

A PHILOSOPHICAL CRITIQUE OF STUDENT ASSESSMENT PRACTICES

by

ROBERT EDWARD SPEAR

B.Ed., The University of Manitoba

M.Ed., The University of Manitoba

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES  
Social and Educational Studies

We accept this thesis as conforming  
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October, 1991

© Robert Edward Spear, 1991

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  
Vancouver, Canada

Date 12/16/91

### ABSTRACT

Standard critiques of student assessment practices (i.e., testing, marking and grading), typically take the form of either a technical critique of assessment instruments, or a socio-political critique of the general enterprise of selection, or both. These approaches can be limiting, however, in that they do not always directly address pedagogical and moral concerns. In an attempt to compensate for their limitations, this study offers a philosophical critique of testing, marking and grading. Drawing on the work of R.S. Peters, Israel Scheffler and Thomas Green, an initial account of 'education' and 'teaching' is offered, which is then used to critically review four common defences of testing, marking and grading. The four defences are 1) that the activities of testing, marking and grading are central to the activity of teaching, 2) that a system of testing, marking and grading motivates students to learn, 3) that a system of testing, marking and grading ensures accountability, and 4) that a system of testing, marking and grading is a necessary and defensible mechanism for sorting people on the basis of academic achievement. As a result of this review, it is concluded that much of what is done within schools in the way of testing, marking and grading undermines the educative project. Some of what is done is also morally suspect. It is hoped that by making the pedagogical and moral objections clear, the way will then be open to redefine the point and purpose of individual assessment within school, and to reshape assessment practices accordingly.

## TABLE OF CONTENTS

Abstract .....	ii
table of contents.....	iii
.....	iii
acknowledgements .....	iv
CHAPTER ONE.....	1
INTRODUCTION .....	1
Education & Teaching - A Conceptual Overview .....	10
CHAPTER TWO .....	19
THE PEDAGOGICAL DEFENCE OF TESTING, MARKING, & GRADING.....	19
The Pedagogical Core of Assessment .....	19
The Pedagogical/Classificatory Distinction.....	23
The Hazards of Classificatory Assessment .....	25
Current Practices: Classificatory or Pedagogical?.....	30
Conclusion .....	39
CHAPTER THREE .....	40
THE MOTIVATION DEFENCE OF TESTING, MARKING & GRADING.....	40
Initial Difficulties .....	40
The Central Critique.....	42
Objections .....	50
Conclusion .....	53
CHAPTER FOUR .....	55
THE ACCOUNTABILITY DEFENCE OF TESTING, MARKING AND GRADING.....	55
Accountability and Education .....	55
Testing, Marking & Grading as Mechanisms of Quality Control .....	59
Conclusion .....	66
CHAPTER FIVE .....	68
THE SELECTION DEFENCE OF TESTING, MARKING & GRADING.....	68
The Problem of Desert Based on Achievement .....	70
The Problem of Selection Within Schools .....	77
Conclusion .....	85
CHAPTER SIX .....	86
CONCLUSION .....	86
BIBLIOGRAPHY.....	118

### ACKNOWLEDGEMENTS

I am indebted to a number of individuals for helping me bring this work to its completion: Kelvin Beckett, for taking the time to talk through the arguments to see where they led; Jean Barman, for providing that sensible third ear which forced me to be more intelligible; and Jerrold Coombs, for having the patience to let me find my own voice in all of this. I would especially like to thank Roi Daniels, my supervisor, for his advice and encouragement, both of which he gave in exactly the right measure, at exactly the right time. I am also deeply indebted to Murray Ross for introducing me to the topic of student assessment practices, and more importantly, for sharing with me something of the point and passion of educational philosophy. Above all else, however, I owe an unpayable debt to my best friend and partner, Dyan Spear.

## CHAPTER ONE

### INTRODUCTION

Most teachers know that there is *something* wrong with the way in which we conduct student assessments within schools. Most know, for example, that there is something peculiar about spending so much time testing, marking and grading students at the expense of teaching them, and something inappropriate about using the threat or promise of marks to get students to attend to their studies. Most teachers know, as well, that there is something not quite right about an administrator pointing to final grades as an indication of a teacher's competence, and something downright indefensible about an education system which convinces at least thirty per cent of the people who go through it that they are stupid. The problem, however, is that while many teachers are aware of these difficulties, they have not had at their disposal the kind of critique which gives voice to the full range of these concerns. The standard critiques of student assessment practices are either too narrow, in the sense of focusing on the instruments rather than the enterprise as a whole, or too pessimistic, in the sense of abandoning altogether the possibility of making authoritative appraisals within an educational context. The aim of this study, then, is to begin to develop a more adequate critique of student assessment practices (i.e. testing, marking and grading), by directly addressing pedagogical and, to some extent, moral concerns. It is hoped, moreover, that the development of such a critique may lead, eventually, to a fundamental transformation of the way we understand and perform individual assessments within schools.

Within the literature, critiques of student assessment practices<sup>1</sup> generally fall into one of two categories: a technical critique of assessment instruments (i.e., tests), or a socio-political

---

<sup>1</sup>I shall make a distinction, right from the beginning, between studies of *evaluation* and studies of *assessment*. I shall take the term "evaluation" to cover the range of judgements made on everything except individual student progress, and "assessment" to refer exclusively to that latter undertaking. Evaluation studies, therefore, include such things as program evaluation, curriculum evaluation, or materials evaluation, while studies of assessment refer to investigations of the instruments and procedures used to identify individual student "ability", "aptitude", or "progress". The literature on evaluation - particularly program evaluation - is broad. Leading commentators include Ralph Tyler, Lee Cronbach, Michael Scriven, Robert Stake, Ernest House, Daniel Stufflebeam, and George Madaus. Although the focus of this dissertation is student assessment, there are two senses in which the work on evaluation sometimes becomes relevant. First, there are a number of philosophical insights discussed within this literature which may be applicable to the problems of student assessment. Second,

(i.e., "radical") critique of the business of assessment as a whole. While some treatments of the subject concentrate on just one type of analysis, many contemporary commentaries of student assessment practices incorporate both.<sup>2</sup>

Technical critiques of assessment instruments make up, by far, the bulk of the literature. These critiques are typically prosecuted in one of two ways: either the technical design of particular testing instruments is called into question or the appropriateness of using certain tests for certain purposes is challenged. In the first case, detractors might argue that certain tests lack construct validity,<sup>3</sup> or have poor predictive power,<sup>4</sup> or are culturally biased.<sup>5</sup> In the second case, the point is sometimes made that all tests serve particular purposes, and that it is a serious mistake to use one kind of test within a context to which it is not properly suited.<sup>6</sup> In both cases the impression is given that if we can only fix our tests,<sup>7</sup> or if we can only become more sensitive about their proper use,<sup>8</sup> the problems of student assessment can be summarily solved.

---

discussions of evaluation sometimes converge with discussions of students assessment at the point where issues of accountability arise. Given that some practitioners seem to want to use student assessment procedures as a vehicle through which to carry our *program* evaluation, the analyses necessarily overlap.

<sup>2</sup>One section of Paul Houts' The Myth of Measurability (New York: Hart Publishing, 1977), for example, concentrates on a technical critique of standardized testing by asking whether or not these tests measure what they are supposed to measure, while another asks the socio-political question as to whether or not these tests are consistent with the central task of democracy - i.e. to reduce inequality. For examples of other commentaries on student assessment practices, see Banesh Hoffman, The Tyranny of Testing (New York: Crowell-Collier, 1962); Andrew Strenio, The Testing Trap (New York: Rawson, Wade, 1981); David Owen, None of the Above: Behind the Myth of Scholastic Aptitude (Boston: Houghton Mifflin, 1985); Frank Smith, Comprehension and Learning, (New York: Holt, Rinehart, and Winston, 1975): 203-245; John Matthews, Examinations (London: George Allen & Unwin, 1985); Don Robertson and Marion Steele, The Halls of Yearning: An Indictment of Formal Education/A Manifesto of Student Liberation (San Francisco: Canfiled Press, 1969).

<sup>3</sup>George Madaus, Peter Airasian and Thomas Kellaghan, School Effectiveness: A Reassessment of the Evidence (New York: McGraw-Hill, 1980): 125-133, 155-158. See also, Deborah Meier, "Why Reading Tests Don't Test Reading," Dissent 28 (Fall, 1981): 457-466.

<sup>4</sup>Charles Willie, "The Problems of Standardized Testing in a Free and Pluralistic Society," Phi Delta Kappan (May, 1985): 626-627. See also, Strenio, Testing Trap, 127-147, 230, 267.

<sup>5</sup>James Fallows, "The Tests and the 'Brightest': How Fair are the College Boards?", The Atlantic, (March, 1980): 39. See also, Matthews, Examinations, 120-137; and Strenio, Testing Trap, 203-206.

<sup>6</sup>John D. Casteen III, "The Public Stake in Proper Test Use," in Charles C. Davis, ed., The Use and Misuse of Tests (San Francisco: Jossey-Bass, 1984).

<sup>7</sup>Paul Houts, Myth of Measurability, 19; Madaus, Airasian and Kellaghan, School Effectiveness: A Reassessment, 113-115; Davis, Use and Misuses (San Francisco: Jossey-Bass, 1984).

<sup>8</sup>Davis, Use and Misuses.

The problem with this approach, however, is that it is too narrow. It is not enough, that is, to ask whether a test does what it purports to do, or whether a given test is appropriately used within a given context. What is needed, rather, is to go one step behind these issues and ask what, if anything, testing (and not just testing, but testing, marking, and grading understood as an integrated set of procedures), might have to do with the central aims of education and teaching. Patricia Broadfoot got to the heart of the matter when she wrote that "If only the efficiency of assessment practices is questioned and not their purposes and effects, ... then the scope for educational change and reform is very limited."<sup>9</sup> She is, I think, importantly correct in making clear that a purely technical analysis of student assessment practices runs the risk of ignoring some of the more fundamental issues implicit within our use of testing, marking, and grading.

Against this rather restricted view of the problems of student assessment is the socio-political (or radical) critique of testing, marking, and grading.<sup>10</sup> Here the scope of analysis is much broader: in addition to, or apart from, examining the technical shortcomings of particular assessment instruments, this level of critique explores the way in which the practices of testing, marking, and grading operate as mechanisms of social control. The basic thesis of the radical critique is that schooling, far from being an institution which liberates, actually serves to reproduce political and economic inequality.<sup>11</sup> The practices of testing,<sup>12</sup> marking and grading,<sup>13</sup> therefore, become instruments of oppression. The point of departure for most

---

<sup>9</sup>Patricia Broadfoot, Assessment, Schools and Society (London: Methuen, 1979): 18.

<sup>10</sup>See especially Broadfoot, Assessment, Schools and Society; Patricia Broadfoot, ed., Selection, Certification and Control (London: Falmer Press, 1984); Ian Hextall and Madan Sarup, "School Knowledge, Evaluation and Alienation," in Michael Young and Geoff Whitty, eds., Society, State and Schooling (Sussex: Falmer Press, 1977).

<sup>11</sup>Jonas Soltis and Walter Feinberg, School and Society (New York: Teachers College Press, 1985): 43; Patricia Broadfoot, Assessment, Schools, and Society, 25-26, 48-50, 76-82; Frank J. Mifflin and Sydney C. Mifflin, The Sociology of Education (Calgary: Detselig Press, 1982): 50-58.

<sup>12</sup>Clarence Karier, "Testing for Order and Control in the Corporate Liberal State," in Roger Dale, et. al., eds., Schooling and Capitalism: A Sociological Reader (London: Routledge and Kegan Paul, 1976).

<sup>13</sup>Ian Hextall, "Marking Work," in Michael Young and Geoff Whitty, Explorations in the Politics of School Knowledge (Driffield: Studies in Education, Ltd., 1976).



radical critiques is a keen appreciation of, and interest in, social equality and justice.<sup>14</sup> Most radical theorists condemn current assessment practices as facilitating the construction and legitimization of "elite" knowledge. This construction and legitimization are reprehensible, they would argue, because they conspire against the possibility of responding to people as equals.

Although there is much within the radical critique of testing, marking, and grading with which I agree, it too has its limitations. In the first place some variations of the radical critique embody certain implications of epistemic relativism - that is, the belief that because all knowledge is, in one (trivial) sense, socially constructed, the strength of propositions rests not in their veracity or inherent logical coherence but instead in the social and political power of those who offer them.<sup>15</sup> Although it is not my intention in this dissertation to address directly the issues of epistemic relativism and the social construction of knowledge, I will say in passing that, insofar as educating students involves making judgements about the adequacy of the way in which students hold beliefs, epistemic relativism is, to say the least, somewhat problematic.<sup>16</sup> Within certain variations of the radical critique there is, as well, a rather depressing sense of inevitability concerning the extent to which a "dominant" class can control its inferiors. What is so troubling and, more importantly, so unrealistic about this view is that it leaves little room for the possibility of resistance or rebellion on the part of those being dominated. Schools, on this view, are *by definition* coercive institutions and, apart from abolishing them completely, there is little that can be done to improve them.

While the technical and radical critiques have serious limitations, there nonetheless exists a third type of analysis which may yet overcome these. The underlying premise of this

---

<sup>14</sup>Michael Apple, "The Process and Ideology of Valuing in Educational Settings," in Michael Apple, Michael J. Subkoviak and Henry S. Lufner, eds., Educational Evaluation: Analysis and Responsibility (Berkeley: McCutchan Publishing, 1974): 3-34.

<sup>15</sup>See, for example, Michael Young, ed., Knowledge and Control (New York: Collier-MacMillan, 1971).

<sup>16</sup>For a thoughtful response to Young's project see Richard Pring, "Knowledge Out Of Control," Education for Teaching 89 (Fall 1972): 19-28.

third approach is perhaps best captured in a point made by educational sociologists Stanley Aronowitz and Henry Giroux. In the context of a discussion on school reform, they have argued for a view of educational research in which a "language of possibility" is taken just as seriously as the more frequent "language of critique."<sup>17</sup> What Aronowitz and Giroux seem to have in mind is a comprehensive analysis of schooling which makes clear not only the many ways in which schools function as institutions of domination, but also the many *possibilities* for schools to operate otherwise. This, it seems to me, is the right approach. In any examination of student assessment in particular, we need an analysis which can reveal not only what is wrong with our current practices, but which also points the way to possibilities for improvement. We need the kind of analysis, moreover, which is capable of directly addressing the full range of pedagogical and moral issues. I take it that the way to conduct such an analysis is to begin with a careful examination of what we seem to have in mind by educating and teaching students, and then to use our findings as a base from which to reflect on the adequacy of our current practices. I take it, in other words, that the way to build a "language of possibility" is to begin with a *philosophical* critique of student assessment practices prosecuted from the point of view of a fully-articulated conception of education and teaching. It is toward this particular end, then, that I would like to make my own contribution.

The strategy, therefore, is to *begin* with a fully-articulated conception of education and teaching, and then to use this conception to reflect on four common defenses of testing, marking and grading. The four arguments I shall review are 1) that the activities of testing, marking and grading are central to the activity of teaching, 2) that a system of testing, marking and grading motivates students to learn, 3) that a system of testing, marking and grading ensures accountability within schools, and 4) that a system of testing, marking and grading is a necessary and defensible mechanism for sorting people on the basis of academic achievement, and thereby for deciding which students should proceed into what occupational niches. Out of

---

<sup>17</sup>Stanley Aronowitz and Henry Giroux, Education Under Siege: The Conservative, Liberal and Radical Debate Over Schooling (South Hadley, Mass.: Bergin & Garvey, 1985): 1-21.

this review a number of philosophical and moral critiques against our current practices will emerge.

Before proceeding, in the last section of this chapter, to a conceptual sketch of education and teaching, it may be useful to say a few brief words about the methodology of analysis attempted here, and the kinds of objections which might be raised against it. Since part of my project is to criticize certain assessment practices from the point of view of what is generally and philosophically regarded to be the point and purpose of education and teaching,<sup>18</sup> the first objection that might be made against this procedure is to suggest that all I am doing is using a stipulated definition to mount the kind of critique I want to mount against these practices. The point of this objection is to say that if someone else had a *different* conception of education and teaching, then it is a simple matter to replace that conception with the one offered here, and thereby reach different conclusions about the efficacy and appropriateness of what we do in schools in the name of testing, marking and grading.

Against this objection, I want to suggest that the conception I am using is not *merely* a stipulation; it is instead a rather precise *account* of what thoughtful language users actually have in mind when using these terms. Jerrold Coombs and LeRoi Daniels refer to the undertaking of providing such an account as concept interpretation.<sup>19</sup> The object is to try and make more concrete those concepts which are usually understood in only the most general and abstract terms. 'Education' and 'teaching' are perfect candidates.

Concept interpretation usually involves two things: a descriptive account of how we use concepts, and a logical analysis of what is implicit within the concepts we use. When Peters tries to explicate the distinction between education and training, he appeals essentially to

---

<sup>18</sup>My aim, that is, is not to construct any new conception of education and teaching, but instead to draw on the considerable work that has already been done in this field. I intend, in particular, to use some of R.S. Peters' insights concerning the concept of education, and much of what Thomas Green and Israel Scheffler have to say about the activity of teaching.

<sup>19</sup>Le Roi B. Daniels & Jerrold R. Coombs, "Analytic Philosophical Inquiry," in Edmund Short, ed., Forms of Curriculum Inquiry (Albany, N.Y.: State University of New York, 1991).

ordinary language use - i.e. to the fact that most of us, if pressed, do in fact differentiate between these concepts. Thus he asks why it is "for instance, that we talk more naturally of educating the emotions than we do of training them, whereas we talk more naturally of training the will than we do of educating it?"<sup>20</sup> (His answer, by the way, is that while emotions are predicated on beliefs, which can be changed by virtue of reasons, the will refers essentially to remaining steadfast in one's purposes - whatever they are.) It is out of the general experience of language users, then, that Peters is eventually able to establish his conceptual point.

The concept of educating people, as distinct from merely training them, has developed and is mirrored in the different words that people use who have developed this differentiated concept. The fact that there may be many people who do not have this concept, or who have it but use words loosely, does not effect the conceptual distinction to which I am drawing attention.<sup>21</sup>

The second technique of identifying what seems to be logically implicit within concepts is also used by Peters. His contention that the concept of education necessarily implies the transmission of something *worthwhile* is meant, at base, to be a claim about what is logically implicit in that term. (The logical link is made, by the way, via the conceptual realization that education, in its standard usage, is essentially an approbative term.) Scheffler's analysis of the conditions of knowledge,<sup>22</sup> likewise, is essentially a logical examination of what is implicit within our use of that term.

What is being offered here, therefore, is not merely a stipulative definition, but instead a set of claims about how the terms 'education' and 'teaching' are currently and generally understood, coupled (in some cases) with claims about what seems to be logically implicit within their usage. Although it is certainly possible to offer alternative conceptions of education and teaching, insofar as these are forwarded as descriptive rather than programmatic

---

<sup>20</sup>R.S. Peters, Ethics and Education (London: George Allen & Unwin, 1966), 32.

<sup>21</sup>Ibid., 30.

<sup>22</sup>Scheffler, The Conditions of Knowledge (Scott, Foresman & Company, 1965).

or persuasive definitions,<sup>23</sup> they are subject to the public test of ordinary language usage. If someone wants to mount a critique of the conceptual account offered here, they must first show how the conception offered does not accord with ordinary usage, and then provide some sort of reasonable substitution. Insofar as I believe that the philosophers which I shall draw on have, by and large, got much of this conceptual work right, it seems to me that the possibilities of providing an alternative descriptive account of education and teaching are limited indeed.<sup>24</sup>

A second objection that might be raised against the sort of project I have in mind could be phrased as follows: even if one accepts the conceptual constellation presented regarding education and teaching, what does any of this have to do with settling the question of whether or not education and teaching should go on in *schools*? That is, even if we were to agree, (as I want to suggest), that education and teaching are, for example, intimately caught up with the development of rationality, it is another, more fundamental, question to ask whether or not we think the development of rationality is of paramount importance within our public schools. The problem, in other words, is to understand how conceptual analysis can provide, or is meant to provide, some sort of justification for particular policies or programs.

Peters' answer to this is very clear: conceptual analysis is not meant to provide a justification for particular policies:

... moral policies cannot be extracted from definitions or conceptual analyses even if they follow fairly closely the lines of ordinary usage.<sup>25</sup>

---

<sup>23</sup>While a descriptive definition seeks only to provide an account of how ordinary language users seem to understand and use a concept, a programmatic definition attempts, in a sense, to persuade language users to understand a concept in a particular way. A stipulative definition is but an invented usage. These distinctions were first drawn by Israel Scheffler in The Language of Education, (Springfield, Illinois: Charles C. Thomas, 1960). See also, Jonas F. Soltis, An Introduction to the Analysis of Educational Concepts, 2nd. ed. (Reading, Massachusetts: Addison-Wesley, 1978): 7-10.

<sup>24</sup>This is not to say, of course, that the work of these philosophers has gone unchallenged. For a critique of R.S. Peters' analysis of the aims of education, for example, see John Woods' and William H. Dray's commentaries in R.S. Peters, ed., The Philosophy of Education (London: Oxford University Press, 1973): 29-39. For objections to part of Thomas Green's analysis of teaching, see David P. Ericson and Frederick S. Ellett, "Teacher Accountability and the Causal Theory of Teaching," Educational Theory 37.3 (Summer 1987): 287-289. None of these objections, however, are fatal to the way I intend to use their work.

<sup>25</sup>Peters, Ethics & Education, 36.

... a detached and clear-sighted view of the shape of issues and institutions is all that conceptual analysis provides. It cannot itself determine the lines of practical policy.<sup>26</sup>

If this is true, then one might well ask, why *do* conceptual analysis (or more precisely, conceptual interpretation)? The best answer, it seems to me, was provided by Austin who said that analysis is not the last word on a subject, but it is the first word.<sup>27</sup> The value of this kind of analysis, in other words, is to get us started. It is, as I suggested, the first step in building a new "language of possibility" for educational reform. With respect to the issue of student assessment, if we can make conceptually clear how many of our current practices operate either apart from, or in direct contradiction to, what is generally meant by education and teaching, and if along the way we can illuminate certain moral hazards as well, then we will have initiated an important debate. The contribution of philosophical analysis, then, is that it forces us to ask first questions first so that our second questions might be more intelligent. The kinds of first questions I intend to raise in this study, then, are as follows:

What do we have in mind by *educating* students, and by *teaching* them?

How might we analyze the concept of assessment and our current assessment practices to discern what within both of these is pedagogically defensible?

What moral considerations, if any, ought to be brought to bear on any discussion of student assessment practices? What grounds, for example, do we have for damaging the self-esteem of students, or for compelling them to have their incompetencies publicly identified and recorded?

What sorts of purposes do we have in mind in sending our children to school? How might our current assessment practices undermine or contradict those purposes?

To begin to answer these questions is to begin to look at the problem of student assessment from a more fundamental, and hopefully more illuminating, point of view.

---

<sup>26</sup>Ibid., 45.

<sup>27</sup>J.L. Austin, "A Plea For Excuses," in Philosophical Papers (Oxford: Clarendon Press, 1961): 123 - 152.

### Education & Teaching - A Conceptual Overview

My aim in this section is to sketch out what I take to be a reasonable account of what is logically implicit within the concepts of "education" and "teaching" as understood in their essential senses. As mentioned, the object is not, however, to construct any new conception of education and teaching, but only to bring together that which has already been done. Very briefly, I want to reiterate how the concepts of education and teaching have built into them procedural and epistemological constraints which limit the range of activities which can or cannot be said to be carried out in their name. I also want to show why the undertaking of education is best captured by the idea of attempting to get students on the inside of a practice, and why the business of getting students on the inside of practices amounts to the development of rationality.

To begin, I take it that analytical philosophers are in broad agreement about a number of very basic features of the concepts of education and teaching. Most analytical practitioners, for example, would agree that the concept of education has within it both an epistemological and what might be called a procedural dimension. The epistemological dimension surfaces because, following R.S. Peters,<sup>28</sup> whatever else "education" might imply, it must imply the transmission of knowledge. Once we concede that the enterprise of education has something to do with transmitting knowledge, it is but a short step, following Israel Scheffler,<sup>29</sup> to emphasize the importance of *evidence* as a condition for saying that someone "knows" something. Once we appreciate the logical centrality of the evidence condition to the concept of knowledge, then we are in a good position to specify the extent to which such a condition should apply *within schools*. Scheffler gets at this issue in his discussion of the contexts in which it is legitimate to say that one "knows" something on the basis of appeals to authority.

We should all, for example, commonly be admitted to know that penicillin is helpful in cases of pneumonia, that there are consistent non-Euclidian geometries, that Washington was the first president of the United States, and that coffee is grown in Brazil even though our evidence for these items consists,

---

<sup>28</sup>Peters, Ethics & Education, 45.

<sup>29</sup>Scheffler, Conditions of Knowledge.

for the most part, in an appeal to authority or the testimony of expert opinion. In the case of the student we have considered, we should, in fact, grant *outside the classroom* that he *does* know the answer to his problem, having based himself on a reasonable appeal to authority. What seems involved here is thus not a blanket exclusion of such appeals but rather a special requirement of the classroom context.

*Inside the classroom*, we want the boy to provide *not* an appeal to authority, no matter how reasonable and how strong; we want him rather to supply evidence from within the subject at hand. We want to determine not simply whether he *knows* the answer, but whether he knows it on, for example, arithmetical grounds. He is expected in the classroom to appeal to the authority of the relevant methods and materials of his subject rather than to the independent authority of persons.<sup>30</sup>

This business of having to supply evidence *from within the subject at hand*, and having to appeal to the authority of the relevant methods and materials - to the authority of the discipline - is important because it suggests that what is sought in attempting to educate students is a state in which they will not only know *about* a discipline, but will also be able to operate *within* it. Scheffler refers to this as knowing in the *strong* sense. He buttresses the point as follows:

In every case where evidence is required for the right to be sure, knowing involves not merely having adequate evidential data but also appreciating their value *as data*, in the light of an appropriately patterned *argument*.<sup>31</sup>

The point about *appreciating* the value of data is significant. Appreciation, in this context, can be understood in two ways: i) as *understanding* how the data fit into the context of the whole argument, and ii) as, in a sense, *approving of*, or *feeling the force of*, in a quite personal way, the power of a fact or proposition to contribute to a larger interpretive framework. The latter kind of appreciation suggests not just knowledge but commitment - i.e., that in addition to being able to describe the procedural rules of a discipline, a student will have also been affected by those rules in the sense of having integrated them into his or her own life. It is one thing, for example, to dutifully recite the "five steps of the scientific method;" it is quite another to actually become convinced of the value of scientific inquiry. What is required of knowing in the strong sense of becoming educated, then, is that knowledge

---

<sup>30</sup>Ibid., 67. (original emphasis)

<sup>31</sup>Ibid., 70. (original emphasis)



not be, to use Peters' phrase, "hived off"<sup>32</sup> - that is, memorized and appropriately regurgitated, but never taken seriously. What is required, rather, is that knowledge be *used*, and that its use be appreciated as a source with which to make sense of the world.<sup>33</sup>

These epistemological imperatives are best captured, I think, by Peters' notion that to be "educated" means, among other things, to be *on the inside of a form of thought*:

It [being educated] must involve the kind of commitment that comes from being on the inside of a form of thought and awareness. A man cannot really understand what it is to think scientifically unless he not only knows that evidence must be found for assumptions, but also knows what counts as evidence and cares that it should be found. In forms of thought where proof is thought possible cogency, simplicity, and elegance must be felt to matter. And what would historical or philosophical thought amount to if there were no concern about relevance, consistency, or coherence? All forms of thought and awareness have their own internal standards of appraisal. To be on the inside of them is both to understand and to care. Without such commitment they lose their point. I do not think that we would call a person 'educated' whose knowledge was purely external in this way.<sup>34</sup>

What is useful about this account is that it brings to the forefront the importance of standards of appraisal, and that students must both understand and care about these. One criticism of Peters, however, is that given the breath of things typically taught in schools, his view can be seen as too narrow. One way to expand the analysis, and to perhaps meet this criticism, is to make use of Alasdair MacIntyre's notion of getting students on the insides of *practices*. As part of a much larger project to reconceptualize moral theory, MacIntyre introduces his view of a *practice* as:

any coherent and complex form of socially established cooperative human activity through which goods internal to that form of activity are realized in the course of trying to achieve those standards of excellence which are appropriate to, and partially definitive of, that form of activity, with the result that human

---

<sup>32</sup>Peters, Ethics & Education, 31.

<sup>33</sup>As Peters makes clear, Alfred North Whitehead has much the same thing in mind when he rails against the transmission of what he calls "inert ideas." See, Alfred North Whitehead, "The Aims of Education," in The Aims of Education and Other Essays (New York: The Macmillan Company, 1929; The Free Press, 1967).

<sup>34</sup>Peters, Ethics & Education, 31.

powers to achieve excellence, and human conceptions of the ends and goods involved, are systematically extended.<sup>35</sup>

What is useful about MacIntyre's notion of a *practice* is that it allows us to admit a broader range of undertakings within the scope of "education," (architecture and farming are practices in MacIntyre's terms), yet still demands that these undertakings be defined in terms of coherent standards of appraisal. As he elaborates:

A practice involves standards of excellence and obedience to rules as well as the achievement of goods. To enter into a practice is to accept the authority of those standards and the inadequacy of my own performance as judged by them. It is to subject my own attitudes, choices, preferences and tastes to the standards which currently and partially define the practice.<sup>36</sup>

What we get with MacIntyre's notion of a practice which goes beyond Peters' notion of a form of thought, then, is the possibility that a broader range of undertakings can now be accommodated beneath the rubric of 'education'.<sup>37</sup>

The key point here is that insofar as we come to understand the business of educating people as the business of attempting to get them on the inside of a practice, we will come to appreciate that our job is not so much one of teaching discrete facts and skills as it is one of showing students how particular facts and skills can be integrated into a comprehensive method of inquiry. The *aim* of education, given such a conception, is to bring our students to a point at which they will be able to, if not initiate, then at least appreciate refinements of the tradition during the course of its evolution.<sup>38</sup>

---

<sup>35</sup>Alisdair MacIntyre, *After Virtue: A Study in Moral Theory* 2nd ed. (London: Gerald Duckworth & Co., 1981; Notre Dame, Ind.: University of Notre Dame Press, 1984): 187.

<sup>36</sup>*Ibid.*, 190.

<sup>37</sup>Perhaps it should go without saying, but we need to be careful, of course, about allowing *too* many activities within the realm of education. The kinds of undertakings we typically think of as educative are those - which Peters rightly points out - in which proof *is* thought possible, and where relevance, consistency and coherence *are* thought important. This means that although astrology, for example, might be thought of as a practice because it has *some* sort of standard of appraisal, it nonetheless need not be labelled an *educational* pursuit precisely because of its apparent disregard for proof, consistency and coherence.

<sup>38</sup>It should probably go without saying, that this therefore represents a condition of *intention*, rather than a condition of *success*, which is to say that we can still claim to be educating students even if, in fact, not all of them *do* come to have such an appreciation.

Another way to capture this aim, and at the same time to specify its scope, is to say that what we are after is *the development of rationality* in students. Although rationality can sometimes be an unwieldy term, it implies, most basically, having both the *competence* and the *inclination* to understand and use modes of inquiry. Scheffler has described rationality as "involving simply the capacity to grasp principles and purposes, and to evaluate them critically in the light of reasons that might be put forward in public discussion."<sup>39</sup> Although he implies this elsewhere, I think it is important to add that not only must students *be able* to critically evaluate principles, they must also be *inclined* to do so. Again, we want to avoid the recitation of inert facts, and even the recitation of reasons which hold no force for the person who offers them. At any rate, Scheffler goes on to suggest how the intention to develop rationality within students might be translated into a program of education:

Rationality is thus, as I view it, the ability to participate in critical and open evaluation of rules and principles in any area of life. To initiate the child into the rational life is to engage him in the critical dialogues that relate to every area of civilization: to science and art, morality and philosophy, history and government. It is to nourish his curiosity and critical judgement as well as his responsibility for choices of belief and conduct. Such a conception goes far beyond the notion of academic mastery of factual subject matter, and far beyond the transmission model [of teaching].<sup>40</sup>

To educate, then, in the sense of intending to enhance the development of rationality, is to attempt to get students on the inside of those practices most generally used to interpret our world. The epistemological imperatives of a fully-articulated conception of education, therefore, are fairly rigorous. Not only must students come to understand and, in a sense, care about the standards which define practices; they must also come to understand and, in a sense, care about the general enterprise of rational inquiry. Not only must students come to understand and appreciate, for example, the point of scientific inquiry; they must also come to understand and appreciate how conforming to the rules and standards of this mode of inquiry allows us to make sense of the world.

---

<sup>39</sup>Israel Scheffler, "Concepts of Education: Reflections on the Current Scene," Reason and Teaching (London: Routledge & Kegan Paul, 1973): 61.

<sup>40</sup>Ibid.

The procedural imperative implicit within the concept of education can be discerned within Peters' third criteria of education:

'education' at least rules out some procedures of transmission, on the grounds that they lack wittingness and voluntariness on the part of the learner.<sup>41</sup>

Part of Peters' argument against these procedures of transmission is a moral one. He writes, at one point, that "'education' suggests the intentional bringing about of a desirable state of mind in a morally unobjectionable manner."<sup>42</sup> The argument seems to be that within the concept of education there resides the principle of respect for persons. What makes conditioning and indoctrinating not count as central cases of educating, therefore, is the fact that neither conditioning nor indoctrinating pays close enough attention to the importance of reasons and evidence - and reasons and evidence *personally* held - in the transmission of knowledge. What makes this rejection of the importance of reasons and evidence *morally* objectionable is the lack of respect for persons it necessarily entails. To presume to educate students by way of anything other than reasons and evidence (by way of, for example, imposed authority), is to fail to respect them as rational and autonomous agents.

Quite apart from any moral argument, it is perhaps more direct to say that certain procedures of transmission are ruled out as cases of educating *because they confound the development of rationality*. If learners lack wittingness and voluntariness within an institutional setting, then the opportunity to develop their rational capacities is surely restricted. On *these* grounds, then, can we point to the hazards of such an institutional situation.

What follows from this characterization of epistemological, procedural, and to some extent moral imperatives is that, in any discussion of *assessment* with respect to an *educational* undertaking, what we will be looking for is some mechanisms or strategies which tell us not so much what facts or skills a student knows or has acquired, but instead what *understanding* and *appreciation* a student has of how those facts and skills combine to form a coherent and

---

<sup>41</sup>Peters, Ethics & Education, 45.

<sup>42</sup>Ibid., 27.

dynamic mode of inquiry. What we will also be looking for is some indication that the student understands and appreciates a mode of inquiry *as a mode of inquiry*, and that the student has come to have these understandings and appreciations more or less on the basis of his or her own volition. Assessments done in reference to *educational* intentions, then, will clearly have to be quite sophisticated.

It follows from all of this, as well, that to *teach* in relation to such a conception of education is, again, to aim at the development of rationality. It should by now be clear that this desire to develop rationality originates from two sources. On one hand there is the epistemological point that we cannot be said to know something (in the context of a classroom) unless we understand and appreciate the reasons which support it. On the other hand, there is the procedural constraint which suggests that we must, at some point or another, shed the protective cover of institutionalized authority and begin to deal with students on the basis of publicly-scrutinized canons of argument. Analytical conceptions of teaching, therefore, are typically framed in reference to these two imperatives. Scheffler's conception of teaching, for example, offers a particularly delicate weave of both of these. He is worth quoting at length:

To teach, in the standard sense, is at some points at least to submit oneself to the understanding and independent judgement of the pupil, to his demand for reasons, to his sense of what constitutes an adequate explanation. To teach someone that such and such is the case is not merely to get him to believe it: deception, for example, is not a method or mode of teaching. Teaching involves further that if we try to get the student to believe that such and such is the case, we try also to get him to believe it for reasons that, within the limit of his capacity to grasp, are *our* reasons. Teaching, in this way, requires us to reveal our reasons to the student and, by so doing, to submit them to his evaluation and criticism.<sup>43</sup>

What distinguishes teaching, as we remarked earlier, is its special connection with rational explanation and critical dialogue: with the enterprise of giving honest reasons and welcoming radical questions. The person engaged in teaching does not merely want to bring about belief, but to bring it about through the exercise of free rational judgement by the student. This is what distinguishes teaching from propaganda or debating, for example. In teaching, the teacher is revealing his reasons for the beliefs he wants to transmit and is thus, in effect, submitting his own judgement to the critical scrutiny and evaluation of the student; he is fully engaged in the dialogue by which he hopes

---

<sup>43</sup>Scheffler, Language of Education, 57. (original emphasis)

to teach, and is thus risking his own beliefs, in lesser or greater degree, as he teaches.

Teaching, it might be said, involves trying to bring about learning under severe restrictions of *manner* - that is to say, within the limitations imposed by the framework of rational discussion.<sup>44</sup>

Thomas Green offers a similar account in his discussion of the concept of *instructing*:

The point [of instructing] is ... that it involves communication *of a certain kind*, and that it is the kind which includes giving reasons, evidence, argument, and so forth, *for the purpose* of helping another to understand or arrive at the truth.

... the important point is that although the purpose of instructing, in one sense, may be to get someone to do something or get someone to believe something, nonetheless the purpose of the *conversation* of instruction is to get him to do it because he thinks he ought to, i.e. because he sees a good reason for doing or believing. In other words, the purpose is to shape someone's belief or behaviour by helping him see that the belief is reasonable or the behaviour justified. This is the sense in which the conversation of instruction has as its purpose the pursuit of truth or the acknowledgement of reasons.<sup>45</sup>

It is crucial to recognize that what Scheffler and Green intend is not only that students recognize the force of evidential proof, but that they feel its force on their own terms. Green has written that "teaching is an activity aimed at transmitting what is reasonable for men to believe, in the object sense of 'reasonable to believe,' *by leading them to assess what is reasonable to believe in the subject sense of 'reasonable to believe'.*" <sup>46</sup> By the object sense of 'what is reasonable to believe', Green means something that is reasonably believed as a result of knowledge, experience, and evidence available to humankind at large. The subject sense of 'what is reasonable to believe', on the other hand, refers to something that is reasonable for a particular person to believe on the basis of the knowledge and experience available to that particular person. Green's point in phrasing his description of teaching in this way is to highlight the importance of getting students to appreciate the grounds for knowledge in their own terms.

Teaching is that activity of education aimed not simply at transmitting reasonable beliefs, but at transmitting them in such a way that they become

---

<sup>44</sup>Scheffler, Conditions of Knowledge, 11-12. (original emphasis)

<sup>45</sup>Thomas Green, The Activities of Teaching (New York: McGraw-Hill, 1971): 29. (original emphasis)

<sup>46</sup>Ibid., 103. (original emphasis)

believable; i.e. that they become reasonable to believe for this or that particular person.<sup>47</sup>

Teaching as understood in relation to a conception of education which emphasizes the development of rationality, then, must necessarily be geared to attempting to have students understand (in their own terms) and care about standards of appraisal. To teach with this purpose in mind is to take seriously the proposition that teaching is a matter of attempting to get students on the inside of practices. To understand this, and what has been said about the concept of education, is to have a basis with which to investigate the appropriateness and adequacy of our current practices of testing, marking, and grading.

In the chapters which immediately follow four common defenses of testing, marking and grading will be reviewed. It is important to note that these defences shall be presented more or less as ideal types; that is, they represent not so much what any particular individual has said, but instead what many individuals have said, and are likely to say, in the course of the debate on student assessment practices.<sup>48</sup> The advantage to referring to ideal types is that in real discourse these can be extracted from the multifaceted (and sometimes contradictory) positions that people typically advance. In serious argument, it is important to move one step at a time. Chapter Two, therefore, will concern itself with the pedagogical defense of testing, marking, and grading, while chapters Three, Four, and Five will examine, respectively, the motivation, accountability and social selection defenses of these practices. Chapter Six will draw the various conceptual and moral arguments together to summarize what I take to be the essential philosophical case against many of our current assessment practices. Some preliminary implications regarding assessment policy will also be suggested.

---

<sup>47</sup>Ibid.

<sup>48</sup>For examples of flesh and blood advocates of testing, marking and grading, however, see James Popham, "The Merits of Measurement-Driven Instruction," Phi Delta Kappan, (May, 1987); Gregory Anrig, "Educational Standards, Testing, and Equity," Phi Delta Kappan, (May, 1985): 623-625; Charles C. Davis, The Uses and Misuses of Tests (San Francisco: Jossey-Bass, 1984).

## CHAPTER TWO

### THE PEDAGOGICAL DEFENCE OF TESTING, MARKING, & GRADING

One argument that might be raised in defence of our current practices of testing, marking and grading is that the business of assessing students is a necessary component to the activity of teaching. The basic intuition is that if we do not get some sort of picture of what students have learned, we will not know how best to proceed. Testing, marking, and grading as mechanisms of assessment, therefore, are of direct *pedagogical* value.

#### The Pedagogical Core of Assessment

There is a kernel of truth in this argument; however, it will take some doing to extract. We should perhaps begin with the obvious: that the activities of *marking and grading* are not, in any logical sense, necessary to teaching. We have all taught people things, and have learned things ourselves, without the benefit of marks and grades. The act of attaching a symbol to a student's work is done *apart from*, not as *a part of*, the activity of teaching that student.

But this may yet seem counter-intuitive. One might object that in giving students marks and grades we convey to them the extent to which they are learning what we want them to learn, and even the extent to which they may be getting on the inside of a practice. This information is of pedagogical value, so the argument would go, in that students would use it to decide what sorts of things they ought to be concentrating on.

The first thing to say in response to such an argument is that it is important to understand the distinction between *grading* a piece of student work and *ranking* it in comparison to his or her peers. To grade is to judge a student's work in relation to a particular standard. It is to make what might be called a criterion-referenced assessment. To rank, on the other hand, is merely to compare a student's work in relation to the work of a larger population. To rank a student high, therefore, is not necessarily to suggest that he or she has come particularly close to the standard: it may be that the student's work merely represents the



best of a bad lot.<sup>1</sup> The difficulty with suggesting that marks and grades yield pedagogical information, then, is that it is often unclear whether they represent judgements about the extent to which a student has achieved a standard, or judgements about the level of performance of students in relation to their peers. Most marks and grades probably represent a combination of the two.

The objection might then proceed by having us consider only those cases in which marks and grades represent *grading* - that is, criterion-referenced assessments. Surely, it would be argued, that, in this case at least, marks and grades can make a pedagogical contribution. But again I would have to say, not necessarily so. The pedagogical contribution implicit here is the delivery, on the part of the teacher to the student, of *substantive feedback* - i.e. information about the extent to which the student has or has not achieved the standard. Notice, however, that we don't need to use *marks and grades* to give students substantive feedback, and that in fact substantive feedback is more effectively conveyed through words, comments, examples, and illustrations. It is one thing to tell a student that her paragraph is logically coherent, has few grammatical faults, and deals with an interesting topic in an original way; it is quite another to give her an "A" or an "85%". To mark or grade a student's work, therefore, is to do something *apart from*, or *in addition to*, providing that student with pedagogical feedback.

But what about a case in which a teacher prespecifies exactly what each mark or grade represents relative to the standards which define the practice? Let us suppose, for example, that in the course of instructing students about what counts as a "good paragraph" a teacher creates an evaluation grid in which a score of 0-4 will be assigned to paragraphs having many grammatical faults, logical inconsistencies, and a lack of style or inventiveness, while a score of 7-8 will indicate few grammatical faults, an acceptable level of logical cohesiveness, and a certain amount of style and/or inventiveness, with other score ranges filled in accordingly. Given such an explanatory grid, so the argument would go, we would then have a case in which marks or scores *would* carry pedagogical information.

---

<sup>1</sup>Paul Taylor, Normative Discourse (Englewood Cliffs, N.J.: Prentice-Hall, 1961), 5-9.

Although we have to concede that the above is true, there are nonetheless two important features about this system of feedback that we need to recognize. In the first place, even where such a grid is provided, problems of translation can arise. To give a mark of seven out of ten on a paragraph is not to tell a student *which* parts of his or her writing are satisfactory, and which parts need work. Now it might be replied that the fault here lies not with the general undertaking, but instead with the design of the grid itself, i.e. that it was not specific enough. But this response suggests a second difficulty: the problem of over-rigorous prespecification narrowing the range of potential feedback. It may be, for example, that a student introduces into his or her writing a problem (or asset) not specified on the grid. It would be unfortunate, and perhaps even pedagogically indefensible, not to address that point merely because it was not thought important at the time the grid was constructed.<sup>2</sup> What we end up with, then, is that although there is (in this particular case) the potential for marks and grades to make a direct pedagogical contribution, there is also the danger that they may end up doing more harm than good. The more general conclusion is that most instances of marking and grading represent an undertaking quite distinct from the activity of teaching.

This leaves, then, the question of the extent to which *testing* might be thought to be central or necessary to teaching. The clearest philosophical argument written in favour of a strong logical connection between the two has been advanced by Anthony Flew in a paper appropriately entitled, "Teaching and Testing."<sup>3</sup> In it Flew suggests that it is a measure of our "sincerity of purpose," as both teachers and learners, that we "be concerned whether, how far, and how well, [we] are succeeding, or have succeeded."<sup>4</sup> What is logically central to teaching, and indeed to learning, therefore, is some sort of periodic investigation of how well we seem

---

<sup>2</sup>For a brief treatment of the pedagogical costs of too much prespecification, see Robert Stake, "Measuring What Learners Learn," in Ernest House, ed., School Evaluation: The Politics and Process (Berkely: McCutchan Publishing, 1973), 201-202.

<sup>3</sup>Anthony Flew, "Teaching and Testing," Proceedings of the Philosophy of Education Society (1973): 201-212.

<sup>4</sup>Ibid., 207.

to be getting along. Flew suggests that the most general word for all such investigations is 'assessment.'<sup>5</sup>

It seems to me that Flew is importantly correct with regard to the *general* undertaking of assessment. Insofar as teaching and learning are understood as intentional activities, it would seem to be a mark of our *insincerity* of purpose to say that we are 'teaching' or that we are 'learning' but that in both cases we never check to see how well, or to what degree, we are succeeding. Teaching and learning do seem to require 'assessment' in the general sense of making periodic checks on our progress.

But notice that none of this commits us to *testing* in the sense of using formal or semi-formal instruments which are to be marked and/or graded.<sup>6</sup> If 'testing' means only checking to see the extent of our success, then this can be managed by using any one of a host of informal strategies, e.g., by simply conversing with a student and asking him or her a few pointed questions, by reading student essays, or by asking a student to demonstrate a particular operation. Testing in the sense of having students write formal tests, then, is unnecessary. There are even reasons to believe, as we shall see, that it is counterproductive to educational purposes.

Another activity of our current assessment activities that is in need of analysis is the practice of reporting grades. If we take reporting to be the business of informing students about how well (in relation to the standards which define the practice) they are doing, then this would seem to be a necessary component of teaching.<sup>7</sup> If, however, reporting is taken to mean

---

<sup>5</sup>Ibid.

<sup>6</sup>This point is made by Jonas Soltis in response to Flew's argument. See Jonas Soltis, "The concept of Assessment: A Response to Mr. Flew," Proceedings of the Philosophy of Education Society (1973): 214. Flew, himself, makes more or less the same point on page 207.

<sup>7</sup>Jonas Soltis, however, makes the important point that it is not a logical requirement of teaching and learning that a teacher *always* inform his or her students how they are doing. (Soltis, "Response to Mr. Flew," 214-215.) There may be times in which teachers want to keep their assessments to themselves. It does seem to be the case, though, that teachers at *some* point need to inform their students of how they are doing (relative to the standards which define the practice). It is difficult to imagine a case of teaching in which such feedback *never* occurs.

the business of compiling and submitting (to parents and administrators) *a public record of achievement*, then this is clearly beside the point of teaching. Creating a public record of achievement has nothing directly to do with getting students on the inside of a practice.

The upshot here is that assessment is central to teaching only in that, insofar as teaching is an intentional activity, we need to check, from time to time, to see how well our students are doing. None of this, however, implies the use of formalized testing, marking and grading, or the mandatory creation of a public record of achievement. Insofar as we are interested in using assessment for strictly *pedagogical* purposes, the most we ever need to do is employ *some* strategies to find out what students know, and at *some* appropriate point inform them of our findings.

#### The Pedagogical/Classificatory Distinction

The question which naturally arises, then, is if formalized testing, marking and grading, and public reporting are not central to *teaching* per se, then why are these procedures so frequently employed within our schools? What purpose do they serve? To answer this it is necessary to take a closer look at the concept of assessment and consider how it can be differentially construed in the light of divergent aims.

I take it that the core idea within 'assessment' is the idea of *appraisal*. What we usually want to do when we assess something or someone is appraise the thing or the person (or more probably the work or performance of the person) in relation to some sort of standard. Our use of the word assessment doesn't become problematical (or interesting), though, until we begin to intimate what an assessment is *for*, i.e. what our purpose is in making the appraisal. Two possibilities come to mind. On the one hand, 'assessment' might be understood as the enterprise of finding out what a student knows for the purpose of designing subsequent strategies to bring that student closer to the standards which define the practice being taught. This I shall call *pedagogical assessment*. On the other hand, 'assessment' could be understood as the enterprise of finding out what a student knows for the purpose of locating that student within a hierarchy; that is, for ranking the academic achievement of that student in relation to

his or her peers. I shall call this latter undertaking *classificatory assessment*. The key distinction between these two is that in pedagogical assessment we make an appraisal of a student's work as part of an overall strategy to get him or her on the inside of a practice, while in classificatory assessment we make the appraisal in order to figure out where the student stands in relation to his or her peers. The point of ranking student performance, moreover, is to be able to make judgements about which students, within a given population, are most worthy of subsequent rewards and opportunities.

Given this distinction, it becomes possible to compare pedagogical assessment to classificatory assessment in reference to a number of key features. Both of them, to begin with, need to make their observations or judgements on the basis of appraisals which have content validity. Both, that is, need to make their appraisals in direct reference to the standards which define the practice. As we shall see, however, there are good reasons to believe that what comes to count as the standards which define a practice within a classificatory appraisal are usually different from - and in fact much reduced from - what comes to count as the standards which define a practice within the context of pedagogical assessment. Second, it seems clear that the kind of information required of pedagogical assessment is subject-specific commentary about how a student is doing in relation to the standards which define the practice. In the case of pedagogical assessment, in other words, we want to know *what* particular faults or virtues a student displays in his or her work. In the case of classificatory assessment, on the other hand, the imperative is not so much to provide precise information about *what* a student knows as it is to provide precise information about *how much* a student knows relative to a larger population. Strictly speaking, the information requirement of classificatory assessment can be met merely by presenting a number which shows where a student ranks in relation to his or her peers. Third, it also seems clear that in the case of pedagogical assessment reports about student progress need only to be made, (at some point), to the student him or herself, while within the context of classificatory assessment these reports will, at some point, need to be made public. If the point of pedagogical assessment is to find out what a student knows for the purpose of enhancing learning, then the only reporting that

need to occur is that between teacher and student. If the point of classificatory assessment, on the other hand, is to rank students for the purpose of distributing subsequent rewards and opportunities, then this rank order will eventually have to become public.

### The Hazards of Classificatory Assessment

With this distinction and these differentiating features in place, we are now in a position to investigate the extent to which classificatory assessment is not only *different* from pedagogical assessment, but also potentially *counter-productive* to the enterprise of getting students on the inside of a practice. The basic thesis will be that the imperative to rank puts certain constraints on what can and can not count as *operationalized* representations of the standards which define the practice, and that this in turn produces both methodological and conceptual deficiencies which, in effect, threaten the integrity of the educative undertaking. This is not to say that classificatory assessment is *necessarily* counter-productive to educational purposes; it is rather to show that as a matter of contingent fact - and for very good reason - it frequently turns out that way.

The imperative to rank might be more generously described as the imperative to get a *fair measure* of student achievement. (The ultimate point of getting this measure, though, is to rank one student against another.) By 'measure', I mean results which are quantifiable. We need such results because, at the end of the day, we need to be able place student achievement within a finely-graded hierarchy. By 'fair', I mean consistent - that is, by way of a process that takes great pains to ensure that no one group or individual will be unfairly disadvantaged in comparison to another. Because the rewards attendant on a classificatory assessment are quite often of significant social value, some effort has to be made to ensure that assessment results are at least consistently derived.

This imperative to get a fair measure of student achievement can, however, bring about certain undesirable methodological consequences. In particular, the implicit or explicit pursuit

of this imperative can lead to an assessment system characterized by curricular narrowing, trivialization, and fragmentation.<sup>8</sup>

Curricular narrowing refers to the practice of excluding from an assessment mechanism certain essential components of a practice on the grounds that these components are not amenable to unambiguous measurement. In a review of the British examination system, for example, John Matthews<sup>9</sup> has observed that even though the Nuffield Foundation had written that "science should be presented to pupils in a way in which *they* can conduct an enquiry into the nature of things," and that "students must approach their studies through experiments designed to awaken the spirit of investigation,"<sup>10</sup> the writers of the Nuffield O level chemistry examinations had decided that it would be impossible to include direct assessment of practical work in the test. One of the reasons, presumably, was that such an assessment would be too difficult to manage, *given the constraints of the examination system*. Interestingly enough, when faced with the choice between changing the scope, as opposed to the mechanism, of assessment, the decision is frequently made to do the former rather than the latter.

We should not necessarily be consoled by the possibility of practical work being assessed via an alternative format. Writing in reference to the field of *program* evaluation, Michael Scriven's discussion of the construction of an evaluative test-question pool illustrates nicely how those features of a practice which are not amenable to measurement quickly become all but ignored.<sup>11</sup> He begins by showing some sensitivity to the problem of test questions becoming the source of curricular goals. He then contends that even though there are certain worthwhile components of a course which cannot be captured in an examination

---

<sup>8</sup>For an analysis of these with regard to the issue of measurement-driven instruction, see Gerald Bracey, "Measurement-Driven Instruction: Catchy Phrase, Dangerous Practice," Phi Delta Kappan (May 1987): 683-686.

<sup>9</sup>John Matthews, Examinations (London: George Allen & Unwin, 1985).

<sup>10</sup>Nuffield Foundation, (1975), 2-3: quoted in John C. Matthews, Examinations (London: George Allen & Unwin, 1985), 70.

<sup>11</sup>Michael Scriven, "The Methodology of Evaluation," in Michael Scriven, Ralph Taylor and Robert Gagne, eds., Perspectives of Curriculum Evaluation (Chicago: Rand McNally, 1967), 39-83.

question, the acquisition of these should in principle be recognizable by way of other measures.

It is a commonplace that in the light of formulating such questions, the conception of the goals of the course will be altered. It is undesirable to devote a large proportion of the time to this activity, but it is typically not "undue influence" to encourage thinking about course goals in terms of "What kind of question would tap this learning achievement or motivation change in the final examination or in a follow-up test?" At times the answer to this will rightly be "None at all!" because not all values in a course manifest themselves in the final or later examinations. But where they do *not* thereby manifest themselves, some indication should be given of the time and manner in which they might be expected to be detectable; as in career choices, adult attitudes, etc..<sup>12</sup>

*Soon after, however, Scriven adds the following:*

If the above procedure is followed throughout the development of a curriculum, we will end up with an oversize question pool of which one should be prepared to say that *any significant desired outcome of the course will show up on answers to these questions* and that what does show up will (normally) only come from the course.<sup>13</sup>

What can happen, in other words, is that the course itself can come to be defined in reference to the constraints of the assessment system. What might have begun, that is, as a comprehensive intellectual undertaking can soon get narrowed down to a much leaner facsimile.

But this is not the worst of it. Given, again, the imperative to get a fair measure of student achievement, what we need are questions which will yield results which are completely unambiguous. The easiest way to do this, of course, is to ask questions which admit of only one answer. This is one of the reasons that multiple choice tests and single-response short-answer quizzes are so popular. From the point of view of wanting to get students on the inside of a practice, however, these kinds of instruments have a number of serious shortcomings. In the first place, to translate the essence of a practice into a series of single-response questions is almost always to trivialize that practice. It is usually, that is, the least significant features of a discipline which can be captured in these sorts of questions. Witness the countless quizzes that

---

<sup>12</sup> Ibid., 57.

<sup>13</sup> Ibid., 58. (emphasis added)



teachers have given students in which the point is to recite dates and names quite apart from any understanding of why those dates and names might be important. This leads to the second objection: that it is not so much the *answers* students give that are important as the *reasoning* the students use to arrive at their answers. The problem with making assessments on the basis of single-response questions, therefore, is that in doing so we get no insight into *why* students believe what they appear to believe. Banesh Hoffman has made this criticism of multiple choice tests in a particularly instructive way:

Any competent person who has ever graded a non-objective mathematics or science examination knows that a correct answer obtained by incorrect methods is worth very little, while a wrong answer obtained by correct methods can deserve a top score; and even that a wrong answer obtained by wrong methods can be indicative of outstanding ability, and merit a bonus score.<sup>14</sup>

More to the point, given the conception of teaching sketched in chapter one, it would seem to be a necessary condition of assessments carried out for *pedagogical* purposes that they at least attempt to ascertain the reasons students have for believing the things they do. If we take seriously Green's notion that the conversation of instruction requires that we attempt to get students to believe and do things because there are *good reasons* for doing so,<sup>15</sup> then it follows that any *assessment* of the extent to which we may have been successful in this goal will necessarily have to focus on the kinds of justifications that students produce to support what they believe and do. The fact that the requirements of a classificatory assessment system are such that single-response instruments are nonetheless perfectly acceptable should give us pause to reflect.

A third and related methodological consequence that can arise from the imperative to get a fair measure of student achievement is the forced fragmentation of an otherwise integrated practice. In an attempt to get a quantifiable measure of a relatively *complex* undertaking, the standard strategy is to break the undertaking down into a series of discrete parts and then assess these individually. The assumption seems to be that in measuring

---

<sup>14</sup>Banesh Hoffman, The Tyranny of Testing (New York: Crowell-Collier, 1962), 66.

<sup>15</sup>see p. 17 above.

students' understanding of the parts, we can get a reasonable picture of their understanding of the whole. This, of course, is seriously wrong-headed, for *one of the things which makes* a complex undertaking complex is the way all its component parts intricately, even mysteriously, fit together to form a comprehensive whole. And what counts as understanding a complex activity is not one's comprehension of a requisite number of bits of information, but instead one's appreciation of an undertaking as an integrated whole.

Take, for example, the activity of writing a good essay. One strategy for teaching such an activity is to break the practice down into discrete parts and then show students how to become proficient at each. We might, for example, show students how to write thesis statements, topic sentences, transitions and conclusions. The truth that such a strategy has serious shortcomings emerges in those cases in which students are successful at writing thesis statements, topic sentences, transitions and conclusions, but are still not able to write what we would be prepared to call a "good" essay. The problem, in other words, is that something *important* can get lost in the translation when we attempt to break down an integrative and organic undertaking into constituent parts. Given, again, the conception, in chapter one, of educating as the business of attempting to get students on the inside of *practices*, this loss is by no means insignificant.

In conjunction with the methodological hazards of curricular narrowing, trivialization, and fragmentation, the imperative to rank can also perpetuate and promote a number of serious conceptual distortions. I have already alluded, in the first place, to the problem of thinking we can say that a student has *learned* something on the basis of that student's success at answering a series of test questions. What gets lost when we begin to equate learning with the business of getting correct answers is a conception of learning in which *reasons* for answers matter more than the answers themselves. What gets lost, in other words, is learning in the sense of developing one's rational capacities. If we take the development of rationality to be central to teaching and educating, then, again, this loss is not insignificant.

Second, I have also alluded to the problem of saying that a student *knows* something on the basis of his or her success at mastering a series of discrete operations. The fatal mistake, it seems to me, is to think that a *compilation* of a student's successes and failures can somehow add up to a reasonable judgement about the extent to which a student is able to operate on the inside of a practice. The fact that a student has scored well on a number of single-response history tests, for example, does not, in itself, indicate that he or she is a good historian. To think that it does is to do a disservice both to history, and to the idea of 'knowing' in the sense of getting on the inside of a practice.

There are, then, both methodological and conceptual hazards associated with classificatory assessment. Although these hazards are not logically implicit within classroom assessments, it is not difficult to see how easily they emerge. Because the imperative within classificatory assessment is, above all else, to get a fair measure of student achievement, the pressure to narrow, trivialize, and break apart the content of the curriculum, and to distort what it means to 'learn' and 'know' in the context of schooling, is considerable. It is precisely because of these strategies and conceptual distortions, moreover, that classificatory assessment is so potentially damaging to the educative project.

#### Current Practices: Classificatory or Pedagogical?

I have been speaking, up until now, of classificatory assessment as an ideal type. The strategy has been to contrast it against pedagogical assessment as a way to make the point that these two undertakings are significantly different. I have also tried to show how and why an assessment system with a classificatory imperative can actually interfere with the process of getting students on the inside of a practice. The next question to ask, therefore, is to what extent do our current practices of testing, marking and grading represent either pedagogical or classificatory strategies. It will be appreciated by now that the way to decide which label to attach is to consider carefully to what *use* the particular instrument or procedure is being put. If the instrument or procedure represents an attempt to find out what students know *for the purpose of enhancing learning*, then it is an example of pedagogical assessment. If the

instrument or procedure represents an attempt to find out what students know *for the purpose of locating their achievement within a hierarchy*, then it is a case of classificatory assessment. Once we establish that a given instrument or procedure is predominantly classificatory, moreover, we can then be put on guard about the possibility of it operating at cross-purposes to education.

A point of clarification is needed before proceeding. In speaking of 'purposes' I do not mean to imply that individuals actually begin with either pedagogical or classificatory intentions, and then shape their assessment practices accordingly. I mean, instead, only that certain practices seem to serve some purposes better than others, and indeed that some practices may serve purposes quite apart from the intentions of those who use them. This point might be better appreciated by noticing the difference between saying that "the purpose of assessment determines how assessments will be carried out," and "the way in which assessments are carried out serves some purposes better than others." My investigation is better characterized by the latter phrase. In judging the extent to which our current practices are either pedagogical or classificatory, that is, I am not attempting to ascribe any particular intentions to individuals. The object, rather, is only to point out how some of our practices *do*, as a matter of fact, facilitate some ends better than others.

With this in mind, it seems clear that much of what we do in schools in the way of testing, marking and grading is fundamentally classificatory. With respect to marking and grading, I have already alluded to the near-impossibility of them having *any* pedagogical value. On a theoretical level, to attach a symbol to a student's work is to do something in addition to, or quite apart from, the pedagogical act of giving that student substantive feedback. On a practical level, our use of marks and grades within actual schools would seem to bear out the classificatory epithet. In the first place few teachers go to the trouble of spelling out exactly how each assigned score or grade can be translated into a subject-specific response. Rather, the common practice is to hand out marks and grades and trust that students will get

the message. The most important message, of course, is not *what* students did right or wrong, but instead *where* they fit in on a comparative scale.

Second, there is the unabashed manner in which marks and grades are compiled so as to arrive at a composite account of a student's apparent level of understanding. It is common practice, that is, for teachers to assign values to all student work, to record these in a mark book, and then to weight and counterweight these according to their perceived importance, in order to arrive, eventually, at a final grade. In a typical English 12 course, for example, the standard assessment strategy is to divide the curriculum into four parts (short stories, poetry, drama, and the novel), assign one or two quizzes and an essay for each section, add a mid-term test and a final exam, then combine all of the scores achieved on each of these items together to arrive at a final judgement about the extent to which students understand and appreciate English literature. The truth is that no such judgement is possible given the data collected. What we get, at best, is a judgement about the extent to which students have successfully completed the assignments of a grade 12 English course.

The general point then is that very little of the time and energy that teachers spend on compiling and calculating grades has anything directly to do with getting students on the inside of a practice. Our current use of marks and grades in schools would seem to suggest, rather, that the classificatory imperative is alive and well.

The same sort of conclusion seems justified with respect to reporting. Most reporting in schools is *public* reporting in the sense that most evaluative exchanges between a student and a teacher eventually get scored, recorded and tabulated into a final grade. It is rare, that is, for a teacher to create an evaluative episode - especially a formal evaluative episode - and then *not* integrate the results into his or her final assessment. This reflects common sense more than anything else on the part of teachers, for if the expectation is that teachers are to generate a final report, then the business of testing without recording and reporting becomes, quite literally, a waste of time. The current emphasis in schools, then, seems to be on public reporting at the expense of private feedback.

As to the question of whether or not the particular *tests* we use in schools are pedagogical or classificatory, this depends, again, on *how* they are used. Although the design of a test can certainly provide some indication of its classificatory or pedagogical potential, design in itself does not tell us conclusively which of these two labels is most appropriate. It might be thought that all criterion-referenced tests, for example, must necessarily serve pedagogical purposes given that they are explicitly designed to provide feedback on a student's apparent understanding of the substantive standards which define a practice. We use a criterion-referenced test in a classificatory fashion, however, if the main point in giving the test is to generate scores which can then be ranked. Suppose, for example, that someone designs a test which truly captures (as much as can be captured) a student's understanding and appreciation of the standards which define a practice. Suppose further that most students do poorly on this test - although poorly in varying degrees. If the main use of the results of the test is to record and report rank order of achievement, then the enterprise has been a classificatory undertaking after all.

It might also be thought that norm-referenced tests can only ever be used in a classificatory fashion. There are good reasons for thinking this. Norm-referenced tests, after all, are explicitly *designed* to yield a normal distribution. The whole point, in other words, is to differentiate people along a predetermined continuum. As testing critic Andrew Strenio puts it in his description of test construction,

we consciously and deliberately select questions so that the kind of people who scored low on the pretest will score low on the subsequent tests. We do the same for the middle and higher scorers. We are imposing our will on the outcome, so that the test scores will fall in the pattern we picked in the beginning.<sup>16</sup>

What is remarkable about the business of building a norm-referenced test is that this pattern is to be achieved whether or not the targeted ability or understanding is, in fact, normally distributed across a population. Educational psychologist Charles Galloway:

---

<sup>16</sup>Andrew Strenio, The Testing Trap (New York: Rawson, Wade Publishers, 1981), 95.

The shape of the distribution we get with a test depends on the way the test is constructed, not on the way the quality being measured is distributed among the individuals of the population. This means, for example, that the scores on general mental ability tests of individuals in norm groups approximate a normal distribution curve because the test items were designed to get this type of curve, not because general mental ability is necessarily distributed in nature according to the proportions reflected by the normal distribution curve.<sup>17</sup>

What matters most in the construction of such a test, then, is not what people know, but rather what kinds of questions need to be asked in order to produce the right statistical results. There *are* good reasons, therefore, for thinking of norm-referenced tests as primarily classificatory mechanisms.

Be that as it may, there is at least one case in which we might be able to say that a norm-referenced test is being used in a pedagogical fashion. Assuming that we use a test which has a reasonable level of content validity, and assuming that the results of the test never become public in the sense of being integrated into a student's final grade, we might be justified in saying that our use of a norm-referenced test is of pedagogical value insofar as it provides an *early warning* that something might be amiss in the case of those students who fall well below the norm. In this restricted sense, then, even a norm-referenced test take can on pedagogical significance.<sup>18</sup> What is most important in characterizing an assessment mechanism, therefore, is not so much how the instrument is designed, but rather how it is used.

This latter point might be better appreciated by introducing a new distinction: the distinction between an assessment *episode* and an assessment *regime*. An assessment episode, as the word implies, refers to a single evaluative encounter. The taking of a particular test, or even a short exchange between a teacher and a student, can be an assessment episode. An

---

<sup>17</sup>Charles Galloway, Psychology of Learning and Teaching (New York: McGraw-Hill Book Company, 1976), 467.

<sup>18</sup>Two comments about this case are nonetheless instructive. It is important to see how *little* of pedagogical value the results of a norm-referenced test tell us. They don't tell us, in particular, *where* a student may be deficient, or *what* he or she doesn't understand. They only tell us where the student's achievement is placed in relation to the achievement of a larger comparison group. Second, insofar as the best pedagogical justification for a norm-referenced test is that it acts as an early warning indicator, one might easily reply that any competent teacher would be able to detect what a norm-referenced test can detect (e.g. that little Johnny seems to be having trouble in math), and with far less expenditure of time and money.

assessment regime, on the other hand, refers to the overall character of an assessment system - i.e. what all the assessment episodes seem to add up to. A *classificatory* assessment regime, then, is one in which most of the assessment episodes require the recall of fragmented (and perhaps trivial) information, and in which the cumulative point of the enterprise is to generate rank order. A *pedagogical* assessment regime, on the other hand, is characterized by assessment episodes which, in the main, attempt to ascertain the reasons students have for making the claims that they make, and by an overall commitment to the development of rationality. It is important to add, by the way, that this does not mean that pedagogical assessment regimes need never ask questions which demand little more than factual recall; it only means that these sorts of questions cannot *dominate* the assessment system, and that when they are asked they must be asked with an explicit understanding of how the answers contribute to the larger object of getting students on the inside of a practice. To point out how a criterion-referenced test can serve classificatory purposes - or even how a norm-referenced test can serve pedagogical purposes - is to pay attention, then, not only to particular assessment episodes, but also to the way in which such episodes can function within the context of an overall assessment regime.

With the episode/regime distinction in mind, it now becomes possible to rephrase the original question: to what extent do our current testing practices - indeed our practices of testing, marking and grading as a whole - meet the conditions of a classificatory regime, as opposed to those of a regime which is predominantly pedagogical? Although more empirical work should, of course, be done on this, there does seem to be enough evidence - both researched and anecdotal - to suggest that the classificatory imperative is ubiquitous. In a study on the impact of large-scale testing on the instructional activity of science teachers, for example, Marvin Wideen and his associates<sup>19</sup> found that the problems of curricular narrowing, fragmentation and trivialization are real indeed. With respect to curricular narrowing, they point to the situation of typical grade 12 science teachers who, because of the

---

<sup>19</sup>Marvin F. Wideen et al., "Impact of Large Scale Testing on the Instructional Activity of Science Teachers," Paper presented at the Canadian Society for Studies in Education, June, 1991.



implementation of large-scale testing, "now teach fewer labs and lecture more," and who "now must concentrate on objectives and reduce the number of side issues students are able to explore."<sup>20</sup> Part of what it means to do science, in other words, has been parred down in an effort to get students ready for the exam. As their study focuses on grades eight, ten, and twelve (with the higher grades being more directly influenced by large-scale testing), they are also able to document the shift that occurs in not only the content, but also the process of instruction.

We saw a narrowing of the instructional pattern as we moved from grade eight to twelve. As observers, it appeared to us that the most vibrant classes took place at the grade 8 and 10 levels. Grade 12 classes, on the other hand, were marked with a palpable desire to come to grips with the material presented. A sense of fun and enjoyment seemed lacking here. This was work and made to seem so. We sensed a strong need to process a great deal of material very quickly. The singlemindedness of the enterprise was underscored by the impatience demonstrated by these students when the teacher withheld answers or ventured into territory that would not appear on examinations. They had no time for anything extraneous. In contrast, classes in grade eight and ten showed a more leisurely pace with more time to explore and attempt different approaches. Though the amount of innovation and use of alternative models of teaching did not appear startling, when we did observe such innovation we typically observed it at the grade eight and ten levels.<sup>21</sup>

With regard to the issues of fragmentation and trivialization, Wideen and his colleagues found that many teachers are seriously concerned about the potential for large-scale examinations to do real damage to the teaching of science.

Many of them felt that science classes had become "content oriented" and that they had been reduced to little more than the presentation of a string of facts to be memorized for the final exam. Even teachers who were ambivalent about the exam agreed that since its reintroduction teachers had been forced to "concentrate on objectives, to concentrate on facts." One grade 12 teacher stated that the government exam had eroded the ideals of good science teaching which included developing a sense of curiosity and a sense of social responsibility. Another felt that in limiting the teacher's ability to cover areas they are keen on, the students miss the enthusiasm and energy that the teacher brings to these topics and which often "rub off" on the students.<sup>22</sup>

---

<sup>20</sup>Ibid., 49.

<sup>21</sup>Ibid., 56-57.

<sup>22</sup>Ibid., 60.

The research group concurs with this perspective, pointing out in their concluding discussion that success on the examinations "depends primarily upon committing to memory the content of the curriculum and algorithms required to solve set problems likely to appear on the exam. This technique underscores the mechanics of test writing rather than promoting any clear understanding of science concepts."<sup>23</sup>

What all of this seems to indicate, in the case of large-scale testing for science courses at least, is that the imperative to get a *measure* of student achievement seems to have overtaken the imperative to get a good grasp of the extent to which students understand science as a living practice. The classificatory imperative, in other words, seems to have usurped the pedagogical one.

The same inference can be drawn from some of the findings of a recent study on the impact of provincial exams, in general, on education in British Columbia.<sup>24</sup> In a section which surveys student perceptions of the strengths and weaknesses of the provincial examination system, the issue of narrowing emerges once more.

Students also expressed some concern for teachers in regard to the reduced flexibility in teaching they now had due to the narrowing of curriculum effect of the exams. These concerns were related to further student comments on their increased workload. The increased workload was viewed as increased amounts of information to memorize. This was expressed directly and also through comments regarding the narrowing of the focus of classroom teaching to exam-specific information which many students viewed as uninteresting, but necessary to ingest in order to pass the exams.<sup>25</sup>

Again it would seem that the imperative to have students "pass the exam" (in order to be appropriately ranked) overrides concerns about maintaining a comprehensive curriculum.

---

<sup>23</sup>Ibid., 65.

<sup>24</sup>John O. Anderson et al., The Impact of Provincial Examinations on Education in British Columbia: General Report (Victoria, B.C.: B.C. Ministry of Education, March 30, 1990).

<sup>25</sup>Ibid., 64.

Again, in other words, the classificatory imperative seems more important than the pedagogical one.<sup>26</sup>

If this is true of large-scale assessment episodes, what about the kind of testing that goes on in classrooms? Anecdotal evidence suggests that in many cases teacher tests are but reduced versions of a large-scale examination format. With questions which focus on the fragmented and the trivial, and with an explicit emphasis on making clear how much credit will be received for how much work (e.g., "section 1 is worth 5 marks, section 2 is worth 15 marks," etc.), the main purpose, again, seems to be that of generating a score which can subsequently be used to contribute to a final ranking.

But the situation may be slightly more complex than this. Robert Wilson and Ruth Rees<sup>27</sup> have done research which suggests that teachers shape their evaluation strategies in reference to what they perceive to be institutional expectations. They believe that teachers make a distinction between evaluations done for "the instructional purpose of feedback and motivation,"<sup>28</sup> and those done to generate a normal distribution curve, and that they mix and differentially weight these two according to the implicit requirements of the system.

The utility of this approach reflects an exquisitely precise understanding by teachers: if they were to meet the somewhat contradictory goals of providing feedback to students on learned materials as well as generating marks for reporting purposes on a common school-wide basis (for example approximately as many F's as A's; medians between 65 and 72), some orchestration of the results would be required. The more frequent, less heavily weighted evaluations, then, served the instructional purposes of feedback and motivation in the classroom while the less frequent, more heavily weighted tests and examinations, with questions whose p-values ranged more widely, brought the final distributions into line with administrative expectations.<sup>29</sup>

---

<sup>26</sup>It is important to note that the Anderson et al. study also offers perspectives and arguments in *favour* of provincial examinations, and in fact concludes that provincial examinations ought to be retained.

<sup>27</sup>Robert Wilson and Ruth Rees, "The Ecology of Assessment: Evaluation in Educational Settings," Canadian Journal of Education, 15, no. 3 (1990): 215-228.

<sup>28</sup>*Ibid.*, 218.

<sup>29</sup> *Ibid.*

What matters, then, is not just the kinds of assessment mechanisms which get used, but also how they are *weighted*. Wilson and Rees' work suggests that for all the pedagogical-inspired assessments that might occur in a class, when all is said and done it is the classificatory imperative to locate students within a normal distribution curve that directs the manner in which those assessments will be made public. It is in this sense, therefore, that it is not unreasonable to suggest that much of what we do in schools in the way of testing - and not only testing, but also marking, grading and reporting - is conducive to the maintenance of a classificatory assessment regime. And, for the reasons presented above, it is precisely because of this that much of what we do in schools in the way of testing, marking, and grading threatens the integrity of the educative project.

### Conclusion

The original proposition was that because testing, marking and grading are mechanisms of assessment, and because assessment is central to teaching, the activities of testing, marking and grading are, therefore, of direct pedagogical value. Against this, I have argued that neither marking, nor grading, nor public reporting is central to teaching, and that *testing* is central to teaching only in the sense that we need, occasionally, to check to see whether or not our students understand what we are trying to teach them. This checking, moreover need not occur via the vehicle of a formalized test. I have argued, in addition, that much of what we do in the way of testing, marking, and grading in schools is implicitly classificatory and, as such, is potentially hazardous to teaching in the sense of getting students on the inside of a practice, or to educating in the sense of being concerned with the development of rationality. Contrary to the original proposition, then, testing, marking and grading are not central to the activity of teaching. There are even reasons to believe that much of what we do in the way of testing, marking and grading actually detracts from, rather than enhances, the basic pedagogical undertaking.

### CHAPTER THREE

#### THE MOTIVATION DEFENCE OF TESTING, MARKING & GRADING

A second major defence of our current practices of testing, marking and grading is that these activities motivate students to learn. The common assumption seems to be that if we did not have tests (and behind them, marks and grades) to goad students into action they would not be sufficiently motivated to apply themselves to their studies. The further assumption is that if students apply themselves to their studies, they will eventually learn. Testing, marking and grading motivate students to learn, so the argument would go, in the sense that they encourage students to take their academic work seriously.

##### Initial Difficulties

A number of preliminary observations are in order. In the first place, even if we accept at face value that our system of testing, marking and grading motivates students, one has to concede that it motivates different students differently. One has to concede, in particular, that for the 30%-35% of the students who (presumably) "feel good" about getting high marks, there are many more students who (presumably) don't feel so good about the responses they receive. Bloom, Madaus, and Hastings point to both the educational and personal damage done to the latter sort of student:

It is not likely that this continual labelling has beneficial consequences for the individual's educational development, and it is likely that it has an unfavourable influence on many a student's self-concept. To be physically (and legally) imprisoned in a school system for 10 or 12 years and to receive negative classifications repeatedly must have a major detrimental effect on personality and character development.<sup>1</sup>

The point, then, is that even if we accept the general proposition that our current system of testing, marking and grading can motivate students to apply themselves, we have to concede that this very same system provides positive feedback to less than 50% of our pupils. This presents two sorts of problems to those who would still defend such a system regardless.

---

<sup>1</sup>Benjamin Bloom, George Madaus and J. Thomas Hastings, Evaluation to Improve Learning (New York: McGraw-Hill, 1981), 4.

First, there is the practical problem of demonstrating how the continuous and systematic receipt of low marks students can nonetheless motivate students to engage in educational activities. I take it that the prospects for such a demonstration are fairly bleak. Second, there is the *moral* issue of whether we are justified in comparing the work of students against the work of their peers, and in making public these comparisons. Given the fact, as Bloom, Madaus and Hastings point out, that public school is compulsory (and therefore that students have no choice about whether or not they are to be marked and graded), I take it that such a justification will be difficult indeed.<sup>2</sup>

In the second place, there is the difficult issue of knowing how students interpret their situation. To suggest that a given stimulus (i.e. testing, marking and grading) unproblematically brings about a particular response (i.e. application to one's studies) is to take a rather naive and outmoded view of school effects. As part of his critique of the *Fifteen Thousand Hours* report, Trevor Pateman pointed out in 1980 that the trend in social science research over the previous twenty years had been to move from behaviourist to hermeneutic methodologies.<sup>3</sup> He characterized that shift as follows:

the whole movement in academic social science as an endeavour aiming at *explanations* of social phenomena has been towards the view that social phenomena can only be understood in terms of the *motives* of the individual participants ('actors', 'agents' or 'members' according to your theoretical orientation) and the *meanings* they give to institutions and events.<sup>4</sup>

If this orientation is correct, then the question of how testing, marking and grading motivates students becomes a question of how any system of testing, marking and grading is *interpreted* by those individuals who are subjected to it. One possible response, of course, is to completely reject or deny the motivational power of marking and testing. Sociologist Paul Willis provides a good example of such a possibility in his 1977 study, Learning to Labour:

---

<sup>2</sup>This is an extremely important issue to which I shall return in Chapter Five.

<sup>3</sup>Trevor Pateman, "Can Schools Educate?" Journal of Philosophy of Education 14, (1980): 139-148.

<sup>4</sup>*Ibid.*, 140. (original emphasis)

How Working Class Kids Get Working Class Jobs.<sup>5</sup> His point, essentially, is that working class children choose not to accept the importance of school authority (marks and grades being part of that authority), as a way to more strongly identify with working class culture.

Even where marks and grades *are* seen as important, though, we can still rightfully ask in *what sense* they are so regarded. It may be, for example, that some students, even those who get high marks, do so for reasons quite apart from the satisfaction of achieving mastery in a given domain. Some students, that is, might take a cynical view of a school reward system, knowing on the one hand that it is reductionist and trivial, and on the other that its rewards nonetheless yield considerable social advantage. The issue of how students *interpret* their evaluative context, therefore, can be very complex. It is an issue which must be addressed, however, before we can unproblematically state that a given regime of testing, marking and grading is motivational.

#### The Central Critique

A more fundamental objection that can be raised against the proposition that testing, marking and grading motivate students to learn is the suggestion that, far from motivating learning, our current use of testing, marking and grading actually *contaminates* the learning process. This contamination can occur in two ways. Insofar as our practices of testing, marking and grading are prosecuted for classificatory purposes, there is a danger, as outlined in the previous chapter, that the subject matter presented to students will be so reduced and distorted as to make it difficult for them to learn in the sense of coming to understand and appreciate the standards which define a practice. Second, insofar as marks and grades represent *extrinsic* goods, their use as motivators can significantly degrade the instructional environment. For both these reasons, therefore, it is ironic to defend testing, marking and grading as a mechanism to motivate students *to learn* in the sense of getting on the inside of a practice. As I hope to make clear, it is far more likely that our use of testing, marking and

---

<sup>5</sup>Paul Willis, Learning to Labour: How Working Class Kids Get Working Class Jobs (Farnborough: Saxon House, 1977).

grading motivates students to learn only in the trivial sense of memorizing answers, and in the cynical sense of figuring out the game that needs to be played in order to achieve success within the context of the institution.

I have already sketched out, in chapter two, the hazards associated with using testing, marking and grading primarily for classificatory purposes. The argument there was that when the classificatory imperative becomes predominant, curricular narrowing, trivialization, and fragmentation can easily follow. What can also follow is the conceptual confusion that learning is merely a matter of getting correct answers, and that judgements about what a student understands and appreciates can be made merely by assembling a composite picture of what he or she has so "learned." The net effect of these methodological and conceptual difficulties is to threaten the integrity of the educative project.

Now what makes it so objectionable to defend and promote testing, marking and grading as a mechanism of *motivation* is that in doing so one necessarily runs the risk of promoting and legitimizing a less-than-complete characterization of the subject matter at hand. Perhaps an analogy will make this clear.

Imagine that, for whatever reason, we decide as a society that we want to promote a sense of art appreciation in our citizens. We decide, therefore, to create schools and to make "art appreciation" classes available to all. We decide, further, that we want to bestow prestige and wealth on those people who seem to have the clearest, best, or most refined sense of art appreciation. We are faced, therefore, with a problem. We need to identify who amongst the total population are "most worthy" of receiving our esteem and our money. We feel a need to make this identification for two reasons: we assume that only a few among the many will be able to reach the (as yet undefined) standard of excellence in "art appreciation", and with regard to both money and prestige we are operating with scarce resources. (We have a practical scarcity of financial resources because our coffers are finite, and a logical scarcity of prestige because if everyone is respected equally, then no one is more prestigious than



another.) Our task, therefore, is to create an assessment system which will identify "the brightest and the best."

Now it might be thought that since art appreciation is a rather complex undertaking, the most reasonable way to find out who is the most advanced in this field is to have recognized experts conduct interviews with students. Two objections, however, will immediately surface. In the first place it will be said that this is far too open-ended, far too *subjective* a process. Interviewers can be biased; irrelevant considerations, such as cultural background, can creep into their evaluations. Secondly, it is unlikely that these interviewers are going to produce for us a *precise* enough indication of where the student stacks up in relation to his or her peers. Given that thousands and thousands of students will be taking our courses, and given that at the end of the day we can only select a small percentage of those to receive our esteem and our money, we need such a precise indication. In the interests of generating a *fair measure* of student achievement, then, we need to design our assessment system accordingly. We need, in other words, to design questions which will yield an unambiguous measure of what students know, and to present these questions in a format uncontaminated by personal bias. The way to solve our problem, therefore, is to break "art appreciation" down into discrete components and then measure students' understanding of these by way of a series of tests. We might get students to name, for example, ten important artists, or to list for us, in proper historical succession, the various artistic periods. Given enough of these questions, we could eventually determine, with a considerable degree of precision, which students within the total group are most worthy of subsequent reward.

The problem with all of this, of course, is that in constructing our assessment system in this way we run the risk of considerably reducing and trivializing the undertaking of art appreciation. It is one thing to be able to name ten important artists; it is quite another to be able to understand and appreciate the contribution that a Rembrandt, or a Picasso, or a Van

Gogh has made to the artistic community. The problem, then, is that the mechanism of assessment contributes to the deterioration of the practice.<sup>6</sup>

What makes it objectionable, then, to defend testing, marking and grading as a mechanism which motivates students "to learn" is that, insofar as testing, marking and grading are prosecuted for classificatory purposes, there is a danger that in doing so one implicitly endorses and promotes the reduction and trivialization of school subject matter. To return to our analogy, to say that testing, marking and grading can be defended as a mechanism which motivates students to learn, is to run the risk of endorsing and promoting a view of art appreciation as little more than the naming of artists and the listing of artistic periods. To defend and promote testing, marking and grading as mechanisms which motivate students *to learn*, then, is little short of ironic.

The second way in which marks and grades, in particular, can contaminate the learning process is by virtue of the fact that marks and grades are goods *external* to a practice. What is meant by "goods external to a practice" is neatly captured in an example of Alasdair MacIntyre's in which he presents a strategy for teaching a child to play chess.<sup>7</sup> MacIntyre has us imagine that he has made an arrangement with a seven-year-old child to play the game once a week in exchange for 50 cents worth of candy. The child has no real desire to play chess *per se*, but plenty of desire for candy. MacIntyre tells the child that he will play in a way that it will be difficult, although not impossible, for his opponent to win, and that for each game the child wins an additional 50 cents worth of candy will be forthcoming. "Thus motivated," he says, "the child plays, and plays to win."<sup>8</sup> MacIntyre then points out that insofar as the child plays for candy alone, he or she is motivated by goods *external* to the practice of chess-

---

<sup>6</sup>The mechanism of assessment may not, of course, be the *only* contributor to this deterioration. The structure of the curriculum itself, and the almost carte blanche acceptance of "teacher effectiveness" strategies of instruction may also be involved. On the latter point, see, for example, George Madaus, Peter Airasian and Thomas Kellaghan, School Effectiveness: A Reassessment of the Evidence (New York: McGraw-Hill, 1980).

<sup>7</sup>Alasdair MacIntyre, After Virtue, 2nd ed. (Notre Dame, Indiana: University of Notre Dame Press, 1984), 188.

<sup>8</sup>Ibid.

playing. When (or if), on the other hand, the child begins to play for the sheer joy and challenge of the game in and of itself, then we can say that the child is playing for the goods *internal* to the practice of chess-playing.

Goods external to a practice, then, are goods achieved quite apart from appreciating the intrinsic value of the practice itself. When we are paid money for engaging in an activity we enjoy in and of itself, the money we receive is an *external* good. The goods *internal* to the practice, on the other hand, are those characteristics of the undertaking which make us want to engage in it quite apart from the promise or receipt of external goods. R.S. Peters offers an insightful elaboration in regard to the particular satisfactions to be had by getting on the insides of particular forms of thought:

Scientists, mathematicians, and philosophers do not just desire to discover the truth; they desire also to devise ingenious experiments, to construct elegant proofs, and to develop clear and cogent arguments. Writers desire to construct neat plots, to make witty remarks, and to fix their feelings in just the right form of words. These joys are intrinsic to the activities and modes of thought in question; they reinforce the more general motivations which urge men on above the level of their 'necessary appetites'.<sup>9</sup>

The problem, in general, with installing *external* goods as the principal motivating feature to get students to learn is that, in doing so, we run the risk of never acknowledging, or even allowing our students to appreciate, goods internal to practices. When *the acquisition of marks and grades* becomes the reason to attend to one's studies (as opposed to the intrinsic interest of the study itself), then the name of the game becomes one getting *marks and grades at any cost* - even at the cost of missing out on the opportunity to find satisfaction in engaging in a practice in and of itself.

In addition, there are at least three more specific ways in which the use of external goods to motivate students to learn can have a deleterious effect on the educative project. In the first place, as sociologists Don Robertson and Marion Steele point out, there is the difficulty of trying to reconcile the role of teacher as inquisitor with the goal of teacher as

---

<sup>9</sup>R.S. Peters, *Ethics & Education* (London: George Allen & Unwin, 1966), 61.

educator.<sup>10</sup> What Robertson and Steele have in mind is that the nature of the pedagogical relationship between the teacher and the student is seriously hampered by virtue of the fact that, at the end of the day, the teacher must assign a value to the responses and work of the student. Insofar as this value is assigned relative to the number of correct answers the student produces, this means that the kind of conversation the student is likely to engage in with the instructor will be rather confined. It means, in particular, that it is unlikely the student will be eager to demonstrate what he or she *does not* know. Students are reduced, instead, to enacting a bit of theatre in which the object of the drama is to create a character more erudite than oneself.

Students want teachers to think they are "getting it" even when they are hopelessly confused. They learn that to reveal their ignorance is to provide evidence that can be used against them. Even the constitution protects against self-incrimination. So the game is to appear wise, to look knowledgeable when the teacher refers to a section of the book you were supposed to have read, to try heroically to bluff when you are hit with questions you're lost on. "Image" is what counts, not substance.<sup>11</sup>

This situation is *pedagogically* troubling for two reasons. First, it systematically encourages students to cut themselves off from one of the richest sources of improved understanding, that is, their mistakes. Second, it tends to ensure that teachers will be constantly kept in the dark about what students really know. Without such an understanding their ability to help students - to *teach* them - is significantly restricted.

A second significant consequence of using external goods to motivate students to learn is that it then becomes possible, indeed even reasonable, for students to cheat. Alasdair MacIntyre makes this clear in his chess playing example. He points out that as long as a child is playing solely for the sake of acquiring candy, he or she "has no reason not to cheat, and every reason to cheat, provided he or she can do so successfully."<sup>12</sup> When external goods

---

<sup>10</sup>Don Robertson and Marion Steele, The Halls of Yearning: An Indictment of Formal Education/A Manifesto of Student Liberation (San Francisco: Canfield Colophon Books, 1969), 47.

<sup>11</sup>Ibid., 53.

<sup>12</sup>MacIntyre, After Virtue, 188.

become paramount, in other words, undertakings become valued not in and of themselves, but only as a means to acquire something else. If that something else can be acquired by fraud or subterfuge, so much the better.

The application to our current use of marking and grading in schools is obvious: when students engage in studies solely for the purpose of acquiring marks, their project can be characterized not so much as a quest to understand and appreciate the canons and standards which define practices, but more as a determination to do whatever needs to be done to receive the requisite numbers. If those numbers can be achieved quite apart from the student paying strict attention to the canons and standards which define a practice, so be it. This merely emphasizes a point alluded to by Robertson and Steele:<sup>13</sup> to emphasize the acquisition of marks as an important undertaking within schools is to virtually invite students to cheat.

It is not that the cheating in itself is objectionable (although some may indeed find this offensive as well); it is rather the fact that it *makes sense to cheat* that is troublesome. The fact, in other words, that students can and do gain academic recognition *without* necessarily having come to understand and appreciate the goods internal to a practice suggests that our assessment system may be seriously flawed. The fact, moreover, that some offer this very same assessment system as a mechanism to *motivate* students to learn is particularly perplexing.

A third negative consequence, following directly from these first two, is the cynicism that can emerge on the part of students with regard to the possibility that what they are studying may have intrinsic value. As Wideen and his colleagues have reported, higher level students, in particular, soon become impatient with instructional strategies that are not explicitly aimed at preparing them for major examinations.<sup>14</sup> The suggestion that one might like to read a chapter of history, for example, for some reason *other* than to prepare for a test

---

<sup>13</sup>Robertson and Steele, Halls of Yearning, 52-53.

<sup>14</sup>Marvin Wideen et al., Impact of Large Scale Testing on the Instructional Activity of Science Teachers Paper presented at the Canadian Society for Studies in Education, 1991, 57.

(say, for example, because it is particularly well written, and because one can learn something that one didn't know before) is , if not openly ridiculed, then at least privately regarded as perhaps quaint, but hopelessly impractical.

It is not surprising that students should take this attitude; many teachers give it tacit endorsement in both word and deed. Given, for example, the emphasis that teachers put on marks and grades, and the way that teachers use the threat or promise of marks and grades to shape student behaviour, it is little wonder that students fail to appreciate that what they are studying might have some value quite apart from the evaluative rewards they receive.

The danger, then, of using goods external to practices to motivate students to learn is that students can soon get the message that goods *internal* to practices are not to be taken seriously. From the point of view of wanting to *educate* students, this again is ironic. If we believe that the aim of educating students is to get them on the inside of a practice, and if part of what it means to be on the inside of a practice is to understand and appreciate *goods internal to that practice*, then it is contrary to one's aim to defend a system of motivation which de-emphasizes and perhaps deflects students away from understanding and appreciating the intrinsic values within practices.

In summary, because marks and grades are external goods, to use them to motivate students to attend to their studies is to run the risk of doing damage to the instructional process. In particular, it is to run the risk of creating an instructional environment in which students cannot be candid with their instructors; in which students will have good reasons to attempt to deceive their instructors about how much they know; in which students can receive academic recognition without necessarily understanding or appreciating goods internal to a practice; and in which students can learn to become cynical about the likelihood that what they study in school is of intrinsic value.

### Objections

Now it might be proposed, in response, that even if there are dangers to using extrinsic rewards as a *permanent* motivational mechanism, they might nonetheless be acceptable - and even necessary - as a way to get students *started* down the road to learning. MacIntyre intimates as much in his example,<sup>15</sup> while R.S. Peters makes the point explicitly:

The aim of the educator is to get others on the inside of such worthwhile activities and forms of awareness so that they will explore them for the ends which are intrinsic to them. But in the early stages he may have to use extrinsic motivations both to get children started on them and to sustain their interest when the stage of precision begins to exert its irksome discipline.<sup>16</sup>

There is a logical point implicit in this suggestion to the effect that one cannot, strictly speaking, be motivated *by* an understanding and appreciation of goods internal to a practice until one truly *understands* and *appreciates* goods internal to a practice. Until such time as this happens, one must necessarily be motivated by *something else*. This logical point, however, sets the problem too rigorously. What we could do to motivate student other than through external goods is to try to motivate them by the *promise* and the *partial* indication that the undertaking is inherently interesting and worthwhile. Peters refers to intrinsic goods of a general nature which might be summoned to motivate students to participate in an intellectual undertaking. These he lists as "the desire just to find out things and to explore the environment, the desire to manipulate things, the sense of competence and of mastery, and the achievement motive."<sup>17</sup> We might attempt to motivate our students to study history, for example, just for the sheer *general* joy of allowing them to find out something they did not previously know. Peters goes on as well, however, to intimate that even goods internal to specific practices can be used to motivate the uninitiated. He characterizes the "greatest educators" as those who "can convey *insensibly* the sense of quality [within practices] so that a

---

<sup>15</sup>MacIntyre, After Virtue, 188.

<sup>16</sup>R.S. Peters, Ethics and Education, 62. The "stage of precision" is a reference to Alfred North Whitehead's essay, "The Rhythm of Education," which can be found in Alfred North Whitehead, The Aims of Education and Other Essays (New York: MacMillan, 1929; The Free Press, 1967), 15-28.

<sup>17</sup>*Ibid.*, 61.

glimmering of what is intrinsic is constantly intimated."<sup>18</sup> What we can also do, in other words, is present if not the full body, then at least something of the flavour of the intrinsic value of the study undertaken. We need not always resort to external goods, that is, as a way to initiate students into a practice.

There are practical difficulties with this counter-proposal, as well. In the first place, if we are going to use external goods to get students *started* on the road to learning, it is always difficult to know *when* to stop. Presumably, students come to appreciate goods internal to a practice at different rates. If we are going to use marks and grades as the external goods to motivate students, then it will be awkward, to say the least, to give out marks and grades to some and at the same time withhold them from others.

Second, it can also be difficult to know *how* to stop giving out external goods, even if one wants to. One reason for this is that once a precedent has been set, it can be difficult to break. In the case of public education, generations of people have been brought up to believe that marking and grading is completely appropriate in school - that, in fact, getting marks and grades in school is what public education is all about. It is unlikely that such ingrained perceptions will easily die. There is, as well, the unmistakable fact that marking and grading are powerful mechanisms of control. Students have learned well to do what is asked of them solely in response to the threat or promise of an adjusted grade. This is a trump card that not all teachers may be willing to give up.

For logical and practical reasons then, it is both unnecessary and inadvisable to use marks and grades - and perhaps even any external good<sup>19</sup> - as a sort of "kick-start" mechanism to initiate students into learning. We can (and perhaps, more importantly, should) motivate students by appealing to the general value of learning something new, and by at least

---

<sup>18</sup>Ibid., 62. (emphasis added)

<sup>19</sup>The use of approbative symbols (e.g. stars and stickers) in the primary grades is an interesting case. One objection against this practice, for example, might be that it sets the precedent for using quantitative symbols (e.g. marks and grades) in later years.



intimating something of the specific value of the practice undertaken. If we choose to use external goods to attempt to introduce students to practices, we will be faced, in the first place, with the problem of knowing when to stop awarding these goods, and, in the second place, with the difficulty of knowing how to stop awarding them.

But the objection might still persist. It might still be thought, that is, that one *can* come to understand and appreciate goods internal to a practice by way of an assessment system which uses external goods to keep a student attending to his or her studies. Insofar as the threat or promise of marks and grades *is* sufficient to motivate a student to work, and insofar as the student *does* (as a result of his or her work) come to understand and appreciate the goods internal to a practice, then there *is* a sense, so the argument would go, in which testing, marking and grading *has* motivated the student to learn. It is conceivable that one could even come to understand and appreciate, for example, philosophy via such a route.

Against this, I want to suggest that if a person does come to understand and appreciate the goods internal to, for example, philosophy it is likely that he or she will have done so not *because* of the motivating effects of testing, marking and grading, but more probably *in spite of* their influence. I want to suggest, in other words, that given the considerable barriers that a system of testing, marking and grading can erect against instruction aimed exclusively at getting students on the inside of a practice, it is more likely that students who nonetheless do come to appreciate such goods do so as a result of *other* motivating factors - the tutelage of an excellent teacher, perhaps, or by way of independent reading, or maybe by both. It is somewhat misleading, therefore, to suggest that a system of *testing, marking and grading* can be given credit for creating conditions which cause students, eventually, to come to understand and appreciate goods internal to a practice.

But the counter-objections can go even deeper than this. There is, as well, something distastefully paternalistic, and even coercive, about using marks and grades as either an initiating mechanism, or a permanent motivational strategy, within the instructional process. It is as though students can't be trusted to see the value of the subject matter itself, and that they

must be perennially treated as children and given candy money instead of being shown why something might be considered intrinsically worthwhile. Given, further, that marks and grades *do* hold significant social value, and that public schooling is compulsory, the use of marks and grades to get students to learn is not so much motivational as it is coercive. It is likely, that is, that many students who make the decision to "work hard" in school do so not necessarily because the subject matter is worthy of their effort, but rather because they understand that the work they do in school is to be translated into an important public record. Students feel compelled to study, in other words, not because of the value of the study itself, but because of the tangible social reward that success in one's studies can bring.

And finally, there is something fundamentally *miseducative* about being "motivated" (i.e. compelled) to attend to one's studies in the interest of getting marks, as opposed to being given subject-specific *reasons* to do so. If we take seriously Scheffler's idea that teaching involves submitting ourselves to the understanding and independent judgement of our pupils, then to "motivate" students by the promise of marks - as opposed to the presentation of good *reasons* as to why the subject matter is important - is to fail to *teach* them. Something has gone terribly wrong, in other words, when the best reason we can give students for attending to a particular study is that they will get marks for doing so.

### Conclusion

The original proposition was that testing, marking and grading can be defended because it motivates students to attend to their studies, and that by attending to their studies students will eventually learn in the sense of being able to operate on the inside of practices. Against this, I have tried to show that there are good reasons to believe that, far from motivating students to learn, the use of testing, marking and grading may actually contaminate the learning process. I have tried to show, first, that where testing, marking and grading are used essentially for classificatory purposes, the kind of learning which results can often be sub-standard, and that to offer these mechanisms as a system of *motivation*, therefore, is implicitly, if not explicitly, to support a restricted view of learning. Second, I have tried to show that

because marks and grades are *goods external to practices*, their use as a motivational mechanism can create serious problems for the educative project. I have tried to show, in particular, that their use can restrict candid communication between teachers and students; can make it sensible for students to cheat; and can encourage students to adopt a cynical attitude about the intrinsic value of intellectual pursuits. The net effect of this critique has been to suggest that it is ironic to defend testing, marking and grading as mechanisms to motivate students *to learn*, where learning is understood as the enterprise of getting on the inside of a practice.

## CHAPTER FOUR

### THE ACCOUNTABILITY DEFENCE OF TESTING, MARKING AND GRADING

A third argument commonly offered in defence of our current practices of testing, marking and grading is that such an undertaking is necessary to ensure accountability within our schools.<sup>1</sup> The proposition here is that if we do not use tests, and with them a system of marking and grading, we will not be able to identify what students know or to what extent they may have progressed. The implication, further, is that by measuring student knowledge and progress, we also get a clear measure of teacher effectiveness. The use of testing, marking and grading as a mechanism of accountability, therefore, is thought to represent an important contribution to the maintenance and promotion of quality schooling.

This proposition can be broken down into two smaller claims: i) that it is both intelligible and appropriate to call for "accountability" in the context of educating people, and ii) that the current system of testing, marking and grading is, in general, a good mechanism of accountability. I should like to take up each of these issues independently.

#### Accountability and Education

The call for *accountability* within the context of educating people is peculiar. Use of the term "accountability" usually implies at least three things: that there is a tangible indicator which will tell us whether or not we have been successful at a given undertaking; that the lines of responsibility for bringing about success or failure are relatively clear; and that there exists a relationship of authority in which it is appropriate that a subordinate report to a superior.<sup>2</sup> A clear case in which it would be appropriate to speak of accountability would be where a

---

<sup>1</sup>Newsweek contributor Tom Morganthau, for example, says of standardized testing that "there is no other way - no equally efficient, relatively objective method - to find out whether American schoolchildren are learning what society wants them to know." ["A Consumer's Guide to Testing," Newsweek, (Fall/Winter, 1990): 63]; Ramsay Selden endorses an increased use of "educational indicators" to provide meaningful information about the quality of schools. ["Missing Data: A Progress Report From the States," Phi Delta Kappan, (March, 1988): 492-494.]

<sup>2</sup>I am indebted to LeRoi Daniels for bringing this third dimension of accountability to my attention.

supervisor informs a worker that judgements about the worker's performance will be made on the basis of how many widgets, say, he or she produces within a specified amount of time, given sufficient materials. What makes it peculiar to use this term in the context of educating people is that none of the above-mentioned conditions seem to apply. In the first place, what counts as success within an *educational* context is not a particularly tangible thing, and certainly not, for example, as tangible as the number of widgets produced in so many hours. Judgements about the extent to which students are operating on the inside of a practice are not at all amenable to unambiguous measurement. Most attempts to make such judgements amenable to measurement, moreover, end up doing damage to the original point of the enterprise.

Second, it is not at all clear that teachers can be held completely responsible for student learning. Given that there can be a number of significant features within a potential learning situation which are beyond the ability of a teacher to control (for example, the state of mind of the student), it hardly seems appropriate, or fair, to blame (or even praise) a teacher for the results which follow. While the issue of who to blame or praise in the case of the worker producing widgets may be relatively straightforward, in the case of teachers bringing about learning, the situation is considerably more complex. This is an issue to which I shall return.

Third, it is also not clear that relationships of subordination are appropriate within a strictly *educative* undertaking. I have in mind, in particular, relationships of bureaucratic subordination in which a teacher is expected to "produce results" for a department head or senior administrator.<sup>3</sup> It seems to me that the only authority which ought to reign within an *educative* context is the "authority" of the discipline being taught, and the practice of teaching. What teachers are obliged to pay attention to, in other words, are the standards which define the discipline being investigated, and the standards, if I can put it this way, of good teaching. If a teacher is beholden to a department head or senior administrator it should only be by

---

<sup>3</sup>A good example is the case of the grade 12 teacher who is expected to ensure that his or her students achieve acceptable results on externally-administered final examinations.

virtue of the fact that the department head and senior administrator share a concern for maintaining the integrity of the discipline and the practice of teaching. But notice, therefore, that this suggests not so much a relationship of subordination (in which it makes sense to speak of accountability), as a relationship of collegiality characterized by a sense of mutual *responsibility* concerning the undertaking at hand. Use of the term "accountability", then, seems out of place within a strictly *educative* context.

The accountability movement is, presumably, offered in part as a response to the question of how we maintain and promote quality schooling. The proposition seems to be that to provide an account of the extent to which students are learning (and, therefore, teachers are teaching)<sup>4</sup> is at the same time to contribute to the enhancement of education within schools. Sometimes this is framed in terms of a free market analogy whereby it is thought that competition amongst schools to bring about the highest level of performance will necessarily yield an overall improvement in the quality of schooling.<sup>5</sup> One problem with this, though, is that the measures which competitors often rely on to make judgements about the performance of a particular school or educational system are not always the kinds of measures which address or represent strictly *educational* considerations. More often than not, for example, the measures used are economic. Walter Feinberg cites the case of a superintendent in 1913 who asked, apparently with great seriousness, "Why is pupil recitation in English costing 7.2 cents in the vocational school while it costs only 5 cents in the technical school?" and "Why does a pupil-recitation in science cost from 55 to 67 percent more in Newton High than in either of the other schools?"<sup>6</sup> On a more contemporary note, George Madaus and Vincent Greaney suggest that behind the Irish implementation of its Primary Certificate examination "lay a

---

<sup>4</sup>Of course neither "teaching implies learning," nor "learning implies teaching" are true. All that holds is that teaching implies *trying* to get someone to learn; in our terminology, *trying* to get someone on the inside of a practice. See pp. 63-64 below; cf. note (?), p. (?).

<sup>5</sup>See, for example, Doug Willms, Monitoring School Performance: A Guide for Educators (London: Falmer Press, 1992).

<sup>6</sup>Walter Feinberg, Reason and Rhetoric: The Intellectual Foundations of 20th Century Liberal Educational Policy (New York: John Wiley & Sons, 1975), 63.

belief that the country was not getting proper value for its educational expenditure," and that similar sentiments have been raised in the United States.<sup>7</sup> The key concern, in other words, is how much bang the public gets for its educational buck. The problem with *this*, of course, is that a lot can be lost in the translation between the imperative to educate students and the imperative to be economically efficient.

A second problem which can arise with the free market analogy of school improvement, even in those cases where economic indicators are not employed, has to do with the legitimacy of inter-school or inter-district comparisons. Former U.S. Secretary of Education Terrel Bell once put together a "wall chart" which offered a state-by-state comparison of SAT scores.<sup>8</sup> The message, presumably, was that states which ranked low on the list would have to improve, while those which were at or near the top should feel justifiably proud.<sup>9</sup> Sociologists Lala Carr Steelman and Brian Powell, however, take issue with Bell's implicit comparison by pointing out that more than four fifths of the difference in SAT scores between states can be attributed to the different percentages of students who take the SAT exam in each state. While Arkansas, for example, ranked 12th in the nation on Bell's "scorecard" in 1982, only 4% of its eligible seniors opted to take the exam. New York, on the other hand, ranked 35th on the scorecard, yet had 59% of its seniors take the test.<sup>10</sup> The kind of unambiguous state-by-state comparison that Bell's chart invited was, therefore, wholly inappropriate.

Even worse than the inappropriateness of these comparisons is the inappropriateness of the very enterprise of educational accountability. Steelman and Powell end their article by

---

<sup>7</sup>George Madaus and Vincent Greaney, "The Irish Experience in Competency Testing: Implications for American Education," American Journal of Education (February 1985): 268-294.

<sup>8</sup>Lala Carr Steelman and Brian Powell, "Appraising the Implications of the SAT for Educational Policy," Phi Delta Kappan (May 1985): 603-606.

<sup>9</sup>In their "Appraising the Implications" article, Steelman and Powell cite the case of an Indiana superintendent of education candidate who unsuccessfully argued that the state's low SAT ranking was a justification for ousting the incumbent. He argued, apparently, that "a football coach whose team was ranked that low would be fired without a moment's hesitation."

<sup>10</sup>Steelman and Powell, "Appraising the Implications," 604.

saying of Terrel Bell's pronouncement to state officials that he will "be disappointed if the increase [in SAT scores] next year isn't twice what it was this year" that Bell is "promoting the devotion to test scores, though not necessarily the devotion to improving education."<sup>11</sup> They pointedly imply, in other words, that the accountability movement in general suffers from the same disease as classificatory assessment: in trying to get a *measure* of quality it can end up distorting the very nature of the educative enterprise. Once again the irony is striking: instead of acting as a force to improve schooling, the accountability project may actually contribute to its deterioration.

### Testing, Marking & Grading as Mechanisms of Quality Control

Apart from the *general* issue of whether or not it is appropriate to speak of accountability in the context of educating people, there is also the more specific question of the extent to which testing, marking and grading, in particular, either contribute to, or detract from the maintenance and promotion of quality schooling. The implicit proposition seems to be that we need testing, marking and grading to provide an account of the extent to which students are learning, and therefore the extent to which teachers are teaching. If we can have such an account, so the thinking goes, we shall then be in a position to make relevant improvements.

With regard to the proposition that a system of testing, marking and grading can provide us with a reliable account of the extent to which students are *learning*, reasons have already been presented to be skeptical about the likelihood of this. If by *learn* we mean the extent to which a student has got on the inside of a practice, I have already shown that the methodological hazards which can accompany such a system (i.e., curricular narrowing, fragmentation, and trivialization) can make this highly unlikely. Evaluation expert Lee Cronbach got to the heart of the matter when he wrote that "learning the answer to a set of questions is by no means the same as acquiring *understanding* of whatever topic the question

---

<sup>11</sup>Ibid., 606.



represents."<sup>12</sup> Similarly, George Madaus and Vincent Greaney have implied that it is not surprising that scores on minimum competency tests dramatically increase when remediation efforts focus directly on the test itself: what else would we expect? They go on to warn, however, that "improvement in test performance does not necessarily signal a concomitant improvement in basic skills."<sup>13</sup> The point, in other words, is that the use of tests and test scores as a mechanism of accountability does not yield an account of *educational* processes per se; rather it yields but an account of how well students have done on tests. The two are not necessarily the same.

What needs to be emphasized here is that it is precisely because test scores *have* become a measure of accountability that teachers have learned to take them seriously, and in fact to shape their instruction accordingly. Gerald Bracey refers to the latter practice as "deflection" - that is, when an over-emphasis on testing deflects the curriculum from its intended purpose. Most people, however, know this by its more common description: teaching to the test. Bracey cites the example of a school system in Virginia which threatened to sue the state for including questions on perpendicular lines on the state-wide minimum competency test because these were not explicitly prespecified. While it was understood that questions on parallel lines would be included in the test, no such warning, apparently, was given about the inclusion of questions about perpendicular lines.<sup>14</sup> What doesn't get tested, doesn't get taught it seems - even though, as in this case, the close logical fit between parallel and perpendicular lines would seem to make the teaching of both mandatory. Andrew Strenio likewise offers the example of New York teachers who decided to stop teaching antonyms to students because their new city-wide reading test didn't place any emphasis on them. In

---

<sup>12</sup>Lee Cronbach, "Course Improvement Through Evaluation," Teachers College Record, 64 (1963): 672-683.

<sup>13</sup>George Madaus and Vincent Greaney, "The Irish Experience in Competency Testing," 288. For an excellent discussion on the folly of guiding instruction according to test results on discrete skill elements, see Frank Smith, Insult to Intelligence (Portsmouth, New Hampshire: Heinemann, 1986), 166-167.

<sup>14</sup>Gerald Bracey, "Measurement-Driven Instruction: Catchy Phrase, Dangerous Practice," Phi Delta Kappan (May, 1987): 683-686.

Strenio's words, "The importance of learning antonyms varies from one year to the next not because of a reasoned change in curriculum adopted by the school board, but because a different reading test has been selected for use."<sup>15</sup> Although Strenio characterizes this situation as representing a "rudderless curriculum," there is little doubt here that the true rudder which guides the curriculum is the test itself.

What is significant for our purposes, then, is that, *precisely because of the importance that test scores have come to have as a measure of accountability*, it makes sense that teachers teach to the test. Given a situation in which judgements about the quality of a teacher's work are likely to be made on the basis of test results, it is in a way foolhardy of a teacher not to try to enhance these. What needs to be remembered, though, is that it is likely that the kinds of tests which yield unambiguous scores for the purposes of accountability are also the kinds of tests which have little or nothing to do with educating students in the sense of getting them on the inside of a practice. Writing in regard to the issue of standardized testing in general, Andrew Strenio characterizes the problem as follows:

The more tests are relied on as an evaluation tool, the more teachers will concentrate on preparing children for those tests. Then what will we have accomplished? If subjects and skills that are too important and too complex to be measured on these tests are excluded from our children's education, we will have lost far more than we have gained. The unhappy prospect of turning out a generation of students who can breeze through multiple-choice tests and little else stares us in the face. We want schools to teach our kids how to read, write, use arithmetic, and then to use those schools in combination with an appreciation of the arts and the duties of citizenship. Constant practice in filling in tiny boxes with pencil marks does not necessarily lead us in that direction.<sup>16</sup>

To use test scores - and test scores from tests which narrow, fragment, and trivialize the curriculum - as a way to monitor the extent to which students are "learning," therefore, is to be faced, once again, with the irony of employing a system of accountability which might do more to detract from the educative project than to enhance it.

---

<sup>15</sup>Andrew Strenio, The Testing Trap (New York: Rawson, Wade, 1981), 26-27.

<sup>16</sup>Ibid., 106.

As for the proposition that a system of testing, marking and grading can shed relevant light on the extent to which teachers are *teaching*, here again there are difficulties. In the first place, there are further reasons to be skeptical that test scores and educational indicators are representative of the extent to which students are operating on the inside of a practice. In addition to the points already mentioned, there is the fact that scores are sometimes deliberately manipulated for political rather than pedagogical reasons. Recently, three school districts in British Columbia had their final exam failure rates decreased by decree of the provincial Ministry of Education.<sup>17</sup> All three districts had experienced labour strife over the preceding school semester, and so it was thought, apparently, that the final marks did not truly represent what the students could do. Abbotsford resident Douglas Hudson, however, took issue with this interpretation of events saying that the attempt to link the higher failure rate with teacher job action was reductionist and that it smacked of political expediency.<sup>18</sup> Given the political climate of British Columbia over the past decade, it is hard not to agree with him.<sup>19</sup> On a similar note, Gerald Bracey has suggested, in response to the fact that a Virginia minimum competency test does not require students to add, subtract, or divide fractions, that one possible reason for the exclusion is that tests which required such operations "may produce a politically unacceptable number of failures."<sup>20</sup> What is needed, apparently, are just enough failures to keep people honest, but not so many as to disrupt the entire system. As Wilson and Rees imply,<sup>21</sup> what seems to be needed is a pattern of results which does not deviate too much from that of a normal distribution curve. The point, in any case, is that if final grades are going to be scaled to avoid a politically unacceptable number of failures, or if the testing

---

<sup>17</sup>The districts were Abbotsford, Peace River North, and Princeton. See Frances Bula, "Exam Marks Adjusted in Strife Areas," Vancouver Sun, 1 August 1991.

<sup>18</sup>Douglas Hudson, "Exam Marks Adjusted in Strife Areas," Letters to the Editor, Vancouver Sun, 3 August 1991.

<sup>19</sup>For an earlier example of the Ministry of Education scaling grades for political expediency, see Bob Simpson, "Grading Those B.C. Tests," Vancouver Sun, 26 March 1984.

<sup>20</sup>Bracey, "Measurement-Driven Instruction," 685.

<sup>21</sup>see p. 45 above.

system is implicitly designed to produce bell curves, then final results can hardly be relied on to represent meaningful indicators of the extent to which teachers are teaching.

Even more fundamental than this problem, though, is the methodological confusion of assuming that something like test scores are a relevant measure of the extent to which teachers are doing their job. To assume this is to adopt what David Ericson and Frederick Ellett refer to as a "causal theory of teaching" - that is, the belief that teaching can be a sufficient condition of learning, and that as long as teachers are teaching *properly* learning will necessarily result.<sup>22</sup> Against this, Ericson and Ellett point out - quite rightly, I think - that teaching and learning are causally connected only in the sense that teaching is but *one* activity *among many* that could impinge upon a student's learning.

As everyone should know, teaching by itself is rarely, if ever, sufficient for learning. If the student fails to attend to the teaching, fails to practice, fails to study or do homework, etc., obviously the student has little chance of learning the subject matter.<sup>23</sup>

It is partly for this reason that Ericson and Ellett argue that teachers cannot be held *completely* accountable for student learning. They make a distinction between "good" teaching (i.e., teaching which reflects an adequate understanding of the material and provides a reasonable instructional sequence) and "successful" teaching (i.e., when the student learns the intended material), and suggest that while it makes sense to speak of accountability in the first case, it does not in the second.<sup>24</sup> This is because, as they put it, "even the best teaching cannot by itself bring about learning."<sup>25</sup> They conclude their article, in part, by calling for a more realistic understanding of the teacher's role in learning:

it is clearly time to quell the clamour for holding teachers mainly responsible for all our educational ills. Teachers should be held properly accountable. But

---

<sup>22</sup>David P. Ericson and Frederick S. Ellett, "Teacher Accountability and the Causal Theory of Teaching," Educational Theory, 37, No. 3 (Summer, 1987): 277-293.

<sup>23</sup>Ibid., 289.

<sup>24</sup>Ibid., 290.

<sup>25</sup>Ibid., 291.

we maintain that they should be held accountable only for that which is in their power to control.<sup>26</sup>

Ericson and Ellett's work is important because it provides good reasons to be skeptical of any claim which would make judgements about the quality of teaching on the basis of educational "outputs" such as test scores. It clearly shows, in particular, that the lines of responsibility for student success and failure are not at all straightforward, and that to build a system of accountability on the assumption that they are is ill-advised.

Besides being ill-advised, the practice of using test scores to measure the quality of teaching can, predictably enough, have a negative impact on classroom instruction. I have already referred, in chapter two, to Marvin Wideen's research on the impact of large-scale testing in which he cites science instructors who are concerned about the possibility of final examinations eroding the ideals of good science teaching.<sup>27</sup> He also found that "many of these teachers felt that 'the luxury of getting side-tracked' had been taken away from them and that a subtle change had taken place toward a more content oriented delivery which discouraged both student and teacher creativity."<sup>28</sup> Part of Wideen's general conclusion reads as follows:

Large-scale testing does not encourage teachers to use different approaches to teaching. It appears to have narrowed the range of instructional practices being used in grade 12 compared to grade eight and 10, and indeed it may also have narrowed the purpose of education itself.<sup>29</sup>

A second example of accountability gone wrong can be found in Frank Smith's review<sup>30</sup> of Harry F. Walcott's book Teachers vs. Technocrats.<sup>31</sup> Smith tells of the University of Oregon SPECS (School Planning, Evaluation and Communication System) project which

<sup>26</sup>Ibid., 292.

<sup>27</sup>Marvin Wideen et al., "Impact of Large Scale Testing on the Instructional Activity of Science Teachers," Paper presented at the Canadian Society for Studies in Education, 1991, 60.

<sup>28</sup>Ibid., 59.

<sup>29</sup>Ibid., 64-65.

<sup>30</sup>Frank Smith, Insult to Intelligence, (Portsmouth, New Hampshire: Heinemann, 1966), 152-154.

<sup>31</sup>Harry F. Walcott, Teachers vs. Technocrats: An Educational Innovation in Anthropological Perspective (Eugene, Oregon: University of Oregon Centre for Educational Policy and Management, 1977).

was foisted on the nearby South Lane School District with the apparent aim being "to help schools to budget, operate and evaluate their total educational system."<sup>32</sup> Accountability, in other words, writ large. This was an extremely comprehensive project which had teachers specifying both general and performance objectives as well as checking and reporting on five levels of mastery and three levels of effort. As Smith explains:

Everything had to be "coded." There was a mass of materials, procedures, and worksheets, so verbose, according to one teacher, that "even an elephant could understand it - but it doesn't say anything."<sup>33</sup>

The teachers did not take too kindly to the program. They resented, in particular, the amount of effort they were expected to put into a process which they regarded as seriously flawed. Some of their comments: "If we did all the pretesting we're supposed to do [when the children arrived at school in September], we wouldn't finish until Christmas." "It's the Mickey Mouse in triplicate that gets me." "This business of testing ... junk, junk, junk. I just throw a lot of it into the wastebasket."<sup>34</sup> They complained, further, that administrators were the primary beneficiaries of the program, not the students.<sup>35</sup> Commenting, in retrospect, on the net effect of the project, University of Oregon administrator Max G. Abbott conceded that, in this case at least, "Accountability stifled spontaneity and creativity."<sup>36</sup> More to the point, I think, is the fact that, in this case, accountability took precious *time* away from the business of educating students.

What we are faced with once again, therefore, is the possibility that, far from accountability systems *enhancing* the quality of classroom instruction, they can actually contribute to its deterioration. Insofar as teachers are compelled to teach to tests which seem to serve administrative purposes far more than pedagogical ones, instruction in the classroom will

---

<sup>32</sup>Smith, Insult to Intelligence, 153.

<sup>33</sup>Ibid.

<sup>34</sup>Ibid.

<sup>35</sup>Ibid., 153-154.

<sup>36</sup>Ibid., 154.

necessarily be compromised. It is, again, ironic that a system defended as a mechanism to ensure that we are getting value for our teaching dollar could at the same time be a system which implicitly devalues quality instruction.

### Conclusion

In this chapter I have tried to show that, in general, it is peculiar to speak of "accountability" in the context of trying to educate people. Where accountability seems to imply tangible indicators of performance, clear lines of responsibility, and an uncontested relationship of authority, the business of educating people seems to involve none of these features. To proceed nonetheless to make judgements about the quality of schooling on the basis of educational outputs (motivated, perhaps, by a free market analogy of school improvement) is to risk the danger of classificatory assessment in general; that is, the danger that the enterprise of getting a *measure* of quality can end up distorting the very nature of the educative project.

This general point becomes concrete when we look at the actual uses of testing, marking and grading as mechanisms of accountability. Building on the previous suggestion that systems of testing, marking and grading tend to promote curricular narrowing, fragmentation and trivialization, the specific concern here has been that it is precisely because these practices have come to be taken seriously as mechanisms of accountability that both "learning" and "teaching" have come to be guided, and in fact virtually defined, by them. Learning, in other words, has come to mean little more than getting correct answers on tests, while teaching seems essentially to be a matter of doing what needs to be done to improve student scores. The irony of the situation is that instead of being a force which would safeguard quality schooling, it is likely that both the accountability movement in general, and the use of testing, marking and grading in particular, contribute significantly to the deterioration of the educative undertaking.

## CHAPTER FIVE

### THE SELECTION DEFENCE OF TESTING, MARKING & GRADING

A fourth possible defence of our current use of tests, marks and grades is that they serve as an effective mechanism of *selection*, which is to say they identify the most capable within a student population and in so doing thereby make it possible for these students to proceed (if they wish) to the kinds of positions in society to which they are most suited. The basic idea, crudely put, is that testing, marking and grading allows us to identify who is "smartest" and therefore who most deserves immediate status and subsequent rewards. This business of selection is thought to be defensible, first, because the criterion used to make distinctions is student *achievement* (rather than, say, social class or ethnic background), and, second, because directing the brightest and the best to positions which require skill and intelligence is in *everyone's* best interest. Those who would defend the use of tests, marks and grades as a mechanism of selection, then, think that the business of identifying and rewarding academic achievement is a perfectly legitimate practice within schools.

Although the selection defence might appear fairly straightforward, it nonetheless carries with it certain assumptions and implications which need to be made explicit. In the first place it is assumed that achievement scores really do represent the extent to which a person is capable in a given area. High mathematics scores, for example, are taken to indicate considerable proficiency in mathematics. Academic achievement scores, that is, are thought to be a reliable indicator of *competence*. Moreover, they reveal not only the extent to which an individual understands an area, but also (perhaps more importantly) the extent to which an individual understands in relation to his or her peers.

Quite apart from the issue of making judgements of competence, it is also assumed that academic achievement scores are a relevant criterion with which to bestow praise, status and rewards on individuals. People who do well in schools, in other words, are to be regarded as more valuable than those who do poorly. When we use tests, marks and grades to select the



'brightest and the best,' we mean not only to identify competence, but also to indicate where respect is due.

Finally, it should be appreciated that the term 'selection' can be misleading. The enterprise is not to identify only the most gifted; it is instead to categorize the entire student population. This should not be surprising, for if the object is to identify the 'best' - and the best in rank order, no less - then it seems necessary to likewise identify the worst, (and everything in between). In the context of our current practices, that is, *selection* has come to mean widespread *classification*. Further, it needs to be noted that the use of testing, marking and grading as a mechanism of selection involves not only *identifying* student competence, but also, in some cases, *directing* an individual's educational future. Students with low physics grades, for example, are sometimes either 'advised' or explicitly restricted from taking higher level physics courses. The *general* imperative to select the brightest and the best, therefore, can sometimes have the effect of forcing students to concentrate on developing their strengths, rather than addressing that which they find intellectually puzzling.

To defend testing, marking and grading as mechanisms of selection, then, is to assume that academic achievement scores are a reliable measure of competence; that praising and rewarding individuals on the basis of these scores is legitimate; and that the business of identifying the brightest and the 'dullest' is both defensible and appropriate. As a way to concentrate on the latter two propositions, let us concede the first. Let us assume, in particular, that our assessment mechanisms *can*, as a matter of fact, reliably identify competence in whatever area examined. Let us assume, in other words, that academic achievement scores really *do* represent achievement. What we shall be interested in for the remainder of this chapter is the legitimacy of using achievement scores as a criterion with which to praise and reward individuals, and the defensibility and appropriateness of the very business of selection in general.

### The Problem of Desert Based on Achievement

The first objection that can be made against the use of testing, marking and grading as a mechanism of selection is that it promotes the fiction that students actually *deserve* the status and rewards which come to them as a function of their academic success, i.e., as a function of their academic achievement. The standard operating procedure, within schools, is to commend, congratulate, and confer high status on those students who "do well," and then, upon graduation, to reward such students with a broad range of educational and occupational opportunities.<sup>1</sup> Students who do not do well in the classroom, on the other hand, are often marginalized and sometimes even ridiculed; their subsequent opportunities at graduation, moreover, are typically less broad. The message in all of this is that students with high levels of academic achievement are to be respected and regarded as worthy of their status and subsequent rewards, while students who do not do well in school are to be scorned and advised to be thankful for whatever opportunities they get.

The problem with this, of course, is that academic achievement is *not* an appropriate criterion to use in deciding whether or not someone deserves the respect and opportunities they receive within school and beyond. The reason is that academic achievement scores come about as the result of a variety of factors, many over which an individual has no control. A student might be academically successful, for example, because he or she has a strong native ability in the subject areas valued in school, or because he or she has access to superior educational resources. Having the advantage of either of these does not entitle a teacher or a student (or anyone for that matter) to then claim that the student *deserves* the benefits which follow.

The problem with trying to suggest that one deserves the rewards which come to him or her as a result of *native ability*, for example, is that native abilities come to us quite by chance. John Rawls points out how peculiar it is to say that someone *deserves* something on the basis

---

<sup>1</sup>I have in mind, in particular, the opportunity to proceed to university, the completion of which provides the opportunity to make more, and better, occupational choices.

of attributes or capabilities they hold quite by accident.<sup>2</sup> The possession of such capabilities and attributes is, in his words, arbitrary from a moral point of view. Writing in reference to the larger question of distributive justice, he neatly summarizes the moral dilemma that would still remain even if the principle of liberal equality - i.e. the principle that all persons should be given the same fair chance to compete - were enforced:

Even if it works to perfection in eliminating the influence of social contingencies, it still permits the distribution of wealth and income to be determined by the natural distribution of abilities and talents. Within the limits allowed by the background arrangements, distributive shares are decided by the outcome of the natural lottery; and this outcome is arbitrary from a moral perspective. There is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune.<sup>3</sup>

The corollary in the case of schooling is that there is no more reason to say that students deserve the rewards they get as a function of their native ability than there is to say that they deserve what they get as a result of their family background or ethnicity. The easiest way to see this is to ask how *fair* it is for one student, who doesn't work hard, yet who manages to get high marks, to be deemed worthy of receiving both status and further educational and occupational opportunity while a second student who works very hard, but who nonetheless has a lower academic standing is not deemed worthy of receiving such rewards.<sup>4</sup>

The problem, likewise, with trying to claim that students deserve the status and opportunities they receive as a function of their achievement scores when it is clear that such scores can be enhanced by access to superior educational resources, is that, again, such access is often not a matter of the student's own doing. The fact that some students just happen to be born into families which value, (and have the resources to appreciate), academic knowledge,

---

<sup>2</sup>John Rawls, A Theory of Justice (Cambridge: Harvard University Press, 1971), 104.

<sup>3</sup>Ibid., 73-74.

<sup>4</sup>The question of how to bestow rewards is, of course, is an issue about which classroom teachers agonize on a day-to-day basis. Some hold rigidly to the view that performance is the only criterion which matters, others measure success (and award credit) almost exclusively in terms of effort, while the majority of teachers probably employ some non-articulated combination of both positions.

or that others students are fortunate enough to run across brilliant teachers or academically-engaging peers, is not reason enough to claim that these students therefore *deserve* the opportunities that are afforded them as a result of these encounters.

It might be argued against all of this that achievement scores come about primarily on the basis of student *effort*, and therefore that when high scores are achieved, students *do* deserve the status and opportunities which follow. The moral assumption implicit here is that individual effort *is* an appropriate criterion with which to bestow praise and subsequent rewards. This assumption conforms to our intuition, I think, for we do typically hold that, insofar as a person works hard, he or she deserves the benefits which follow.<sup>5</sup> The case of "working hard" to generate a high level of student achievement, however, is problematical for a number of reasons. In the first place, it is not likely that what a person *achieves* (in school, or otherwise) is achieved *solely* on the basis of his or her effort, and *in no relation* to his or her native ability. If I were to paint a masterpiece, for example, it may be that *some* of my accomplishment could be attributed to my "efforts" (i.e. my determination to study the great masters, and to try to replicate their styles), but it is unlikely that *all* of what I had accomplished could be attributed to my "hard work." As Rawls implies, native ability would seem to be a core determiner of what people achieve after all. If this is true, it means that no measure of achievement can ever be a measure of effort in and of itself; it can, at best, represent some mixture of effort and native ability, (and presumably many other things). Second, it is not clear, in any event, what part a student's "efforts" play in his or her understanding something, or in his or her getting on the inside of a practice. It may be, in fact, that "effort" is a relatively unimportant variable with regard to these processes. To demonstrate persistence in one's studies while tackling a new intellectual endeavour might be roughly similar to resolving to be unfailingly punctual for a job that one nonetheless does poorly. Both persistence and punctuality, in other words, might be regarded as admirable traits

---

<sup>5</sup>One important exception, of course, is the case of the person who works hard at activities we would normally consider immoral: bank robbery, murder, etc. We are not inclined, that is, to say that a bank robber deserves to keep stolen money simply because it took a lot of effort to get.

in and of themselves, but may have little or nothing to do with one's performance of the central task at hand. The problems, then, are twofold: effort may have little to do with scholastic achievement, and even if it does, it is unlikely to be the *sole* determinant of academic success.

The point, then, is that academic achievement scores come about as the result of a variety of factors, many over which a student has no control. To promote the idea, therefore, that students *deserve* to be praised and rewarded (or punished) for their academic achievements is objectionable on at least three counts. In the first place it offends an outside observer's basic intuitions about fair treatment. It is morally upsetting, that is, to see students rewarded or ridiculed on the basis of outcomes over which they have little control. Second, such a system demonstrates to students themselves that principles of justice and fairness are not to be taken seriously in schools. Praise, status, and rewards go to those who are capable and lucky; those who are incapable and unlucky must be content with what is left over. Third, and even worse, students may learn to regard this state of affairs as legitimate; they may come to believe, that is, that individuals *do* deserve that which comes to them as a function of their academic achievement. They may learn to adopt, in other words, a view of desert which is fundamentally irrational.

These objections can perhaps be made clearer by way of an analogy. Imagine that a teacher decides to play a "game" with ten students.<sup>6</sup> The teacher has built a platform having six or seven levels on each of which sits a chair with a seat belt. There are thirteen or fourteen spaces on the platform, i.e. some of the levels are the same. The teacher first randomly assigns the ten students to ten chairs. To each student she then randomly distributes an envelope having a score inside. On the basis of the scores, students are either to move up or down, or to remain where they are. After the students have been resettled, the teacher instructs them to put on their seat belts. The seat belts have been designed in such a way that, once on, they are almost impossible to get out of. Only those who are very strong and agile will be able to break

---

<sup>6</sup>For the sake of this discussion, we will assume the teacher is female.

free of them. Once the students are buckled in, the teacher then declares that the "game" is over. She then informs the students, in complete seriousness, that where they sit on the platform determines what sort of praise and respect they will get for the rest of the school year, and what sort of rewards and opportunities they will receive upon graduation. Students sitting on the uppermost levels will receive much praise and many rewards, while those on the bottom will be denigrated and given almost nothing. Seeing that the teacher is serious, those students at the bottom, in particular, start to squirm. Two manage to break free of their seat belts and take up better positions on the vacant levels. The rest either cannot, or will not, move. The teacher proceeds to praise and discourage, reward and punish according to the students' placement on the platform. The students have no choice but to take what comes to them.

Now it is clear that any outside observer would find this situation morally outrageous. Given that students take up positions on the platform by way of criteria which are so completely arbitrary, it is nothing less than unconscionable to then bestow praise, status and rewards on the basis of these placements. It is also clear that the students themselves would likely find their situation intolerable. It is likely, in fact, that many of them would openly rebel. And finally, it should be clear that were the teacher (or any student) to attempt to present this state of affairs as legitimate, he or she would rightly be accused of attempting to defend a proposition which is patently irrational.

If this is clear for the analogy, so too should it be clear with regard to our current state of affairs. If we take platform levels to be representative of native ability, and envelopes with scores to be representative of access to educational resources or luck, and the ability to break free of seat belts to be representative of effort, then we should be able to see that our current system of testing, marking and grading - which implies that individuals *deserve* that which comes to them as a function of academic achievement - is no more defensible than our

imaginary teacher's game of musical chairs.<sup>7</sup> We should be able to see, as well, that anyone who thinks it *is* defensible is advocating a position which is fundamentally irrational.

The tragedy here is that it is likely many teachers and students *have* become convinced that this game is appropriate. It is probable, for example, that many teachers see no problem in giving the impression to students who do poorly that they should feel guilty and ashamed of their performance. It is likely the case, as well, that many students have come to believe that they actually deserve the denigration which comes to them as a function of their academic abilities. It is likely, moreover, that both teachers and students have learned to adopt this peculiar view of the nature of desert as a result of participating in a system which makes it seem legitimate. All of this is particularly grievous, of course, when promoted within the context of an institution which is supposed to be devoted to the development of rationality. To promote within the context of a school system, that is, the idea that people *deserve* that which comes to them as a function of their academic achievement scores, is to demonstrate rather directly to students just how irrational adults can be.

Now it might be objected that it is not the *system of assessment* which teaches or promotes the idea that people deserve that which comes to them as a result of their achievement, but instead only the confusion of those who would misinterpret the significance of assessment results. The argument might be, in other words, that the fault lies not with the system of testing, marking and grading itself, but rather with the way in which people choose to understand or interpret the results of that system.

The force of this objection might be more fully appreciated by considering the difference between the following two phrases:

You are very bright, and you've worked very hard; you *deserve* to go to university.

---

<sup>7</sup>It should also be clear that even though it may be possible for *some* students to "break free of their seat belts" (i.e., advance on the basis of effort), this is not reason enough to justify the system as a whole. The fact that all initial placements are arbitrary, and that most students cannot get out of their chairs indicates that the entire system is fundamentally flawed.

You seem to have some ability in history and politics, and you seem willing to expand your understanding in these areas; it is *appropriate* that you investigate political science courses at university.

In the first phrase, (which I shall call a 'judgement of desert' phrase), the implication seems to be that the person has a number of virtuous qualities (intelligence, perhaps, being one), that these qualities were somehow cultivated by the person, and that therefore, because of all of this, the person is *worthy* of subsequent rewards. In the second phrase, (which I shall call a 'judgement of competence' phrase), the point of the message is to comment on the apparent interest and abilities of a person with regard to a specific subject area, and to say that given that interest and ability it *makes sense* to pursue further studies in this area. In the second phrase there is, as well, no implication that the person is a particularly virtuous character, nor that the person came to have his or her ability and interest as the result of a lot of hard work. The point of the objection, then, is to say that our current system of testing, marking and grading, in particular, does not, in itself, promote or issue judgements of desert. The only judgements that our current practices of testing, marking and grading do promote and issue, so the argument would go, are judgements of competence. If people go ahead and infer judgements of desert from judgements of competence, then that is the fault of the people making the inference, not the fault of the assessment system itself.

Against this, however, I think it is clear that judgements of desert are *meant* to be read off our current assessment practices. The proof for this is to be found in the *attendant* practices of assessment: that is, in the multitude of ways in which teachers and educational administrators communicate to students that people who do well at school are to be valued, while those who do poorly are to be regarded as second rate. This message gets communicated, for example, when students with high achievement scores are given special privileges or are assigned to the 'Dean's List' or 'Honor Roll', while students with low scores have their privileges removed, and are referred to in unflattering terms. It likewise gets communicated when successful students are encouraged to feel 'proud' about their academic results, while unsuccessful students are expected to feel 'ashamed'. It is not, therefore, that



people mistakenly infer something they shouldn't from our current assessment system; it is rather the case that they are hearing a message which is all too loud and clear.

### The Problem of Selection Within Schools

Let us assume that someone either finds the previous argument unsatisfactory, or that he or she accepts the previous argument, but offers up an assessment system which does not promote the idea that people deserve the rewards they receive as a function of their academic achievement. Let us assume, further, that this person still wants to defend the general undertaking of selection within schools on the grounds that it is both practical and defensible to sort people on the basis of achievement. It is practical, so the argument would go, because, given scarce resources, it makes good sense to direct students into studies at which they are likely to excel. Such tracking represents a good use of human resources. The general undertaking of selection is defensible, moreover, in that it is in *everyone's* interest that only the most capable come to assume positions in society which require skill and intelligence. *Everyone* benefits, that is, when the doctors, engineers, and other such professionals are competent at what they do. This latter justification can even be made in reference to a Rawlsian conception of justice in that the allocations implicit within it (i.e., subsequent educational opportunities distributed unequally in accordance with native ability), conform to his difference principle; that is, they are thought to enhance the condition of the least well off.<sup>8</sup> Provided we can circumvent the problem of desert, the *general* undertaking of selection still has much to recommend it.

Walter Feinberg, however, takes issue with at least one part of this argument. He doubts, in particular, that the unequal distribution of opportunities does, in fact, bring about a better life for the least well off.

If "the widest social good" is taken to mean not only the good of the people who have the financial resources to take advantage of the level of skills and social resources available but also those of lesser means then the argument is questionable. .... Many of our most gifted engineers spend their time designing super-speed highways, comfortable automobiles, and safe jet-airplanes so that

---

<sup>8</sup>Rawls, Theory of Justice, 75-83.

middle-class and wealthy people can travel from coast to coast in quiet, safety, and convenience while many poor people consider themselves fortunate if there is a bus available to take them to work. Many of our most competent doctors are engaged in medical practice and research that benefits the rich disproportionately to their numbers, while overseas medical research sponsored by American institutions often spends the greatest amount of resources examining and researching diseases that the few rich can be expected to die from - heart disease, or ulcers - while allocating a much smaller percentage of their resources to the diseases that poor people die from, such as dysentery.<sup>9</sup>

Unfortunately, Feinberg's critique does not work. His mistake is to confuse the system which decides *who gets to be a professional* with the system which directs *what professional do* once they get their credentials. It is one thing, that is, to allow people to become doctors, lawyers, and engineers as a function of their academic achievements; it is quite another to use the promise of status and money to entice doctors to work on wonder drugs, and engineers to build airplanes. The fact that the latter may be objectionable does not, in itself, undermine the justification of the former.<sup>10</sup>

There are, however, other sorts of critiques which might be made against the general enterprise of selection. In the first place it is not clear that Rawls himself would be committed to defend the position above. A central proposition in his work, after all, is that self-respect (or self-esteem) is to be regarded as a primary social value, and as such must be distributed according to the constraints of the difference principle.<sup>11</sup> Self-esteem, that is, must be distributed equally unless unequal distribution of self-esteem is to the advantage of the least well off. Now one obvious feature of an assessment system which puts some people at the top and others at the bottom is that self-esteem is *not* equally distributed. Some individuals have their self-esteem enhanced by the receipt of high marks, while others have it eroded as a function of getting poor academic results. What should also be obvious is that this state of affairs does *not* improve the position of the least well off - insofar, that is, as the primary good of self-esteem is concerned. It is assumed, in other words, that the average self-esteem of

---

<sup>9</sup>Walter Feinberg, Reason & Rhetoric: The Intellectual Foundations of 20th Century Liberal Educational Policy (New York: John Wiley & Sons, 1975), 276.

<sup>10</sup>I am indebted to Jerrold Coombs for making this counter-objection clear.

<sup>11</sup>Rawls, Theory of Justice, 62.

individuals at the bottom of a differential assessment scale would be significantly lower than the average self-esteem of individuals who participated within an educational system where no such differential assessments were made. It is in this sense, then, that Rawls need not defend any assessment practices which systematically erode the self-esteem of those who are subjected to them.

But there is another, perhaps even more powerful, objection that can be raised against the general undertaking of selection. It begins by asking the question of whether or not we are justified in selecting people without their explicit consent. One thing we need to remember about public schools is that they are *compulsory* institutions. Students, as such, have no choice about whether or not they will have their achievements recorded and reported. Although few people would likely protest against having their successes made public, it is doubtful that many individuals would be as eager to have their failures permanently recorded. The real issue, then, is whether or not there are sufficient grounds to subject students, without their consent, to a system of assessment which publicly identifies their competencies *and incompetencies*.<sup>12</sup> Notice that this is primarily a moral question for it seeks a justification to treat people other than they would most likely choose to be treated.

The first place to search for an answer is to look to the standard justifications that are given to compel someone to do something quite apart from his or her explicit consent. The most obvious case that comes to mind is safety: that is, we routinely restrict people - particularly children - from doing things which are likely to harm them. But this justification seems inapplicable. There is little evidence that the removal of a compulsory selection system within schools is liable to cause individuals harm. There are, in fact, good reasons to believe just the opposite: that the maintenance of an explicit selection system within schools has the potential to do real damage. Such a system, for example, can put considerable stress on

---

<sup>12</sup>I will not address the more general issue of whether or not compulsory schooling is itself justified. For a particularly insightful treatment of this issue - especially the issue of the justificatory grounds for interference with the liberty of children - see Roland Case, "Pulling the Plug on Appeals to Irrationality, Immaturity and Expediency," Proceedings of the Philosophy of Education Society (1985): 445-454.

students. Gilbert Ryle has pointed, for example, to the kinds of psychic discomfort which can occur around examination time, including, in his words, "excessive sleepiness, which represents a kind of 'frozen rabbit' opting out of horrid reality."<sup>13</sup> University researchers John Anderson and David Bateson similarly cite increased student stress as one of the major consequences of the British Columbia Provincial Examination program.<sup>14</sup> But this tension is not confined, apparently, to the higher levels of schooling. Amity Buxton, of the Oakland, California, Teacher Shelter has written of "children in kindergarten who were crying and wetting their pants when forced to take a standardized test used to determine eligibility for federal funds." <sup>15</sup>

Apart from the stress of anticipation, there is also the damage done when students receive assessment results. As mentioned in chapter three, Bloom, Madaus and Hastings have suggested, not unreasonably, that 10 or 12 years of negative classifications bestowed upon a student are likely to have a significant detrimental effect.<sup>16</sup> Similarly, school principal Arthur Laughland has written that "testing can mark innocent children with stigmata for the rest of their lives, ... [Testing] is not fun for the child who freezes; it is not fun for the child who struggles for small successes to be labeled a 'failure' time after time; and it is not fun for the child whose one bad day yields results that must be lived with until the next testing time." <sup>17</sup> For some students this labelling is unbearable - severe depression and even suicide can result. Insofar, then, as the basic well-being of individuals is concerned, it would appear that

---

<sup>13</sup>Gilbert Ryle, Student Casualties (Hammondsworth: Penguin Books, 1969), 100; quoted in J.C. Matthews, Examinations (London: George Allen & Unwin, 1985), 88.

<sup>14</sup>John Anderson et al., The Impact of Provincial Examinations on Education in British Columbia (Victoria: British Columbia Ministry of Education, 1990), 62, 66, 79.

<sup>15</sup>Fall 1979 NCT Conference Summary, 43; quoted in Andrew Strenio, The Testing Trap (New York: Rawson, Wade Publishers, 1981), 121.

<sup>16</sup>Benjamin Bloom, George Madaus and J. Thomas Hastings, Evaluation to Improve Learning (New York: McGraw-Hill, 1981), 4.

<sup>17</sup>Arthur Laughland, "Two Principals Look at Standardized Tests," in Paul Houts, ed., The Myth of Measurability (New York: Hart Publishing, 1977), 332.; quoted in Andrew Strenio, The Testing Trap (New York: Rawson, Wade, Publishers, 1981), 121.

subjecting students to an assessment system designed to identify the brightest and the best is, at the best, questionable and, at the worst, unconscionable.

A second standard justification for compelling someone to do something against his or her will, or without his or her consent, is the clear demonstration that it is in the individual's *own interest* that the thing be done. In the case of an assessment system which emphasizes selection, the basic intuition is that we do students a favour if we steer them clear of (or restrict them from) areas to which they are not suited. We save these students, so the argument would go, from exasperation, lost time, and failure. Against this proposition at least three counter-arguments can be forwarded. First, as already implied, it is hardly in an *individual's* best interest, other things being equal, to be labelled as incompetent. Second, it is at least questionable whether we have the technical ability to predict with confidence the future success of *individual* students. While there may be some - although not much<sup>18</sup> - reason to be confident about making predictions for *groups* of students, there is little cause to have any such confidence with regard to predictions made about *individuals*. There are many reasons why a student may do poorly in a course which have nothing to do with his or her ability to grasp the central concepts presented. It may be, for example, that the student dislikes the teacher; or that he or she is preoccupied with extra-curricular concerns; or even something as simple as the fact that the person can't see the blackboard. There are all sorts of reasons, then, why a student might eventually come to demonstrate a level of competence which far surpasses that of a previous assessment. He or she might, for example, have the luck of coming across a gifted tutor; or have resolved his or her extra-curricular preoccupations; or even have purchased a pair of glasses. To use an *individual's* past performance, then, to make a prediction about his or her future success is morally dubious.<sup>19</sup>

---

<sup>18</sup>SAT scores, for example, are valid predictors of academic performance for only the first year of college. See, Charles V. Willie, "The Problems of Standardized Testing in a Free and Pluralistic Society," Phi Delta Kappan, (May 1985): 626-627.

<sup>19</sup>It is important to add that it is also morally questionable to make such predictions on the basis of even such professionally-constructed measures as SAT scores. Even if it is true that 80% of those with SAT scores of 450 will likely fail first year university, this is not in itself a strong enough reason to deny university entrance to an *individual* with a 450 SAT score. The reason should be obvious: we never know *which* twenty individuals, out of the imaginary one hundred, can and will succeed.

And finally, even if we were able to make valid predictions of an individual's success, two further objections can be raised. In the first place, there is something disturbingly paternalistic about not letting an individual make his or her own mistakes. To steer students in the "right direction" is, in a sense, to fail to respect them as rational and autonomous human beings. Second, this kind of paternalism is also potentially *miseducative* in the sense of denying a student the opportunity to make planning decisions based on a frank appraisal of his or her apparent strengths and weaknesses. To deny a student the opportunity to make informed decisions is ultimately to deny him or her the opportunity to act *rationally* with regard to his or her future plans.

It would appear that the standard justifications for compelling someone to do something apart from his or her consent are not applicable in this case. Subjecting students to an assessment system which publicly identifies their competencies and incompetencies doesn't seem to be a matter, in the first place, of ensuring their well-being. Nor would it appear that this system can be justified by claiming that it is in the students' best interest that it be maintained. There do *not* seem to be adequate grounds, then, for compelling students to run the gauntlet of an assessment system which publicly identifies their strengths and weaknesses.

None of this is to suggest, by the way, that the state has no right whatsoever in identifying, for the purpose of resource allocation, which people seem to be the most capable in what particular areas; it is only to suggest that such an emphasis on selection should not occur *within schools*, and in particular within schools which are *compulsory*. This leaves open the possibility that selection might be appropriate within educational institutions which people *voluntarily* attend - say, for example, colleges and universities. Patricia Broadfoot, as part of her critique of formal assessment procedures, intimates as much in the following:

Certainly the liberation of the compulsory schooling stage at least from formal assessment would enable a much sharper distinction to be drawn between general mass education - the old elementary education in many ways or the 'basic' school in Scandinavia - and the more specialized, voluntary courses embarked upon after this stage.<sup>20</sup>

---

<sup>20</sup>Patricia Broadfoot, Assessment, Schools and Society (London: Methuen, 1979), 127.

It is important to distinguish, then, between institutions which enforce compulsory attendance and those which people attend voluntarily. While it is difficult to justify the activity of selection within compulsory institutions, it should be less difficult to do so in the context of a college or university. Given that students attend these latter institutions more or less of their own free will, they are in a sense agreeing to subject themselves to the selection system implicit within them.

In addition to the problem of selection mechanism bringing about low self-esteem, and the moral dubiousness of publicly identifying incompetencies without an individual's consent, a third objection which can be raised against the general undertaking of selection is that it can undermine, and even contradict, part of the very purpose of educating students within schools. This can happen on two levels. In the first place, where processes of selection get translated into practices of classification (and it is hard to see how they would not do so), the enterprise is subject to many of the same criticisms reviewed in the preceding three chapters. Selecting 'the brightest and the best,' in other words, can, paradoxically, create an instructional regime characterized by curricular narrowing, fragmentation and trivialization. Insofar as our purpose in *educating* students is to get them on the insides of practices, the introduction of mechanisms of selection can detract from that aim. On this first level, then, the very *substance* of education is endangered.

On a second level, it is likely that the general enterprise of selection contradicts some of our most fundamental beliefs about what we think schools are for. To use testing, marking and grading as a mechanism of selection is to create categories of exclusion on the basis of student competence. These categories are created both within schools (when, for example, students are directed away from studies in which they do not excel), and after school, when low achievement scores are used to restrict students from entering various career paths. Part of what we seem to have in mind by having students attend school, however, is that *all* students be introduced and exposed to a *broad* range of knowledge - introduced and exposed, that is, to

a broad range of practices.<sup>21</sup> The fact that a student is not competent or proficient within a particular area is not, therefore, reason enough to direct a student away from a study. In fact, insofar as our goal is to *educate*, this is probably a good reason to encourage the student to pursue the study with that much more vigor.

These points can best be captured, I think, by way of another analogy. Consider the difference between team try-outs and practices in competitive as opposed to recreational baseball leagues. In competitive leagues, the point of the try-out and the practice is to find out which players are the best, and then to enhance and develop their particular talent. Less able players get cut, while those who stay concentrate on enhancing their area of expertise. In a recreational league, on the other hand, the idea of a "try-out" is an anathema: the whole point of a recreational league is to ensure that anyone who is interested gets a chance to play. Indeed it would be considered bad form, and irrelevant to the point of the enterprise, to kick someone off a team simply because he or she wasn't very good at pitching or batting. The point of the enterprise, rather, is better served by giving that person some exposure to pitching and batting so that he or she might, if not excel at these activities, then at least not be intimidated by them. The emphasis, in other words, would be on *inclusion* rather than exclusion.<sup>22</sup>

Now it seems to me that much of what we have in mind as the purpose of schooling is better characterized by way of the recreational baseball analogy. It seems to me, that is, that one of the most important reasons for sending children to school is the expectation that schools

---

<sup>21</sup>This is particularly the case when we see the purpose of schooling as either that of preparing people for responsible citizenship, or of providing a liberal education to students. See, Jerrold Coombs, "Accessibility" [unpublished essay, 1988].

<sup>22</sup>It is important to add, by the way, that in a recreational league there is certainly no disregard for the rules, strategies, and skills which make up the game of baseball. It is, after all, the understanding of these rules, the skillful application of these strategies, and the eventual acquisition of these skills which define the nature of the activity undertaken. It is not, therefore, that recreational baseball players have less appreciation for the game than professional baseball players; it is rather that recreational players find the game so inherently worthwhile that they want anyone who is interested to benefit from participating in it. Similarly, it is not that recreational baseball is but a poor imitation of professional baseball; it is rather that recreational baseball and professional baseball are the same game played with two different sets of intentions. It is also important to note that there is nothing in principle within the recreational baseball analogy which would restrict a player from becoming very good at one particular aspect of the game. The only requirement implicit within the system is that a "player" not devote all his or her attention to that one area, and likewise that a "coach" not concentrate on that aspect to the exclusion of others.



will be places where students - all students - are exposed to the unfamiliar, and in being so exposed, will eventually come to broaden their horizons. The problem with the general enterprise of selection, then, is that in creating categories of exclusion it undermines and contradicts this most fundamental aim. Insofar as we are serious, therefore, about getting students on the insides of practices, the time is perhaps ripe to ask whether or not the general enterprise of selection is appropriate within schools after all.

### Conclusion

Two arguments have been advanced against the selection defence of testing, marking and grading. In the first place it has been pointed out that the use of testing, marking and grading as a mechanism of selection tends to promote the idea that people actually deserve the praise, status, and rewards which come to them as a function of their academic achievement scores. Since such a view is patently irrational (given that academic achievement scores come about as the result of factors over which an individual has little control), its promotion within schools was regarded as particularly grievous, seeing as part of what we want to do in schools (insofar as we want to *educate* people) is to enhance, rather than impede the development of rationality. In the second place, the very defensibility and appropriateness of the general enterprise of selection was called into question. It was argued, first, that the selection process can threaten self-esteem which is a primary good; second, that the business of publicly identifying an individual's incompetencies without his or her consent is morally hazardous; and third, that the general enterprise of selection undermines and contradicts some of what we have in mind by educating students within schools. For all these reasons, then, the selection defence of testing, marking and grading was found wanting.

## CHAPTER SIX

### CONCLUSION

Most people who have taught within a formalized educational system become uneasy, sooner or later, with the business of testing, marking and grading students. They know - perhaps only intuitively - that to engage in these activities can be pedagogically irrelevant and, in some cases, morally objectionable. The problem, however, is that, up until now, educators have not had at their disposal the kind of critique which makes the pedagogical and moral objections clear. The standard responses to the issue (which I have called the technical and radical critiques of testing, marking and grading), have either been too narrow or too pessimistic. The technical critique is too narrow because it accepts at face value that the general enterprise of testing, marking and grading is sound, and then devotes itself almost exclusively to the building of better assessment instruments. The radical critique, although broader, is too pessimistic in the sense of implying that all mechanisms of appraisal must necessarily be instruments of social control.<sup>1</sup> What is needed, therefore, is the kind of critique which is able to raise serious questions about the nature of assessment as a whole, without at the same time giving up on education and schooling. What is needed, in other words, is an analysis which will directly address pedagogical and moral concerns while at the same time offering up a language of possibility for school reform. The purpose of this study has been to present just such a critique.

The methodology has been to survey four common defences of testing, marking and grading and out of this survey to generate an explicitly philosophical critique of our current assessment practices. It is intended, moreover, that out of the negative critique, positive policy implications might eventually emerge.

---

<sup>1</sup>For an example of just such a complaint made against a major proponent of the radical approach, see Richard J. Bates, "Educational Versus Managerial Evaluation," in Patricia Broadfoot, ed., Selection, Certification and Control (London: Falmer Press, 1984): 132-133.

I said in chapter one that the point of philosophical analysis is to ask first questions first so that second questions might be more intelligent. Within the context of this dissertation, the first questions have been as follows:

What do we have in mind by *educating* students, and by *teaching* them?

How might we analyze the concept of assessment and our current assessment practices to discern what within both of these is pedagogically defensible?

What moral considerations, if any, ought to be brought to bear on any discussion of student assessment practices? What grounds, for example, do we have for damaging the self-esteem of students, or for compelling them to have their incompetencies publicly identified and recorded?

What sorts of purposes do we have in mind in sending our children to school? How might our current assessment practices undermine or contradict those purposes?

In answer to the first of these questions I have said that to *educate* is to strive to get students on the inside of a practice, by which I meant bringing them to a point at which they both understand and in a sense care about the standards which define the undertaking. This business of trying to get students on the inside of practices, moreover, represents a general commitment to the development of rationality. To learn about the standards which define a practice, and to understand when and how to apply those standards, (and to *care* that they be properly applied), is to appreciate the general value of being able to operate within a coherent discourse. I have also suggested - or rather have concurred with Scheffler and Green in suggesting<sup>2</sup> - that to *teach* in reference to such a conception is, again, to aim at the development of rationality within students. Because we care *that* students understand the standards which define a practice, and that they understand them in a particular way - i.e. as a function, in some sense, of their own reasoning powers - teaching, on this view, is not so much a matter of getting students to replicate facts as it is a matter of getting students, as Green says, to assess what is reasonable to believe.<sup>3</sup> Getting students to assess what is

---

<sup>2</sup>see pp. 18-20 above.

<sup>3</sup>Thomas Green, *The Activities of Teaching* (New York: McGraw-Hill, 1971), 29; see p. 20 above.

reasonable to believe, moreover, is just another way of saying that teaching is a matter of helping students become rational agents.

In answer to the next question, I have suggested, first, that certain assessment practices serve some purposes better than others; second, that a distinction can be drawn between appraisals done for the purpose of enhancing learning (pedagogical assessment), and appraisals done for the purpose of ranking achievement (classificatory assessment); third, that the classificatory project, (particularly when it is understood as the imperative to get a fair measure of student achievement), tends to represent itself within assessment regimes characterized by curricular narrowing, fragmentation and trivialization; and finally, that because curricular narrowing, fragmentation and trivialization make it difficult to get students on the inside of practices, there is a danger of classificatory assessment regimes threatening the integrity of the educative project.

In answer to the third question, I have suggested that we *do* need to look *very* closely at the justification for eroding self-esteem and for publicly identifying people's incompetencies without their explicit consent. One way to do this is to review the standard arguments used to compel persons to do something against their wishes to see if they hold up in this case.

In answer to the fourth question, I have suggested that, insofar as we think the purpose of schools is to educate, we need to consider closely the extent to which our current assessment practices either impede or enhance the educative project. I have also suggested that we need to consider the extent to which we think schools are places to identify and develop "the brightest and the best" or places to broaden the horizons of all who attend them.

It is in reference to these prior philosophical considerations, then, that the specific arguments against certain practices of testing, marking and grading have been advanced. The first of these - which might be called the baseline argument - used the pedagogical/classificatory distinction to identify many of our current assessment practices as classificatory, and therefore as potentially destructive of the educative project. Insofar, that is,

as many of our current assessment efforts are aimed primarily at locating students within a hierarchy of achievement, we run the risk of narrowing, fragmenting, and trivializing the curriculum. And to narrow, fragment, and trivialize knowledge is to hold out little hope of getting students on the inside of a practice.

This same argument provided one of the reasons for suggesting that it is ironic to defend testing, marking and grading as a strategy to motivate students "to learn", and likewise one of the reasons to be skeptical of using any system of testing, marking and grading as a mechanism to ensure quality schooling. With respect to the motivation issue, the problem, again, is that insofar as a system of testing, marking and grading tends to narrow, fragment and trivialize the curriculum, the kind of learning it promotes is far removed from that which might be recognized as getting students on the inside of a practice. The irony is that the system which motivates students "to learn" is also the system which can bring about the deterioration of learning. The same sort of problem arises with regard to accountability in that, again, the system defended as a mechanism to ensure quality schooling can also be a system which, in a sense, sabotages the educative project. Because testing, marking and grading is regarded as a mechanism of accountability, teachers and students take it seriously and align their efforts accordingly. Because, however, the view of knowledge implicit within such an assessment mechanism can be so impoverished, the goal to which students and teachers aim is not necessarily that lofty.

The second argument which emerged, within the context of the analysis of motivation, was that the use of *extrinsic* rewards (i.e., marks and grades) as a strategy to educate people is highly problematic. In the first place, insofar as *the acquisition of marks and grades* becomes the aim of schooling (as opposed, say, to gaining an appreciation of goods internal to practices), then we have fallen short of the goal of *educating* students. Second, even if the giving out of marks and grades is meant only to be a *strategy* to *eventually* have students come to appreciate goods internal to practices, this is still problematic for two reasons: it can bring about an instructional environment which is detrimental to learning, and it sets in place a

less-than-ideal incentive system which can be difficult, if not impossible, to dismantle. Finally, the use of marks and grades to coerce students to believe certain things (or to say they believe certain things), and to behave in particular ways is to betray a fundamental tenet of the educative project, i.e., that of striving to have students think and act on the basis of reasons which are relevant to the subject matter at hand.

The third argument which emerged, this time within the context of the discussion on accountability, was that the use of marks and grades to make judgements about the extent to which students are learning and teachers are teaching is, again, not without its difficulties. In the first place it is somewhat peculiar to use the term "accountability" within the context of attempting to help people to learn. As suggested in chapter four, where "accountability" seems to imply a situation in which there is a tangible indicator of success, an unambiguous source of responsibility, and a clear relationship of authority, none of these conditions seem directly applicable in the case of attempting to get students on the inside of practices. Second, insofar as individual tests, and assessment systems as a total regimes are norm-referenced, the final judgement is not so much a definitive statement of what students know, as it is a precise statement of where students fit in relation to their peers. Third, and in any case, final results are often manipulated for political purposes - to get, for example, a politically acceptable number of failures. We should be cautious, therefore, about reading into them very much of pedagogical import. Last, and even more fundamental, is the fact that relying on student achievement scores to make judgements about the quality of instruction presupposes a causal theory of teaching, which itself is in need of precise clarification. On this point I agree with Ericson and Ellett's analysis which suggests that teachers ought to be held accountable for only those features of a learning situation over which they have control.

The fourth argument, presented in chapter five, represented an attempt to analyze the merits of *selection* quite apart from the hazards associated with a classificatory regime (i.e. the hazards identified in the baseline argument). Two points emerged. First, promoting the idea that people deserve subsequent opportunities as a result of achievement scores amounts to the

promotion of an irrational conception of desert. Since academic achievement scores derive, in part, from factors over which a student has no control (the most significant, perhaps, being his or her native ability), it is downright nonsense to promote the fiction that people deserve the benefits that come to them as a result of these scores. It is, as I suggested, particularly grievous to promote such a story within the context of an undertaking aimed at the development of rationality. Second, the general undertaking of selection does not seem to be particularly defensible or appropriate. There are no adequate reasons, that is, for threatening the self-esteem of students, or for compelling them to have their incompetencies publicly identified and recorded. In addition, by making it difficult to deliver on the promise of providing an education (i.e., by promoting fragmentation, trivialization and curricular narrowing), and by emphasizing categories of exclusion rather than inclusion, the general undertaking of selection undermines at least some of what we have in mind by the purpose of schooling.

Negative arguments are, in a sense, easy enough to make. Any teacher who has worked within a formal educational system will probably be familiar with, if not all, then at least most of the issues raised here. The real challenge is to find a way to get beyond the current presuppositions about what constitutes appropriate assessment practices in schools. The task, in other words, is to begin to build Aronowitz's and Giroux's "language of possibility" - to begin, that is, to understand and reconceptualize "assessment" from the point of view of a serious commitment to educate. Although it is not within the scope of this dissertation to offer a complete reconceptualization of assessment, preliminary possibilities can, I think, be sketched. The way to proceed is to see what positive implications might be extracted from the negative arguments reviewed thus far.

Anthony Flew has pointed out the important truth that insofar as we are sincere about our intention to teach someone something we shall need, at some time or another, to check to see the extent to which our charge is learning what we intend.<sup>4</sup> This business of *checking to*

<sup>4</sup>Flew, "Teaching and Testing," Proceedings of the Philosophy of Education Society (1973): 210-212.

*see to what extent a student has learned* might be described as the *pedagogical core of assessment*. Insofar as we are interested in *educating* students in the sense of having them operate on the inside of practices (that is, having them understand and appreciate what constitutes good reasons within a practice), then the kind of "checking" we must do will be of a fairly sophisticated nature. We shall need, in particular, to attempt to ascertain the extent to which students understand an intellectual inquiry *as* a practice - that is, as a system of inquiry having core concepts and linked by a distinctive methodology.<sup>5</sup> What follows from this is that our assessments will have to be comprehensive - more comprehensive, in fact, than our current practices. What we're looking for, following Scheffler and Green, is not just what facts students seem to have at their disposal, but how they *hold* those facts: that is, why they believe what they believe. The kinds of assessment strategies conducive to *this* kind of inquiry are likely to be different than the ones which are traditionally respected.

The most straightforward procedure to discover a student's apparent level of *understanding* of the subject area under study is to conduct a one-to-one interview between the teacher and the pupil. It is important to notice that the kinds of questions which a teacher would ask are not to be prespecified in detail because one would never know in which direction the conversation might go. The whole point of the interview, rather, would be to see where the student's thinking leads.

A second assessment strategy for discerning a student's apparent level of understanding would be to make more use of student essays and/or major projects. What would be sought in reviewing this work would be some sort of evidence of the extent to which students are capable of linking together the disparate parts of a study into a cohesive whole. Ernest Boyer said it well when he defended the use of essays for assessment purposes as follows:

The single most important way to measure student progress is to ask them to write a serious essay on a consequential topic. And that, more than any single measure, indicates whether they can take knowledge from across disciplines, put

---

<sup>5</sup>A close examination of Paul Hirst's work on "forms of knowledge" would not be out of place here. See especially "Liberal Education and the Nature of Knowledge" in Paul Hirst, Knowledge and the Curriculum (London: Routledge & Kegan Paul, 1974.): 30-53.



it together in a coherent way, and develop persuasively an independent idea of their own.<sup>6</sup>

The thing to keep in mind about these strategies - be they interviews, or the close examination of essays or major projects - is that they would be pursued not for the purpose of ranking students, but instead for the purpose of discovering how students think so as, in turn, to decide how best to complement their understanding. To be an "educator" seems to require that one be educated (that is, be able to operate on the inside of the practice being taught), and that one have an active interest in bridging the gap between the explanatory framework of a student and the subject matter at hand. To make an "assessment" in the role of an educator, then, is essentially to probe the explanatory framework of a student. Once this framework is revealed, teachers - as educators - are then in a position to attempt to amend it as required. The *pedagogical* point of assessment, therefore, is to find out how students think so as to bring them around, eventually, to a broader, perhaps more disciplined, way of understanding.

Turning to the issue of motivation, it seems clear that to *educate* someone is, at some point, to attempt to have them appreciate goods *internal* to a practice. There is something paradoxical in our insistence on using goods external to practices as mechanisms to motivate students "to learn." Frank Smith alluded to the problem when he wrote that "Learning is its own reward. A child does not have to be bribed to learn, only to stay in a situation where learning is impossible."<sup>7</sup> The paradox, then, is that the very system we use to motivate students may so contribute to the degradation of the educational experience that *it becomes necessary* to maintain a system of subtle coercion to keep students attending to their studies.

Consider the alternative: that is, attempting to motivate students by appealing to the intrinsic value of the subject matter itself. Two advantages are immediately apparent. In the first place such an approach would force teachers to make clear (or at least to be clear about), the goods internal to a practice which might, in turn, prompt the removal of much of what is

---

<sup>6</sup>Ernest Boyer, quoted in Gerald Bracey, "Dangerous Practice," Phi Delta Kappan, (May, 1987): 683-686.

<sup>7</sup>Frank Smith, Comprehension and Learning (New York: Holt, Rinehart and Winston, 1975), 244.

trivial within contemporary instruction. Focusing on goods internal to practices, in other words, might force teachers to get to the heart of subject areas. Second, such an approach would be more conducive to the development of rationality. It would be an important advance, that is, to say to a student that he or she ought to attend to a particular study not in order to get marks, but instead because the particular study is inherently worthwhile, *and here are my reasons for thinking so*. To submit one's judgements, as Scheffler says, to the critical scrutiny and evaluation of the student is to bring that student into the arena of rational discourse. To demand, then, that motivational strategies carried out within an *educational* context concentrate on making clear those goods internal to practices would, it seems fair to say, enhance the entire undertaking.

As for the issue of accountability, a number of improvements suggest themselves. All of these, however, can be summed up in one strategic move: that is, to shift the focus from accountability to *responsibility*. What is needed, in other words, is a renewed understanding of what teaching and educating demand, and a renewed commitment to bring about the conditions which make teaching and educating possible. What teaching and educating demand, in part, is that teachers fully understand and appreciate goods internal to practices, and that they *take* the responsibility to try and convey these to students. One of the conditions necessary for making teaching and educating possible is that teachers be given the freedom to focus on this task. Any system which seeks to provide an account of the extent to which students are learning, therefore, should support this general undertaking, not impede it. By shifting the focus from accountability to responsibility, it is more likely, in my view, that this would occur.

This shift might be manifested in a number of ways. In the first place it is unlikely that teachers would be required to submit academic reports to administrators. When accountability gets de-emphasized in favour of responsibility, the onus is on the teacher to ensure that students are exposed to a worthwhile educative experience. This responsibility, moreover, is

owed to the student, not to the central administration.<sup>8</sup> Second, the kind of communication that teachers would be expected to have with parents would also be somewhat different. Parents, of course, have a right to know how their children are doing, but this does not necessarily imply that they must be given a comparative analysis which ranks their child in relation to his or her peers. The most that a teacher would be obliged to tell a parent is how well a child is doing in relation to the subject matter at hand and, perhaps, how the student seems to be getting along *in general*. Insofar as teachers have a responsibility to convey pedagogical information only, the business of giving parents a precise measure of where their child ranks in relation to a comparison population would be inappropriate. And finally, the kinds of reports which teachers would be expected to make for subsequent instructors is also open to interpretation. Given the fact that such reports can sometimes prejudice a new teacher's view of a student, and given that competent teachers can usually find out what is pedagogically important about a student during the first three weeks of classes anyway, it is not at all clear that these kinds of reports need necessarily be recommended.<sup>9</sup>

With regard to the issue of *teacher* accountability, the proposed shift toward responsibility again suggests some interesting possibilities. Rather than rely on assessment instruments which focus on student achievement scores to provide an account of the extent to which teachers are teaching, for example, it might be more appropriate to turn to the profession itself to generate this information. I have in mind, in particular, a program of in-class observation and peer review designed not so much for the purpose of sanctioning and

---

<sup>8</sup>It will be remembered that from a strictly *pedagogical* point of view it is irrelevant how students are ranked in comparison to their peers. The most that a central administration might legitimately demand of a teacher is a judgement about whether or not a student should proceed to a higher level class.

<sup>9</sup>This calls into question, then, even such recent initiatives as "learner profiles" and "records of achievement." For an example of learner profiles defined and then redescribed with contradictory purposes, see, respectively, Information Circular #432 (Victoria, B.C.: Ministry of Education) July 16, 1990; Changes in Education: A Guide for Parents (Victoria, B.C.: Ministry of Education, 1991), 18; Year 2000: A Curriculum and Assessment Framework for the Future (Victoria, B.C.: Ministry of Education, 1989), 30. For a comprehensive review of both learner profiles and records of achievement, see Patricia Broadfoot, ed. Profiles and Records of Achievement: A Review of Issues and Practices (London: Holt, Rinehart and Winston, 1986). For a critical analysis of records of achievement, see especially Andy Hargreaves, "Record Breakers" in Patricia Broadfoot, ed. Profiles and Records of Achievement: A Review of Issues and Practices (London: Holt, Rinehart and Winston, 1986).

evaluating teachers as for the purpose of initiating and sustaining a constructive dialogue about the nature of teaching, and how best to improve the practice. I assume that such a dialogue would be comprehensive in the sense of investigating not just methodological issues, but also the sociological, psychological, and philosophical dimensions of teaching. If one of the questions that an accountability system is meant to answer is "what is the current state of teaching?" then it seems to me that a serious on-going investigation of current practices is the best way to provide not only an account of what is presently happening, but also some analysis of the steps which might be taken to make improvements for the future.<sup>10</sup>

The fourth issue which can be re-examined with a view to extracting policy implications is the question of the selection function of schooling. Two imperatives seem clearly to follow from the points which were raised above. In the first place we need to get out of the business of promoting the idea that people deserve the rewards and opportunities that come to them as a function of their native abilities or their fortuitous access to superior educational resources. We need to get out of the business, that is, of ascribing general worth to individuals as the result of factors over which they have little control. And more than this, we need to abandon, within compulsory schools, the entire enterprise of publicly identifying the "smart" and the "stupid," and of implying in our every word and deed that the former are to be respected, while the latter are to be either pitied or reviled. If we are going to make appraisals of students (which we must do if we intend to educate them), then we must make them in a way which acknowledges differences in ability, without at the same time suggesting differences in personal worth. Since it is doubtful that a system of *testing, marking and grading* can do this, we need to look seriously at overhauling not only our current practices, but indeed our very conception of educational assessment.

Second, the time and energy spent on selecting the brightest and the best in schools needs to be diverted to the task of providing the best possible education for *all* students. An

---

<sup>10</sup>I assume, further, by the way, that such an arrangement would yield more useful information than, for example, Canada's recently-announced National Indicators Tests.

indication of a district's success need not be measured in terms of how many students win, for example, National Mathematics contests; success might be better judged in terms of the overall quality of the educative experience of the student population as a whole. What needs to be removed, in other words, is the emphasis on selecting out the few, at the expense of educating the many.

Both these imperatives point to the same policy implication: the elimination, at some point or another, of the practice of generating, recording and reporting, for each student, a precise measure of academic achievement. The elimination, in other words, of final grades. This, of course, presents a problem for colleges and universities: how to decide which students to admit into programs? Two strategies come to mind; one, in my view, immeasurably better than the other. We might, in the first place, allow such institutions to administer entrance examinations as a way to decide who should be admitted and who should not. The drawbacks of this policy are immediately apparent. In addition to all the technical issues that can be raised about the validity of any such test, and all the moral issues that can be raised about its inherent fairness, there is, again, the institutional problem of public schools inevitably teaching toward this examination. The only solace that might be taken in recommending such a strategy is that it ensures that most (but of course not all) social selection assessments will be deferred until after the period of compulsory schooling.

A far better strategy would be to implement an open admissions policy, with the first year of college or university being a probationary or qualifying period. Students would decide for themselves whether or not they wanted to attempt a college or university education. First year university instructors would be rigorous in their demand that students meet first year requirements in their particular field of study. The point of the qualifying year, after all, would be to decide who should proceed. The value of such a policy is that it would allow admissions personnel to decide who is most capable of benefitting from a college or university education on the basis of the most valid indicator possible - i.e. a year of college or university work. An open admissions policy would, moreover, be on better moral ground in that,

although classificatory assessments would be made, they would be done at the *voluntary* consent of the student. One additional benefit to such a policy, by the way, is that it would represent, if not exactly a guarantee, then at least a broader commitment to the principle of equality of educational opportunity.

Open admissions at colleges and universities might, admittedly, be regarded as a rather extreme solution to the problem of social selection within primary and secondary school. It seems, however, that the severity of the solution merely reflects the severity of the problem. Insofar as we sincerely want to educate students within our schools, sooner or later we shall have to face the issue of how our assessment systems can either enhance or impede that purpose. And once we begin that latter investigation then we shall likewise have to face squarely the consequences which might follow. This holds true not only for the issue of selection, but also for questions of pedagogy, motivation, and accountability.

To conclude, it seems that the way we think about assessment has come to affect the way we think about education. In assuming that the acquisition of a *practice* can be adequately caught in the successful completion of test questions, and in conceiving of teaching as simply a matter of preparing students for such tests, the tail, unfortunately, has come to wag the dog. The point of this dissertation has been to attempt to right that inversion - to *begin* with a conception of education and teaching, and to use that to reflect on the nature of assessment within the context of schooling. The object has been to critique our current practices, and in so doing to begin to reformulate how assessment might be understood were we to take education and teaching seriously. The intent, in other words, has been to lay the groundwork for the development of a new language of possibility - a language in which our conception of assessment would *follow from*, not predetermine, our conception of education and teaching. The task that now awaits, of course, is to go out and build on that foundation.

## BIBLIOGRAPHY

- Anderson, John O. et al. The Impact of Provincial Examinations on Education in British Columbia: General Report. Victoria, B.C.: B.C. Ministry of Education, 1990.
- Apple, Michael, Michael J. Subkoviak and Henry S. Lufner, eds. Educational Evaluation: Analysis and Responsibility. Berkeley: McCutchan Publishing, 1974.
- Aronowitz, Stanley and Henry Giroux. Education Under Siege. South Hadley: Bergin & Garvey, 1985.
- Austin, J.L.. Philosophical Papers. Oxford: Claredon Press, 1961.
- Bloom, Benjamin, George Madaus and J. Thomas Hastings. Evaluation to Improve Learning. New York: McGraw-Hill, 1981.
- Bracey, Gerald. "Measurement-Driven Instruction: Catchy Phrase, Dangerous Practice." Phi Delta Kappan (May 1987): 683-686.
- Broadfoot, Patricia. Assessment, Schools and Society. London: Methuen, 1979.
- Broadfoot, Patricia, ed. Selection, Certification and Control. London: Falmer Press, 1984.
- Broadfoot, Patricia, ed. Profiles and Records of Achievement: A Review of Issues and Practices. London: Holt, Rinehart and Winston, 1986.
- Case, Roland. "Pulling the Plug on Appeals to Irrationality, Immaturity and Expediency." Proceedings of the Philosophy of Education Society (1985): 445-454.
- Casteen, John D.. "The Public Stake in Proper Test Use." In Charles C. Davis, ed. The Use and Misuse of Tests. San Francisco: Jossey-Bass, 1984.
- Changes in Education: A Guide for Parents. Victoria, B.C.: Ministry of Education, 1991.
- Cronbach, Lee. "Course Improvement Through Evaluation." Teachers College Record 64 (1963): 672-683.
- Dale, Roger et al., eds. Schooling and Capitalism: A Sociological Reader. London: Routledge and Kegan Paul, 1976.
- Daniels, Le Roi B., and Jerrold R. Coombs. "Analytic Philosophical Inquiry." In Edmund Short, ed. Forms of Curriculum Inquiry. Albany, New York: State University of New York, 1991.
- Davis, Charles C., ed. The Use and Misuse of Tests. San Francisco: Jossey-Bass, 1984.
- Ericson, David P., and Frederick S. Ellet. "Teacher Accountability and the Causal Theory of Teaching." Educational Theory 37, No. 3 (Summer, 1987): 277-293.
- Fallows, James. "The Tests and the 'Brightest': How Fair are the College Boards?" The Atlantic March 1980, 37-48.
- Feinberg, Walter and Jonas Soltis. School and Society. New York: Teachers College Press, 1985.

- Feinberg, Walter. Reason & Rhetoric: The Intellectual Foundations of 20th Century Liberal Educational Policy. New York: John Wiley & Sons, 1975.
- Flew, Anthony. "Teaching and Testing." Proceedings of the Philosophy of Education Society (1973): 201-212.
- Galloway, Charles. Psychology of Learning and Teaching. New York: McGraw-Hill Book Company, 1976.
- Green, Thomas. The Activities of Teaching. New York: McGraw-Hill, 1971.
- Hargreaves, Andy. "Record Breakers." In Patricia Broadfoot, ed. Profiles and Records of Achievement: A Review of Issues and Practices. London: Holt, Rinehart and Winston, 1986.
- Hextall, Ian, and Madan Sarup. "School Knowledge, Evaluation and Alienation." In Michael Young and Geoff Whitty, eds. Society, State and Schooling. Sussex: Falmer Press, 1977.
- Hextall, Ian. "Marking Work." In Michael Young and Geoff Whitty Explorations in the Politics of School Knowledge. Driffield: Studies in Education, Ltd., 1976.
- Hirst, Paul. Knowledge and the Curriculum. London: Routledge & Kegan Paul, 1974.
- Hoffman, Banesh. The Tyranny of Testing. New York: Crowell-Collier, 1962.
- House, Ernest. School Evaluation: The Politics & Process. Berkely, California: McCutchan Publishing Corporation, 1973.
- House, Ernest, ed. Philosophy of Evaluation. San Francisco: Jossey-Bass Inc., Publishers, 1983.
- Houts, Paul. The Myth of Measurability. New York: Hart Publishing, 1977.
- Information Circular #432. Victoria, B.C.: Ministry of Education, July 16, 1990.
- Laughland, Arthur. "Two Principals Look at Standardized Tests." In Paul Houts, ed. The Myth of Measurability. New York: Hart Publishing, 1977.
- MacIntyre, Alisdair. After Virtue: A Study in Moral Theory. 2nd ed. London: Gerald Duckworth & Co., 1981; Notre Dame, Ind.: University of Notre Dame Press, 1984.
- Madaus, George, Peter Airasian, and Thomas Kellaghan. School Effectiveness: A Reassessment of the Evidence. New York: McGraw-Hill, 1980.
- Madaus, George, Michael S. Scriven, and Daniel L. Stufflebeam. Evaluation Models: Viewpoints on Educational and Human Services Evaluation. Boston: Kluwer-Nijhoff, 1983.
- Madaus, George and Vincent Greaney. "The Irish Experience in Competency Testing: Implications for American Educators." American Journal of Education (February 1985): 268-294.
- Matthews, John. Examinations. London: George Allen & Unwin, 1985.
- Meier, Deborah. "Why Reading Tests Don't Test Reading." Dissent 28 (Fall 1981): 457-466.



- Mifflin, Frank J., and Sydney C. Mifflin. The Sociology of Education. Calgary: Detselig Press, 1982.
- Morganthau, Tom. "A Consumer's Guide to Testing." Newsweek, Fall/Winter 1990, 63.
- Owen, David. None of the Above: Behind the Myth of Scholastic Aptitude. Boston: Houghton Mifflin, 1985.
- Pateman, Trevor. "Can Schools Educate?" Journal of Philosophy of Education 14, no. 2 (1980): 139-148.
- Peters, R.S. Ethics and Education. London: George Allen & Unwin, 1966.
- Pring, Richard. "Knowledge Out Of Control." Education for Teaching 89 (Fall 1972): 19-28.
- Rawls, John. A Theory of Justice. Cambridge: Harvard University Press, 1971.
- Robertson, Don, and Marion Steele. The Halls of Yearning: An Indictment of Formal Education/A Manifesto of Student Liberation. San Francisco: Canfield Press, 1969.
- Ryle, Gilbert. Student Casualties. Hammondsworth: Penguin Books, 1969.
- Scheffler, Israel. The Language of Education. Springfield, Ill.: Charles C. Thomas, Publisher, 1960.
- Scheffler, Israel. The Conditions of Knowledge. Scott, Foresman & Company, 1965.
- Scheffler, Israel. Reason and Teaching. London: Routledge and Kegan Paul, 1973.
- Scriven, Michael, Ralph Taylor and Robert Gagne, eds. Perspectives of Curriculum Evaluation. Chicago: Rand McNally, 1967.
- Selden, Ramsay. "Missing Data: A Progress Report From the States." Phi Delta Kappan (March 1988): 492-494.
- Smith, Frank. Comprehension and Learning. New York: Holt, Rinehart, and Winston, 1975.
- Smith, Frank. Insult to Intelligence. Portsmouth, New Hampshire: Heinemann Educational Books, Inc., 1986.
- Soltis, Jonas and Walter Feinberg. School and Society. New York: Teachers College Press, 1985.
- Soltis, Jonas. "The Concept of Assessment: A Response to Mr. Flew." Proceedings of the Philosophy of Education Society (1973): 213-216.
- Steelman, Lala Carr, and Brian Powell. "Appraising the Implications of the SAT for Educational Policy." Phi Delta Kappan (May 1985): 603-606.
- Strenio, Andrew. The Testing Trap. New York: Rawson, Wade Publishers, Inc., 1981.
- Taylor, Paul. Normative Discourse. Englewood Cliffs, N.J.: Prentice-Hall, 1961.
- Walcott, Harry F. Teachers vs. Technocrats: An Educational Innovation in Anthropological Perspective. Eugene, Oregon: University of Oregon Centre for Educational Policy and Management, 1977.

- Whitehead, Alfred North. The Aims of Education and Other Essays. New York: The Macmillan Company, 1929; The Free Press, 1967.
- Wideen , Marvin F. et al. "Impact of Large Scale Testing on the Instructional Activity of Science Teachers." Paper presented at the Canadian Society for Studies in Education, 1991.
- Willie, Charles V. "The Problem of Standardized Testing in a Free and Pluralistic Society" Phi Delta Kappan (May 1985): 626-627.
- Willis, Paul. Learning to Labour: How Working Class Kids Get Working Class Jobs. Farnborough: Saxon House, 1977.
- Willms, Doug. Monitoring School Performance: A Guide for Educators. London: Falmer Press, 1992. [in press]
- Wilson, Robert, and Ruth Rees. "The Ecology of Assessment: Evaluation in Educational Settings." Canadian Journal of Education 15, no. 3 (1990): 215-228.
- Year 2000: A Curriculum and Assessment Framework for the Future. (Draft Document) Victoria: Ministry of Education, 1989.
- Year 2000: A Framework for Learning. Victoria, B.C.: Ministry of Education, 1990.
- Young, Michael, ed. Knowledge and Control. New York: Collier-MacMillan, 1971.