second order volitions as being those characteristics of personhood which are distinctly human.

As Dennett points out, personhood is not simply a sort of status that we bestow on an entity once we recognize in it a number of key attributes, like rationality or verbal ability. It is partly because we treat an entity in a particular way that we call it a person and are likewise entitled to be called persons ourselves. But what does it mean to treat a person as a person, or to treat him with respect?

The answer to this question, as Richard Peters has emphasized, has a lot to do with the way of life we have been born into and in general are committed to maintaining. (Peters, 1966) Central to this way of life is a commitment to the idea of the individual as a rational, self-determining and assertive self. People in our society are encouraged to have this concept of self, and, it seems to Peters, are only likely to develop this concept fully if they are so encouraged.

At the heart of the type of encouragement Peters has in mind is the treatment of individuals as if they are rational and self-determining to begin with. We should treat our fellow humans as persons not only because it is hateful "to be treated as a moron or merely as an instrument of the purposes of other men", but also because to be treated as a person allows and encourages us to be persons.
People only begin to think of themselves as persons, as centres of valuation, decision and choice, in so far as the fact that consciousness is individuated into distinct centres, linked with distinct physical bodies and with distinctive points of view, is taken to be a matter of importance in society. And they will only really develop as persons in so far as they learn to think of themselves as such. (Peters, 1966, p. 211)

An additional aspect of being a person, already recognized by Kant, Liebniz, Locke and numerous others, is that to be a person one must be aware of oneself as distinct from others. The sense of being an entity distinct from other entities comes not only from an awareness of other bodies, but as Peters has pointed out, from our awareness of other points of view.

Our society is also committed to the idea that the truth is important, that the search for truth is important, not simply for prudential reasons, but also because it is a fundamental characteristic of human beings in the modern world to want to know the truth. Life is for human beings a context where the demands of reason are inescapable, where daily we are faced with the question: What ought to be done? The answer to this question will most often, if not always, involve considerations of what might be the truth of the matter. And one of the chief means we have for arriving at the truth is, in Peters' view, serious rational discussion. By rational discussion Peters means that certain principles must govern the actions of participants. Participants must not interfere in an unreasonable way with the wishes and interests of other participants; they must not be
partial or prejudiced in considering what is being said; they must take others' views into account. If one accepts these principles as binding on oneself, then they must serve "to safeguard the experiences which we most intimately associate with being a person, i.e. not being arbitrarily interfered with in respect of the execution of our wants and decisions and not having our claims and interests ignored or treated in a partial or prejudiced manner." (Peters, 1966, p.214)

We value persons, therefore, and let the respect for persons principle serve as a guide because by doing so we maximize the likelihood it will be possible for us to pursue our interests in a world governed by rational discourse, and therefore in one more predictable, than would be possible in a world dominated by the whims and wishes of those whose influence rests solely on their status or institutional position. Unfortunately for most of us, the world, even the parts of it where rational discourse and the respect for persons are encouraged, is a place where people are dominated by the whims and wishes of those whose influence rests on their status and institutional position. Indeed, very often, it is the whims of those who are held to be most rational that lead to the domination of others, since status, institutional position and power are commonly the rewards held out to those most adept at practical reasoning. Moreover, Peters' view of things notwithstanding, there is more to why we value persons than this. As Dennett has noted, despite the diminished capacities of people who are severely retarded or comatose, we continue to treat them
as if moral considerations are still relevant. In cases where an individual human being is less a self-determining agent than say a dog or a chimpanzee, we are still unwilling to treat him as an object or as a means to another's ends in the way we routinely do with animals in medical experiments. We wouldn't think of killing human beings so we could eat them, even if they in no way could be said to be capable of second order volitions. The ethical theories like Peters', which try to justify respect for persons on the grounds that persons are estimable because they are rational, self determining agents, fall well short of the ethical rules we commonly adhere to already. If we look to Darwall's notion of recognition respect in an attempt to find an object of respect that doesn't vary across individuals, we can do no better than find the one feature all living human beings have in common: namely, that they are living human beings. The objects we need to take into account in our deliberations concerning what to do are individual human beings, not some attribute they may or may not possess. One may grant that the sort of deliberations required in our dealings with the comatose are of a much narrower range than those which we will have with self determining, rational individuals. And that only in a few cases will the comatose or brain damaged represent an obstacle to our reaching our goals or pursuing our interests, since it is a relatively simple, albeit expensive, matter to accommodate the needs and wants of the severely retarded, senile or comatose. There seems no disputing, however, that sympathy for fellow human beings is a persistent human trait when circumstances allow the adoption of the moral
point of view. Circumstances of want, fear, scarcity and mistrust make it supremely difficult for people to adopt such a point of view, although despite such hardships men and women have in the past and continue today to act as if the people around them are as Kant said "ends in themselves and sources of value in their own right." That it is easier to think of people in this way when they lack the ability to form intentions or the power to realize them as in the case of the comatose, etc. underlines the point that asymmetries in power and the fear of scarcity serve to keep us from treating other human beings in ways that, were the circumstances reversed, we would like them to treat us. In a way that surprises no one, then, a person, as we said at the outset, is a human being; a human being and a person are the same thing. And the concept of a person, although it does not pick out with any precision just what we ought to do, does act as a reminder that we ought at least to adopt a moral point of view in our dealings with human beings. That we will do so is less likely when we routinely conceive of human beings as passive objects. It is this sort of conception of human beings that typifies the depiction of students in process-product research.

Recently Brophy and Good have reported in the Third Handbook of Research on Teaching that "the fund of available information on producing student achievement has progressed from a collection of disappointing and inconsistent findings to a small but well established knowledge base." (Brophy & Good, 1986, p.337) Typically, the knowledge base they refer to developed from a progression of studies that first sought to
identify promising teaching methods, then test them in controlled experimental situations. Much of the research has not yet reached the experimental stage, with a good deal of it based on correlational studies. Nevertheless, N.L. Gage, speaking of these correlational studies, suggests that they are just beginning to give promise of quantitative knowledge... We are beginning to have evidence that the correlations betoken causal relationships, so that changing teaching practices cause desirable changes in student achievement, attitude and conduct. (Gage, 1984, p. 93)

Evident in this remark is the belief that the relationship between teaching and learning is a causal one, and that not only is it desirable to produce higher levels of student achievement, it is desirable also to "manipulate variables" to produce in human beings certain attitudes and certain types of conduct. A belief such as this leaves little room for a conception of the student as a purposive being with intentions of her own. If one were to conceive of the student as a rational person possessing an independent will, the billiard ball analogy used earlier by Tom would look quite different.

Instead of the cue ball (teacher behaviour) striking the target billiard ball (the student) at the precise angle needed to deflect that ball into a pocket (a learning goal), we now see that the target billiard ball is a cause of its own movement in a particular direction. Student motives, for example, can act as an irregularly shaped barrier located just in front of the target billiard ball; this barrier absorbs the impact of the cue ball, but does not necessarily transmit the force of the cue ball exactly in the way this force was received. In fact, the billiard ball completely overrides the intended effect of the cue ball, if, for instance, the student decides to drop out of school. (Tom, 1984, p. 59)
An additional feature of the causal relationship said to hold between teacher behaviour and student outcomes is the assumption of a one way flow of influence. It is taken for granted that influence flows from the teacher to the student, and that the student is the more or less passive recipient of inputs from the teacher. Yet, as we shall see, a major component of teacher effectiveness research, classroom management, is predicated on the idea that students can have enormous influence on the tone of the classroom and the ease with which teachers can conduct their lessons.

In its various manifestations over time teacher effectiveness research has conceived of the direct relationship between teacher behaviour and student learning in one of three ways. At first researchers attempted to establish a linear relationship between what teachers do and what students learn. Later researchers then came to conceive of the relationship as either curvilinear or situational. The linear conception took for granted that the frequency of a teacher behaviour had a direct bearing on the extent to which learning occurs. Several prominent researchers in this field have conducted studies in which they counted the number of teacher behaviours thought to be of a type linked to effective teaching, then determined the correlations between these behaviours and a criterion of effectiveness like pupil gain scores or supervisory ratings. (Gage, 1963; Medley, 1973)

David Berliner (1976) admits he began to doubt the linearity of the relationship between teacher behaviour and
outcomes when during the Beginning Teacher Evaluation Study observers noticed the operation of individual intentions and interests competing with the teacher's efforts to reinforce positively a particular student behaviour.

In our classroom observations...we have seen positive verbal reinforcement used with a new child in the class, one who was trying to win peer-group acceptance, and whose behaviour the teacher chose to use as a standard of excellence. We watched silently as the class rejected the intruder, while the teacher's count in the verbal praise category went up and up and up. (p. 372)

It seemed evident to Berliner that the teacher praise in this case could not have had the intended effect, even if it were taken as genuine praise by the new student, since what counted for this student at the moment was being accepted by his new classmates. Berliner's response to this apparent flaw in the process-product view of learning was to suggest that the relationship between a certain teacher behaviour and student behaviour was in some cases curvilinear, or perhaps situational. That is, that after a certain point or under certain conditions a particular teacher behaviour will no longer be effective. As Tom points out however, this suggestion moves us into the realm of individual student perception and thinking, a move that leaves little room for a crudely behaviourist explanation of learning. Moreover it hardly seems adequate to imply that certain teaching behaviours cause learning, only to add that these may not work under all conditions, all of the time, since they depend on a measure of congruence between student intentions and purposes which cannot be fully specified in advance. The teacher's
effectiveness in Berliner's example depended on her insight into a student's motives. Unfortunately for teachers, a student's motives do not announce themselves unambiguously, and beyond a banal level of generality it is impossible to predict accurately what a youngster's motives will be. Even though certain teachers seem more able than others to interest their students in their lessons, we should not let this fact lead us into denying that individual interests, intentions, and preoccupations can at any time compete with a teacher's efforts at instruction. That there may be a different set of purposes and intentions for teachers and students is, of course, no secret. The large body of literature on the effects of student culture and school climate testifies to the influence of broadly held beliefs, and further points to the inadequacy of causal explanations of teacher influence.

Doubts of the sort expressed by Berliner are by no means rare in educational research. In fact, according to Doyle (1977) there is a large body of research devoted to such instances where "certain processes ... intervene in the relationships between teacher variables and student learning outcomes" and which clearly stand in opposition to this sort of causal conception of learning. (p.165) These include student motives in general, or student rejection of schooling (Anderson, 1970; Dole, 1977; Glick, 1968); the demands of the classroom environment (Cusick, 1973; Doyle, 1977); student attention to instructional tasks (Anderson, 1970); degree of student familiarity with the
intellectual operations employed by the teacher (Berliner, 1976); the classroom peer group (Schmuck & Schmuck, 1983; Zeichner, 1978); student perception of the fairness of the teacher in her dealings with students (Smith & Geoffrey, 1968); teacher ability to interpret for the child the significance of objects, events, and ideas (Feuerstein, 1980); and the structure of a subject matter (Bantock, 1961; Resnick & Ford, 1981; Shulman, 1974). While this is a finite list, it seems likely that the list could go on forever, given the conceivably vast number of choices a person is capable of. So despite the existence of a research literature that documents what already seems self-evidently true — that teachers can't cause learning in the way envisioned by Gage and his colleagues in the field of teacher effectiveness research, the notion that they can persists. And while it can be admitted that the question of whether and to what extent learning is caused by the actions of others remains an open one, it is clear nevertheless that teacher behaviour alone cannot be sufficient to cause learning.

One of the effects of our accepting this simple causal theory of learning is that a language of manipulation begins to seem reasonable. Tom notes that once the student is conceived as the passive recipient of "treatments" he becomes little more than a receptacle of inputs or an object to be manipulated. This carries a number of consequences. The student is to a large extent freed of any responsibility for learning; the responsibility therefore shifts elsewhere. In discussions of teacher effectiveness it shifts primarily to the teacher. In
discussions of social policy it frequently shifts to the state and its agencies, or to the family. Responsibility for learning, and therefore achievement, once it has been taken away from the student, can be moved about according to the context in which it figures. With each context a set of obligations faces different individuals, and with each set of obligations, opportunities for the exercise of power. What might explain a good part of the initial attractiveness of causal explanations of learning is that they expanded the area of influence for educators. If it can be shown that what teachers do is more important to student learning than was thought to be the case, teachers gain a certain amount of power. For instance, if it can be demonstrated that teaching behaviour accounts for learning more than family background does, then teachers gain, at the expense of others who wish to administer one intervention or another, a broader area of jurisdiction.

Foucault notes a similar state of affairs in the history of medicine and the emergence of criminal psychiatry. Throughout most of European history, murderers were routinely put to death once their guilt had been established. From the moment they were apprehended until their deaths, murderers belonged to the courts and to those who operated the penal system. Foucault, as we have indicated elsewhere, saw in the ceremonies of the penal system a ritualistic demonstration of this power over bodies. In the case of Pierre Riviere, an eighteenth century peasant who murdered his mother and sisters, Foucault noticed a dramatic difference in the
way he was treated. Instead of being sentenced to death as soon as his guilt was determined, Riviere became the object of a struggle between jurists and those we would recognize today as psychiatrists. In the end the medical men won out; Riviere was confined to an institute where he could be studied for the rest of his life. For these early human scientists, murderers whose motives were incomprehensible to ordinary people were too valuable a source of knowledge to be put to death. These "scientists" insisted "there are crimes which are our business, these people belong to us." (Foucault, 1980, p. 205) In these insane criminals doctors had a source of knowledge and power. Foucault can't help suspecting that the emergence of a discourse on homicidal mania was as much dependent on the fabrication of a pathology that would justify criminal psychiatry as it was on a genuine desire on the part of these early physicians to discover the truth about the darker side of human nature. The postulation that there was indeed a darker side to human nature, one which is beyond the apprehension of ordinary people, is in Foucault's view, one of the fundamental conditions of possibility for the emergence of psychiatry.

I would be tempted to say that there was, in fact, a necessity here ... linked to the very existence of a psychiatry which had made itself autonomous but needed thereafter to secure a basis for its intervention by gaining recognition as a component of public hygiene. And it could establish this basis only through the fact that there was a disease ... for it to mop up. There had also to be a danger for it to combat, comparable with that of an epidemic ... Now, how can it be proved that madness constituted a danger except by showing that there exist extreme cases where madness, even though not apparent to the public gaze, without manifesting itself beforehand through any symptom except for a few minute fissures, miniscule murmurings, perceptible only to the
highly trained observer, can suddenly explode into a monstrous crime. This was how the diagnosis of homicidal mania was constructed. Madness is a redoubtable danger precisely in that it is not foreseeable by any of those persons of good sense who claim to be able to recognize it. Only a doctor can spot it and thus madness becomes exclusively an object for the doctor, whose right of intervention is grounded by the same token. (1980, p. 205)

So power (parasitically drawn from the law) enabled doctors to isolate and examine murderers who would otherwise be executed, and the patient, grateful perhaps for a reprieve and a sympathetic ear, would by his discourse give forth the raw material for a new truth of homicidal mania, a truth which would justify the life of the patient, the activity of the doctor, and the new science of psychiatry.

The case of process-product researchers is clearly not so dramatic. They are not claiming they understand students or teachers better than anyone else; they are saying merely that they are gaining a clearer idea of the relationship between certain teacher behaviours and student learning which enables them to generalize about the best way to teach. That there is a best way to teach which is still somewhat of a secret opens up the possibility that whatever shortcomings there have been in student achievement might be due to ineffective teaching. It is this latter idea that school effectiveness advocate Ronald Edmonds (1979) has seized upon and which has proven so popular. If Edmonds is right about the reason for low achievement levels among the poor, then interventions by others - state legislators and bureaucrats concerned with equity, for example - may not be
as necessary as they once seemed.

It is worth noting, however, the particular way in which the causal theory of learning employed in process-product research leads to a paradoxical effect. On the one hand educators are given a broader field of operation than might be granted by those who see achievement in school closely related to family background characteristics. On the other hand, educators can then be held in a larger net of responsibility for any failures that might result from their efforts. It is for this reason that it would be more than prudent for educators to be sure of the adequacy of the research findings. For if the teaching practices which have emerged from the literature as being most effective lead to more dramatic declines in achievement than already exist, teachers will again face severe criticism, while children from low income families will once again look unteachable.

In much the same way that students are viewed as objects to be manipulated by teachers, the logic of this conception of learning calls for the manipulation of teacher behaviour, which of course means a manipulation of teachers. Tom points to this in a comment designed to show the passive role in learning given students in this research.

The billiard ball analogy also helps highlight the passive role attributed to the student by teacher effectiveness researchers. Just as pool players naturally think of how they can use the cue ball to move the other billiard balls to desired locations, so too do teacher effectiveness researchers think in terms of how teacher behaviour can be manipulated in order to obtain the desired results, that is, student learning.(p.56)
Insofar as this research is viewed with approval it carries a certain force. It sets norms for what constitutes effectiveness in teaching. And to disregard these norms may leave teachers and administrators in the position of appearing indifferent or out of date, an image they would likely be anxious to avoid. We can see from this, another instance of the manner in which power and knowledge operate. To the extent that teacher effectiveness research is respected, it is taken, at best, to be made up of defensible propositions regarding how to teach, or at least, recommendations about how to teach, that are worth a try. In between these two extreme views is the more common one that teacher effectiveness research offers prescriptions for teaching which are superior to anything else currently available. The need for educators to show that reform efforts in education are underway and effective, militates against a careful consideration of approaches to teaching that are less capable of pinpointing the successes and failures of teachers and their students. To the extent the research comes to represent the basis for teaching orthodoxy it expresses certain "truths" about learning and the best ways to teach. A teacher who ignores these truths is left appearing negligent. The teacher who refuses to accept at least some of the research findings can be seen to have an obligation to justify her position; a formidable task given the mountain of data and the prestige afforded empirical research. Research findings, insofar as they are taken to be true, have powerful effects.

Process-product research, then, has largely been the search
for a one best way, a set of generic teaching skills or behaviors that will serve equally well in all subjects at all grade levels. If such a set of behaviors were to be found it would then be possible to persuade teachers to teach in more or less identical ways. Under such conditions of formal equality it would also become easier to compare one teacher with another, especially if identical means of judging teaching success could be applied. When the means of judging teaching success are identical with the measures of student success (test scores) then it becomes a simpler matter to rank both students and teachers. It is in this way that the conception of learning found in process-product research begins to show its political usefulness. With teachers and students caught in the grid of visibility provided by test scores, they are vulnerable to coercion and control. They can conform to the norms of schooling that begin to emerge from test scores, or try to evade the panoptic gaze. Since test scores define effective teaching in the research, it is only a matter of time before they come to define effectiveness for school supervisors. By the same token, since test scores define acceptable standards in classrooms, schools, and districts, and are the outcome of the best way to teach, students who do poorly will, by definition, be abnormal.

It is instructive to trace the development of teacher effectiveness research since it reveals the manner in which pedagogy has evolved and the steps it has taken along the way. It is the evolution and development of this search for a one best
way that gives us a sense of the character of this search. In it a good many people have been willing to sacrifice a good deal of common sense in order to develop superior means of measuring reponsibility, in order to find ways of stripping away all the obstacles that separate the researcher from his ability to say with confidence -- "This is the cause of that."

Medley (1979) has identified in the history of teacher effectiveness research five successive conceptions of the effective teacher: one who possesses desirable personal traits; one who uses effective methods; one who is capable of creating and maintaining a good classroom atmosphere; one who possesses a repertoire of teaching competencies; and lastly, the professional decision-maker who possesses the requisite competencies plus the sense of when to apply them.

The very earliest research into teaching in America attempted to describe the characteristics of the best teachers by asking pupils what they thought distinguished good teachers from poor ones. One study from the mid 1890s asked pupils to list the characteristics of teachers they had found to be most effective. From the data researchers distilled four characteristics. The effective teacher was one who "makes greater demands of students," "has more teaching skill," "has more knowledge of subject matter," and "has better discipline." (Medley, 1979, 13) The implications for teaching practice seemed clear enough. Teachers should be strict, demanding, and knowledgeable. How they might also become skillful remained a problem, however. Despite
their commonsense appeal, there were doubts as well about the validity of these subjective judgements. It seemed likely that school age children would lack the maturity and objectivity needed to assess their teachers competently. It would be necessary to look elsewhere.

By the 1920s investigators looked to the opinions of expert judges. In one study from 1929, considered by Medley to be one of the best from that period, six characteristics of an effective teacher were identified: good judgement, self control, considerateness, enthusiasm, magnetism, and adaptability. (Medley, 1979, p. 13) By the 1930s characteristics such as these had been incorporated into teacher rating scales which were in widespread use as teacher evaluation devices. From this study emerged a picture of the characteristics school inspectors considered most important in teachers. These were "cooperation (helpfulness, loyalty), personal magnetism, personal appearance, breadth and intensity of interest, considerateness, and leadership." (Medley, 1979, p. 13) The chief problem with these characteristics, like those that preceded them, was their lack of specificity. Terms like personal magnetism, good judgement or teaching skill were hard to pin down, and too often meant different things to different people. As well there could be little room for such normative language in a discipline aspiring to be a science. So despite the widespread use of teacher rating scales, there was a general and growing sense that more exact ways of identifying the characteristics of an effective teacher were needed. As early as 1912, E.L. Thorndike saw the promise of
such precision in science:

Education, like history, economics, sociology, and the other sciences of man, is just beginning to give promise of quantitative knowledge, of descriptions of facts as numerically defined amounts, and of relations of laws in terms of rigid unambiguous equations. The changes that take place in intellect and character are coming to be measured with the same passion for clearness and precision, which has served the physical sciences for the last two hundred years. (Tom, 1984, p. 13)

Since the 1960s educational research has increasingly come to be expressed in such terms: quantitatively, where facts are presented as numerically defined amounts and relationships are seen in terms of rigid, unambiguous equations. Where it has not been possible to establish relationships of this sort promoters of these methods have called the relationships curvilinear or suggested that their effectiveness is restricted to specific contexts.

Medley notes that the first studies to compare various teaching methods proved inconclusive. These studies failed because the experiments in each case were designed to use the pupil (rather than the teacher) as the unit of analysis... It became apparent that sound research in teacher effectiveness must focus both on teacher behaviour (what the teacher does) and on pupil learning (teacher effectiveness). (p.15)

Hence there developed an interest in tracing from superior results the things teachers do when they teach, and from the relationships between teacher behaviour and achievement, a list of
competencies that could provide a standard by which other
teachers might be judged and to which pre-service teachers could
aspire. Such a project required the observation of teachers
while they taught, the recording of what methods they used, and
an examination of the relationships that seemed to hold between
certain teaching practices and test scores. It took a
surprisingly long time for researchers to realize this. Medley
(1979) notes that

systematic observation of behaviour in the
classroom as a means of studying the nature of
effective teaching was uncommon before 1960... As early as 1930 Barr and others advocated
the use of criteria of teacher effectiveness
based on measured pupil gains rather than on
expert opinion, but nobody seems to have paid
them any heed. (p. 15).

By 1970, however, there were more than 100 classroom observation
systems. Most of these had been developed for teacher training,
but some like that by Flanders (1970) were used extensively in
research. From this research it was possible for Rosenshine and
Furst to report in 1971 that certain teaching behaviours were
consistently correlated with student achievement gains.
Nevertheless, Rosenshine and Furst admitted these correlations
were not always significant and most often were only marginal to
moderate in strength. Rosenshine and Furst's review drew upon 50
correlational studies and from these emerged five variables for
which they felt there was strong support and six more for which
the support was less strong. These eleven variables were:
clarity of teacher presentation; variety of instructional
procedures and materials; teacher enthusiasm; task-oriented or business-like teacher behaviour; student opportunity to learn what is subsequently tested; teacher recognition and use of student ideas; criticism of students (negatively related to achievement); teacher use of structuring comments; varied types of questions; teacher probing of student responses; and student perception of the difficulty of instruction. By 1979, Rosenshine's analysis of subsequent research on the eleven variables showed that only two variables correlated satisfactorily with gains in student achievement. These were task-oriented teacher behaviour and student opportunity to learn what will later be tested. (Tom, 1984,43) From the 50 studies reviewed by Rosenshine and Furst in 1971, Medley in 1979 could find only 14 significant studies and these showed correlations of only .39 or lower. Gage in 1984 reported that correlation coefficients in this research typically have values between .2 and .5. (p.89) In the most recent review of the literature on teacher effectiveness, Brophy and Good (1986) admit that "even the most generally replicated findings tend to be based on low to moderate correlations, and are not always strong enough to reach statistical significance" (p.360) It is difficult to see why such research is so influential.

What is also remarkable about these findings, expressed in terms of inputs, outputs and variables, is their normative nature. Virtually none of the variables can be said to be objective or stable over time. What, for example, is meant by task oriented teacher behaviour? Clearly not any old task will qualify as one
of the sort of tasks we might want to associate with scholastic achievement. What sort of activity or lesson do the researchers wish to disqualify when they speak in this way?

Despite the inadequacies of these studies, there has emerged from them a model of instruction known as direct instruction. From two variables -- task oriented teacher behaviour and teaching to the test -- Barak Rosenshine (1979) developed a model of instruction which has in the literature increasingly come to stand for the single best instructional method. (Peterson, 1979, p. 58; Good, 1979, p. 60; Good, Grouws, & Ebmeier, 1983; Stevens & Rosenshine, 1981; B.O. Smith, 1980)

Rosenshine defines direct instruction in this way:

Direct instruction refers to academically focused, teacher directed classrooms using sequenced and structured materials. It refers to teaching activities where goals are clear to students, time allocated is sufficient and continuous, coverage of content is extensive, the performance of students is monitored, questions are at a low cognitive level so that students can produce many correct responses and feedback to students is immediate and academically oriented. (1983, p. 38)

It should be noted that many of the words in this passage are purely evaluative. It is not clear what is meant by clear goals, sufficient time, or extensive coverage of content. Despite these problems, the message is clear enough in other ways. Put simply, the teacher is to keep a close eye on her students to see that they pay attention to their school work, which is usually identical and restricted to what is called in the literature
"basic skills acquisition", but which may in fact be something else altogether, something like excessive practice and rote memorization.

Rosenshine emphasizes the teacher-centered focus of direct instruction and points to the main reason for limiting the class to a single activity at any one given time.

Although it would still seem important for students to have a variety of ways to learn the same material, classrooms in which a variety of activities are occurring simultaneously are usually more disorderly, and students are less attentive because the teacher is unable to monitor all activities. Students, as we shall see, are more task oriented when they are being supervised. (Rosenshine, 1979, p. 30)

Rosenshine further reports that researchers came to this conclusion after they had carefully recorded the variations in the amount of "academically engaged time" different teaching practices seemed to produce. According to the findings of phase III of the Beginning Teacher Evaluation Study:

In the second grade, an average of 122 minutes (42 percent) are allocated to all academic subjects. The remainder of the time is spent in nonacademic activities (story time, breaks, lining up, taking seats, quieting down, collecting and distributing money or material, making arrangements for events, free time, music, art). If we take the engagement rate into account, the average student in the second grade is engaged in academic activities for about eighty-nine minutes. (Rosenshine, 1979, p. 37)

Such an interest in academically engaged time echoes Foucault's emphasis that in a disciplined environment "time
measured and paid must be time without impurities or defects; a
time of good quality, throughout which the body is constantly
applied to its exercise". And Rosenshine's insistence on class-
wide activities which permit easier supervision recall Foucault's
notion of the panopticon and its normalizing gaze. In fact a
considerable amount of the emphasis of teacher effectiveness
research has to do with classroom management and the control of
students.

Brophy in 1983 began his review of the literature on
classroom management techniques with the assumption that "good
classroom management implies good instruction".(p. ) At first
it seems Brophy is making the commonsense claim that a noisy
classroom, full of distractions, does not represent an ideal
learning environment, and that to the extent teachers can
convince their students of the need to be co-operative and
diligent they should make every reasonable effort to do so. But
as Brophy continues it becomes clear that he is making a
different and much larger claim, namely that poor instruction,
because it is poor, promotes unruliness among students, and that
good instruction, because it is effective, promotes co-operation
and "task-oriented behaviour". Students who disrupt the class or
defy the teacher are doing so, in this view, because the poor
quality of instruction leaves them bored or confused, or simply
insufficiently engaged in their work. It is evident that for
Brophy good instruction consists in, among other things, a set of
organizational principles which permits easy supervision of
students, easy tasks, and a steady pace of instruction which
leaves students little time to behave inappropriately. "Because successful classroom managers maximize the time their students spend engaged in academic tasks, they also maximize their students' opportunities to learn academic content, and this is exhibited in superior performance on achievement tests."
(Brophy, 1983, p. 266)

It is difficult to know what sort of claim Brophy is making. If he is making an empirical claim then his evidence in support of it is, as we have noted, quite weak. Because the chief empirical support for this view is the correlation between the amount of academic engaged time and test scores, the generally weak to moderate correlations that typify this research raise some doubts about the claim's validity. And again, the conceptualization of learning that undergirds this account is, as it stands, completely unconvincing. Gains in learning, it seems clear, do not stem from increases in the amount of time students are confronted with books, films, or other materials; nor are they the product of the number of minutes for which teachers can require students to sit quietly during lectures. In this literature, a student's being engaged in an academic task requires little more than that he not behave in a way that seems overtly unproductive.

In his description of how researchers came to this conclusion, Brophy mentions the landmark study by Kounin (1970) that serves as the foundation for the view that good teaching

74
minimizes discipline problems in the classroom. In this study Kounin videotaped two types of classes -- "smoothly functioning classrooms" where students were attentive to presentations and chaotic classrooms where they were not. Brophy's comments on this study are instructive.

Kounin began by analyzing the teachers' methods of dealing with misconduct and disruption. Considering the great differences in management success displayed by these two groups of teachers, the researchers expected major differences in methods of dealing with misconduct. To their surprise, they found no systematic differences at all. Good classroom managers were not notably different from poor ones when responding to student misconduct. The two groups of teachers differed in other ways, however. In particular, the effective classroom managers systematically minimized the frequency with which students became disruptive in the first place.

(1983, p. 266)

What sort of evidence would enable an observer to conclude that differences in the way students behaved in orderly, as opposed to disorderly classes, are linked to observable differences in the actions of their teachers? For Brophy it is enough that there be a coincidence of teacher "behaviour" in orderly classrooms. Let us examine Brophy's explanation of why one set of classes was better behaved than the other.

Brophy observes that "effective classroom managers systematically minimize the frequency with which students [are] disruptive in the first place." (p.267) He identifies a number of Kounin's management techniques as being central to the effective classroom: "withitness", overlapping, signal continuity and lesson momentum, group alerting and accountability in
lessons, and variety and challenge in seatwork. The effective teacher, he suggests, is one who possesses "withitness". Teachers who are "with it" prevent problems before they can become disruptions, with the result that when students see their teachers are "with it" they are less likely to become fractious. Effective teachers, then, are those who prevent small problems from escalating into major disruptions by monitoring the classroom regularly, "stationing themselves where they could see all of the students continuously". (p. 267) Similarly, "overlapping", another of Kounin's techniques, consists in the teacher being able to do more than one thing at once, such as being able to monitor the whole class while conferring with a single student. (Brophy's example) Signal continuity, on the other hand, is that quality found in lessons which are characterized by a steady pace, free of interruptions or distractions. Thus the teacher who maintains signal continuity is one who comes to class prepared and organized and who refuses to let student disruptiveness disturb the lesson. In Brophy's words, signal continuity is effective because students tend to be attentive when they are presented with a continuous academic "signal" to attend to. Problems begin when students have no clear "signal" or task to focus on, and these problems will escalate the longer the students are left without such a focus. (p. 267)

Effective teachers also use presentation and questioning techniques to keep students alert and accountable. These techniques include "looking around the group before calling on someone to recite, keeping the students in suspense as to who
would be called on next by selecting randomly, getting around to everyone frequently..." (Brophy, 1983, 267) The idea here is for the teacher to distribute questions in such a way that no student can be certain he will not be called upon next. In this way the student will pay attention in order to avoid the unpleasant consequence of being found out. Lastly, effective teachers, according to Kounin, are those who provide their students with varied and challenging seatwork. It is assumed here that "the appropriateness and interest value of the assigned work will influence the quality of task engagement". (p. 268) Appropriate seatwork in this instance is that which is "easy enough to allow successful completion but difficult or different enough from previous work to challenge each student". (p. 268) Brophy notes that more recent research suggests that challenging seatwork does not lead to an orderly learning environment as much as does easy seatwork. He adds that when the teacher is available to give immediate feedback, teachers should aim at a level of difficulty that ensures success rates of 70%-80%. When students are to work independently the level of difficulty should be lowered to ensure success rates of 95%-100%. (p. 268) Given that an additional feature of this literature is an emphasis on teaching to the whole class it seems likely that whole class instruction and class wide sets of seatwork - worksheets and textbooks - will characterize activities. It seems equally likely that instructional materials or methods that ensure success rates of this magnitude for the least able and most able at the same time are bound to be addressed to the least able. Brophy stresses
that "confusion about what to do or lack of even a single important concept or skill will frustrate students' progress and lead to both management and instructional problems for teachers". (p.268) This is effective teaching?

It should be noted in his account that one must accept beforehand that the teachers in the orderly classroom were responsible for the class being well-behaved, attentive, etc. One must accept in advance what this account is intended to demonstrate, namely that attentive, hard working students are the product of certain teacher behaviours. Given the same set of circumstances: disruptive, rebellious students in one set of classes and attentive, cooperative students in the other, different observers could have just as easily drawn the conclusion that one set of classes was made up of difficult students who for one reason or another did not want to cooperate, and another made up of students who for one reason or another had decided to be cooperative and attentive. When for one reason or another these students became distracted or restive, a mere glance from the teacher, or the knowledge that this teacher asks everybody questions would be sufficient to ensure their attention, or perhaps only their passive cooperation. Brophy's account, while it does have a surface plausibility, is only truly convincing when a conception of students as active, purposive beings is absent. His students are not given credit for acting in ways that teachers, at least, would deem commendable. It is as if they had little choice, for if these students did wish to defy the teacher and do as they pleased it is hard to imagine how the...
following techniques used by successful classroom managers could have subdued their wills. Those teachers who displayed "greater management success" were those who employed teaching practices that did little more than keep students busy and allow for more or less continuous teacher supervision. Even so, the students would need to decide to keep busy and stay out of trouble in order for this teacher to appear effective. In classrooms where several students view getting in trouble and bothering the teacher as being more interesting than their schoolwork, Brophy's "effective classroom managers" would have their hands full, one suspects.

These prescriptions appear to boil down to this: closely supervised classrooms are effective classrooms, especially when students are aware their teachers are capable supervisors, and secondly, busy classrooms are effective classrooms. One of the costs of keeping students busy, at least during periods of seatwork, is a level of difficulty that can only be described as trivial.

This account of the well managed classroom depends for its intelligibility on a conception of the student as one whose tendency is to resist instruction at every turn; hence the importance of signal continuity and trivial levels of difficulty. There can little excuse for the student who is acting out, since the work isn't too difficult and there is always something to do. Nevertheless, this picture of the reluctant learner is contradicted by the image of the student nervously paying
attention to his teacher's every word for fear of being asked a question for which he is unprepared. It seems quite reasonable to suppose that a student could be completely indifferent to a lesson many of his classmates found interesting, or to suppose that another student would want to be seen defying her teacher. In such circumstances it is hard to see how teacher eye contact, randomly distributed questions, or any other form of teacher "withitness" are likely to make the sort of difference Brophy claims they will. Unless, of course, a range of more powerful penalties and rewards than currently exist in the schools are introduced. Without significant penalties and gratifications the disciplinary apparatus that these techniques amount to lack much force.

It has been observed that one result of the wide-spread acceptance of the prescriptions contained in teacher effectiveness research is that

repetitive low-level intellectual skills are now surrounded by a halo of legitimacy. Filling in blanks, getting test-wise to multiple choice items and completing exercises elevate tedious tasks to the status of effective instruction.... Learning [has become] a series of repetitive tasks that need to be completed, placed in folders, and marked by the teacher. (Cuban, 1984, p. 131)

In response to questions concerning his historical interest in prisons, poorhouses, military barracks and the like Foucault has observed that historically these, and schools as well, have been sites where work has had a largely symbolic function.
Unlike work in factories and offices, schoolwork is not productive of goods or services that can at some future time be purchased. And while there is in the capacities and abilities that are "produced" in school an element of value similar to the exchange value of goods and services, much of what students do in school, Foucault insists, can be viewed as a form of dressage, like military drill; justified in practical terms that seem unconvincing, but nevertheless widely taken as evidence of efficiency and effectiveness. The real point of this sort of work is to make obedience automatic, to remove independent judgement and will, and to make, in the case of military drill, the movements of individuals identical to one another. That such a thing is possible and even common in armies seems likely to be due to the compelling reasons soldiers have for following orders. No such compelling reasons exist in school, however.

A system of instruction which depends on supervision and inspection, broadly conceived, that is in some ways little more than a system of supervision and inspection, is a system which is unlikely to promote literacy, numeracy or any other of the sorts of capacities or abilities that commonly stand as the goals of education. In a manner of speaking, the classroom of the "effective" teacher is, in the Foucauldian sense, a completely disciplined environment. The central characteristics of such an environment -- hierarchical observation, normalizing judgement and examination -- define the effective classroom. Classrooms are organized, and activities are selected, on the basis of the degree
of visibility they allow. The teacher must be able to see what everyone is up to. This "seeing" takes various forms: simple observation of students at work to the detailed examination that comes from moving questions around the class. Reading kits and mathematics notebooks consist of questions which can be quickly answered and easily marked. Ranking and ordering students becomes a relatively simple matter. Scores and progress can be recorded and where based on results from standardized tests can be used by teachers and their superiors to make comparisons and evaluations. The school that contains "disciplined classrooms" is "a sort of apparatus of uninterrupted examination that duplicate[s] along its entire length the operation of teaching. It [is] less and less a question of jousts in which pupils [pitch] their forces against one another and increasingly a perpetual comparison of each and all that [makes] it possible both to measure and to judge." (Foucault, 1984, p. 198)

Deborah Meier believes that in reading instruction the current recipes for effectiveness are turning the schools into "coaching institutions which are preparing their subjects for the tricky world of tests (the "real world")." (Meier, 1981, p. 460) In her view, test coaching masquerading as reading instruction is one effect of the conviction that test scores are the best indicators of effective instruction. As this conviction grew in the late 1970's she felt teachers were being pushed into programs aimed at ever narrower and more trivial subskills that, it was hoped, would show up quickly on paper-and-pencil tests. The heart of a good education -- respect for subject matter and
intellectual inquiry -- was sidetracked in favor of prescribed kits filled with disguised reading tests, hundreds of one paragraph reading "tasks" followed by multiple choice questions. (Meier, 1981, p. 460)

Reading instruction was beginning to become a process of nearly continuous testing, insofar as prescribed reading programs were adhered to and student work graded. What troubles Meier about this trend is the likelihood that as a result of teaching reading in this way students will not learn to read. In her view reading requires searching for meaning in the text in such a way that errors of a certain type are likely. But these errors, she suggests, are noticed by capable readers because of their impact on meaning.

Redundancy and involvement in the meaning of the material generally corrects the errors that good readers make as they proceed fairly rapidly through the written page. But, reading tests, which involve fairly short and generally pointless passages followed by trick questions and answers, require being constantly alert to precisely the kinds of errors that are unimportant in most real reading. Also, the heavy emphasis on phonics, syllabification, and pronunciation in most tests for young children require schools to overemphasize slow oral reading, the conscious mastery of phonic rules and word perfect, pronunciation-perfect reading. (p.460)

The irony of our dependence on test outcomes in reading instruction, says Meier, is that this dependence is leading us further away from the promised end of better readers and toward ever more deepening concern over illiteracy. Given the logic of effectiveness as it is currently defined, it seems likely that the proposed solution to the literacy crisis will be more of the same: an emphasis on reading subskills, standardized and norm referenced reading materials, and recall questions. Perhaps of
even greater significance is the likelihood that along with this sort of emphasis will be an increased reliance on tests as measures of student performance and instructional effectiveness.

Although the direct instruction model is plainly derived from research in primary level reading and computation "skills" of the most basic kind, Good, like Rosenshine, claims that "in comparison to other available treatments (or at least those conventionally present in classrooms), direct instruction may have superior general effects for all types of students". (Good, 1979, p. 60) Tom (1984) rejects the idea that direct instruction is the best instructional method currently available. Because the direct instruction model emerged from research on a particular classroom context, that of reading and mathematics achievement in the primary grades "there is neither a conceptual nor an empirical basis to support the generic potential of direct instruction". (p.48) While there may be a number of low level empirical generalizations that can be developed from this research it seems hardly adequate to greet what is little more than a cluster of factual correlations as if it were a model of the one best way to teach.

Duncan and Biddle (1974) in their review of the process-product literature complained of the near universal tendency of researchers "to make educational prescriptions based on untested theoretical commitments rather than convincing empirical data." (cited in Brophy and Good, 1986, p. 332) In addition they noted
that in too many cases the research did not show effects that were strong or independent of other effects; nor did the research show that effects applied over a wide range of teaching contexts. More importantly, it was clear that in a great many cases researchers did not understand why the effect took place.

(p. 332) Although it is not always evident in the reviews of the literature, especially those that conclude with advice for teachers, the bulk of the research findings apply only to very few populations and circumstances. Brophy and Good caution that the findings are drawn almost exclusively from basic skills instruction in mathematics and reading at the primary level and not surprisingly tend only to apply in similar circumstances. In addition they note that the instructional practices seen as most effective in the literature, do not relate to achievement gains for high SES students as much as they do for low SES students, especially in the case of students beyond the third grade level. (p. 337) One wonders, therefore, what could have led Good to say that direct instruction has "superior general effects for all types of students."

Management techniques identified in the literature as most effective in producing gains in achievement are similarly restricted in their applicability. "Within any particular study, gains on lower-level objectives were associated primarily with recitation, drill, and other low-cognitive-level, high teacher focus activities, and gains on tests of higher-level skills were associated more with discussion and other activities offering more pupil freedom". (p.337) Nevertheless in the same article
Brophy and Good proclaim

the fund of available information on producing student achievement (especially the literature related to the general area of classroom management and to the subject areas of elementary reading and mathematics instruction) has progressed from a collection of disappointing and inconsistent findings to a small but well established base. (p. 316)

Brophy and Good are here going beyond the data they take considerable pains to describe in their review. From their remarks it seems as if the research has uncovered a number of superior methods for producing student achievement in general, and that these methods are particularly effective in the areas of classroom management and elementary reading and mathematics instruction. What their review has shown however is that there are relatively weak relationships between the teaching and management practices favoured by teacher effectiveness researchers and a particularly narrow definition of achievement, and that these relationships hold primarily, not especially, in the areas they mention: basic operations in reading and mathematics in grades one to three. Researchers believe it is likely that these methods will be as effective in other contexts but at this point lack the data to demonstrate this.

What appears to be at work here is the justification of a particular way of dealing with students in school by means of looking for some definition of learning, achievement, reading, etc. that will make this approach look "effective". For this attempt to begin to appear successful a great many conceptual
weaknesses must be ignored and the spurious nature of the evidence overlooked. Do these oversights amount to anything? Do they point us to any conclusions other than the usual complaints about appalling intellectual standards in faculties of education? They do seem to lead us to conclude that a great deal more work of better quality needs to be done before we can say we have discovered effective teaching methods. It may very well be that the search for certainty in education can only lead us to this sort of conclusion. The search for Thorndike's rigid unambiguous equations has led us to remove intention and will from a conception of learning and thereby make objects of students; it has encouraged us to redefine achievement, success, effectiveness in ways that serve the research paradigm more than it does the enterprise of educating children. And as a result we are left with teaching methods which while they grow in popularity diminish the likelihood our children will become literate, thoughtful adults. The benefactors of this state of affairs may well be the social scientists who, using the crisis in education as the grounds for their research, are given an opportunity to refine their methods and bring people more fully into view and therefore more subject to control. A research program as flawed as teacher effectiveness seems to be is likely to generate years of research. Virtually every report, review, or assessment of the literature called for more work in one way or another. Very few (Tom is among them) suggested abandoning the current approach.
CHAPTER FOUR: SCHOOL EFFECTIVENESS RESEARCH

As we have seen, the tradition of teacher effectiveness research is a long and durable one. Only recently has it been overshadowed and absorbed by larger scale analyses that commonly fall under the category of school effectiveness research. While to a large extent current research into school effectiveness is concerned with improving all schools, it was originally the case that school effectiveness research grew out of the conviction in the late 1960s that for the poor and socially disadvantaged, schools did not make enough of a difference to be able to alter income distribution patterns or existing social categories. This conviction had its source in a number of studies: some large scale empirical ones like the Coleman report; later analyses by Christopher Jencks, Averch et al, Mosteller and Moynihan; and even later work by Bowles and Gintis. Arthur Jensen's work on the hereditary basis of intelligence and the questions which his work raised regarding the educability of blacks added to the already pessimistic atmosphere. But mainly it was the Coleman report that upset the applecart.

In 1954, in Brown v. The Board of Education, the Supreme Court of the United States repudiated the separate but equal doctrine contained in the 1886 decision from Plessy v. Ferguson. Plessy v. Ferguson had upheld the constitutionality of a Louisiana law which required railroad companies to provide "equal but separate accommodations for the white, and coloured races." Plessy, until Brown, had provided the necessary legal
Relying on research by sociologists and child psychologists, Chief Justice Warren in Brown observed:

Segregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of law; for the policy of separating the races is usually interpreted as denoting the inferiority of the negro group. A sense of inferiority affects the motivation of a child to learn. Segregation with the sanction of the law, therefore, has a tendency [to retard] the educational and mental development of negro children and to deprive them of some of the benefits they would receive in a racially integrated school system. (Strike, 1983, p. 192)

Because the stigma attached to segregated schools for blacks was felt to destroy the motivation of black children to learn, Justice Warren saw that segregated schools led to inequality of educational opportunity and, by extension, inequality of opportunity in life. Justice Warren concluded:

In these days, it is doubtful that any child may reasonably be expected to succeed in life if he is denied the opportunity of an education. Such an opportunity, where the state has undertaken to provide it, is a right which must be made available to all on equal terms. (Strike, 1983, p. 192)

The Court's decision -- that segregated schools are inherently unequal and produce an unequal distribution of social and educational goods -- determined that segregated schools were illegal constitutionally. This decision in turn legitimated the view found throughout social science literature that many if not
most, of the incapacities for learning that plague children are the result of the child's background. It is worth noting that the Brown decision made segregation illegal on the basis of controversial sociological and psychological claims. From the standpoint of those interested in promoting equity the power of this particular body of sociological and psychological discourse had positive effects that overturned the effects of the earlier legal decisions found in the Plessy v. Ferguson ruling.

Despite the apparent finality of the Supreme Court ruling in Brown, the case was not considered closed by everyone. Legal authority Edmond Cahn noted in a law review at the time of the Brown decision that "I would not have the constitutional rights of Negroes -- or of other Americans -- rest on any such flimsy foundations as some of the scientific demonstrations in these records." (Strike, 1983, p. 193)

In the Brown case, despite the protests of skeptics, social science knowledge combined with the power of the law in such a way that the result profoundly altered existing relations between the races, and between the schools and their primary clients. Moreover, this decision added impetus to the already growing conviction among liberal social scientists that the state had a legal and constitutional obligation to overcome the sort of incapacities identified in Brown as being products of circumstances over which the state had some control. Alongside this conviction, perhaps even the source of it, was a relatively new sociological perspective that had emerged as America was
transformed from a set of internally focussed, local communities to a nation in which the focus was national, significantly so in the marketing, distribution, and advertising of goods. Coleman (1980) observes of this period:

Washing machines, made by a few national firms, [had come to replace] washtubs and washboards, made and sold locally. Breakfast cereals, heralded by entrepreneurs with persuasive public relations skills, [came to replace] locally cracked wheat and oats cooked by mothers each morning ....One of the consequences of this change was that a new set of sociological problems emerged. These were problems related to the national markets and national audiences -- in short, problems of market research and audience research. (p.335)

From these problems arose new methods -- random samples, national samples, "questions that were appropriate to the population as a whole". (p.336) In this way a sociological perspective was born, and along with it, sociologists with an interest and some expertise in measuring various aspects of the national body. The conviction that the state had an obligation to intervene in the matter of equal opportunities, plus the existence of the social science apparatus capable of gathering and sorting vast amounts of data made it possible to conduct the sort of mammoth study undertaken by Coleman in the mid-sixties to determine the extent of a "lack of availability of equal opportunity" in American public schools.

In Foucauldian terms, we have, with the development of these new sociological tools, "the production of effective instruments for the formation and accumulation of knowledge -
methods of observation, techniques of registration, procedures for investigation and research, apparatuses of control".  
(Foucault, 1980, p. 102) In this manner developed sites for the production of certain type of knowledge. Not only would this knowledge pertain to characteristics of the social body in general, it would pertain to the general characteristics of the social body, the norm. The establishment of norms, as we have noted, can allow a process of normalization to begin. Rabinow (1984) interprets Foucault's sense of normalization in the following way:

By normalization, Foucault means a system of finely gradated and measurable intervals in which individuals can be distributed around a norm - a norm which both organizes and is the result of this controlled distribution. A system of norms is opposed to a system of law or a system of personal power....Normative order is an essential component of the regime of bio-power, for "a power whose task is to take charge of life needs continuous regulatory and corrective mechanisms.... Such a power has to qualify, measure, appraise, and hierarchize...it effects distributions around the norm.  
(Rabinow, 1984, p. 20)

Foucault has noted of the relationship between social science discourse and the law that a process of systematic normalization of the law took place in the 19th century as medicine, psychiatry, and other social sciences entered legal debates. In these debates there was a tendency to rely on statistical measures and judgements about what was normal (or healthy) and what was abnormal (or unhealthy) for a given population. Foucault observed that this process of normalization served a key role in the creation, classification, and control of
anomalies in the population. The "purpose" of these anomalies becomes clear when one considers the claims of their promoters. The first claim is that certain disciplines serve to isolate anomalies; the second that one can then normalize or cure the anomalies by means of the corrective or therapeutic procedures developed by these disciplines. In this way are dangerous or counter-productive tendencies checked.

In both *Discipline and Punish*, and *The History of Sexuality* Foucault goes to great lengths to show that "the advance of bio-power in the nineteenth century is in fact contemporary with the appearance and proliferation of the modern categories of anomaly - the delinquent, the pervert - which [the disciplines] are supposedly designed to eliminate but never do." (Rabinow, 1984, p.21) It is not that people in the 19th century cared for delinquents, or madmen, or perverts only to the extent they could be used to refine methods for controlling them. Rather Foucault is arguing that one of the effects of identifying these people as types who needed to be isolated, studied and treated was the development of ways to control people on grounds that could be seen as socially acceptable and morally defensible.

The mechanisms of the exclusion of madness, and of the surveillance of infantile sexuality, began from a particular point in time, and for reasons which need to be studied, [began] to reveal their political usefulness and to lend themselves to economic profit, and that as a natural consequence, all of sudden, they came to be colonized and maintained by global mechanisms and the entire State system....The bourgeoisie is interested in power, not in madness, in the system of control of infantile sexuality, not in that phenomena itself. The
bourgeoisie could not care less about delinquents, about their punishment and rehabilitation, which economically have little importance, but it is concerned about the complex of mechanisms with which delinquency is controlled, pursued, punished, and reformed etc. (Foucault, 1980, p. 101-102)

So then it is in the development of technologies of discipline within the social science disciplines themselves that Foucault sees the workings of power/knowledge.

In 1964, the US Congress commissioned University of Chicago social scientist James Coleman to determine the extent of restrictions to equal opportunity in American schools. Despite Coleman's expectation that he would find evidence of glaring inequalities in the quality of school buildings, class sizes, texts and trained teachers, he found few such inequalities. More surprising was the finding that differences in school-related variables did not account for the differences in scholastic achievement as much as did students' socioeconomic background. Also, the school variable which seemed to have the most effect on student achievement was the socioeconomic status of the peer group. The unequal distribution of educational resources as they had hitherto been conceived did not, then, seem to contribute significantly to inequality. Rather it was the socioeconomic status of students and how these students were distributed within the school system that seemed significant. The policy implication seemed clear. In order to ensure equality of opportunity in education, and to a large degree in the competition for jobs outside of schools, socioeconomic and racial integration would be
necessary. In the words of philosopher Kenneth Strike, "students, therefore, [were] considered an educational resource to be equitably distributed." (Strike, 1983, p.197)

In 1972, Christopher Jencks, in a follow-up study of the Coleman data, concluded that schooling accounts for little of the variation in cognitive skills among students, and that moreover, differences in cognitive skills are not so closely linked to future income as was commonly believed. As he put it "equalizing the quality of high schools would reduce cognitive inequality by one per cent or less;...additional school expenditures are unlikely to increase achievements, and redistributing resources will not reduce test score inequality." (cited in Rutter, 1983, p.1)

Whereas Coleman had argued that socio-economic status was a better predictor of school achievement than educational resources, thus casting some doubt on the efficacy of schools to develop cognitive abilities, Jencks' study led many to suppose that "schools don't make a difference". What faith in schools that remained in the wake of the Coleman report was shaken even further by Jencks' conclusion that:

as long as egalitarians assume that public policy cannot contribute to economic equality directly but must proceed by ingenious manipulations of marginal institutions like the schools, progress will remain glacial. If we want to move beyond this tradition, we will have to establish political control over the economic institutions that shape our society. This is what other countries usually call socialism. (Jencks, 1972, p.265)
The implication that educational opportunities had little
to do with attaining equality of condition struck right at the
heart of one of the state's primary reasons for mass public
schooling. Given the prevailing belief that education was the
fundamental means by which a democracy could provide equal access
to good jobs and a decent standard of living, the findings by
Coleman, Jencks et al, represented a serious blow to the prestige
of educators and brought into question the orthodox view that
providing equal access to educational opportunities was an
effective way to promote equality. Jencks, as we have seen, drew
the conclusion that only income redistribution was likely to
prove an effective means for promoting equality. In his view,
continuing to rely on schools to redress imbalances in wealth and
influence could only be viewed as utopian since "you cannot have
equality of opportunity without a good deal of equality of
that a main effect of these investigations into the relationship
between schooling and individual prosperity was the growing
conviction that social science findings supported the abandonment
of social reform. He feared that "this new skepticism which is
eroding the confident liberal assumptions could be distorted and
used to rationalize a second period of indifference in a nation
once again weary of the stress of reform." (p. 43) Significantly,
however, both Coleman, and to a lesser degree, Jencks suggested
that there was some point to tinkering with the school system in
an effort to improve social conditions affecting the poor. Both
admitted that racial integration in schools could contribute to reducing achievement differentials between higher income whites and poor blacks.

It was not long before competing bodies of educational research would attempt to show that schools were not the marginal institutions Jencks claimed they were; there was too much at stake for it to be otherwise. Removing from education its traditional promise of social mobility for those with merit upset liberals and conservatives alike. Hodgson (1973) notes that top level bureaucrats in the Nixon administration such as Patrick Moynihan advised the president "that enormous expectations had built up [in the US] that you could achieve racial equality through compensatory education, and it was not working. Point two: a proposition had been put forward by Dr. Jensen which democracy could not live with. Therefore, point three: you had to move directly to income distribution." (p.44)

This was not the sort of news President Nixon wanted to hear. Policy-makers were just beginning to feel the pressure to lower spending levels that has since become a near permanent feature of the political scene. Growing concern over the size of government and the tax burden associated with this growth called for a different sort of research finding. What was needed was research which would suggest that what failures there were in the educational system could largely be placed at the feet of indifferent or incompetent teachers, and that an effective school system for the poor needn't be expensive or politically divisive.
Certainly, this suggestion did not appear as plainly as this. If it had it is unlikely that educators would have embraced the school effectiveness literature as they have. It was the gloomy implications of Coleman's, Jencks' and Jensen's work that paved the way for the enthusiastic reception school effectiveness research would enjoy. As well, Coleman's conclusions were simply contrary to the experience of thousands of teachers working in schools. Despite the mountain of data produced by the Coleman report, the conviction that Coleman had to be wrong would not go away.

As we have noted, the bleak appraisal of the schools' ability to "make a difference" typified by the work of Coleman et al was not accepted willingly by those who viewed the schools as one of the few avenues of social mobility available to the poor and disadvantaged, and whose own futures were tied up with the maintenance or expansion of the public school system. To a number of prominent educators it seemed likely that, if unopposed, the new social science orthodoxy with respect to school effects would ensure systematic indifference to low achievement standards in the schools of the poor. Educational researchers in England and in the US, wondered, in light of the existence of effective schools already serving low SES populations, whether or not those working within the school system might not be simply failing to do the job with which they had been charged and of which they were capable. They may have, as one of these researchers, Ronald Edmonds, has suggested, been blinded by the "pernicious social
science notion that family background is the principal cause of pupil acquisition of basic school skills." (Edmonds, 1979, p. 16) Working on just such an assumption, Weber, in his 1971 study of inner city schools, explicitly sought to counter the work of Coleman and Jensen, who, according to Edmonds, "had satisfied themselves that low achievement by poor children derived principally from inherent disabilities characterizing the poor." (p. 16) Edmonds' uncharitable view of teachers can be summed up by his often asserted charge that "schools teach those they think they must and when they think they needn't, they don't." (Edmonds, 1979, p. 16) Since it is the people who work in schools who do the teaching and not the schools themselves, it was clear from Edmonds' remark that he thought the low levels of achievement among black students were attributable to those teachers whose thinking had become muddled by "pernicious social science notions" and who as a result felt themselves to be off the hook with respect to low income students. With school personnel blinded in this way, it seemed that educators had relieved themselves of their responsibility to be instructionally effective, and had concluded that the children of the poor were more or less ineducable. Jean Anyon discovered just such an attitude from a teacher in an inner city school who bemoaned the fact that "you can't teach these kids anything". (Anyon, 1979, p. 36) With such low levels of expectation it seemed clear to this new generation of researchers why poor children fail at school: teachers aren't trying to teach these children because they don't believe in their ability to learn. In turn, children come to believe that they cannot learn and a vicious circle is
completed. Instead of taking responsibility for teaching basic skills, teachers, following the advice of educational researchers or the lessons learned in teachers' college, engage their students in trivial "vocational" activities or ill-advised therapeutic efforts at enhancing self esteem. "Playing in the sand becomes science, and games become mathematics...[Teachers] physically touch little children and display great positive affect toward potentially dissident bigger ones." (Meyer, 1980, p. 55)

In an attempt to establish a sound empirical basis for the belief that schools do make a difference, effective schools researchers, since the early 1970's, have re-examined the issue of school effects on student outcomes in the hope of identifying the characteristics of schools that are effective. Since then this research tradition has grown in influence to the point where it is sometimes recognized as the school effectiveness movement, attractive not only because it reaffirms for teachers the belief that they really do have more than a custodial role, but also because it seems from the research that to have an effective school is within everyone's grasp, given the right attitudes.

Perhaps the best known summary of the literature is that by Edmonds (1981). Based on his work and that of others, Averch et al. (1972), Brophy and Good (1970), Brookover, Beady, Flood, Schweitzer, and Wisenbaker (1979), Mayeske, Wisler, Beaton, Weinfeld, Cohen, Okada, Proshak, and Tabler (1972), and Weber
Edmonds lists five ingredients of an effective school: strong administrative leadership, high expectations for children's achievement, an orderly atmosphere conducive to learning, an emphasis on basic skill acquisition, and frequent monitoring of pupil progress. (Purkey and Smith, 1983, p. 429)

Other reviewers of this literature (Murnane, 1981; Clark, Lotto and McCarthy, 1980; Bridge, Judd and Moock, 1979; Rutter, 1983) have concluded along with Edmonds that schools do make a difference and that we now know what we must do to establish effective schools. First, we need good teachers.

Children learn more when they are taught by talented, highly motivated teachers who believe that their pupils can learn and who structure the school day so that pupils spend large amounts of time "on task", working at basic skill development. (Murnane, 1981, p. 27)

Second, as Coleman found, the research indicates that pupils from disadvantaged backgrounds perform better in schools with a significant number of children from advantaged backgrounds (Brookover et al., 1979; Reynolds et al., 1980, Rutter et al., 1979, Willms, 1983).

Third, even though class size, school size, instructional strategies, and school expenditure do not appear to have direct links with cognitive achievement, they may have indirect effects by providing the sort of conditions that facilitate effective teaching and the smooth operation of the school. Meyer (1977) and Meyer and Rowan (1977) have considered factors beyond those found at the classroom, school, or district level. They see education
as consisting of "a network of rules creating public classifications of persons and knowledge" (Meyer, 1977, p. 55). More recently, researchers examined within-school processes in an attempt to link pupil inputs to schooling outcomes rather than concentrate as they have in the past on aggregated data. At the school and classroom level, researchers are examining the social organization of classrooms and schools, and the types of learning environments to be found there (McLaughlin, 1978; Rutter et al., 1979). They are also examining detailed teacher behaviours and teaching styles (Brophy, 1980, 1982; Evertson et al., 1980). Some are considering the effect of school climate (Anderson, 1982), while others focus even more closely on the interactions between teachers and pupils (Moos, 1979).

In sum, one can say the school effectiveness research literature rests on two assumptions. One, that there exist exemplary schools which serve low income students. Two, there are specifiable characteristics of these effective schools which account for their success in teaching poor, urban minority children. However, as Ralph and Fennessy (1983) note of the reviews of the research literature, "these reviews give the impression that the two empirical propositions mentioned above are established facts. That is, they imply that there are high-performing schools which serve the urban poor, and that some five to seven characteristics distinguish these effective schools from the rest." (p. 690) Ralph and Fennessy, along with Purkey and Smith (1983), accuse school effectiveness researchers of having
what they call "an unusual disregard for what constitutes evidence." (p.690) They claim it is common practice for reviewers to make baseless or inadequately supported empirical claims and to refer to earlier reviews as if they were evidence. Moreover, they claim, most reviewers "fail to distinguish between those studies that are conducted on representative data bases, using specific measurement tools and incorporating control variables, from others that are largely impressionistic and employ no statistical controls." (p.690)

Rowan, Bossert, and Dwyer (1983) have noted three main sorts of weaknesses in research on effective schools: those to do with the measures of effectiveness, those relating to research design, and the inherent weaknesses in global comparisons. In their view, research has been narrowly focussed on a single dimension of school effectiveness -- basic skills acquisition. They cite studies by Tikunoff (1981) and Frechtling (1982) that point to the fact that school personnel and their constituencies assess school effectiveness not simply on the basis of test scores, but by reference to the sort of administrative, social, and cognitive goals these schools are working toward, and the programs that have been instituted to reach them. Frechtling's study (1982) of schools in Montgomery, Alabama found that when quantitative measures assessing instructional effectiveness were used to evaluate schools the measures correlated negatively to the assessments made by district personnel.
Rowan et al. (1983) argue that there are enormous problems with the four main ways school effectiveness researchers have assessed instructional outcomes. They note four types of assessment:

1. Absolute measures of instructional outcomes such as the proportion of students within the school who are at or above the national median in achievement. (e.g. Weber, 1971)

2. An analysis of trends in test scores at a particular grade level. Such an analysis might reveal rising or falling test scores in grade four over the last few years. (e.g. Phi Delta Kappan, 1980)

3. An analysis of gain scores for pupils in a particular age group throughout the year compared to the rest of the nation. (e.g. Wellisch et al., 1978)

4. Various regression-based techniques that are used to generate residuals.

In general, Rowan et al. fault school effectiveness research on methodological grounds. They argue, as others have, that studies of effective schools seldom measure the instructional performance of an entire school. Usually schools were identified as effective on the basis of assessments of test scores at only one or two grade levels, in only one or two
curriculum areas. (p.27) However, Rowan et al.'s main focus is on the contradictory nature of the research. The four approaches employed in the research tended to have low correlations with one another, and identified different schools as effective. As well, reliability studies have revealed that the measures are extremely unstable over time. (Forsythe, 1973; Jencks et al., 1972, p.124; Rowan & Denk,1982). Approaches which compare an absolute standard like national norms against school means (approach #1) or those which use trends in achievement (approach #2) are biased insofar as they prevent low SES schools from being labelled effective. Bossert and Dwyer's study of 405 California schools (using approaches 2 & 4) during the period 1975-1977 showed that increases/decreases in scores from year to year were significantly correlated to changes in the socioeconomic composition of a school's population, thus casting doubt on the validity of approach #2. Moreover this approach lacked stability. The correlation between gains made from 1975 to 1976, and from 1976 to 1977 was -.45. Measures using approach #4 were unreliable; the correlation of residuals for the period between 1975-1976 was .24 and for 1976-1977 was .19. Following standard practice in effective schools research, Bossert & Dwyer classified as effective schools in the top quartile of the residuals distribution. Those in the bottom quartile were classified as ineffective. By this system only 10% of the schools in the sample of 405 were effective for two consecutive years and only 5% were effective for three years. This finding was only marginally better than what could be expected from chance alone. Bossert and Dwyer estimate from their study that school level
characteristics account for roughly 4.5% of the variance in individual student achievement. They note that this figure is very close to the figure arrived at by Jencks et al. (1972). The Jencks study termed such variance as trivial in comparison to variables such as family background and social class. This finding lead Rowan et al. to conclude "many of the results from research on effective schools may be spurious...Most of the studies we reviewed failed to adequately control school demographics and prior school achievement in their analyses. As a result it is difficult to determine whether the relationships between school organization and school effectiveness found in quantitative studies are unbiased or if they are inflated by specification error and thus are largely spurious." (p.29)

Lawrence Stedman, in his critical review of the school effectiveness literature (1985), identifies what he says are the six most commonly cited factors connected with school effectiveness. These are: strong principal leadership in instructional matters, high teacher expectations, an emphasis on basic skills, order and discipline, systematic evaluation, and increased time on task. He notes, however, that in a good many cases, researchers went looking for these factors in their visits to schools and not surprisingly found what they were looking for. "Many researchers did not use systematic procedures to observe schools, but relied upon the impressions of observers, who knew, in advance, which schools were effective and ineffective. Findings biased in favor of the six factor formulation were inevitable." (p.306) In the
preface to the Phi Delta Kappan report on school effectiveness research it is admitted that Wilbur Brookover and other outside presenters were asked to prepare the observers in the study before data collection was begun. Their job was to "stimulate and inform" the observers. The danger of systematic observer bias seems obvious. Indeed, Ralph and Fennessy (1983) note that "the subsequent emphasis on certain aspects of educational practice [in the study] should come as no surprise; it is roughly the recipe that Edmonds and Brookover have advocated for making schools effective." (p.691) Subsequent studies which have used a "blind" design in which observers and interviewers didn't know which schools were the effective ones came up with findings that challenge the six factor formulation. One such study in New York found that in both high achieving and low achieving schools the principal was not the schools' instructional leader. As well, the emphasis on basic skills was greater in the ineffective schools.(p.306)

Another study in Massachusetts found that 6 out of 10 effective schools were low on leadership, atmosphere, and reading emphasis, whereas 5 of the 8 ineffective schools rated highly on these factors.(Ellis, 1975, p. 306) Lezotte (1981) reported that ineffective schools in Lansing, Michigan tended to have principals who were more involved in monitoring instruction than principals in the effective schools. As well he found that teachers in the ineffective schools had higher expectations for grade level achievement than did their counterparts in the effective schools. (p.306)
In addition to researchers appearing to "find" what they were looking for, Stedman noticed also that in several of the more prominent reviews of school effectiveness literature studies were cited that don't support the six factor formulation. In a frequently cited study conducted by the Maryland State Department of Education (1978) there were no clear differences in the amount of time principals devoted to instructional supervision between effective and ineffective schools. There were no differences in teacher expectations between these two types of schools either. The emphasis on basic skills was roughly the same for all the schools, and no significant difference in amount of time on task distinguished the more effective schools from the least effective.

Rutter et al.'s very influential study from 1979 also fails to support the six factor formulation. They found that the head teacher's emphasis on instruction or the lack of it was unrelated to academic outcomes. They also found that teacher expectation was only weakly related to achievement and that only half of the variables associated with academic emphasis related to achievement. Among these was the displaying of student work on classroom walls. In addition, only one out of six punishment variables and one out of five reward variables were associated with achievement. Total teaching time and staying on the topic were not related to achievement. (p. 308)
Part of the problem of fitting Rutter's findings with the six factor formulation is that his study was not designed with the six factors in mind. Insofar as teaching time and staying on topic may pick out the same thing as time on task, or as much as the punishment and reward variables correspond to the order and discipline factor, it seems fair to compare Rutter's work with its American counterparts. When we do, we find little support in Rutter's study for the six factor formula held to be the recipe for effectiveness in the United States.

A third problem with this formulation is that many of the schools which displayed these factors were ineffective schools. Wellisch et al. (1978) concluded that instructional leadership, school-wide coordination of the curriculum, and, an academic emphasis, produced effective schools. Yet none of the schools in the study ranked above the 30th percentile in reading and only one did in mathematics. Phi Delta Kappa (1980) stated that principal leadership, high expectations, discipline, and an emphasis on basic skills characterized the eight effective schools studied. In one school, 75% of the students were below grade level; in another, the students were over two years behind grade level. New York researchers (New York Department of Education, 1974) claimed that strong instructional leadership was responsible for its effective schools, yet two-thirds of its 6th graders were 2 or more years behind grade level.

Lastly, most reviewers of this literature, and especially some influential ones (Austin, 1981; Edmonds, 1979; Glenn, 1981)
ignored a number of studies that were unable to link effectiveness to school characteristics. Philadelphia's "Successful Schools Study" (1979), for example, could not identify any characteristics that distinguished 15 effective schools from other schools. Talmage and Rippey (1974) in their case studies of four Chicago elementary schools were unable to determine why two were among the best and two among the worst in Chicago. Echternacht (1977) could not explain the differences in achievement between high schools with stable SAT scores and those which had had drastic declines. He found that in both types of school truancy, discipline problems, and teacher permissiveness had increased along the same lines. There was only a slight difference in the number of academic courses taken by students in both types of school; in some cases schools with declining SAT scores had a higher proportion of students taking academic courses. Moreover, the schools with decreasing scores had increased homework and expanded basic skill instruction, but seemingly to little or no avail.

More striking still are the doubts raised by the manner in which the data on leadership and teacher expectation were gathered. In a number of cases the data on principal leadership and teacher expectation were self-reported; that is, principals answered questionnaires which required them to estimate what amount of their overall time they devoted to instructional matters, while teachers similarly answered questionnaires regarding the likelihood of their students reaching grade levels.
of achievement throughout the year. Needless to say, such questionnaires are a flimsy basis on which to erect a theory of effective schools. Discussions of research into the pygmalion effect and psychological explanations of motivation based on this research are in many faculties of education a staple of pre-service teacher training. It seems unlikely that teachers would not realize that having faith in the ability of one's students is considered part of being a good teacher. Likewise, principals cannot help but know that traditionally one of their main responsibilities is instructional leadership. The weakness of these self-reported findings is simply that in the case of teachers and principals we cannot know how many respondents were merely paying lip service to ideas that are a major part of educational orthodoxy, but which are often neglected for one reason or another.

Taken together, the research does not support the idea that the six factor formulation, if introduced into schools, will lead to school effectiveness. Nor does it support the idea that the effective schools identified in these studies are effective because of these factors, at least five of them. The one factor that did repeatedly relate to effectiveness was systematic evaluation, especially where this meant focussing on objectives and teaching to the test.

In the Brookover and Lezotte study (1977), school G used statewide criterion referenced tests to inform discussion of what, and, how to teach. Teachers relied on dittoed assessment
questions based on the tests and drilled the students in the use of these tests until the test content was mastered. A study by Glenn (1981) of effective schools for poor black children in Massachussetts notes that teachers at one of the effective schools gave their regular classroom quizzes in standardized test format and made an effort to coach their students in test taking techniques. The principal was reported as administering practice tests on a regular basis and as time went by raising the passing grade. Four times a year, students in this elementary school sat formal exams based on city-wide tests and standardized test batteries. Those who passed the test at the end of the year were rewarded with popsicles. (cited in Stedman, 1985, p.310)

A similar situation obtained in a number of effective high schools investigated in a study by Thomson and DeLeonibus in 1978. The focus in these schools was in keeping their SAT scores high. Over half the principals of their schools admitted their English curricula focused on areas that would prepare students for the SAT: traditional grammar and vocabulary exercises. (p.7-10, 18-20, 37) The superintendent of the Richmond school system set a system-wide goal of gaining the equivalent of one month's achievement measured on standardized achievement tests for each month of school. Test results by classroom and teacher were made available to the superintendent. Teacher job security and student promotion were directly tied to performance on these standardized tests. Students who were assessed as being behind two grade levels were automatically retained, and placed in remedial
programs whose main emphasis was preparing pupils to pass tests
that came with the kits such students were required to work on.
The progress of these students was monitored by a central testing
department. To ensure a greater degree of uniformity the reading
curriculum was standardized district-wide by the adoption of a
single reading series. (cited in Stedman, p.310)

A teaching manual associated with teacher effectiveness, Teachers Do Make A Difference recommends an instructional
approach known as "active participation". This approach is one
which, in effect, transforms teaching into an almost continuous
form of testing. Teachers who employ "active participation" are
encouraged to teach didactically, but at the same time they are
to check for understanding by asking questions frequently and in
a manner that enables them to see who has the correct answer to
questions. Teachers are to ask their questions in such a way
that all students will be able to signal an answer. For example,
in mathematics lessons students will be expected to write their
answers on individual slates and hold them up for the teacher to
see. Questions which would otherwise involve written responses
can be phrased as multiple choice questions. Thus students can
respond by writing the correspondingly correct number on their
slates or by holding up one or more fingers. In this way the
teacher at a glance can tell who was successful in getting the
correct answer. (Cummings, 1980)

There is some reason to believe that teachers may resort to
such forms of teaching simply to protect themselves. According to
Allan Ornstein, 40 states in the US have begun to administer standardized tests for the purpose of making decisions about student promotion or graduation. In Michigan there are plans to withhold up to 5% of funding from elementary schools which do not meet state norms in reading and mathematics. In New York state, students must pass examinations in reading, math, and writing in order to graduate. New York city has introduced "promotional gates" at the end of the 4th and 7th grades. Students must pass achievement tests to pass on. The standard for fourth graders has been fixed at the 3.7 grade level and at the 6.2 grade level for seventh graders. Twenty-five per cent fail each year. (Ornstein, 1984, p.96) Discussions regarding the city-wide implementation of a promotion policy in Philadelphia which would require all elementary students to pass grade level skills tests floundered when the committee proposing this move learned more than 40,000 pupils were expected to fail or need special remediation classes. (Toch, 1984, p.174) In California, the Stull Act requires that the competence of certified personnel be measured in part by reference to student performance. In Texas HB 72 calls for a ban on social promotion, and links promotion from grade to grade, as well as the right of students to participate in extra curricular activities, on their achieving course grades of at least 70%. In 1984 the states of Arkansas, Kentucky, South Carolina and Texas introduced education reform bills which would allow state governments to take over the control of local school systems if their test scores drop below certain levels. (Anderson & Pipho, 1984, p.211) As Darling-Hammond has said of circumstances like
If there is one thing social scientists have discovered that has not yet been disproved, it is that if bureaucrats are evaluated by a performance measure, they will seek to maximize that metric at the expense of other areas of performance that are not measured. (p. 5)

As one might guess there is some concern over what might be the negative, unintended consequences of this emphasis on testing and supervision. The centralization of school governance is a major concern of those opposed to the emphasis on standardization, as is the suspicion that teachers will respond to this centralization and control in ways not intended by the advocates of teacher accountability. Madaus and Greaney (1985) reported that one effect of installing competency tests for grade to grade promotion in Ireland between the 1940's and the 1960's was a phenomenon they called "the tradition of past exams". Once the exams were set and teachers, parents, and students became familiar with them, a number of things began to happen. Teachers predictably enough began to teach to the test, and so ignored other subjects or topics not likely to appear on the test that the Department of Education became concerned. A department report stated that "some of the inspectors of the opinion that it would be appropriate to include History and Geography in the examination as well, since these subjects are neglected in some of the schools in the term after Easter." (cited in Madaus and Greaney, 1985, p. 284) As well publishers saw the opportunity to prepare books of questions similar to those found in the exams. Most of these, according to Madaus and Greaney, were sold to
teachers to help them coach their students for the exam. Senior officials in the Department of Education finally admitted they could not alter the exam in any major way since the "inspectors setting them feel they must adhere to the expected pattern which they have helped to create." (p.283)

Meier (1981) has noted the deleterious effects on reading ability that stem from an over-reliance on testing, or more typically from the use of standardized reading kits which in her estimation are no more than "disguised reading tests". Such kits, she observes, usually consist of "hundreds of unrelated paragraphs followed by multiple-choice questions and reams of ditto sheets... We are being pushed into programs aimed at ever narrower and more trivial subskills that...show up quickly on paper and pencil tests."(p.460)

As part of a Minneapolis school system improvement program in mathematics, district officials mandated an emphasis on the teaching of fractions when they discovered from standardized test results that their junior high students were below national norms. It was pointed out at the time by the district's math supervisor that given the widespread use of calculators and the prevalence of the metric system such an emphasis on fractions could only be justified on the grounds that test scores were important in themselves. (Tyler and White, 1979, p.8; cited in Stedman, 1985, p.313) Similarly, Brookover has given this advice to teachers regarding tests:
It is important that teachers cover the material in class in the same form it is to be used for testing... If the test uses a horizontal format for addition, students must be taught in this format. There should be no surprises in test format. This has been found to cause unexpectedly poor results on the part of students, even those who the teacher knows have mastered the material. (1982, p. 256)

We have here a clear example of how tests begin to influence teaching. There is also explicit in Brookover's remarks the recognition that tests often fail to assess students as well as teachers can, yet there is a complete acceptance of the legitimacy of tests. It is teachers and how they teach that must change.

The Ellis study (1975) of effective Massachusetts schools found these schools had "a very strong orientation to drill and practice. Much of class time is spent having the entire group go over worksheets, or generally drilling students on basic skills." (p. 19) In 1983, Darling-Hammond and Wise studied three school districts whose curricula were tied to testing programs. In their interviews with teachers in these districts they found that two thirds of the teachers had changed what they taught and one-third admitted that they were either teaching to the test or training their students in test taking techniques. (p. 5-6) Teachers who were interviewed admitted to abandoning essay assignments since the curriculum was geared toward multiple choice exercises and tests. One teacher reported she was forbidden to use creative writing as part of her reading program, while another described how teachers in her school were required by the principal to
rewrite standardized tests, teach these to her students so they would score well when the real test was administered. She reported that two teachers who refused to do this were threatened with disciplinary action. (p.5-6) In this survey nearly one half of the teachers in these districts reported they were considering leaving teaching as a result of the emphasis on standardization and testing. (p.11)

Resistance to this trend has taken other forms as well. Several commentators have noted a dramatic rise in cheating and falsification of test results. The New York State Education Department study of effective schools noted in its conclusion that "in at least one school -- the school whose scores were rising dramatically -- at least half a dozen teachers indicated that objective test administration was violated, making reading achievement scores probably higher than expected." (1974,p.9) David Armour and his colleagues in a study of 20 elementary schools spanning four years found that in six of the ten high scoring classrooms students had been coached for the test. In the other four classrooms the results suggest tampering as well: data from previous and subsequent years deviated sharply from that which was gathered during the study. (Ralph & Fennessy, 1983, p.692) As a result Armour et al. divided their data into three categories. One category is made up of data for which there is no reason to doubt the validity of the data. The second category consists of data which the research team had strong suspicions about. The third data set contained clearly fraudulent data. In
the view of Ralph and Fennessy, only if one accepts the first two sets of data can one say the study gave any evidence of there being exceptionally high achieving schools. (Ralph and Fennessy, 1983, p. 693; Meier, 1981.)

It also seems to these critics that one of the chief effects of increased supervision will be the standardization of what is taught and the manner in which it is taught. Cuban speaks of the "strong irresistible tug" toward a standard curriculum and the system-wide use of the same textbooks, and expresses concern over the fact that "adopting the school effectiveness research will drive the curriculum and school management toward uniformity". (Cuban, 1984, p. 148) The question arises, then, in light of this pointed criticism, of why it is that school effectiveness research has taken such a hold, and why it is able to exert a "strong, irresistible tug" that draws people to adopt its prescriptions.

The answers to this question are many, some of them having to do with the purported low cost of this sort of school reform, others to do with the genuine enthusiasm of teachers and administrators who see in school effectiveness research a degree of certainty and simplicity concerning what should be done to improve schools. There is also the sense that this research is scientific and therefore reliable. Ralph and Fennessy reject this idea that school effectiveness studies have been conducted scientifically. In their view school effectiveness research is influential because "it lends the mantle of science to what
educators are already committed to believing about schools... In the guise of positive science, what we find is a set of normative principles." (p.693)

There are surely political reasons as well, reasons that have to do with the nature of power/knowledge. Cuban speaks of the pressure for results that "pinches" school boards and superintendents. He notes also the "boosterism surrounding direct instructional methods [which] ...presses teachers toward these practices" and of school boards and superintendents being "driven by the inexorable logic of the research findings on effective schools". (p. 142) The knowledge produced and circulated by this research has, in a Foucauldian sense, produced and circulated the effects of power as well. These effects need not always act on individuals in a coercive way, at least not completely. Superintendents, principals and teachers are empowered to take a more overtly supervisory role as the research findings act to legitimate closer supervision of students, teachers and principals. And they may freely want to do so, for any number of reasons that have nothing to do with their being coerced. Yet everyone can become caught in this "field of visibility". There is in the research on effective schools a heavy emphasis placed on the importance of the "educational leader", an emphasis no doubt enjoyed by many current principals and superintendents who have had to live with the suspicion that, as their subordinates might confirm, they do little more than shuffle papers and get in the way of teachers doing their jobs. But as Cuban has noted,
"buried in the language of principals as instructional leaders and effective teachers... is a crisp accountability for student performance - a steel fist encased in velvet". (Cuban, 1984, p.137) State departments of education may find themselves in the same bind their Irish counterparts found themselves with "the tradition of past exams". Local school boards may come to realize there are enormous political and financial costs connected with retaining students who fail competency tests.

Linked together in all this are the people in the schools -- children, youths, teachers, and supervisors. They are linked by more than a common purpose and an overlapping set of obligations, however. They are linked together in a disciplinary grid of visibility made possible by tests. Students are brought into the domain of power by means of examination, broadly defined. Teachers rely on teaching methods which turn teaching into a process of uninterrupted examination. A major part of this process consists of the sort of teaching practices emphasized by Brophy, Good, and Rosenshine, and criticized by Meier and Cuban. Reading kits, workbooks, and classroom tests which allow for more or less continuous testing form the basis of instruction. Teachers are encouraged to organize their classrooms and teach in a way that maximizes their capacity for supervising their students. The use of standardized norm-referenced tests permit multiple comparisons: student to student, teacher to teacher, school to school, principal to principal and so on. Schools and school systems come to be organized along principles of supervision with teachers and principals being judged on the
basis of such panoptic relays as test scores and other forms of inspection. Still another part of this process involves the creation of new norms. Foucault observes that each individual test score individualizes.

It refers individual actions to a whole that is at once a field of comparison, a space of differentiation, and the principle of a rule to be followed. It differentiates individuals from one another, in terms of the following overall rule: that the rule be respected, or as an optimum toward which one must move. It measures in quantitative terms and hierarchizes in terms of value the abilities, the level, the nature of individuals. (1984, p.195)

At some point arbitrary limits are imposed on this hierarchy, and stigmatization and exclusion can be justified on the grounds that students with deficiencies need remedial help, retention, etc., while fear of failure will encourage students to do their best. Madaus (1985) provides some telling examples of the form this process may take: the student who has done average work for 11 years fails the state mandated functional literacy test by a point or two and is told she can't graduate until she scores above the cutoff; the child who is pulled out of a regular classroom is assigned to a remedial class following a poor showing on a statewide test; the high school athlete with the SAT score of 690 who as a result of this score is ineligible to play in his freshman year. (p.612) When it appears that evaluation is biased in an arbitrary way: that it favours one group over another because of one group's familiarity with a particular manner of speaking or of their acquaintance with facts seldom
encountered in school, we usually object. Moreover, if it can be shown that standards are a product of mapping a normal or bell curve which will always ensure a certain percentage of failures, we are again concerned.

Madaus et al. (1980) point out the manner in which standardized tests of achievement are devised; it is clear from his account that because of the way tests are constructed it may be that they are measuring little more than the social background characteristics of students.

Test constructors, when they go about constructing tests, have no clear national objectives to refer to, so they must establish their own. They do this by reading curriculum guides and textbooks, and by talking to teachers. Because of the constraints on time and money, test designers seldom can afford to do more than this. Madaus et al. note that

evidence based on direct observation of instruction is never provided to demonstrate that the goals and standards which are embodied in a test reflect the actual goals or standards emphasized in schools. Further there is no independent check on the congruence between common objectives inferred from an examination of leading textbooks and curricula and what, when, and how the subject matter is actually taught in classrooms throughout the country. (p.135)

The next step in test building is to develop a table of specifications in which the content to be covered in the test and the skills needed to answer questions correctly are arranged in a grid. Each cell in the grid is given a weighting. Then, test
items are developed with the table of specifications being used to determine how many questions of a particular sort are included in the test. Panels of teachers and subject area specialists are subsequently consulted to see which items are most appropriate.

From this point on a number assumptions made by makers of standardized tests determine the procedure of test building. First, it is assumed that the purpose of standardized tests will be to measure inter-individual differences in student achievement. The tests are devised so as to maximize the differences between individuals. Second, it is assumed that achievement is normally distributed in the population about to be tested. The tests are accordingly designed to distribute students on a normal curve. Madaus et al. observe that "one consequence of building standardized achievement tests on the assumption that achievement is normally distributed is that the test results derive their meaning from the distribution of the scores of the population who take the test, or more likely from the representative reference population on which the test was 'normed'." (p.138) An individual's score tells us nothing directly about what the student has actually learned, only that in relation to others he has learned more or less than they have. When the pool of test items has been developed, the items themselves are field tested. In order to get a normal distribution of scores, test designers screen items according to their difficulty and what is called their discrimination index. Since it is important to discriminate between students, items
that are either too easy or too difficult are thrown out. Madaus et al. note that

items which most or all students answer correctly or incorrectly are discarded or revised, since difficulties of near 0 or 100 percent provide no differentiation between students...As early as 1936, Hawkes, Lindquist, and Mann pointed out that one consequence of discarding easy items is that if instruction has been adequate, very important or very fundamental items may have been so thoroughly taught that they have been mastered by all pupils. Such items would be eliminated in item-screening procedures used in the construction of standardized tests.

(p.143)

It bears emphasizing that the "easy" items in these tests are those which most students answer correctly. In this way the manner of test construction leads to the exclusion of those items which reflect the instructional emphasis of schools. Yet these sorts of instruments are precisely those which are most commonly used to assess the adequacy of school programs.

An additional feature of the test item selection process is that because items of intermediate difficulty are those that end up on the tests, changes in the achievement levels of weak students are not likely to show up in test results. Most tests have items clustered around the 60% level of difficulty, while fewer than 5% of the items are at the difficulty level which will permit weak students to make correct responses. (Madaus et al., p.145) Madaus et al. conclude

As the result of psychometric screening procedures there is such a small sample of items of suitable difficulty for disadvantaged children (who tend to score lowest on such tests) that reliable estimates of changes in the
children's performance cannot be obtained, unless the compensatory program effects an enormous improvement in the children's responses. Traditional standardized tests simply were not designed to provide reliable measures of change at the extremes of the distribution. (p.145)

Ironically, it was tests designed in this way that provided the basic evidence for the claim that compensatory education programs weren't working. Moreover these tests fail to guide educators who are primarily interested in improving, not just special programs for the poor, but entire schools whose students come from predominantly low-income families. Edmonds made this same point in 1984 when he argued that

norm-referenced, standardized achievement testing constitutes a formidable obstacle to resolution of the issues associated with greater achievement for low-income students. First, the tests measure students in relation to each other and, therefore, do not produce results that establish confidently whether or not individual minimum academic mastery has occurred. Second, mean or average aggregate school scores obscure whether and to what extent all students are progressing as they ought or might. (p.38)

Not only are tests of this sort insensitive to improvements and declines at the extreme ends of the tested population, they can very easily fail to distinguish between effective and ineffective schools in general. If, during field testing of test items, most students in one school answer a particular item correctly, while most of the students in another school answer incorrectly, the test designer will likely throw the item out since it will be considered too difficult. Conversely we would have the same result if all the students in one school answered
incorrectly, while most students in four other schools answered correctly. Because the item would appear too easy it would be discarded. Yet standardized tests have been the usual basis for distinguishing between effective and ineffective schools, despite the likelihood that the test items which would be most useful in distinguishing between schools are those which never appear on the final test.

Test publishers are only too aware of the limits of their measures, and according to Coffman (1974), Dyer (1972), and Fitzgibbon (1975) they traditionally point out these limits to their customers. Yet, despite these cautions, 29 states which enroll two-thirds of school-age children in the US already use or are considering the use of these tests to determine mastery of basic school objectives. (Madaus et al., 1980, p.166)

It is not clear what trait is being measured by these tests. There is reason, as we have seen, to doubt that a close fit exists between what is taught in schools and what is measured in standardized tests. Madaus concludes that "the general constructs these tests actually measure are so heavily loaded on and confounded with home background and general ability as to render impossible a verdict about the differential effectiveness of different kinds of schools and school resources on pupil learning." (p.166)

It is also ironic that these tests, which are now being criticized on the grounds they deny low income students equal
opportunities to education, were in the 1960s, the main weapon used by liberals in the fight for compensatory education programs. Standardized tests and test scores were used in Congressional battles to push through liberal reform measures: Title IV of the Civil Rights Act 1964, and Title I & Title III of the Elementary and Secondary Education Act of 1965. These programs were initially justified on the grounds that "an achievement gap", defined in terms of standard test performance levels, existed between disadvantaged and other children. Anthony Celebrezze, Secretary of Health, Education, and Welfare in 1965, argued that

You will find that by the end of the third year this student [in central Harlem in New York City] is approximately 1.2 grades behind the national average and 1.1 grades behind the New York City average. By the time he gets to the sixth grade, he is 2.1 grades below the national average and two grades below the New York average. And by the time he gets to the eighth grade, he is [two and a half] grades below the national average and approximately 2 grades below the New York average.... The students continue to get further and further behind in terms of standardized test norms...

(cited in Madaus et al., 1980, p.117)

In this way it was possible to supply what appeared to be solid documentation for the view that groups which were recognized as being economically deprived were also those who seemed to benefit least from regular educational programs. Over a short period of time test scores became accepted as the outward standard of school performance and encouraged the view that school effectiveness should be judged on the basis of outputs and not inputs. Madaus adds
Paradoxically, many individuals and groups who accepted standardized test results as indicative of a pressing need for remedial programs became critics of these tests when they were used in studies which revealed little improvement in pupil performance resulting from compensatory programs. (p.118)

Since the mid 1970's when Patrick Moynihan informed President Nixon that compensatory programs weren't working, that Dr. Jensen's thesis was politically unacceptable, and that income distribution might be necessary, standardized test scores have come to be the foundation for the view that American education is in crisis and that the solution to this crisis rests with a greater reliance on testing and inspection. Under these circumstances even the possibility that standardized tests measure a trait or traits which have more to do with social class than student achievement should prevent us from placing our faith in such tests, and should lead us to view with suspicion this general trend toward an even greater reliance on test scores. If we do not proceed cautiously in this area, we risk deepening the gulf that already separates the socially disadvantaged from the rest of American society. By permitting a school system to determine its successes and guide its instruction on the basis of standardized test scores, Americans risk institutionalizing a system for evaluating students, teachers, and schools that will not only redefine curricula, and centralize control of education, but will promote further crises over illiteracy, teacher competence, and the capacities of the poor to benefit from education.
CHAPTER FIVE: CONCLUSION and FURTHER RESEARCH

The research literature dealing with teacher effectiveness and school effectiveness has been criticized on a number of grounds.

1. The teaching-learning relationship is conceived to be a mechanical process acting on an object. Students are seen as being little more than conduits into which flow teacher behaviours and from which emerge evidence of learning. (Tom, 1984; Murnane, 1981; Meyer, 1977)

2. A severe narrowing of the curriculum will result from the adoption of teacher and school effectiveness research. (Cuban, 1984; Stedman, 1985; Stedman & Smith, 1983; Darling-Hammond & Wise, 1985)

3. Teacher effectiveness research is characterized by weak to moderate correlations and limited generalizability. (Tom, 1984; Garrison & Macmillan, 1984)

4. School effectiveness research is characterized by flawed research design, and contradictory findings. (Purkey & Smith, 1983; Ralph & Fennessy, 1983; Stedman, 1985; Rowan, Bossert, & Dwyer, 1984)

All of these critics object to the excessive reliance in this body of research on outcome measures as indicators of
effectiveness. The foregoing is meant to be more than a catalogue of mistakes and muddled thinking, however. It has been demonstrated that the conceptual weaknesses, the methodological shortcomings, and general willingness of advocates to overlook these weaknesses and overstate the applicability of the research combine to produce certain effects. Foucault is useful in that he provides a theory which makes it possible to see that all the shortcomings in this literature amount to something, lead to a particular state of affairs; namely, disciplined bodies.

Without the many mistakes that can be seen as constituting a "systematic blindness", it would not be possible for the disciplinary mechanisms which have been discussed in these pages to be introduced into classrooms or refined and extended. In some cases, no doubt, these mistakes are simply the product of human error. But that these lapses of judgement should be so systematic, that the findings of this research should be so enthusiastically received, even by some of its harshest critics, and that they should form the basis for so many of the recent changes in educational policy, suggests a widespread suspension of critical acuity.

As we have seen, these errors and oversights lead to a number of problems. Firstly, if these mistakes go unrectified, particularly those related to the inadequacies of tests, we may expect a severe narrowing of the curriculum and an emphasis in instruction that is not likely to develop capacities for critical
thinking, let alone promote basic literacy and numeracy.

Nine reports on the state of American education were published in 1983, an event which provoked the editors of the Harvard Educational Review to term 1983 "the year of the reports." (1984, p.1) Four of these reports were either commissioned by legislative bodies or spoke to the contribution federal and state governments could make to school improvement.

1. **A NATION AT RISK: THE IMPERATIVE FOR EDUCATIONAL REFORM**

   The report of the National Commission on Excellence in Education. This commission was appointed by the Secretary of Education Terrel H. Bell to examine the educational system and recommend reforms.

2. **ACTION FOR EXCELLENCE**

   The report of the Education Commission of the States' Task Force on Education for Economic Growth. This report was designed to help inform state governors of the relationship between education and economic growth and the sort of educational reforms needed to promote such growth.

3. **MAKING THE GRADE**

   The report of the Twentieth Century Fund Task Force on federal elementary and secondary education policy.

4. **EDUCATING AMERICANS FOR THE 21ST CENTURY**

   A report, commissioned by the National Science Board, which is solely concerned with the state of science and mathematics learning in American schools.

   While these reports vary in their particulars and the stridency with which they condemn American schools, they share a
common perspective derived from human capital theory. They see
the American economy, or more precisely American dominance in
international trade, as being largely dependent on the quality of
its educational system and the quality of the scientists,
technicians, engineers, etc., it is capable of producing.
According to this view one can explain the decline of American
dominance in applied science, trade and industry in terms of the
declining quality of education and the erosion of standards in
schools. A Nation at Risk begins with the following:

Our nation is at risk. Our once unchallenged preeminence
in commerce, industry, science, and technological
innovation is being overtaken by competitors throughout
the world... If an unfriendly foreign power had
attempted to impose on America the mediocre educational
performance that exists today, we might well have viewed
it as an act of war. (p.12)

A Nation at Risk points to the absence of high standards in
schools and the proliferation of easy elective courses which
permit ignorant, illiterate students to graduate as being among
the chief reasons for the decline in achievement. It fails to
recognize, however, that the shortcomings in student achievement
might be attributable to factors other than the indulgence of the
school system toward its clients. Despite high levels of
unemployment, it claims that "All children, by virtue of their
own efforts...can hope to attain the mature and informed
judgement needed to secure gainful employment." (p.12) That such
employment has eluded so many must, by implication, be due in
large part to a school system that fails to develop the potential
of its clients. Students all along, it seems, have had the ability to succeed in school; they have simply been too lazy to take challenging courses or their teachers have either been too lazy as well, or too incompetent to teach them properly. The task force responsible for Making the Grade begins its report with the statement that "the nation's public schools are in trouble. By almost every measure -- the commitment and competency of teachers, student test scores, truancy and dropout rates, crimes of violence -- the performance of our schools falls far short of expectation." (1983, p.3) Action for Excellence notes that individuals planning to be teachers score well below the national average on the Scholastic Aptitude Test. The members of the Task Force find this "a disturbing fact. But it is probably what we should expect, given the low levels of pay and esteem that we accord teachers in America." (p.27) Educating Americans states that:

the quality and style of elementary and secondary teaching constitute the most obvious and immediate source of the problems facing mathematics and science education...A substantial number of our Nation's 1.17 million elementary school faculty members lack sufficient knowledge, training and, in many cases, interest to teach mathematics and science effectively. (p.29)

More and more students are turning their backs on algebra, French and other respected academic subjects, and are instead:

enrolling in physical and health education, work experience outside the school, remedial English and mathematics, and personal-service and development courses, such as training for adulthood and marriage...
Given this freedom to choose the substance of half or more of their education, many students opt for less demanding personal-services courses, such as bachelor living. (National Commission, 1983, p. 14)

Not only are teachers "drawn from the bottom quarter of graduating high school and college students", they are, once in college, required to take "courses in "educational methods" at the expense of courses in subjects to be taught." (National Commission, p.14) Making the Grade also points to the "trade union mentality" that is responsible for transforming "what had been a noble though poorly compensated profession into a craft led by collective bargaining organizations with a focus on bread and butter issues -- wages, working conditions, and job security (for which read seniority)." (Making the Grade, 1983, p.5) Action for Excellence implies that because there are no financial incentives for superior teaching, the result is an alarmingly high number of "unmotivated teachers".(p.26) It adds that "40% of secondary school science teachers have not attended a course or workshop in their subject area since they began teaching."(p.26)

The message from these reports seem clear enough: life has been too easy for students. Given the opportunity to take the easy way out they are doing so. Teachers are not likely to be competent given they are drawn from the least academically promising students in the system and spend a large proportion of their time in courses that do little to give them a solid grounding in their subject areas.

It is interesting to notice the similarities between these
sentiments and those of the Irish prime minister who, over forty years ago, called for a system of nation-wide competency testing in Irish primary schools. He insisted in 1941 that students and teachers be subject to standardized tests on the grounds that human beings are fundamentally lazy, and therefore need to be held in check by regular inspection.

Human tendency would be to make things easy if we were not all the time kept up to concert pitch. The same is true of teachers... I know they are an excellent body, but I say that they are not more than human. Inspection, therefore, is necessary and, more than inspection, the most important things of all are the tests at the end. (Madaus & Greaney, 1985, p. 274)

If we assume that the recent commissioned reports on the state of education in the US give us a sense of the direction in which public schooling is moving, an examination of these reports can provide a glimpse of the future. It might prove worthwhile to explore, within a similar Foucauldian framework, these reports which define a crisis in education in terms of declines in American economic performance. Since it is beyond the scope of this thesis to explore these reports fully, the following analysis will only sketch the general outlines of what could be the basis for further research. We shall consider the similarities between the recommendations contained in these four reports and those found in teacher and school effectiveness research. As well, we shall examine the way in which a "systematic blindness" has once again led to conclusions which make regulation and coercion seem a reasonable and promising way to improve American education.
The reports provide considerable statistical data, mostly in the form of mean test scores at the national level, to support the assertion that the United States is at risk because of the decline of its educational system. As previously mentioned, 23 million American adults are functionally illiterate,... 13% of all 17-year olds in the US can be considered functionally illiterate,... and Scholastic Aptitude Test scores have been steadily declining since 1963 - a decline of 50 points on average for verbal ability scores and one of 40 points in mathematics. (The National Commission, 1983, p.12)

Teachers fare no better. Action for Excellence notes that "in 1982 SAT scores for students preparing to be teachers were 80 points below the national average".(p.27) Educating Americans for the 21st Century states that "shortfalls in math and science education at the pre-collegiate level exist primarily because of a shortage of competent and dedicated teachers".(p.33)

To counteract the effects of years of poor schooling A Nation at Risk recommends that more academic courses, conducted with more rigor, be required for graduation from high school; that more time be spent on learning, first by better use of the school day, then by either lengthening the school day or school year; that better teachers be produced by the training institutions (poor teachers should be fired, while good teachers be given incentives); and lastly that citizens provide more
fiscal support for education while holding educators and politicians responsible for the effective leadership of education. In general the other reports frame the problem in more or less the same way. Low standards and easy graduation requirements have combined with an indifferent and incompetent teaching profession to produce a serious shortage of skilled manpower and intellectual talents.

This emphasis on standards has led some commentators to worry about the possible negative effects of raising standards. It seems to some that the effect of higher standards, particularly as they are enforced through standardized testing, will be to exclude the new breed of "failures" from full participation in the school system, while at the same time making it appear as if the quality of education has risen. Deborah Meier comments:

The focus of normative testing has played a major role in the trend toward defining all those in the bottom portion of any particular curve as "deviant", in need of "special education", thus justifying the systematic removal of ever larger groups of children who are not "making it" on the normative scales. As they are removed from the rolls of "regular" schools we have an illusory feeling that standards have gone up... But the losers do not disappear, except from our immediate view. The proliferation of "special education" enrollments is partly a triumph. Some kids do need to be provided with special services. But it has also become a way of trying to look more successful by merely redefining our population. (Meier, 1984,p.70)

The workings of power/knowledge and discipline seem evident in these reports. By means of statistical methods and enormous
and sophisticated testing apparatuses, students and teachers have been brought into a "field of visibility". Once in view, as it were, they are, by means of these tests and other measures distributed in relation to a norm. The talk of firings for incompetent teachers, merit pay for the good ones, and varieties of exclusion for students who fail to meet one standard or another reveal the extent to which these reports call for the adoption of dividing practices. It seems equally clear that the tactic of legitimating educational policy changes through the use of statistical research data has been adopted by a quite different group from that of the influential investigators of the Coleman – Jencks era, a group, it seems whose goal appears to be improving the material well-being of the social body by means of supervision and regulation. Now, as distinct from 1966, the target has changed. Paradoxically, regulation is being advocated at a time when deregulation is generally most favoured, and freedom from bureaucratic supervision most highly valued.

Of course the visibility which increasingly characterizes the situation of students and teachers is not due simply to statistical surveys and standardized tests; it is due to the way the data are interpreted, how and to whom the issue is reported. The interpretations of the data, and the reports, newspaper and magazine articles which reach a large audience are also characterized by the sort of "systematic blindness" that is so typical of teacher and school effectiveness research. A Nation at Risk is addressed to "the American people", Action for Excellence is printed on glossy paper in a four colour format. Sidebars with
colorfully bold and provocative quotations from the main text are scattered throughout its pages. The title page of *Educating Americans* promises that Americans will be the best educated people in the world by 1999 if its recommendations are adopted. These are more the techniques of advertising journalism and demagogery than those which characterize sober and responsible scholarship.

Stedman and Smith (1983) in their response to these reports note that "these reports are political documents; the case they make takes the form of a polemic, not a reasoned treatise." (p.87) They furthermore observe that the authors of these reports, rather than presenting carefully reasoned arguments in support of their position, "present a litany of charges without examining the veracity of their evidence or its sources." (p.87)

*A Nation at Risk* provides 13 indicators of a disastrous decline in educational standards in American schools. One indicator contrasts achievement in the US with other nations; five other indicators describe contemporary achievement, while seven others contrast past achievement with that of the present.

Two of the five generalizations given concerning contemporary achievement point to high levels of illiteracy. The first stated that 23 million adults are functionally illiterate (*A Nation at Risk*, 1983,p.8), while the other referred to an illiteracy rate among teenagers of 13%. (p.8) Though the reports
suggest they are giving a view of the current situation, the data are nearly 10 years old, collected by the National Assessment of Educational Progress (NAEP) in 1974 and 1975. Many of the same items used by NAEP to assess literacy in those two years were also used in 1971. The findings from this earlier study raise some interesting questions. Fisher (1981) and Gadway & Wilson (1975) found that the 1974 and 1975 cohorts scored higher than the 1971 cohort, yet there is no mention in these reports of progress, only of decline.

The third generalization concerning the current situation states that "over half the population of gifted students do not match their tested ability with comparable achievement in school." (p.8) This is taken as evidence that bright students go unchallenged in school since their scores on standardized tests of ability indicate higher levels of ability than is suggested by their school grades. It is assumed that standards, as they are reflected by grades, have fallen so badly that the top students fail to maximize their potential. Because they are not challenged by demanding coursework bright students lose interest and become so apathetic their grades in school drop below what one would expect. It is also assumed that in cases where there is a difference in reported abilities, a standardized test is a more reliable indicator than a teacher's assessment.

In making the case for a decline in standards, the reports draw on data provided by SAT scores. Three of the seven indicators of a decline come from these scores. One indicator
cited a drop in SAT scores over the last 20 years. The second indicator found "consistent achievement test declines in recent years in such subjects as physics and English". (p.9) A third indicator drawn from SAT data found the number and percentage of very high test scores had dropped considerably over the last 20 years. However, there are very good reasons to doubt that SAT scores can be used to judge the effectiveness of the school system. First of all, SATs are not designed to test school achievement. Secondly, there is reason to believe SATs do little more than measure familiarity with upper middle class culture, and predict degree of success in the first year of college. (Schrader, 1971, p.119) Willie (1985) asserts that research has shown that SAT scores are valid predictors of academic performance for only the first year of college, that these scores explain less than half of the variance in first year college grades, and that academic performance during the first year of college does not predict performance during the fourth year of college for minority students, many of whom must make significant adjustments to adapt to the college environment. (p.626)

Slack and Porter (1980) demonstrate how SAT scores are not significantly better at predicting college grades than are high school grades alone. A study by Nairn and Nader (1980) concludes that the SAT is a poor predictor of college performance.

The Educational Testing Service (ETS), the designers and administrators of the SAT since 1929, have been insinching that the SAT is not and never was intended to be a measure of student
achievement. It is designed to be a scholastic aptitude test. It is meant to measure a student's capacity to learn in the future, not what he has previously learned. To demonstrate this point the ETS has, over the years, commissioned numerous studies to demonstrate the SAT is not measuring the extent of students' success in school. In fact the ETS insists that "the previous experience or training on the part of the individual is assumed either to be lacking or to be constant for all individuals comprising the population considered." To ensure that school experiences do not account for variations in test scores the ETS has intentionally excluded test items that would correspond to what is typically learned in school. Several independent analyses confirm that the ETS has succeeded in its ambition of ridding the SAT of questions which might relate to school curricula. According to such reports, SAT test items emphasize "little used vocabulary and algorithms that are rarely presented in high school courses." (p.163) For example, consider the following typical questions taken from the sample test published by the College Board in 1978.

25. In his, Parton reprints two comments upon her written by society people, one a man, the other a woman. These are not the putrescence later vomited upon her name by ordinary political buzzards of the newspaper press. These are of a different order -- comments presumably realistic by people of education professing friendship for her. They do not slander. They only sneer. Through every paragraph there runs under the main theme as sort of a contrapuntal melody the spitting of a cat. Every line is the mark of feline claws.

It can be inferred that newspaper articles written about Rachel were

(A) perfunctory and apathetic
(B) realistic and detailed
(C) insipid and ambiguous
(D) amusing and flippant
(E) vile and abusive

17. If p, q, and r are integers and q/p and r/q are both integers greater than 1, which of the following is NOT an integer?

(A) p/r
(B) r/p
(C) rp/q
(D) rq/p
(E) qr/rp  (quoted in Slack & Porter, 1980, p.163)

It is not that the test is measuring what should have been taught in schools, it is measuring exactly what the ETS expects will not have been taught in school and which it believes cannot be transmitted by "intensive last-ditch tutoring". (p.164) And yet the reports use the decline in SAT scores over the last 20 years as evidence that schools aren't doing their jobs.

As for the idea that the SAT measures ability or capacity for learning rather than prior learning, other studies have shown that intensive tutoring does indeed raise test scores. In one case, short-term tutoring produced a mean gain of 98 points, in another case, of 109 points, both considerably greater than the 50 point over 20 year national decline so widely publicized. (Slack & Porter, p.158-159) Fallows (1980) has noted of SAT preparation programs: "if courses can be designed for the specific purpose of increasing scores on the tests, does that not suggest that the tests reveal, rather than "aptitude" or "achievement," only mastery of an unusual or specialized system of thought." (p.45) Not only does all this raise doubts about
what is being measured in the SAT, it introduces the question of whether or not it is even sensible to talk about innate capacities or aptitudes which are sufficiently independent of experience to be free of the influence of culture and family background.

Among those who do poorly on the SAT are those least familiar with the world of those who do well. According to Porter and Slack (1980) those who do poorly are "students from poor families and minority groups whose experience with language, literature, and math are likely to differ from those of the test designers... With little opportunity to learn the little used vocabulary and tricky math of the SAT, they are unlikely to do well on the test." (p.163)

Data for 1974 released by the ETS show the relationship between family income and test performance. Those students who scored highest on the verbal ability section, those with scores between 750 and 800, are the same students who come from the wealthiest families; those with middling scores come from families in the middle range of income and so on. Fallows suggests that "what the tests measure is exposure to upper middle class culture -- perhaps even the culture of the professional class of the east coast." (p.47) The College Board has published some of its data on the relationship between economic standing and test performance.

<table>
<thead>
<tr>
<th>STUDENT SCORE</th>
<th>MEAN FAMILY INCOME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>145</td>
</tr>
</tbody>
</table>
He also points out how the manner of test construction and the procedures used by the test-makers to ensure the test's validity ensure that it is the same sort of people who year after year do well on the SAT. On each test there are a number of questions that have not appeared on other versions of the SAT and are there for the purpose of testing their reliability. Those who evaluate the new test items look at the number of times high scoring students chose the right answer. If high scoring students give the wrong answer, the test item is revised or abandoned. By means of this process, it seems plausible to maintain that new versions of the SAT are made up of the questions which were answered correctly by members of a particular social and economic stratum. As the table above suggests, students who happen to come from the wealthiest families have the most influence in determining the sort of questions which appear on the test. Fallows concludes that "if talents are diverse, if different groups display their abilities in different ways, this process [of testing] will never reveal it, because the standard set in the beginning is the standard it retains." (p.42)

An important effect of devising tests in this manner is the
way in which the results from these tests can serve to enshrine a kind of privilege. Insofar as access to the best colleges is regulated by reference to SAT scores, those who do well on these tests are given an important advantage over those who do less well. Harvard Business School produces far and away more successful and influential CEOs than do any of the colleges which would accept applicants with mediocre SAT scores. (Gifford, 1984, p.40) Future members of the upper middle class, it seems, are assisted in their pursuit of the most prestigious and highest paying jobs by means of a device which favours the upper middle class and which limits the extent education can provide social mobility to the less fortunate. It appears from SAT results that it is mainly the rich who are smart; it may be, however, that SATs simply define the rich as smart, while at the same time providing, as Fallows suggests "a scientific basis for the order of lords, vassals, and serfs." (p.48)

Another indicator of a decline in standards cited in A Nation at Risk rests on data taken from the National Assessment of Educational Progress. According to A Nation at Risk there has been a "steady decline in science achievement scores of U.S. 17-year-olds as measured by the NAEP in 1969, 1973, 1977." (p.9) The Commission fails to mention, however, that the decline is very small, as it is in other areas such as math and writing. (Stedman and Smith, 1983, p.89) As well, the Commission fails to mention that reading performance improved for young students, and remained more or less the same for teenagers during this period.
In fact, the Commission suggests the very opposite from these findings: that teenagers are less literate than they used to be and that the trend toward illiteracy is only getting worse.

The reports, especially the National Commission's, emphasis that standards are down because of a lack of academic emphasis in high school programs. Too many students take too many easy elective courses. There has been "a cafeteria-style curriculum in which the appetizers and desserts can be mistaken for the main course." (A Nation at Risk, p.14) The main evidence for this charge comes from a study by Adelman (1983) which analyzed changes in high school transcripts of two samples, one from 1964-1969 and another from 1975-1981. (Stedman and Smith, 1983, p.89) The first sample was of only 27 high schools. Very few of these were from the South and none of them were in cities with populations over 1 million. The second sample was a national survey of households. These are not comparable populations.

The Commission concludes, on the basis of data which show an increase in the number of general track and academic students electing to take driver training and the like, that there is therefore a de-emphasis of academics in American high schools. The same data, however, make it clear that there has not been a significant decline in the number or proportion of academic courses students take in high school. In some of the areas the Commission is most interested -- in math, science and foreign languages -- the differences are miniscule for academic students and represent an increase for general track students.
The Commission identifies other indicators of a lack of academic focus: declining amounts of homework, lax discipline, grade inflation. No attempt is made to demonstrate a link between the mainly anecdotal evidence for declines in amount of homework etc. and declines in achievement; the link is merely assumed. The strongest evidence they do provide consists of work by Coleman, Kilgore, and Hoffer (1981) that shows private schools, which as a group tended to "produce" higher test scores, also assigned more homework than public schools. The Coleman study raises the possibility that homework contributes to achievement gains. However, it can hardly be said on the basis of this study that declines in the amount of assigned homework account for a drop in achievement levels, or that increasing the amount of homework will contribute to achievement gains. Indeed, a study on effective schools by Echternacht (1977), cited earlier, found that many ineffective schools assigned homework, stressed discipline and emphasized academic subjects, while effective schools did not.

The point is not so much that one study is better than the next, but that the authors of the reports either fail to notice or choose not to report the full implications of the data. It may well be that there have been serious declines in standards and achievement. What seems undeniable is that some people have been willing to abandon, or for one reason or another have failed to
employ, ordinary standards of judgement in drawing conclusions from the data. Peterson (1983), in the position paper on the state of American education that accompanied Making the Grade, concluded that "nothing in these data permits the conclusion that educational institutions have deteriorated badly." (p.59) Yet as Stedman and Smith point out, the authors of Making the Grade argue as if the data do permit such a conclusion. (p.90)

There are similar problems with the way in which the reports make international comparisons. The reason for making these comparisons in the first place is to reinforce the idea that America's declining economic performance is attributable to its comparatively weak school system. The reports suggest that students in Europe and Japan are better educated and that if Americans are to be competitive the American school system should be modelled on the Japanese and western European systems. This would mean increasing the time schools devote to academic subjects by increasing the length of the school day and academic year, and emphasizing certain subjects such as mathematics, science, English, and foreign languages.

The bald assertion that current levels of achievement in the US are uniformly lower than in Europe and Japan is simply unsupportable. It is defensible to assert that achievement levels expressed as national averages show American students ranked lower than their European and Japanese counterparts during the years 1964 to 1971. The only available data tell us this much, and it this data, collected by the International Assessment of Educational Achievement (IEA), that the reports rely on. It does
not follow from the data, however, that this difference accounts in any way for variations in international levels of industrial productivity, or that it supports the view that American youths are less well educated than their foreign age-mates.

First of all, the data are 15 to 20 years old. Secondly, very small, select groups of students in foreign countries are being compared to a more comprehensive group of American students. Husen (1979) argues

There is no point in comparing the standard achieved in mathematics among upper secondary school students in a system with practically universal upper secondary schooling with that in a system where a small select fraction of the relevant age group proceeds to that level. The achievement standard in, say, mathematics of high school seniors in the United States is far below that of their age-mates in England who sit for the GCE or of the Oberprimaner in the Federal Republic. The pointlessness of such a comparison stands out when we find that 75% of the relevant age group in the United States is compared with only 15-25% in England or Germany. (p.95)

A more useful comparison, one between elite students, has shown that "the elite among US high school seniors did not differ considerably in their performance from their age-mates in France, England, or Germany." (Husen, 1983, p. 456) According to Stedman and Smith's analysis of more recent (1970-1971) IEA data the top 9% of US students do better than their foreign counterparts. It does not appear, then, that the matter of comparative levels of achievement is quite as simple as the reports make it sound.

As we have already noted, it is assumed in the reports that the low levels of achievement in American schools can be
attributed in large part to the poor quality of teaching. The way to improve the quality of teaching, the reports suggest, is to improve the quality of teachers. The criticism of teachers rests on a number of claims: their low test scores compared to other professional groups reveal an intellectually mediocre profession; teacher preparation programs emphasize pedagogy and ignore subject matter; low salaries fail to attract competent people to teaching; shortages of qualified math and science teachers have led to the use of unqualified teachers.

There are several problems with these claims. Given that education majors traditionally have had lower test scores than other majors, one would expect uniformly low levels of achievement among schoolchildren rather than the alleged sharp decline. Since this factor appears to have been constant throughout the years of the so-called decline in educational attainments, it cannot account for it. Furthermore, Jencks (1972) was unable to find any significant relationship between teachers' scores on standardized tests and the achievement of their students. (p.96,127)

The comparatively low SAT scores teachers are said to have earned in their high school days is taken as evidence that teachers as a group are likely to be incompetent. However, at best, SAT scores are evidence that prospective teachers are likely to have difficulty with university coursework in their first year of college. Beyond that there is very little that SATs
can predict about the future. Therefore, the inference that those who once did poorly on an aptitude test will turn out to be incompetent teachers five or ten years later is not valid.

The reports imply that teachers, as a group, have low SAT scores. In fact, the low SAT scores belong to individuals who at 17 or 18 thought they might like to be teachers and said so on the forms provided with the test. Even if one cavalierly assumes that those who indicated they would major in education did so, one could not assume that most teachers are represented in the SAT results of would-be education majors. The bulk of secondary school teachers, with the exception of physical education specialists, do not major in education, they major in their subject specialities of English, math, science and the like. (Clark & Marker, 1983, pp.55-56; Hacker, 1984; Stedman & Smith, 1983, p. 101)

There is recent evidence, also, that the majority of those who report their intention to become teachers do not end up as teachers. Lyson and Falk (1984) using data from the National Longitudinal Study of the High School Class of 1972 found that "seven years after high school graduation over three-fourths of the high school seniors who reported teaching plans were not teaching." (p.191) Of special interest was their further finding that "for high school students, the teaching option appears most feasible and attainable for white women, from rural areas, who performed well in high school. Nonwhites, those less gifted academically, and those from lower social class backgrounds are
more likely to abandon their original teaching plans in favor of some alternative employment." (p.191) The study identified six groups among those who reported their occupational plans as seniors: achievers, defectors, dropouts, climbers, converts, and professionals. Achievers reported plans to teach when they were seniors and were teaching seven years later. Defectors had planned to teach but instead entered professional or managerial positions seven years later. Lyson and Falk used the US census codes 001-196 to determine which jobs were professional, and codes 201-246 to determine managerial jobs. Dropouts expected to become teachers but later came to hold nonprofessional jobs (census codes 260-995). Those who aspired to professional and managerial jobs in high school but later came to hold teaching jobs were termed converts. Professionals were those who as seniors intended to become professionals and did so. Lyson and Falk found that among these groups SAT scores were highest for the defectors and professionals and lowest for the dropouts. Those who followed up their plans to become teachers, the achievers, were "more likely to be white and female, to have performed well in high school, but not to have scored high on the SAT." (p.190) This study by Lyson and Falk raises some important questions about the assumed relationship between SAT scores and academic ability, and clearly casts doubt on the assumption that those who say they want to be teachers do in fact end up in the classrooms.

Unfortunately the complicated question of the SAT's
usefulness goes largely ignored in these reports. They are more intent in laying out the crisis in education and presenting recommendations for reform. A Nation at Risk reminds us that "History is not kind to idlers". (p.13) Educating Americans likewise warns us that "human history becomes more and more a race between education and catastrophe." (p.3) If the American standard of living and way of life are to survive the schools will have to improve. If the recommendations of these reports are acted upon, improvement will mean different things to different people. For students it will mean tougher grading and courses; no longer will students be allowed to have "a cafeteria style curriculum." Instead students will be required to take more academic courses, especially in math, science and foreign languages. They can expect more homework, more time in school and more tests, especially standardized tests. A Nation at Risk recommends a national system of standardized tests of achievement be implemented to guide certification and identify needs. Those who excel at school, the top 2%, will be eligible for the 2000 proposed magnet schools that would specialize in math and science. (Educating Americans, 1983) Those students who fail to meet the new demands are to be placed in special programs designed to give them "basic employment and economic competencies". (Action for Excellence, 1983, p.50)

The recommendations of the reports mean something else for teachers. They will be subject to the same longer days and longer years their students face. They will need to assign more homework and grade it. But most of all they will be subject to a degree of
examination and inspection unlike anything seen to date in the schools. Their effectiveness and that of the school as a whole will be judged by the results of standardized tests. Principals will be expected to increase the number of inspections and will be held accountable for the performance of teachers. Teacher salaries and those for principals will be tied to measures of their effectiveness. With a clearer idea of who is "effective" and who is not, teachers and principals can be ranked and fitted into the expanded career ladders all the reports called for. Pre-service teachers will be required to pass a standardized test in order to become certified; not a new circumstance, but one the authors of the reports think should be universal.

The result of the sort of reforms recommended by *Educating Americans* will be "education for all American elementary and secondary students so that their achievement is the best in the world by 1995." (title page) Having the best educated children will, it is argued, bring many rewards: "improved productivity; sustained economic growth; job and career opportunities for all people; the economic wherewithal to provide adequate public services; a secure defense." (p. 16)

One of the chief concerns expressed by critics of these reports is that a likely consequence of these measures is that many students will drop out of school. Unless the assumption made in the reports that students are not being challenged is correct, students already struggling with school will face even greater
difficulties. John Lawson commenting in the Boston Globe said in response to calls for higher standards: "if a kid can't clear four feet, it doesn't do much good to raise the bar to four feet, six inches." (cited in Duckworth, 1983, p.15) More of the same schooling is not likely to do more for children who are already struggling in school than convince them to drop out. Boston schools in 1983 had a dropout rate of 50% according to Rosemarie Rosen, deputy superintendent for finance and administration for Boston Public Schools. (Rosen, 1983, p.28) Higher standards will not likely reduce the reasons these students had for dropping out, though paradoxically, they may ultimately reduce the drop-out rate in cities like Boston simply by means of attrition.

McDill, Natriello, and Pallas (1986) indicate that the drop-out problem is a serious one. Their summary of the literature on dropouts indicates there are three main factors which are associated with teenagers dropping out of school. Poor academic performance is the most commonly given reason students have for dropping out.

The second most powerful predictor of dropping out is "family formation"; more simply, teenage pregnancy and/or marriage. According to Neill (1979) "about one million adolescent girls -- nearly one in ten -- conceive each year, and 600,000 young women carry their pregnancies to full term.... An estimated 400,000 pregnant teens are under 17 years of age." (p.32) Eight out of ten of these mothers under age 17 never
finish high school. For these students, keeping up with school becomes impossible, and many female dropouts report pregnancy as a reason for dropping out. Since approximately 60% of all teenage mothers end up on welfare, it seems reasonable to assume many of them come from poor families. (cited in McDill et al., p.141, 171)

The third category of conditions associated with dropping out is economic factors. According to Peng et al. (1983) more than 1/4 of all male drop-outs report they left school for a job. Roughly 1/5 of minority male dropouts and 1/10 of all other dropouts report leaving school because they had to support a family. A 1983 study by Michael and Tuma revealed that 25% of all 14 year olds and 50% of all 17 year olds held part time jobs while attending school. D'Amico (1984) indicates that working 12th graders average 15 to 18 hours of work per week. It seems reasonable to assume that those who left school for a job, or to support or raise a family are, in a great many cases, those with the least amount of money to begin with. Likewise many of those teenagers who hold part time jobs surely do so out of necessity. Longer school days, a longer academic year, and more hours of homework will make it increasingly difficult for students in poor families to continue in school.

Meier (1983) has already pointed out that one way of making schools look more effective is to redefine the population of students in schools. While her remarks were directed to the
proliferation of "special education" programs which remove students from regular programs, they could just as easily be applied to those efforts to raise standards which are predicated on the assumption that teachers and students are working well below capacity. Several critics of this literature have noted this tendency already, especially where tests will be used to determine promotion to higher levels of education, or when they govern the awarding of valued diplomas.

Madaus and Greaney (1985) in their analysis of the effects of competency testing in Ireland between 1943 and 1967 note that one of the chief effects of the testing program was a widespread attempt by teachers to limit the number of students who wrote the tests, or to slow down the rate of progress of the weaker students. The most important test students faced in Irish schools during this period was the Examination for Primary School Certificate, set at the end of grade six when students were 12 or 13. The certificate was valued by students in much the same way high school diplomas are today. They provided a qualification for those who would continue into postprimary schooling, and were very useful in helping students obtain jobs once they left school. Madaus notes:

The evidence suggests that when the primary certificate became compulsory, Irish teachers adopted a number of strategies to control the failure rate. There was an overall decrease in promotion rate, with a pronounced decrease in promotion rate from third to fourth grade and, more important, from fifth to sixth grade.
Similarly, Steelman and Powell (1985) express concern over the way in which states which have few students write the SAT appear to have higher standards in its schools, an impression at least one state governor has capitalized on in an attempt to draw investors to his state. South Dakota tried, in a 1982 national advertising campaign, to attract business investment by publicizing its relatively high SAT scores. It didn't announce, however, that only 2% of its students write the test compared to 69% in Connecticut or 59% in New York, two states which scored lower. When considerations such as percentage of students who write the SAT are taken into account, New York's ranking among all the states goes from 35th to 5th, while Utah's went from 8th to 46th, and North Dakota's from 3rd to 30th. (1982 data)

Steelman and Powell are concerned that given the political reasons for keeping SAT scores high or having them improve, educators may discourage some students from writing the SAT. Indiana, which ranked 46th in former Secretary of Education Terrel Bell's state "school scoreboard", tried unsuccessfully to reduce the number of students taking the SAT. Georgia, which ranked 49th, has discussed plans to eliminate the SAT as an entrance requirement for state colleges. Alabama, between 1972 and 1983, enjoyed an average increase in SAT scores of 114 points. At the same time the proportion of students in Alabama who wrote the test declined by one third. (pp. 603-607) There is a danger, which these statistics only hint at, that the appearance of success will be bought at the expense of the least able.
Even though *Educating Americans* insists that "academic excellence...does not mean the provision of high quality education to only a small group of highly talented youth" and that *A Nation at Risk* declares "that all children by virtue of their own efforts...can hope to attain the mature and informed judgement needed to secure gainful employment", there is little evidence in these reports to suggest we should not fear that the stage is being set for "once again abandoning the disadvantaged in the name of "excellence". (Meier, 1983, p.63) Former US Commissioner of Education Harold Howe is similarly concerned about the manner in which equity is ignored in these reports:

Fairness to both students and taxpayers in the funding of education, along with continued attention to issues of discrimination on the basis of sex, race, and national origin, constitute a continuing equity agenda that is ill attended in the reports and studies or responses to them. (cited in McDill et al., 1986, p.139)

Teacher effectiveness and school effectiveness research, along with the policy implications of the commissioned reports represent a far-reaching, though partially developed system for controlling teachers and students. The school system is, as Meyers has noted, a nested system, with classrooms nested in schools, schools nested in districts and so on. These three bodies of research have produced effects throughout the various nested layers, connecting them together in what Foucault would term a disciplinary matrix. One might wonder what the school system would look like if such a matrix were to operate
unopposed. The stage for hierarchical observation, normalizing judgement, and examination has already been set. It is a simple matter to predict what the school system would look like in an imaginary world where domination was complete, and accountability procedures fully determined the actions of teachers and students. Teachers would be pressured to test and rank students. Standardized tests, teaching materials and methods of teaching which amount to the same thing as testing would represent the foundation of instruction. Principals would be expected to ensure teachers teach the standardized curricula in a standard way. The performance of teachers and their supervisors would be evaluated by scores on city-wide, district-wide or state wide norm-referenced exams. Promotion and salary increases for both teachers and school administrators would be tied to raising test scores. The ability of states to ensure acceptable standards would be judged by means of measures like the SAT, or regulated through a reliance on minimum competency testing. Overall educational effectiveness would be judged by international comparisons and continuing assessments. As we have seen, to varying degrees, and in a piecemeal fashion, this is already being done.

If we could be confident that Foucault is correct in thinking disciplinary power produces docile, useful and productive bodies, we might only object to these practices on the ground they represent a form of power that is "counterlaw" and which operates largely unrecognized and beyond the reach of ordinary people. And many of us would likely be very
uncomfortable with any coercive social or political arrangements, regardless of how effective they might make our school system. We would surely also want to object to the way disciplinary power tends only to operate alongside the production of failures, anomalies, problems to be solved. As Foucault has demonstrated in *Discipline and Punish*, people classified in this way have tended to act as sites for the further operation of power and knowledge.

Political technologies advance by taking what is essentially a political problem, removing it from the realm of political discourse, and recasting it in the neutral language of science... The language of reform is...an essential component of these political technologies. Bio-power spread under the banner of making people healthy and protecting them. When there was resistance, or failure to achieve its stated aims, this was construed as further proof of the need to reinforce and extend the power of the experts. (Dreyfus & Rabinow, 1983, p.196)

Foucault noted in *The History of Sexuality* that the restricted social role for women, and the ensemble of coercive practices which were associated with the control and regulation of sexual urges, drives, etc., during the Victorian era were both justified on the grounds that the health of the nation and its future welfare were at stake. In a similar, if less dramatic, fashion, the commissioned reports on education justify increased supervision and regulation of teachers and students on the grounds that the nation is at risk. It is particularly disturbing that the means of regulation and control in schools are likely to lead to circumstances that will be
"construed as further proof of the need to reinforce and extend the power of the experts", and will serve to justify the imposition of ever finer forms of supervision and inspection.

Be that as it may, a factor which Foucault tends to downplay is the human capacity for escaping or otherwise resisting coercion. Resistance is likely to have its own deleterious effects. We have seen that at-risk students are likely to drop out of school, as are teachers who, feeling they have lost a good deal of their autonomy, resent being managed by remote control. Teachers and administrators are encouraged by the emphasis on testing to teach to the test, to tamper with test results, or to redefine the population of test-takers in order to appear successful. We have also seen that children are likely to be given fewer and fewer opportunities for sustained reading and critical discussion, and to be subject to hours of repetitive drill and practice at subskills. The "tradition of past exams" will come to define the curriculum. And as Madaus has warned, educators will be encouraged to "hand over crucial decision making to a surrogate teacher [the test] with a narrow range of multiple choice questions, measuring an equally narrow range of abilities, with the final mechanical decision based on a still narrower point on a numerical scale..." He further adds: "the agency that controls test content through the tradition of past exams...controls what is taught and how it is taught."

(1985,p.615-616) Those who do not find themselves stigmatized and excluded, those not identified as "ineffective" teachers, "problem" students, or "special" students, will find themselves...
in classrooms where established norms of teaching pressure teachers to teach in ways that do little to encourage independent rational judgement, and which may in fact inhibit youngsters' efforts to learn to read. Students will find themselves in classrooms where norms of achievement defined by standardized tests determine the content of education. And where, in time, normal distributions of grades are the only ones which appear to be normal. In this way a certain percentage of failures is inevitable, as it is now.

There is every reason to believe that those who fall below acceptable levels of competence will, in the majority of cases, be poor whites and visible minorities. Since their exclusion will not appear to have come at the hands of the politically motivated, they will appear to deserve the limitations on their freedom that must surely result from this narrowing of their opportunities. Paradoxically, Edmonds, in his advocacy of teacher effectiveness research will have contributed to the adoption of practices which will further disenfranchise blacks and the disadvantaged. By their reliance on quantitative measures in the struggle to secure equal opportunities for the disadvantaged, the liberal reformers of the Coleman-Jencks era will have secured just the opposite. It was the Coleman and Jencks studies that did so much to legitimate the use of social research in educational policy formation, and which revealed to successive generations of social scientists the possibilities such research possessed.
Taken together, these three bodies of educational discourse exemplify the workings of bio-power. Teacher effectiveness and school effectiveness research provide the means of disciplining individual bodies; students are increasingly pushed to exhibit a narrow range of competencies in the same ways, while teachers are pressured to standardize their teaching practices. The targets of power in the literature of effectiveness are individual teachers and students. The research on the state of education at the national level provides the justification for the regulation of the social body in the interests of its preservation and well-being. Yet, this research, if implemented and accepted by educators, will likely contribute to declines in literacy and numeracy so long as the tests which drive the curriculum and teaching practices in general remain inadequate. While they serve the purpose of monitoring instruction and ensure a degree of accountability, they do not represent the goals we have in mind when we speak of education as the initiation of the young into worthwhile forms of knowledge and understanding. These tests are not likely to lead us any closer to attaining those goals.

Since it is evident that teachers will resist -- they are doing so now -- their resistance will likely have a deleterious impact on students. Whether teachers accept or reject the system of schooling that is beginning to emerge, the system will have the capacity to produce failures in record numbers, which of course will be the responsibility of the school system, or at least some body of learned individuals, to correct.
The picture needn't be as gloomy as this, however. It may be granted that whether or not we accept Foucault, the direction in which the American school system appears to be drifting is one which we should view with grave concern. This picture, however, is only truly bleak if we accept with Foucault the supremacy of disciplinary power over democratically constituted power. As Michael Walzer has noted, Foucault very often speaks as if there is no center of power to which we have recourse, no way in which we can directly influence the course of events. Yet his own work would be pointless if this were the case. The service that Foucault has done for us is to bring to light the operation of discipline in institutions that "appear to be both neutral and independent: to criticize them in such a manner that the political violence which has always exercised itself obscurely through them will be unmasked, so that we may fight them." But he underestimates the effectiveness of democratic forms of resistance in modern society and this society's "universal juridicism ...[which] seems to fix limits on the exercise of power." In the end, for Foucault, society's "universally widespread panopticism enables [discipline] to operate, on the underside of the law, a machinery that is both immense and minute, which supports, reinforces, multiplies the asymmetry of power and undermines the limits that are traced around the law." It is fortunate that he has warned us in the way he has. But if Foucault truly felt disciplinary power was as irresistible as he sometimes makes it seem, there would have been little point in
his trying to "unmask" the way it operates. Clearly there is some point to illuminating the ways in which the American school system is being transformed. The purpose of this thesis has been to provide just this sort of illumination, just the sort of "analytics of power" that might unmask the potential for political violence which the literature on effective school reform holds.
REFERENCES


Averch, H., Carroll, S., Donaldson, T., Kiesling, H., & Pincus, J. How effective is schooling? A critical review and synthesis of research findings. Santa Monica: The Rand Corporation, 1972


169


Fallows, J. The tests and the "brightest": How fair are the College Boards?. Atlantic Monthly, 1980,


Foucault, M. Language, counter-memory, practice: Selected essays

170

Foucault, M. Orders of discourse. Social science information, 1971, 10, 7-30.


Howe, Harold. Let's have another SAT score decline. Phi Delta Kappan, 1985, 66, (9), 599-603.


James, T. & Tyack, D. Learning from past efforts to reform the high school. Phi Delta Kappan, Feb. 1983.


Madaus, G.F. Test scores as administrative mechanisms in educational policy. *Phi Delta Kappan, 1985, 66*, (9), 611-618.


Simon, B. *Can education change society?*. In J. Donald Wilson (ed.), *An imperfect past: Education and society in Canadian