THE ROLE OF AUTHENTIC ASSESSMENT IN EVALUATING

CRITICAL THINKING

by

David Nicholas Neufeld

B.Ed., University of British Columbia, 1978

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ARTS

in the Faculty

of

Education

© David Nicholas Neufeld 1994

SIMON FRASER UNIVERSITY

July 1994

APPROVAL

NAME   David Nicholas Neufeld

DEGREE   Master of Arts

TITLE   The Role of Authentic Assessment in
Evaluating Critical Thinking

EXAMINING COMMITTEE:

   Chair   Philip H. Winne

Roland Case, Associate Professor
Senior Supervisor

Sharon Bailin, Associate Professor
Member

Dr. Giselle O. Martin-Kniep
Adelphi University
School of Education
Garden City, New York
External Examiner

          Date: _July 8, 1994_

PARTIAL COPYRIGHT LICENSE
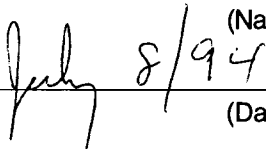
Title of Thesis/Project/Extended Essay

The Role of Authentic Assessment in Evaluating Critical Thinking

Author:

(Signature)

David Nicholas Neufeld

(Name)

July 8/94

(Date)

ABSTRACT

This thesis examines methods of evaluating critical thinking and finds that key aspects of critical thinking not addressed by current assessment practices can be evaluated using authentic assessment tasks. Initially, several conceptions of critical thinking are discussed. The notion of critical thinking "skills" is examined and found misleading, but other central aspects of critical thinking are identified. Critical thinking a) is rational, b) meets both general and subject- or discipline-specific standards, c) contains both evaluative and generative components, and d) implies the existence of certain dispositional factors.

Three methods of evaluating thinking -- multiple choice, essay, and short answer -- are analyzed, but each fails to address vital aspects of the conception. Multiple choice tests give little direct information about the reasons behind students' thinking, fail to take the generative nature of critical thinking into account, and give no information regarding critical thinking dispositions. Neither the short answer nor the essay test measures "judgments-in-action" nor do they adequately address the dispositional aspect of critical thinking.

Authentic assessment is defined as any measure designed to

be less artificial and more connected to and patterned after the challenges people face outside the school context. It is found to fall into three categories: a) naturalistic assessment, where the teacher collects information about student thinking during the regular routine of the classroom; b) performance assessment, where the student is expected to complete a set of complex tasks through performing a feat or creating an object; and c) portfolio assessment, where a collection of the student's work is used to demonstrate her abilities in a certain area. Authentic assessment is characterized as requiring judgment, tending toward "real-world" tasks, is likely to be holistic in nature, and attempts to hit at the core of particular fields of study.

The thesis concludes by examining authentic assessment in light of the demands of critical thinking. It finds that while authentic assessment should not be expected to replace all existing forms of assessment, it is ideally suited to evaluate those aspects of critical thinking not addressed by current tests, namely, the reasons behind students' judgments, as well as the generative and dispositional components of critical thinking.

TABLE OF CONTENTS

1. INTRODUCTION

Evaluation is an integral part of instruction. Effective evaluation is essential to determining which elements of a subject have been learned, which remain to be taught, and which need reinforcing. Evaluation tells us whether or not instruction is having an effect on learning and helps determine which programs and methods are the most effective in bringing about our educational aims. In this thesis, I examine current practices in the evaluation of students' critical thinking, and make recommendations for the use of a range of assessment devices which capture more completely the essential features of such thinking.

While there are many important questions about large-scale critical thinking assessments to be raised, I have chosen to focus my attention on critical thinking evaluation of individual students in a classroom setting. Although large-scale assessments and program appraisals have an important place in critical thinking instruction, day-to-day decisions about student progress depend on more immediate assessments, and it is these which are the subject of this thesis. If teachers are to tailor their instruction to the needs of particular students, a systematic, ongoing program of assessment will have to be put in place. This thesis makes recommendations regarding the components of such an assessment plan.

An essential ingredient in examining and revising any
evaluation is a clear conception of that which is to be
evaluated. For this reason, the first major question to
be answered is, "What is critical thinking?" This thesis
establishes a context for evaluating critical thinking by
first analyzing the concept itself. The answer to this
question quite clearly has vast implications for both the
teaching and evaluation of critical thinking. Various
answers to this question have been offered and there are
claims that up to 35 definitions have been identified
(Romanish, 1986). Numerous conceptions emphasize the
process(es) of thinking (Beyer, 1985b; Norris, 1986;
Norris & Ennis, 1989; Raths, Wassermann, Jonas, &
Rothstein, 1986), thereby characterizing critical thinking
as a particular mode of thinking or thinking done
according to certain procedures. The conception adopted
in this thesis characterizes critical thinking not as a
descriptive concept but as a normative one. Critical
thinking is not determined by the steps which are followed
but by the purpose and quality of the result; it is
thinking which meets certain standards according to
particular criteria, and it is the products of thinking
which are to be evaluated.

A second characteristic of critical thinking identified
here is its generative nature. While the word "critical"
implies fault-finding and evaluation, the bringing about

of products which are critical involves a generative component (Paul, 1992). Phrases like "creating an argument" or "producing a defense" betray the close association between the critical and the creative, and critical thinking assessment instruments should be sensitive to this quality. Finally, critical thinking as conceptualized here goes beyond the ability to think well, and includes a dispositional element made up of character traits, attitudes, and tendencies which drive the individual to be critical when the circumstances warrant. A critical thinker is a particular type of person, that is, someone who has acquired a critical spirit (Siegel, 1980). While it could be argued that this is not a quality of critical thinking per se, but rather a quality found in critical thinkers, it is adopted here as being essential to the instruction and evaluation of critical thinking. We are not particularly interested in people who can think critically but choose not to, in the same way that we are not particularly interested in firemen who have the skill and strength necessary to save lives, but don't have the courage to do so. Critical thinking is a vital part of what it means to be an educated person (Chipman, Segal & Glaser, 1985; Hirst & Peters, 1970; Paul, 1987; Romanish, 1986), and the actual practice of thinking critically demands the inclusion of each of the elements mentioned.

Once a defensible conception is adopted, the challenge
becomes one of finding assessment instruments which
capture the full range of features characterizing critical
thinking. This is not an easy task, and Wiggins (1989a)
speaks for many when he deplores the state of assessment
in education generally. Criticizing the current emphasis
on statistical accuracy and economy at the expense of true
testing of human abilities, he calls for tests which
assess intellectual ability by requiring the performance
of exemplary tasks (p. 703). It is a call for reform in
assessment, and has been joined by many others (Johnson,
1992; Martinez & Lipson, 1989; Redding, 1992). Currently,
the evaluation of critical thinking takes several forms,
and three popular methods loosely characterized as the
multiple choice test, the essay, and the short answer
exercise, are examined in Chapter 3 of this thesis. While
each method is found to have a place in critical thinking
assessment, none meets the complex demands of the rich
conception of critical thinking adopted here.

In response to calls for assessment change, Wiggins
(1989b, 1992) outlines criteria for what he calls
"authentic" assessments. Designed to replicate the
challenges and standards of performance of experts in
given fields, authentic assessments become an essential
part of instruction. Authentic assessments reflect more
closely the varied situations that are presented to

students in the normal course of Instruction or real life (Bateson, 1992). Three categories of authentic assessment -- naturalistic, portfolio, and performance -- are discussed in Chapter 4.

In the final chapter of this thesis, I examine authentic assessment more closely and hold it up to the concept of critical thinking. While authentic assessments show a great deal of promise for providing a more thorough evaluation of critical thinking, they are not going to be the miracle cure some may hope for. They do, however, broaden the scope of assessment Instruments available to those who wish to assess critical thinking more fully. The fit between critical thinking and authentic assessment is not a perfect one, but authentic assessment does capture features of critical thinking not addressed by traditional methods of evaluation.

2. CRITICAL THINKING

This chapter is devoted to an examination of the concept of critical thinking. A number of different conceptions are reviewed and contrasted in order to outline the territory claimed for critical thinking by different theorists. This overview of current conceptions is followed by a closer look at the notion of critical thinking "skills", a notion common to many conceptions of critical thinking. I argue that though many attempts at identifying such skills have been made, the thinking skills idea itself is seriously flawed and leads to unwarranted conclusions about the teaching and evaluation of critical thinking.

Based upon a number of conceptions reviewed in the early part of the chapter, I adopt a conception of critical thinking which includes several vital elements. I argue that critical thinking is a rational enterprise judged "critical" not on the basis of the types of steps followed, but by the degree to which certain standards are met. Critical thinking is thus not simply the application of a series of moves, but is a normative enterprise subject to standards which characteristically identify good thinking within a particular domain.

I point out, however, that it is not enough that educators develop individuals who have the ability to meet such

standards. Critical thinkers must actually have the tendency and determination to think critically when the situation warrants. Thus, I argue that critical thinking dispositions are an integral part of what it means to be a critical thinker, and that this dispositional element must also be addressed in critical thinking evaluation.

Finally, I show that incidences of critical thought are not characterized exclusively by evaluative judgments, but frequently include a generative component. The answers to many of life's problems are not readily apparent, and the array of possible solutions are not often listed for critical appraisal. The critical thinker thus frequently generates her own answers and for this reason, the tasks used to evaluate critical thinking must require this productive aspect of the student.

2.1 A SURVEY OF CURRENT CONCEPTIONS

Since Robert Ennis published his influential discussion of critical thinking in 1962, numerous conceptions have been posited in the educational literature. Ennis has expanded his original definition, "the correct assessing of statements" (1962, p. 81) to include any thinking which is "reflective and reasonable" and "focused on deciding what to believe or do" (1985, p. 45). His "multi-aspect"

approach identifies abilities and dispositions he claims are vital to the critical thinker (1985, p. 46).

John McPeck's epistemologically based account defines critical thinking as "the appropriate use of reflective skepticism within the problem area'under consideration" (McPeck 1981, p. 7). The purpose of this reflective skepticism or suspension of assent is to allow the thinker to construct alternatives which might resolve the problem at hand (McPeck 1981, p. 9). Unlike Ennis, McPeck argues that while there may be a limited set of general thinking skills, their usefulness decreases the more general they are, and the most useful thinking skills are largely limited to specific domains or narrow areas of application (1990, p. 12).

Paul and Nosich (1991) define critical thinking as "the intellectually disciplined process of actively and skillfully conceptualizing, applying, analyzing, synthesizing or evaluating information gathered from, or generated by, observation, experience, reflection, reasoning, or communication, as a guide to belief and action" (p. 4). This definition focuses on the process or activities associated with thinking. Subsequently, Paul, Fisher, and Nosich (1993) redefine critical thinking as "that mode of thinking - about any subject, content, or problem - in which the thinker improves the quality of his

or her thinking by skillfully taking charge of the structures inherent in thinking and imposing standards upon them". This latter definition introduces standards of thinking and centers on the products of thinking.

Paul (1984, 1987) has strongly argued in defence of two senses of critical thinking, differentiated partially by the use to which one's critical thinking is put. Critical thinking in the "weak" sense is a self-serving type of thinking in which critical thinking is used primarily to defend previously held beliefs and to defeat any ideas which may threaten one's egocentric views. This sense of critical thinking is not sufficient to create the ideal thinker, yet it is a precursor to what Paul calls "strong" sense critical thinking.

In the "strong" sense, critical thinking is regulated by what Paul earlier called the "rational passions" (1987), also referred to as the affective dimension of critical thinking (Paul & Nosich, 1991). The drive for clarity, accuracy, fair-mindedness, independence of thought, as well as dispositions such as intellectual courage, perseverance and intellectual curiosity all lead the thinker to examine her own prejudices, biases, and misconceptions in her "search for a fuller grasp of what is so" (Paul 1987, pp. 140-142).

Siegel (1980) views critical thinking as the embodiment of the ideal of rationality. This, in turn, is understood as being "coextensive with the relevance of reasons" (p. 8). Critical thinking is therefore rational, or reasoned thinking and involves knowledge of, and a commitment to principles governing such an activity: for example, impartiality, nonarbitrariness, and objectivity (p. 8).

Siegel believes that two conditions must obtain before an individual can be considered a critical thinker. First, she must have the ability to assess claims and make judgments on the basis of reasons while conforming to standards and principles governing the evaluation of those reasons. Second, she must have certain attitudes, dispositions, habits, and character traits which Siegel labels the "critical spirit" (p. 9). The critical thinker must be disposed to think critically, have a willingness to conform judgment to principle, and have a character inclined to seek reasons. She rejects partiality and arbitrariness and is committed to objective evaluation of relevant evidence (p. 9).

Other conceptions of critical thinking have led to the process approach (Raths, Wassermann, Jonas, and Rothstein, 1986) and the logic skills approach. In the former, critical thinking primarily involves the application of a series of general abilities. Used as a basis for some

school thinking programs, it is grounded in the assumption that there is a series of processes, learned through extensive practice which can then be applied to any number of contexts. These processes, frequently called skills or operations, vary depending on the program, but might include such things as "comparing", observing", "decision making", or "classifying" (Raths et al., 1986, pp. 6-20).

Elements of formal and informal logic have also found their way into many critical thinking programs. Ennis (1985) lists judging inductive and deductive inferences among other critical thinking abilities, while Paul and Nosich (1991) include "discrimination of necessary from probable and improbable consequences" and "evaluating inferences" in their list of critical thinking skills (p. 18). Moore and Parker's text (1986) is an example of a popular program based primarily on the rules of formal and informal logic. While arguing for a fuller conception of critical thinking,[1] their program deals primarily with the evaluation of arguments and the relationships among reasons and claims.

While these conceptions are not mutually exclusive, major differences in emphasis place some of them poles apart in practice. McPeck's views in particular place him in conflict with many other critical thinking theorists. His argument that it makes no sense to talk about teaching

critical thinking 'simpliciter' (1985, pp. 296-297) since
one cannot think without thinking about something, places
him in direct conflict with those, for example, who
advocate a "direct approach" to teaching thinking (Beyer,
1984a, 1984b, 1985c; De Bono, 1984).

McPeck (1985) opposes the notion that critical thinking is
either a general ability[2] or a set of specific skills
which, once learned, can be applied to a number of
different areas (1985, pp. 296-302). This places him in
opposition to Ennis (1962, 1985), Paul and Nosich (1991),
Beyer (1985b), and numerous others (Raths, Wassermann,
Jonas, & Rothstein, 1986) who have devised lists of such
skills and also contradicts those who have devised tests
which ostensibly measure those same skills (Watson &
Glaser, 1980; Ennis & Millman, 1985, for example).

In spite of these and other differences among the
conceptions, many of them highlight important aspects of
critical thinking, and it is possible to gather key
insights into the concept from them.

2.2 CRITICAL THINKING SKILLS

Critical thinking, from most people's perspective, implies
thinking which conforms to high standards and utilizes
sound reasoning. Critical thinkers are considered to be
more able thinkers than those who are non-critical.

Critical thinkers are identified by their ability to draw more valid inferences and solve more difficult problems than those not considered so. Critical thinking is, in a sense, expert thinking, and critical thinkers, some conclude, must therefore be in possession of a set of abilities or skills which can be taught to those who are not such capable thinkers (Beyer, 1985b; Young, 1992).

Others (Barrow, 1990; McPeck, 1981) debate the existence of a general set of critical thinking skills, questioning both the validity and usefulness of the many lists of skills put forward by skills proponents.

The answer to the debate clearly has significant implications for the teaching and assessment of critical thinking. If such general skills exist, and if they are as discrete and as generalizable and transferable as advocates claim, then it would be imperative that students are introduced to skills-based critical thinking programs, or at the very least that such skills be infused directly into content-based programs. It would follow that individual student assessment would include substantial skill-testing components. If, however, these "skills" are not generalizable and thus not transferable, then other approaches to instruction and subsequent evaluation must be taken.

Barrow (1990) points to undisciplined use of the term 'skill' as a culprit in allowing skills talk to enter critical thinking discussions, and claims that this results in an approach to teaching critical thinking which borders on nonsense (p. 89). He points out that if critical thinking abilities are indeed to be called skills, they operate in a manner quite different from what we normally call skills since the latter can be divorced from context while the former are context-sensitive.

Barrow (1990) argues that the word 'skill' normally denotes a discrete, physical ability improved through practice and that understanding is neither necessary nor sufficient for its use (p. 88). In his view, conceptualizing critical thinking as a skill-based enterprise is not only incorrect, but it results in instruction which fails to appreciate the vital place that understanding has in such an enterprise. Dribbling a ball or standing on one's head are skills, thinking is not.

Griffiths (1987) argues for a broader conception of skill, pointing out that ordinary usage dictates this more diverse use of the term. She rejects Barrow's contention that the word singles out physical abilities and claims for it the broader territory encompassed by Ryle's concept of 'knowledge how' (Griffiths, 1987, pp. 209-210).

In response to Barrow's discussion of skill, it might be pointed out that those we consider to be the most skilled do have a considerable amount of understanding of their craft. Even in activities which clearly involve a substantial physical component such as painting or carpentry, the skill shown is not divorced from understanding.

The skilled painter is not simply one who has cultivated the ability to draw perfect lines, and the skilled carpenter is not one whose skill is limited to purely physical abilities such as hammering nails without denting the wood. Their skills are demonstrated in the appropriate application of those abilities - a knowledge of when to exercise caution or when to proceed with abandon; how to use a certain brush to achieve the desired effect or how to go about constructing a unique piece of work; and whether to continue with a project or whether to abandon it.

While the narrow conception of 'skill' focuses on the physical dimension, it would appear that the ultimate exercise of a skill is actually contingent on the degree of understanding of the user. There are in all likelihood skills which are so simple that understanding in no way enhances their application, but Barrow's (1987) example of

dribbling a ball (pp. 190-191) does not provide us with
such a case.

In the controlled atmosphere of the practice field, the
skill of dribbling a soccer ball may possibly be reduced
to a physical activity absent of any understanding, yet
during an actual game an understanding of the implications
brought on by such factors as field conditions, score,
opponent, proximity to the sideline or penalty area,
fatigue, and many others all have a direct bearing on the
exercise of the physical dimension of that skill.
Regardless of the speed, dexterity, quickness of feet, or
any other physical characteristic contributing to the
player's dribbling skill, if he should lose control of the
ball due to a lack of understanding of other factors
involved in the game such as field conditions, opponent's
skill, proximity to other players, or team strategy, his
skill would be called into question. Thus, to omit the
element of understanding from a conception of 'skill'
would appear to be unnecessarily restrictive and to call
critical thinking abilities 'skill' is, from that
perspective at least, not an illegitimate use of the term.
While understanding may not be necessary to the exercise
or practice of certain skills, the need for understanding
does not of itself disqualify something as a skill.

Despite the previous discussion, however, the notion of critical thinking 'skills' is still not particularly satisfying. Barrow (1990) shows concern that the discrete, uncomplicated nature of some of life's less complex skills (standing on one's head, perhaps) will lull educators into thinking that the skills of critical thinking are equally discrete and uncomplicated. His concern is not without foundation since Griffiths (1987), for one, does in fact suggest that thinking can be improved through practice in much the same way that physical skills are. This raises a number of questions including "Exactly what is being practiced?" and "What will the result of such practice be?"

When a basketball player is asked by her coach to "go practice your dribbling", it is quite clear what she has to do. If she is physically capable of playing the game, and understands the concepts "practice" and "dribbling", she will be able to repeat actions which will help her learn this skill for application outside of the practice context. When a student is told by his teacher to "go work on 'distinguishing relevant from irrelevant information'[3]", what is it that the student is to do? Under the assumption that such thinking skills are generalizable, a teacher might give the student a series of exercises containing data both relevant and irrelevant to certain questions or problems, and have the student

actually make such distinctions. Questions arise, however, about whether the student is rehearsing the general skill mentioned, or whether she is doing no more than improving his ability to answer questions identical to the ones on which he is practicing. In other words, will she be able to apply his newly-refined skill to questions and problems in other contexts?

To use Beyer's skill as an example, a microbiologist may spend a lifetime distinguishing relevant from irrelevant information in his search for a disease cure, but it is unlikely that this will in any way help him to distinguish relevant from irrelevant information were he suddenly to start devising a game plan for a basketball team. His failure on the basketball sidelines would likely not be blamed on his failure to have mastered that particular skill, but on his failure to have accumulated enough knowledge about the finer points of basketball.

The question of the generalizability of critical thinking has been debated at length (Ennis 1989, 1990; McPeck 1981, 1990), and it is unlikely that this thesis will shed a great deal more light on the subject. It is an important question, however, and a stand will here be taken on the issue.

First of all, as Barrow (1990, p. 89) and McPeck (1981, p. 7, 1990, p. 10) point out, it is impossible to apply

critical thinking to different subjects without substantial knowledge of the field in question. Barrow (1990) observes that critical thinking "is not an ability that can be divorced from any context; it has to be instantiated in, and takes a different form in relation to, various subject matters" (p. 89). In other words, "recognizing unstated assumptions" in physics involves not only substantial knowledge of physics, the assumptions themselves are of a different kind and the skill of finding them significantly different than in, for example, the moral domain.

Second, it is unclear what the notion of "skill" adds to one's understanding of critical thinking. If a student wants to distinguish relevant from irrelevant information while deciding which prime ministerial candidate to vote for, chances are that she will collect as much information as possible about the candidates and then decide which information is relevant based on criteria which are peculiar to that specific task. She will note, for example that candidate A is male and that candidate B is female. Now, is there a skill involved in determining the relevancy of gender or does she simply consider the job which a prime minister is expected to do and determine on the basis of those criteria that gender is largely irrelevant in this case?

She may also note that candidate A has filed for personal bankruptcy on two occasions while candidate B has managed to keep a large corporation afloat through difficult times and make a determination on those grounds. One should notice, however, that in each case reference is made not to a general skill, but to specific criteria by which to judge the activity. Her ability to make a reasonable judgment in this case would not help her in making judgments about relevant or irrelevant information in determining the cause of a space shuttle failure, for example, unless she had a substantial amount of knowledge about space shuttles.

The notion of critical thinking skills is therefore misleading in that it suggests that such skills, practiced in isolation, or at least in a narrow range of contexts, can lead students to become critical thinkers in a variety of areas. This is simply not the case. This is not to conclude, however that the lists of 'skills' have no utility whatever, for these 'skills' pick out tasks which critical thinkers regularly perform and strategies which they frequently use. It is the case that microbiologists and basketball coaches distinguish relevant from irrelevant information on a regular basis and that it does make sense to use the same phrase to describe what each of them is doing. It does not follow, however, that the task is independent enough of context to allow educators to

assume that success in one context implies success in another.

Good critical thinkers do, as a matter of fact, make judgments about relevant and irrelevant information; they do determine the credibility of a 'speaker; and they do evaluate observation data. When students are working on tasks in the classroom and are failing to think critically about what they are doing, it is useful for the teacher to instruct the student to "make sure you keep the situation in mind", or "clarify your main point", or "go back and select the criteria against which your solution will be judged[4]". The lists of 'skills' thus become useful checklists by which teachers can begin to direct students' thinking.

Skilled writers make appropriate and sometimes even creative use of punctuation, yet it would be odd to refer to one as having the skill of 'punctuation use', or even the skill of 'comma placement'. The comma is a vital part of a writer's arsenal in the same way that fallacy identification is vital to a critical thinker, yet the appropriate use of neither one needs to be explained in terms of a particular skill.

Each of these critical thinking tasks and strategies should be introduced to the student in a variety of contexts and should be infused into the material as a

regular part of instruction. The names of these tasks and strategies should therefore become as familiar to the students as are the comma and period, but they should always remain subject to the criteria, standards, and concepts of each particular context.

McPeck (1985) claims that what is needed is a return to the fundamental principles inherent in a liberal education (pp. 305-306) -- an interesting comment considering that the reason educators are grasping at critical thinking programs is their observations which indicate that such teaching was simply not doing the job. McPeck's claims may be correct by definition, that is, when one includes critical thinking as part of what it means to have a liberal education, then a liberal education will indeed suffice. In that case, students have not been receiving a liberal education and we still need to determine which missing elements will provide students with appropriate types of instruction.

While the notion of thinking 'skills' carries with it too many misleading ideas of how critical thinking can be taught in classrooms, the tasks and strategies listed as 'skills' are useful in directing students as they take their investigations in new and more critical directions. Both a substantial understanding of the field in question

and a knowledge of the criteria and concepts employed by good thinkers are necessary elements of critical thinking.

2.3 CRITICAL THINKING CRITERIA AND STANDARDS

One shortcoming of the flawed "skills" conception of critical thinking is that it ignores the extent to which critical thinking is a normative enterprise (Selman, 1989). The judgments made by critical thinkers are not evaluated simply in terms of the steps which are followed or the moves that are made, but by how well those steps are taken and how those moves meet relevant criteria. The previous section of this thesis has acknowledged that there are certain tasks which critical thinkers characteristically engage in, yet the simple completion of such tasks is not enough. The critical thinker is not just someone who does certain things, but is someone who does certain things well (Selman, 1989, p. 36). For this reason, the evaluation of critical thinking involves not only a search for particular behaviours which the thinker may perform, but involves comparing the products of his thinking with relevant standards.

Much has been written about the "processes" of critical thinking (for example, Beyer, 1985b; Norris & Ennis, 1989, p. 5; Norris, 1986, p. 136; Raths, Wassermann, Jonas, & Rothstein, 1986). The problem with this notion is that it

seems to imply that simply engaging in certain types of mental activities is sufficient; that the most important question to be answered in critical thinking is "How does such thinking take place?" rather than questions like "What kinds of standards will be used as bench-marks of good thinking?"

As was stated in the previous section, there are certain tasks which critical thinkers characteristically engage in, but simply being involved in such a "process" is not sufficient. It is not enough, for example, that one engages in the activity of "judging the credibility of a source" (Norris & Ennis, 1989, p. 10); one must, in addition, correctly establish credibility according to relevant standards. It is not enough that one engages in assumption identification, one must, in addition, accurately identify the relevant ones.

The normative nature of critical thinking is going to be discussed here with reference to two notions: criteria and standards. In this discussion, 'criteria' refers to the types of things by which judgments are made, and 'standards' refers to the extent or degree to which these criteria need to be present. For example, the criteria for entrance to most universities would include high school grade point average while the standard might be set at a G.P.A of 3.0. The criteria for employment as a

firefighter might include strength, endurance, and knowledge of first aid, and the standards might be set respectively at a dead-lift of 200 pounds, a 60-pound carry over 40 metres in 25 seconds, and successful completion of a first aid course.

It is frequently easier to establish the criteria for a particular judgment than to establish the standards. For example, in selecting a marriage partner, the criteria might include physical attractiveness, sense of humour, and wide range of interests, but it is difficult to set clear standards within those criteria. Physical attractiveness is highly subjective; a sense of humour can possibly be measured by how easily the person laughs, but at some ill-defined point laughter becomes annoying; and establishing when someone's interests are wide-ranging is unclear. In spite of this, people do make judgments about prospective partners -- although with varying degrees of success.

Critical thinking, as a normative enterprise, is determined on the basis of criteria and standards, but due to its complexity and the number of contexts within which it is exercised, the criteria and standards are difficult to establish. It is possible to state that someone is not thinking critically enough, and it is also possible to judge someone as being sufficiently critical, yet where

that critical point between insufficient and sufficient
is, is not easy to determine.

Numerous criteria for judging critical thinking have been
suggested. Paul and Nosich (1991) list what they call
'intellectual standards' which they argue apply to
thinking in every subject (p. 16). They include clarity,
precision, specificity, accuracy, relevance, plausibility,
depth, and breadth in their list and claim that both
abilities and traits of mind can be evaluated by these
criteria. What is important about a list such as this is
that while critical thinkers do indeed strive to be clear,
precise, accurate, etc. in their thinking, the degree of
clarity, precision, and accuracy depends upon the context.
If the writer of a letter to the editor were to point out
that the paper was inaccurate in reporting that violent
crimes had increased by 12 % over the previous year and
that the number was actually 12.3%, one would not judge
her as being a highly developed critical thinker, but as a
person with too much time on her hands. Conversely, when
calculating the degree of re-entry for a returning space
shuttle, rounding off to the nearest whole number is
probably not good enough. In each of these cases the
standards are not static or absolute since the purpose and
context of the enterprise determine the grounds for
judgment.

Paul and Nosich's criteria would seem to be universal in that some amount of clarity, precision, etc. needs to be present in every thinking activity. The actual meaning of the terms varies from context to context, however, and the means by which they are achieved and established are highly subject-specific. Accuracy in mathematics, for example, is achieved by appropriately applying certain prescribed algorithms with precision, and is measured by comparing the result against a particular answer. Accuracy in medical diagnosis is achieved by weighing and utilizing a complicated and extensive number of factors including patient reports, past experience, accepted medical practice, and medical tests, and is measured by the degree to which accepted procedures for promoting recovery are successful. While it makes sense to use "accuracy" in both of those contexts, the similarity between the two uses is almost completely at a figurative level. Substituting the term "hitting the mark" might just as effectively be used instead of the word "accuracy" even though in most cases there is no mark and nothing is really being hit.

What the preceding discussion points out is that a) critical thinking is not determined solely by what thinkers do but also by how they do it; and that b) the criteria and standards judged relevant and sufficient vary from context to context. For this reason, without

substantial contextual detail the criteria and standards by which the thinking is to be measured are unclear. The implications this has on the evaluation of critical thinking is discussed in Chapter 5.

2.4 CRITICAL THINKING DISPOSITIONS

Critical thinking is an activity different from some other activities such as television viewing or reading in that one's participation requires a fair degree of commitment. It is often difficult to think critically and involves a certain amount of risk, discouraging to some individuals who might otherwise be quite capable of engaging in critical thought. To be a critical thinker, it is not enough to have the requisite knowledge, strategies, or ability. One must also have the tendency and inclination to actually use this ability.

Ennis's earliest account of critical thinking (1962) made no mention of any critical thinking dispositions, but increasingly, accounts of critical thinking acknowledge their existence. The dispositional aspect of critical thinking is characterized in many different ways and has acquired a variety of labels. Moore and Parker (1986) make vague reference to a "desire" to bring information to bear on decisions, but focus primarily on the 'skills' aspect of critical thinking. McPeck (1981), who divides

the teaching of critical thinking into cases of 'teaching how' (procedures or skills), and 'teaching to' (dispositions, propensities, or tendencies), sees it primarily as the will to use one's critical thinking capacity (p. 18). Thus we have his definition of critical thinking as "the propensity and skill to engage in an activity with reflective skepticism" (p. 8).

While McPeck (1981) rejects the notion of generalizable critical thinking skills, he acknowledges that a person might have the disposition to be critical in all disciplines. Such a person would still not be a critical thinker unless he had the requisite knowledge and understanding of the area in question (1981, p. 155).

Hawes (1990) calls the dispositional element of critical thinking the critical attitude and sees it not only as a willingness to engage in critical thinking, but as a regulative device which determines the appropriateness of undertaking such thinking (p. 59). The critical thinker displays an attitude of intellectual humility: an acknowledgement that what she presently believes may not be the final word and that those beliefs are open to reasonable revision (p. 59). Unlike McPeck, Hawes does not perceive the critical attitude as being generalizable, since to become critical in a new area requires a certain amount of skill and confidence (1990, p. 60). His

difference with McPeck is a conceptual one since the inclusion of the normative term "appropriate" in his conceptualization of the critical attitude would rule out instances where an individual might feel compelled to be "critical", but know nothing of the subject at hand.

Siegel (1988) refers to the dispositional component as the critical spirit and characterizes it as a willingness, desire, or disposition to base one's actions and beliefs on reasons (pp. 23-24). He sees this component of critical thinking as a general, global disposition to pay attention to reasons and to regard them as important, or to be "appropriately moved" by them (p. 8). The danger in ignoring this key element of critical thinking lies in the possibility that educators might regard their efforts at promoting critical thinking as successful if students passed critical thinking "skills" tests but didn't think critically outside of the testing context.

While some (Ennis, 1985; Paul & Nosich, 1991) have attempted to identify specific tasks and achievements as exemplifying the dispositional component of critical thinking, Siegel's (1988) critical spirit is a composite of attitudes, dispositions, habits of mind, and character traits.

> A critical thinker must have a willingness to
> conform judgment and action to principle, not
> simply an ability to so conform. One who has

> the critical attitude has a certain character as
> well as certain skills: a character which is
> inclined to seek, and to base judgment and
> action upon reasons; which rejects partiality
> and arbitrariness; which is committed to the
> objective evaluation of relevant evidence; and
> which values such aspects of critical thinking
> as intellectual honesty, justice to evidence,
> sympathetic and impartial consideration of
> interests, objectivity, and impartiality. (p.
> 39)

Despite Siegel's rational approach to critical thinking,

he rejects the notion that the dispositional element of

critical thinking is an emotionless tendency. It involves

a love of reason -- a set of "rational passions" which

become part of the character of the individual and lead

her to seek good reasons even when this runs counter to

pure self-interest (p. 39). These rational passions

include a drive for clarity, accuracy, and

fair-mindedness, sympathy for opposing views, a drive to

seek out evidence, an aversion to contradiction, sloppy

thinking, and inconsistency, and a devotion to truth. The

critical thinker is therefore not simply a person who acts

rationally, but is a certain sort of person (pp. 40-41).

Paul (1987), an advocate of the skills notion of teaching

critical thinking cautions that these skills can be taught

for the wrong purpose. Borrowing from C.W. Mills, Paul

classifies individuals into three groups according to how

they relate to their acquired belief systems (1987, p.

138). 'Vulgar' believers are those who operate with

slogans and stereotypes to support a point of view they

identify with. 'Sophisticated' thinkers are interested in learning about other points of view, but only for the purpose of refuting their arguments. Critical thinkers are willing to enter sympathetically into opposing points of view recognizing the weaknesses in their own. They don't shy from criticism, but learn from it, understanding that a point of view is something to be continually refined and developed as new evidence and better reasoning are considered.

It is the character of the person which will determine whether her ability as a critical thinker will be used in the weak or the strong sense, whether it will be used to protect and rationalize existing perspectives or whether her ability will be used to carefully consider alternative points of view. Paul (1984) contends that strong sense critical thinking is typified by the disposition of open-mindedness which begins with the willingness to consider the validity of contending points of view and the recognition that it is the "principles of comprehensive reason and evidence, not eternal authority, ego-identification, or technical expertise" which are the ultimate court of appeal (p. 12).

Paul (1985) also accepts a disposition of mind which parallels McPeck's sense of reflective skepticism. He rejects the notion of knowledge as a commodity

pre-packaged by experts and acquired in its complete form by the learner, but sees it more as an achievement of a mind which is slow rather than quick to believe (p. 38). Knowledge is a distinctive construction by the learner who weighs evidence critically before advancing to belief. The importance placed on the learner being able to test as well as actually testing propositions before assimilating them as knowledge[5] has clear implications for evaluation. The repetition of memorized information on a test does not assess whether this reflection has actually taken place.

Paul's strong sense of critical thinking captures essential characteristics of critical thinking missed in some narrower approaches. He construes critical thinking as relevant to the practice of everyday living, and thus characterizes it as being substantially more than a set of skills, techniques or other 'bag of tricks' of relevance only when applied formula-fashion to textbook questions in the classroom.

Critical thinking should bring the individual closer to the truth, or nearer to the solution of a specific question, and this is why Siegel's and Paul's concept of the critical spirit is so vital. It is clear that ability is an integral part of critical thinking, but without a critical spirit, critical thinking becomes little more than an effective defense of previously acquired beliefs.

Without open-mindedness, and the commitment to values such as clarity, accuracy, precision, consistency, relevance, sound evidence, good reasons, depth, breadth and fairness, one becomes entrapped by prior commitments to intellectual positions, and any kind of movement toward truth or resolution is clearly made more difficult.

This position neither ignores nor denies the importance of skillfulness in critical thinking. It would seem, however, that the critical spirit must in some sense develop in concert with such an ability. The critical spirit not only sets the direction in which one's thinking is to be focused, it gives the purpose for such thinking. Intellectual integrity must be a part of any program in education and thus should not only guide the actions of the teacher, but also be a part of the content.

There is a danger inherent in programs such as Moore and Parker's (1986) that the formulas for critical thinking promoted in such courses are accepted without question. Where emphasis is placed on acquiring the critical spirit, one is more likely to avoid the irony where critical thinking programs are taught in the absence of any critical thought. Once again, this aspect of critical thinking will have an impact on how students are to be evaluated.

2.5 THE GENERATIVE COMPONENT OF CRITICAL THINKING

Even before Ennis (1962) first defined critical thinking as the correct assessing of statements, educators had recognized the evaluative nature of the concept. Black (1952) characterizes critical thinking as an exercise in reasoned judgment. Critics are individuals who voice approval or disapproval based on relevant reasons and therefore a critical thinker is one who reasons for the purpose of passing favorable or adverse judgments (pp. 6-7).

Moore and Parker's (1986) definition of critical thinking as "the careful and deliberate determination of whether to accept reject, or suspend judgment about a claim" (p. 4) characterizes critical thinking as a reactive, second-order type of enterprise. From this perspective, the critical thinker is engaged solely in reacting to previously formulated statements, beliefs, or calls to action generated largely by others. Critical thinking is thus thinking focused on evaluating claims and is vital "since the way we conduct our lives depends on what claims we believe -- on what claims we accept" (p. 4).

Increasingly, however, theorists are broadening their conception of critical thinking to include a generative component. While oversimplifying the relationship between the generative and evaluative aspects to critical thinking

(Ballin, 1993), McPeck (1981) observes that critical thinking takes place in two contexts: a context of justification and a context of discovery (p. 15). Hawes (1990) also questions whether critical thinking denotes solely those particular phases of thinking that are specifically evaluative, or whether "critical thinking" denotes more complex activities which have important phases of reasoned or reasonable evaluation or judgment within them (p. 48). He concludes that it is used in the second way, and that activities such as problem-solving, decision-making, reading, writing, and interpretation, require the exercise of creative acts such as the formation of problems and purposes, the invention of possibilities and tests, and the making of new connections (p. 48). Critical thinking comprises not only the evaluation of ideas, sentences, arguments, alternatives, reasons and actions, but their production as well.

Ennis (1985), whose earlier evaluative conception has already been referred to, has subsequently broadened his account to include an element of creativity. His more recent definition -- reflective and reasonable thinking focused on deciding what to believe or do -- makes room for creative activities such as the formulation of hypotheses, questions, alternatives and plans.

As he notes, this broadening of the conception makes critical thinking a much more practical undertaking. Circumstances in real life which demand critical thought characteristically require a healthy dose of inventiveness. Parenting, for example, is an enterprise which constantly requires decisions about what to believe or do, and despite the availability of countless books on child-rearing, parents continue to struggle with questions concerning which actions to take with their children. Simply assessing self-help manuals as inadequate is itself not adequate. It is not enough, for example, to recognize that advice concerning what to do when your son has a tantrum in the supermarket does not apply to tantrums at funerals. A solution must be generated, and that end will be achieved only by means of thinking which is both evaluative and creative in nature.

Paul (1992) acknowledges the existence of a close, natural relationship between creative and critical thinking. He observes that excellence in thinking is achieved when we succeed in designing and producing results and outcomes appropriate to our ends, and that this involves both the creation of intellectual products and the judicious application of intellectual standards -- or critical and creative thinking (p. 1).

While critical thinking is still intimately involved in making judgments, even the making of a judgment itself is a creative enterprise. Thus the artificial dichotomy drawn between creative and critical thinking has provided us with an inaccurate conception of thinking in general and contributed to the acceptance of logic programs like Black's (1952) and Moore and Parker's (1986) as complete critical thinking programs.

The conception of critical thinking as an exclusively evaluative enterprise has also influenced the types of instruments accepted as being adequate for measuring student critical thinking competence. This issue will be addressed in Chapter 5.

2.6 SUMMARY

Critical thinking is an educational ideal of great interest to educators. While most conceptions provide insight into critical thinking, none of them on their own captures completely the richness of the concept. A combination of features from each gives a fuller account of the term.

Critical thinking is rational thinking concerned with a deliberate examination of the reasons supporting our beliefs and actions. It is good thinking and is often thought of as skilled thinking. While the notion of

skills In thinking was rejected as misleading, the lists of skills offered by many theorists were found to be useful in describing the kinds of tasks which critical thinkers perform, or the strategies which they use. While the tasks and strategies cross subject boundaries, a substantial amount of subject knowledge is necessary for critical thinking to take place.

Critical thinking is a normative enterprise and thus is judged according to a set of standards or relevant criteria. Critical thinking is more than the detached application of a series of prescribed steps, but involves the use of judgment, particularly in cases where there are questions about what is to be counted as evidence or what degree of precision is required for a particular case.

Critical thinking is both reasoned -- that is, its methods are such that good reasons come into play -- and it is reasonable -- that is, the result stands up to evaluation against relevant standards. The criteria and standards used can be either subject-specific or general, but must be referred to when evaluations take place.

Critical thinkers are not only capable of such thinking, they have a propensity to actually think critically and do so at appropriate times to appropriate degrees. This propensity to be critical is variously known as a tendency, disposition, attitude, trait, habit of mind, or

spirit, but generally refers to an individual's willingness to base her actions and beliefs on good reasons. Critical thinkers are inclined toward clarity and accuracy in their own arguments, and are open-minded about opposing views.

Finally, critical thinking is generative as well as evaluative in nature. It has a strong creative element which manifests itself in activities such as the creation of hypotheses and solutions as well as the formulation of judgments. The complex, multi-faceted problems characteristic of life outside the classroom are not solved in a context of justification alone, but must be addressed in a context where evaluation and generation each has a significant role. This account of critical thinking has implications for evaluation, but first the notion of evaluation itself will be examined.

3. EVALUATION

Numerous issues need to be addressed when examining the evaluation of critical thinking. The first and most obvious question is "What is critical thinking?" The effectiveness of any tool used for the purpose of evaluation depends on its being based upon a clear, complete and defensible conception of the term. Such a conception of critical thinking has been proposed in the opening pages of this thesis and will be used to judge existing methods of critical thinking assessment.

It is equally important to be clear about the types of evaluation referred to. This thesis focuses primarily on classroom assessment of individual students' critical thinking, although some issues concerning large scale assessments are addressed. Different methods of evaluation are appraised by comparing them with aspects of the concept itself.

I begin by discussing the object of thinking evaluations. I point out that inferences about students' thinking are made by observing and assessing the products of thinking, including such products as arguments, justifications, and discussions of our thoughts. The adequacy of a critical thinking evaluation is determined by the degree to which its inferences about a student's thinking are justified.

In the second section of this chapter, I examine three current methods of evaluation, the multiple-choice test, the essay, and the short answer question. The multiple choice format is the most popular form of objective test and is represented in the discussion by two of the most widely-used tests, the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980; hereafter cited as the Watson-Glaser) and the Cornell Critical Thinking Test (Ennis & Millman, 1971; hereafter cited as the Cornell). These tests are found to have several advantages over other methods of evaluation, yet inadequately account for differences in background assumptions among test-takers, fail to provide opportunities for test-takers to give reasons for their conclusions, and fail to address the generative and dispositional elements of critical thinking. Both the Cornell and the Watson-Glaser, each based on the skills notion of critical thinking, assume that to demonstrate proficiency in a particular area of the test allows inferences to be made about the exercise of similar proficiencies in other subject areas and contexts, an assumption found invalid earlier in this thesis. I also find problematic the notion that the quality of students' thinking is adequately measured by their success in matching the test-designers' answers and point out that this dilemma is unavoidable in multiple choice tests. Lack of complexity and cohesiveness are

two other Inadequacies of the Cornell and Watson-Glaser respectively and result from attempts at neutralizing the problem of the varied background beliefs of students.

In my examination of the essay format, I find that it affords several advantages over the multiple choice test in evaluating critical thinking. It allows students to explicitly state reasons for the judgments they make, and it allows the test-taker to justify and receive credit for answers not anticipated by test designers. It is difficult, however, to separate poor thinking from poor writing, and this penalizes students who may think critically, but are not able to express their thoughts in written form. The essay is found to accommodate the examination of more complex issues than multiple choice tests, but fails to assess students' ability to make critical "judgments-in-action" and group deliberations.

Finally, I examine the use of short answer critical thinking activities and find that while some of the activities purporting to demand critical thought require nothing more than recall of textual information, several of the other assignments are quite useful. The better activities are critically challenging and open-ended, with the demand for reasons implicit in the task. I find, however, that they lack the complexity and untidyness of

challenges found outside the classroom, and I move on to an examination of authenticity and authentic assessment.

3.1 EVALUATING PRODUCTS

We care about thinking because good thinking increases the probability of producing something of value. As a matter of fact, good thinking and good products would seem necessarily to be linked since it would be odd to refer to thinking as good if it resulted in products which were not good. Good thinking is synonymous with thinking which generates good products. While advocates of the process approach to critical thinking (Raths, Wassermann, Jonas, and Rothstein, 1986) imply that it is the "process" itself which is the key, any process is considered valuable only to the extent that it consistently produces useful, desirable products. A food processing plant, for example, is set up in such a way that the fruit arrives at the consumer's table in the best condition possible. Procedures are set up which have been proven likely to achieve that end, but adherence to those procedures is rigid only so long as the process produces the desired result. The criteria for determining a good process in this case are determined by what constitutes a good product. The value of the process is strictly instrumental to the product, and the primary concern is

the quality, not the process, moves, or steps of the thinking.

We regard critical thinking as an end or product in education due to its continued effectiveness in producing ideas we find useful. With this established, we can look to certain strategies, procedures, activities or tasks likely to produce critical thinking, but the adherence to these strategies, procedures, etc. is no guarantee that critical thinking has taken place. Thinking can be referred to as "critical" when the product of such thinking meets the standards referred to in section 2.3. It should be noted that the word "product" here does not just refer to those products we might normally think of in reference to completed assignments. Arguments are also products of thinking as are discussions of our thoughts and reasons for taking certain actions. When we ask, therefore, for a student to tell us to justify a particular belief, that justification is a product of her thinking. A critical thinking evaluation is adequate, therefore, to the extent that it makes justified inferences about a student's thinking based on the products of that thinking.

When assessing a student's thinking, various kinds of products can be evaluated. One can, for example, pose questions and assume that if the student produces the

correct answer, his thinking is good. While this may be adequate under conditions where there is agreement over what the correct answer is, there are many types of questions where the correct answer is under debate, or where slightly different conditions or contexts change what constitutes a correct answer. It might also be the case that the student arrives at the correct answer using methods which are unlikely to produce acceptable conclusions to future questions. A student might have achieved the correct answer by looking at his friend's paper, but the conditions which allowed him to do this are unlikely to be duplicated in very many avenues of life where he is required to think well.

In the next section of this chapter, three different approaches to evaluating the products of student thinking are examined. While the multiple choice test, the essay, and the short answer exercise are not exhaustive of all methods of critical thinking assessment currently in use in the classroom, they are the most common. The strengths and failings of each approach will be examined, beginning with the multiple choice test.

3.2 THE MULTIPLE CHOICE TEST

Frequently, critical thinking evaluation is done by means of objective tests which commonly follow a multiple-choice

format. Students who are taught the rules of formal or informal logic, for example, are normally evaluated in this manner.[1] Different fallacies and rules of logic are emphasized using samples from a number of contexts and subject areas. The student is then required to evaluate the form and structure of the argument by identifying assumptions or fallacies, making deductive or inductive inferences, and demonstrating an understanding of related vocabulary. In each case the correct answer must be selected from the choices given.

The Watson-Glaser and the Cornell are two of the most widely-used tests, and both use the multiple-choice format. The Watson-Glaser is made up of five sub-tests:

1. Inference

2. Recognition of Assumptions

3. Deduction

4. Interpretation

5. Evaluation of Arguments.

The sub-tests are named after and based upon five skills which the authors believe are central to critical thinking.

The Cornell test is based on a conception of the critical thinker as someone who has proficiency in making judgments about whether:

1. a statement follows from the premises

2. something is an assumption

3. a reliable observation is being made

4. an alleged authority is reliable

5. a simple generalization is warranted

6. a hypothesis is warranted

7. a theory is warranted

8. an argument depends on an ambiguity

9. a statement is overly vague or specific

10. a reason is relevant.

The Cornell test has two levels,[2] and not all the above proficiencies are covered in each, but they do give fairly comprehensive coverage to tasks widely accepted as demanding critical thought. While these tests will be examined in the following pages, the purpose is not to draw conclusions about any particular test, but rather to point out advantages and shortcomings of the multiple choice format in general.

3.21 Characteristics of Multiple Choice Tests

There are definite advantages to using multiple-choice questions in critical thinking tests. The most obvious one is ease of marking, since multiple-choice exams can be scored by machine or by individuals unfamiliar with critical thinking skills, thus producing results which are more reliable than those gained from more subjective evaluations (Norris & Ennis, 1989, p. 28). This format is

attractive from a financial perspective since reduced time and training result directly in reduced administrative costs while maintaining a high degree of reliability.

Multiple-choice questions also require less time to complete, therefore allowing a larger number of questions to be asked, enabling the evaluator to be more specific in his questioning and increasing the statistical reliability of the results. Multiple-choice tests are effective in singling out specific elements to be tested in critical thinking -- elements which might become lost or confounded in more complex types of measurement. Knowledge of concepts such as "fact" and "value", or "cause and effect" can most easily be tested in the multiple-choice format. As such, multiple choice tests are valuable as a first-line method of evaluation in that knowledge of basic concepts and an understanding of the appropriate use of simple strategies is often a vital first step to thinking critically about an issue.

Despite the utility and popularity of these tests, there are numerous problems associated with their use -- many of those difficulties stemming from their being based on limited conceptions of critical thinking. The next section looks at some of the limitations characteristic of multiple choice critical thinking tests.

3.22 Problems With Multiple Choice Tests

As has been pointed out, the concept "critical thinking" implies that certain qualities or characteristics of thinking are present, and these need to be addressed in the instruments used to evaluate the critical thinking of students. If there are certain qualities which the products of critical thinking should reflect, then assessment devices should be designed to pick up on those qualities. As will be indicated in this section, there are limitations inherent in multiple choice tests which make them inadequate for doing such evaluating.

The notion that critical thinking can be reduced to a series of general skills has been shown to be seriously flawed. A substantial amount of subject knowledge is required for someone to be able to think critically in any context, and success in thinking critically in one subject area is not indicative of success in any other. It is doubtful that one can test critical thinking without also testing the students' knowledge of the particular content used in the actual question.

Recognizing the importance of subject knowledge in answering any questions requiring critical thought, test designers stick to topics and structure questions which in most cases do not require specialized background knowledge. Thus, incorrect selections can be ascribed to

poor thinking rather than lack of requisite prior knowledge. Level X of the Cornell test, for example, places questions in the context of a party of explorers searching a newly-discovered planet for a previous group of adventurers. In the first of three sections, test-takers are asked to make judgments about whether certain facts support, deny, or fail to support or deny the hypothesis that the first group is dead. Different background assumptions, based on students' prior knowledge, however, can legitimately result in varying judgments about the direction of weight of evidence.

The first sample question on Level X of the Cornell states that everything in the first hut of the community is covered by a layer of dust. According to the keyed answer, this is to be taken as evidence in support of the hypothesis that the first group has died. A student, living in an environment where dust removal is part of family life, might think critically, and conclude that the presence of dust does indeed provide evidence of the absence of human activity. He would thus see the dust as evidence supporting the hypothesis that the first team had died, and be rewarded for his answer.

It is conceivable, however, that a student who lives near a construction site views the accumulation of dust as a sign of activity rather than inactivity and thus concludes

that the dust provides evidence against the hypothesis that the first group is dead. Her conclusion would not be the result of poor thinking, but would be based on the legitimate contention that under certain circumstances -- new construction, for example -- active people stir up a great deal of dust. She might even recognize that dusting and keeping a clean house is an activity which pales in comparison with matters more closely associated with exploring a new planet, and conclude that the explorers were doing far more important things away from the house.

A third student, thinking critically, might recognize that dust accumulation can be a sign either of activity or the lack of it, and decide that the accumulation of dust therefore could provide evidence both supporting and opposing the hypothesis. This option is not available among the three choices given, leaving the student in the difficult position of having to decide which of the three available choices most closely approximates her own. She might choose the option "neither" since the evidence gives neither clear support for nor clear denial of the hypothesis.

These students, each of them thinking critically, might therefore come to three different conclusions about what the evidence indicates. The failure of two of the students to select the keyed answer in this case would be

due not to their lack of critical thinking, but to differing prior experiences related to dust, and to differing views about the tasks which explorers are likely to perform. Despite thinking critically, these students are penalized for going beyond what the test-designer had in mind.

Unfortunately, in a multiple-choice format, there is no room for individuals to indicate the reasons for their selections -- information vital to anyone attempting to make inferences about the students' critical thinking ability. McPeck (1981) argues that in critical thinking tests good answers should not be predicated on their being right, but on the quality of the justification given for the response. There is no room for such justification to be given on standard multiple choice exams.

Norris (1989) suggests incorporating requests for justification into multiple choice exams as one method of controlling for extra-critical-thinking empirical beliefs and the Cornell Level WJX (Ennis & Millman, 1993) does just that. Test-takers are given two lines after each question on which to give the reason for their selection. The justification gives the test-designer direct information about the test-taker's thinking, but in doing so, removes the advantages of using multiple choice tests in the first place. According to the manual (Ennis,

1992), points awarded for each item range from a high of 5 for "an answer at the level that a good critical thinking teacher would give, when trying to make fully explicit the justification of an answer" to a low of 1 for "no answer or a badly justified unkeyed answer" (p. 2).

Sample justifications are provided for each of the four sections, but the advantages which multiple choice tests have over other types of tests -- namely: (a) speed of marking and (b) no reliance on the expertise of the marker -- are lessened in this format. Instructing markers to give full marks for answers like those given by good thinkers is question-begging and useful only to good critical thinkers. In effect, what Ennis and Millman have done is traded the efficiency of the multiple choice exam for some of the effectiveness of an essay -- although the two lines provided to justify test-takers' choices allow only limited opportunity for reasons to be discussed. The Cornell Level WJX does give information about student thinking, but has neither the efficiency of true multiple choice tests nor the extensive room for justification that essay tests have.

The virtual impossibility of testing critical thinking without at the same time testing student knowledge of a particular subject -- whether that subject is considered common sense, general knowledge, or anything else -- has

already been noted. As McPeck (1981, 1985) has pointed out, one must always be thinking about something, and it is impossible to isolate the thinking from that something.

A further assumption, also based on the flawed skills notion of critical thinking, is that an individual's ability to apply these skills in the testing context means that she will be able to apply them to other subjects in other, less artificial contexts. This is to some extent an empirical question, yet the earlier discussion on skills has questioned whether, for example, identifying assumptions in literature even means the same thing as identifying assumptions in mathematics. Once again the utility of multiple choice tests is called into question since the subject around which the test question is formed is likely to be quite different from the subjects about which the students will be thinking critically once they leave the examination room.

The hypothetical nature of the content on the Cornell test, for example, means that unless there is a significant level of transfer, the test may tell us that the student has acquired the ability to make correct (match the test-makers) judgments about what the presence or absence of dust means in a space colony, but little else. Ironically, the hypothesis that the colonists have died is ultimately found to be incorrect. A later section

of the test finds that the first group of explorers have simply left the original community.

Norris and Ennis (1989) find similar problems with the Watson-Glaser. They point out that the subtest designed to judge the student's ability ' to make justified inferences might in fact penalize individuals who have more knowledge, creativity and interest and who are willing to spend more time reflecting on the item (p. 59).

These problems are not unique to the Watson-Glaser and Cornell tests. They are problems inherent in multiple choice tests in general. The quality of the individual's thinking is measured by his success in coming to the same conclusions as the test designer, yet the very nature of the activity of drawing inferences demands that different prior knowledge and varying background assumptions will lead to different conclusions regardless of the person's ability to make justified inferences.

Of note is the fact that in many contexts outside the classroom, inferences which go beyond the norm are not penalized, but rewarded. Investigative reporters and police detectives earn a living by piecing together evidence and making inferences not immediately recognized by others. What sets them apart is not their ability to come to the same conclusions as everyone else, but their

ability to construct possible resolutions which others may have missed.

It is this generative component which objective measures such as multiple choice tests fail to address. One creates an argument, makes an inference, and produces reasons, yet multiple choice tests focus completely on the evaluative rather than the productive aspects of critical thinking. A critical thinker should be judged not only on her ability to select the best answer from a series, but on her ability to create solutions to real-life, complex questions. By providing choices for the student, test-makers have done all of the creative work, and have left the examinee the much narrower task of attempting to recognize the merits in each of the choices already conceived. The tests effectively discourage test-takers from generating any kinds of unique responses to these thinking challenges.

Novelty is also discouraged by the simplistic nature of the questions used on multiple choice tests. Petrie (1985) argues that the technical requirements of standard objective tests render them virtually unusable as tests for critical thinking since they require an unambiguous right answer. In order to do this, he contends, test constructors have to hold constant all but one of the ways in which adaptive thought can occur. What this does, he

states, is force test-designers to: (a) limit the test to recall items; (b) make the accepted answer a matter of well-accepted deductive techniques; (c) provide enough information so that simple comprehension rules out the unwanted alternatives; or (d) put trick qualifiers on the wrong choices (p. 11). One can see that in taking such measures, the conception of critical thinking has been narrowed significantly.

Multiple choice exams are characterized by a lack of complexity which is far removed from what Paul (1987) refers to as the multilogical problems people face outside the classroom. The need to design questions with a single answer eliminates the possibility of assessing the student's ability to cope with the complex and open-ended challenges which confront most people in their every-day lives. Sternberg (1985) contrasts the problems found in critical thinking programs with those which people actually face. Among observations about the ill-structured nature of everyday problems, he finds that recognition that a problem exists and identifying exactly what the problem is are often more difficult than actually figuring out how to solve the problem (pp. 195-196). Such demands are not placed on students writing objective tests where the problems are clearly identified and the options simply spelled out.

Sternberg also contends that everyday problems are complicated, messy and stubbornly persistent, and their solutions have consequences that actually matter in contrast to the simple, neat textbook problems encountered in critical thinking programs. Students taking the Watson-Glaser are confronted with questions like the following:

> Rice and celery must have a good deal of moisture in order to grow well, but rye and cotton grow best where it is relatively dry. Rice and cotton grow only where it is hot, and celery and rye only where it is cool. In Timbuktu, it is very hot and damp. Therefore --
>
> 44. Neither the temperature nor the moisture conditions in Timbuktu are favorable for growing a celery crop.
>
> 45. The temperature and moisture conditions in Timbuktu are more favorable for growing rice than for growing a celery, cotton, or rye.
>
> 46. Conditions in Timbuktu are not altogether favorable for growing a cotton or a rye crop.

One is immediately struck with the question's lack of relevance not only to the lives of most people answering the question, but also its lack of connection to the previous question, which is about symphony orchestras, and the following question, which concerns people's smoking habits.

Such tests are not without any use at all (they do after all test individuals' ability at making warranted deductions within severely limited contexts), but they are

clearly quite inadequate in determining an individual's ability to think critically about complex issues where judgments actually matter. They are also incapable of determining whether an individual is actually disposed to think critically about issues outside of the testing context. Critical thinking is hard work. To think critically when the situation warrants it does not naturally follow simply from being able to do so. There is little in the structure of multiple choice tests to indicate that an individual who has shown that she has sufficient understanding of what it means to make a valid deductive inference about rice and celery on a two hour exam will use that understanding at appropriate times in her everyday life. There is also no indication that the test-taker will turn her critical thinking on her own ideas, nor that she will act in a manner consistent with the results of her critical thinking.

Norris (1986) points out that if evaluation is to give an indication of critical thinking in use, then critical thinking tests are not adequate. It is reasonable to assume, given the substantial difference between the types of questions posed in objective critical thinking tests and the problems facing people outside the classroom, that a large number of people who do quite well on informal and formal logic exams as well as on tests such as the Cornell

and the Watson-Glaser could in fact be poor critical thinkers in their everyday lives.

Multiple choice critical thinking tests such as the Watson-Glaser and the Cornell do not suffer from poor design. It is the fundamental structure of such tests which makes them inadequate for the purposes of critical thinking evaluation. The one-answer format, their heavy reliance on the flawed skills notion of critical thinking and its assumption of the generalizability of such skills, the simplistic structure and irrelevant content resulting from attempts to neutralize problems of background beliefs, and their inability to assess critical thinking dispositions all contribute to their limited utility in assessing the critical thinking of students.

3.3 THE ESSAY

The essay remains a viable alternative to multiple choice tests in evaluating critical thinking. It provides a more comprehensive examination of critical thinking than the multiple choice test, but it loses some of the advantages. Since responses are not limited as in multiple choice tests, essays are more difficult to mark and therefore require a significant amount of expertise on the part of the marker.

Commercial essay tests require the development of a comprehensive, detailed scoring rubric. One commercial product is the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985; hereafter cited as Ennis-Weir). It is geared toward high school or college level students and uses a fictitious letter to the editor of a newspaper as the basis for the test. The eight-paragraph letter argues that overnight parking on city streets should be eliminated. Test-takers are asked to evaluate the thinking in each of the eight paragraphs and then evaluate the letter as a whole. Norris and Ennis (1989) report that raters of responses to the Ennis-Weir were relatively successful in ranking examinees in similar order, but varied significantly in the level of score they assigned to the responses.

Essays are also far more time-consuming to mark. Referring to the Ennis-Weir (1985), Norris and Ennis (1989) report that each of the paragraph responses to the questions can be marked in under 10 minutes and in as quickly as six minutes, or 10 per hour. As they point out, a machine that scores multiple-choice tests at ten thousand per hour operates one thousand times as fast, a considerable saving in time and money. In essay exams, problems which examinees may have in expressing their thoughts in written form are difficult to separate from poor thinking per se, and therefore conclusions about the

individual's ability to think become entangled with writing abilities. Students who do poorly on such tests may not be poor thinkers, but poor writers.

Despite these difficulties, essay tests do offer advantages over multiple choice tests. Perhaps the most significant is that graders are given more direct information about the thinking students do while answering the questions. The reasons students provide for coming to certain conclusions give the marker a much clearer indication about the quality of the individual's thinking than does the selection of the correct answer on a multiple-choice test.

Essay tests also give graders a greater degree of flexibility in giving credit for good thinking. Unforeseen answers based on different background beliefs will not hide good thinking as they do in multiple-choice tests, and graders have the ability to give credit for novel answers based on sound thinking.

Essay tests allow for a greater degree of complexity in the questions they raise. While multiple-choice tests must eliminate all but a single variable in order to support the single-answer format, essay tests can include a number of relevant factors and require the test-taker to take each of them into consideration. The Ennis-Weir, for example, asks examinees to make reasoned judgments about

the strength of the arguments made by a writer in a
fictitious letter to the editor. The letter makes
empirical as well as meaning claims, is guilty of numerous
fallacies, yet also makes some sound claims. While the
test stresses the evaluative side of critical thinking,
students are asked to generate sound arguments in defense
of their claims about the letter. The complexity of this
type of question approximates more closely the complexity
of many of the issues individuals face outside the
classroom.

Although essay exams have a broader utility than multiple
choice tests, they are still limited in the types of
challenges they can present. Bailin, Case, Coombs, and
Daniels (1993) point out, for example, that
judgments-in-action -- challenges which occur while
producing an object or performing a feat -- and group
deliberations are not adequately measured by this type of
format. Most performances require this type of
on-the-spot critical thought, and it is impossible for
essay tests to demand this of students unless the actual
performance is one of writing an essay. Simply put, there
is a big difference between writing an essay explaining
how you would stop a fight in the playground, and actually
going out and doing it.

3.4 SHORT ANSWER APPROACHES

In this section two textbook series which claim to teach
and evaluate critical thinking will be reviewed to find
whether methods used in these programs evaluate students'
thinking more fully than the methods just examined. The
two series, Patterns of Civilization (Beers, 1984;
hereafter cited as Patterns) and Towards Tomorrow --
Canada in a Changing World (Bartlett, Craig, & Sass, 1989;
Dunlop, 1987; Morton, 1988; hereafter cited as Towards
Tomorrow with a designation indicating whether the text
focuses on history, geography, or government where
applicable) are prescribed for grades 8 & 9 and grade 11
respectively. The Patterns series includes two history
textbooks commonly used in junior high school Humanities
classrooms while the Towards Tomorrow series is made up of
three texts focusing on Canadian twentieth century
history, Canadian government, and world geography. These
series have been chosen because they are in common use
throughout British Columbia and because they claim that a
significant number of their activities require critical
thought.

To characterize the short answer critical thinking
activities found in these and other textbooks as an
'approach' is actually misleading since there is a wide
range of activities represented in current texts.

Patterns, for example, uses both multiple choice questions and the short answer format to assess students' critical thinking about historical events. Towards Tomorrow avoids the multiple choice format, but provides the classroom teacher with a variety of critical thinking activities ranging from short answer questions, to essay topics, to activities requiring an oral defence of a student's position.

The two Patterns texts each includes two full sets of chapter tests in the teacher's manual. The majority of the questions in these tests are multiple choice, but several short answer questions are included as well. A majority of these tests include a multiple choice section labeled "Critical Thinking" but a closer examination of the questions indicates that very little critical thinking is actually required in order for students to select the keyed answer on the test. Far from requiring critical thought, the questions demand little more than the recall of information explicitly laid out in the student text. A question for the chapter entitled "A New Age of Exploration", for example, asks the following critical thinking question:

> The growth of museums since 1945 has been largely due to
> a) computer technology.
> b) reduced literacy rates.
> c) the population explosion.
> d) greater interest in the past. (p. 190)

Little thinking is involved since the student text states:

> People are also taking a new interest in the past.
> Many records of the past have been stored in museums.
> (Beers, 1984, p. 181)

The answer to every question listed as requiring critical thinking actually demanded no such 'thing as long as the student had read the text and could recall what was written in it. While the failings of this one series is not enough to condemn textbook critical thinking in general, it does illustrate that generally accepted and widely used commercial series cannot necessarily be relied on to provide classroom teachers with adequate critical thinking materials.

As it happens, the section- and chapter-ending questions in the student text provide much better activities for assessing critical thinking than the tests do. Students are frequently expected to analyse and interpret quotations, charts, cartoons, and graphs and to formulate and justify opinions about a number of historical events. Activities ask students to identify qualities which would have helped a ruler get along with Parliament (p. 21); explain how the success of the American Revolution might still affect people in the world today (p. 21); evaluate Robespierre's claims that the Reign of Terror was necessary to save the revolution (p. 41); and study a map

to determine reasons why Austria's influence was greatest over southern German states (p. 81).

There are several reasons why these tasks can effectively be used to evaluate the critical thinking of students. One is that the tasks are actually 'critically demanding. The answers are not readily apparent nor will the simple application of a particular strategy provide a satisfactory answer. Answers are not explicitly stated in the text and thus the questions demand more than recall from students who normally rely solely on their memory to produce responses. The questions assume that students have acquired substantial knowledge of subject content, but they require the student to make critical judgments about the material.

Another characteristic of these activities is that demands for reasons are implicit in the tasks. The nature and wording of many of the questions imply that what is sought is a justification of the student's conclusions rather than any particular answer. The de-emphasis on correct answers is not only implied in the student text, but is acknowledged in the teacher's guide where answers to questions like those listed above generally begin with the reminder that "Answers will vary". The open-ended nature of the questions means that a number of answers can legitimately be justified in most cases.

The Towards Tomorrow series does not provide any tests as
such, but the teacher's guides provide numerous activities
identified as developing "Critical Thinking Skills". They
have been divided into three groups: (a) interpretive
skills which include analyzing maps, cartoons, graphs, and
statistics; (b) reasoning skills which include
summarizing, determining cause and effect, predicting,
creating questions and problem solving; and (c) evaluative
skills which include role playing, testing
generalizations, making comparisons and values
clarification. A broad spectrum of challenges is offered
ranging from interpreting cartoons (Craig & Perry, 1989,
p. 54), or developing procedures for settling labour
disputes (1989, p. 97), to creating support for a position
on the nationalization of a forest company (Cresswell,
1989b, p. 224).

Although some activities listed as requiring critical
thinking actually demand little more than recall since the
answers once again are written in the student text, the
majority of the tasks represent true critical challenges
for the reasons given in the previous discussion of the
Patterns series. The tasks are demanding, open-ended, and
focus on justification rather than on producing the
"right" answer. Several are also useful in assessing
dispositional factors in critical thinking. For example,
one activity (Cresswell, 1989b, pp. 123-125) centers on

the issue of Third World debt and asks students to propose plans for debt management on behalf of a debtor nation. It then asks students to devise a course of action on behalf of a Canadian bank owed money by a developing country. The activity assesses students' commitment to open-mindedness and willingness to look at a problem from contending points of view.

What distinguishes the two programs is that the Towards Tomorrow series appears to have placed more emphasis on having students represent their ideas using a variety of media. The activities in the Patterns series are centered almost exclusively on the traditional "pen-and-paper" mode, while the Towards Tomorrow program has students present their ideas in essays, paragraphs, editorials and charts as well as speeches, debates, mock interviews, bumper stickers, and role play. This latter approach not only gives poor writers an opportunity to produce ideas in another format, it more closely imitates the types of challenges students will face outside the classroom.

The goal of education is not to produce individuals who think well in contexts peculiar to the schoolroom, but to think well in their every-day activities. To this end, the Towards Tomorrow series has made some headway. Increased emphasis on generation and justification rather than on right-answer inculcation is an important step.

The fact that the activities in many cases presuppose substantial subject knowledge acknowledges that critical thinking is done in particular contexts and that the goal of teaching thinking isolated from any particular context is a hollow one. Some improvements, however, still remain to be made.

First, the activities in virtually all cases can be "wrapped up" in a relatively short period of time. Written activities are broken down into "mini-issues" so that a paragraph or two will be enough to resolve the question. If the room provided on worksheets is an indication, an evaluation of a New York Times comment on the battle at Vimy Ridge is to be completed in eight lines (Cresswell, 1989b, pp. 126). The previously mentioned questions on Third World debt are to be completed in nine lines each, and a question regarding automation and its effects on third world labour is given four lines (Cresswell, 1989a, p. 151-152).

The obvious solution of providing students with extra paper won't solve the problem, though, since the questions are constructed in such a way that in most cases a few lines will actually resolve the issue. Complex problems are broken down into simpler components so that students only have to deal with small parts of the whole issue. The history text of the Towards Tomorrow series, for

example, introduces the issue of foreign investment in
Canada and an accompanying worksheet traces the history of
foreign investment in this country (pp. 144-145).
Following a series of questions requiring the student to
interpret the graph comes the question: "Why, in your
opinion, does large-scale investment in Canada by
Americans make some Canadians uneasy?" The question, to
be answered in six lines, is part of the more complex
issue of the benefits and drawbacks of living next door to
a large, wealthy, influential neighbour, but the larger
question is never addressed. It is true that there are
reasons why Canadians are legitimately uneasy about
American investment, and to identify these drawbacks
requires critical thought, but it would seem that the
larger, more important and also more complex question
might be, "Are the drawbacks to being so closely tied to
such a wealthy neighbour compensated by the advantages?"

Breaking larger issues into smaller parts causes students
to focus more directly on key aspects of major questions,
but it certainly doesn't mirror complex, multi-faceted
issues faced outside of school. If textbook tasks are to
be used to evaluate students' critical thinking, the
challenges will have to be less neatly packaged and
contain a greater degree of complexity than Towards
Tomorrow currently exhibits. The questions will have to
deal with issues which cross traditional subject

boundaries and deal more with the "big picture" than with isolated parts of the whole.

A second short-coming of both programs is that the overwhelming emphasis is on critical thinking about beliefs rather than critical thinking about "doing" or "judgments-in-action". As was mentioned, there is a big difference between writing about "What you would do if..." and actually going about doing it. For example, several assignments in Towards Tomorrow have students writing editorials or letters to the editor of fictional newspapers. While these are valuable critical thinking activities since generating a reasoned viewpoint on an issue is implicit in the assignment, there are aspects to actually writing a letter which are missing from such an assignment. Were students to write real letters about related issues to the editor of a local paper, it would bring in dynamics not present in the "mock" letter. A goal of the assignment might be to actually get the letter published, so students would need to consider their arguments in a different light than if they were writing solely for their teacher. This would bring with it the added responsibility of convincing an editor of the importance of the issue in the first place -- a real-life problem not duplicated when the assignment is being completed for the person giving the assignment.

Realizing that they are possibly writing for a real audience of thousands rather than an audience of one brings additional considerations not present when the letter is only an assignment. Willingness to lay one's opinions on the line in front of people you might know assesses a student's commitment and willingness to take risks much more clearly than when there is less at stake. The possibility of having other readers respond to the same issue and possibly to the students' letters puts the debate in an entirely different light than when the issue is discussed only among classmates. While the debates, role play, and mock interviews are excellent activities demanding critical thought, they lack significant contextual richness due to their artificial nature and therefore fall short of having students meet the demands of true judgments-in-action.

The critical thinking activities represented in the two programs are too varied to be characterized in any single way, but they help highlight certain qualities which need to be present in critical thinking assessments. The activities in the Towards Tomorrow series, although not without shortcomings, resemble more closely the kinds of tasks people face outside the classroom. These tasks appear to be more realistic or more authentic -- a term which has been coined to refer to assessment activities designed to imitate contexts found in real-life

situations. The activities from the <u>Towards Tomorrow</u> series highlighted above begin to move toward authenticity -- a notion examined in detail in the next chapter.

3.5 SUMMARY

This chapter began by re-emphasizing the normative nature of critical thinking. Critical thinking is good thinking, and therefore to have made the right moves or followed the right "process" is not enough. To evaluate critical thinking, the products of thinking, including such things as arguments, reasons, and discussions of thoughts need to be evaluated against good thinking standards.

The focus then turned to three common methods of assessing students' critical thinking. The multiple choice test was found to be useful due to ease of marking, reliability, and effectiveness in singling out specific critical thinking terms and concepts. It doesn't, however, pick up on the generative and dispositional elements, nor does it account for differences in background beliefs, since reasons for choices can not be stated directly on most tests. The content of the questions is also simplistic to the point where the validity of the test has to be questioned.

The essay is useful since it allows students to justify their beliefs and thereby gives the marker direct

information about the student's reasoning. Teachers are able to give credit for good thinking even when the students' answers are novel and unforeseen. Essays are difficult to mark, however, and writing problems are difficult to distinguish from poor thinking. While essays allow for greater complexity than multiple choice tests, they do not test judgments-in-action nor do they allow for group deliberations.

An evaluation of short answer critical assignments produced mixed results. While some activities were found to be critically challenging others required nothing more of the student than recall. The two textbook series evaluated have a wide range of critical activities, and their variety and open-endedness capture both the productive and dispositional elements of critical thinking in different tasks. The straightforward, uncomplicated nature of the activities, however, fail to reproduce the complexity of the real-life tasks faced outside of school, and students are not required to think critically in action.

4. ALTERNATIVE AND AUTHENTIC ASSESSMENT

The previous chapter looked at critical thinking assessment from a variety of perspectives beginning with the multiple choice test and the essay and ending with a collection of tasks more effective than the first two, but also more difficult to characterize. While the multiple choice question and the essay each have a place in critical thinking assessment, a number of the activities found in the Towards Tomorrow series were found to imitate more closely the kinds of tasks performed by individuals in situations outside the classroom. Individuals seldom (if ever) are asked to take part in activities resembling a multiple choice exam, and while the essay is extremely useful in judging students' reasoning, it too is limited in imitating the contexts in which students will find themselves outside of school. The activities highlighted in the final section of the previous chapter are varied and difficult to characterize, yet it is this variety which is the first step toward imitating the kinds of critical challenges we face in our lives.

This chapter looks at the characteristics of these more authentic tasks and assesses their usefulness in evaluating student critical thinking in the classroom. In the first section of the chapter, I define authentic assessment as a collection of measures designed to be less

artificial and more connected to the challenges students face once they leave school. Standards for performance are known in advance and the students generally have increased control over assessment data. The context often determines the authenticity of an assignment and giving students control over contextual factors often increases the authenticity of the activity. I find no clear line between authentic and inauthentic tasks, but conclude that evaluation instruments range along a continuum from those which are highly authentic to those which are completely artificial.

Section 2 identifies four principles under which authentic assessment proponents operate. The first is that evaluations will be more accurate if they mirror real-world tasks, the second states that the effectiveness of an assessment will increase in direct relation to the meaningfulness the task has for the student, the third principles underscores the relationship between curriculum, assessment, and instruction, and the fourth principle stresses that assessment should be diagnostic, formative, and summative. Based on these principles, I then propose a set of criteria by which the authenticity of assessments can be evaluated.

Authentic assessments generally fall into one of three categories: naturalistic assessment, performance

assessment, and portfolio assessment. These are defined and compared in the fourth section. Each approach holds distinct advantages over others, but these strengths are sometimes balanced by certain disadvantages, both of which are discussed. Several cautions need to be expressed with regard to authentic assessments, and these are mentioned in the concluding sections of the chapter.

4.1 DEFINING AUTHENTIC ASSESSMENT

Dissatisfaction with traditional methods of evaluation is not peculiar to critical thinking advocates. Writers from a number of fields within education are calling for alternative methods of evaluating student performance (Hebert, 1992; Herman, 1992; Reithaug, 1992; Wiggins, 1989a, 1989b). Solutions to the assessment problem are varied and include methods such as "naturalistic assessment" (Reithaug, 1992), "portfolio assessment" (Wolf, 1989), "performance assessment" (Nuttall, 1992), "authentic assessment" (Mitchell, 1989), or simply "alternative assessment" (Maeroff, 1991). While each of these is unique in certain respects, they do share common elements: one of these elements is the common purpose of creating evaluative measures which more closely follow the real-world contexts for which schools are attempting to prepare students.

The term "authentic assessment" is generally applied to any measures designed to be less artificial and more connected to and patterned after the challenges students will face once they leave school. Wiggins (1989b) contrasts typical tests with authentic tasks in the following manner: (a) whereas typical tests require recall or plugging in of algorithms, authentic tasks require judgment, heuristics and style; (b) whereas typical tests tend to be disconnected and abstract, authentic tasks are "real-world; (c) whereas typical test items tend to include atomistic "bits", authentic test tasks focus on a complex "whole"; and (d) whereas typical test items are tangential and trivial, authentic tasks are "essential" -- that is, they strive to hit the core challenges of a particular field of study (p. 2).

An authentic test is thus any one the purpose of which is to match the conditions under which students will eventually have to make decisions about what to believe or which actions to take. Ideally, we might wish to have students produce products and performances which match, as closely as possible, those of recognized experts in certain fields, but decisions about beliefs and actions frequently are made by individuals who are not experts, but are simply consumers of knowledge. Decisions about issues ranging from the environment to child-rearing are made after considering the statements of experts, and the

individual must think critically about the degree of expertise of the speaker or vested interests she might have. Authentic tasks thus must account for critical thinking in contexts which match those under which experts might work as well as contexts where the thinking is directed toward assessing the veracity of contending claims of experts.

Authenticity often refers to the context in which a task is completed. Thus, restrictions placed on a student in terms of the time allotted for completion of a task, the number of resources he is allowed to refer to while writing, and the actual topic of his piece could all detract from the authenticity of the assignment. In authentic assessment, the locus of control regarding the topic, time, pacing, and conditions under which the assignment is completed remain as much as possible with the student and with constraints normally associated with that type of task. There are no clear lines separating authentic from non-authentic assessments; authenticity is spread along a continuum ranging from those tasks which are highly authentic to those which are completely contrived and artificial.

Numerous educators have wrestled with the notion of authenticity in assessment in hopes of both crystalizing the concept and justifying its importance. Foremost among

these is Grant Wiggins (1989a) who argues that a true test of intellectual ability requires "the performance of exemplary tasks" (p. 703). When he goes on to state that "authentic assessments replicate the challenges and standards of performance that typically face writers, business-people, scientists, community leaders, designers, or historians" (p. 703), he doesn't offer this simply as a definition of authentic assessment, but as an indication of what all assessments should do. This focus on real-world activities is one of four principles guiding the development of authentic assessment tasks.

4.2 PRINCIPLES GOVERNING AUTHENTIC ASSESSMENT

As Bateson (1992) points out, the purpose of educational measurement is not to trick, fool, or trip up students, but to provide the most accurate estimates and descriptions possible of the students' abilities (p. 5). Whether we are interested in critical thinking ability or running speed, the more artificial the conditions under which the measurements are made, the more tenuous are the conclusions drawn from the results. A basic principle underlying authentic assessment, therefore, is that a more accurate evaluation of students' competence will be achieved if the assessments mirror the real-world tasks they are being prepared for.

A second principle under which proponents operate is that the effectiveness of the assessment will increase in direct relation to the interest and meaning of the tasks students are expected to perform. Wiggins (1992) states that students should be so engaged in mastering the challenge that they lose sight of the extrinsic factors and motives at stake -- namely that evaluation is taking place (p. 28). Putting out a school newspaper for a journalism course may, from the teacher's perspective, be a valuable means of assessing student writing ability, but for the students it is a real task with challenges and rewards beyond its evaluative utility.

Wiggins (1992) cautions that the desire to engage students' interest should not overshadow the fact that matters of importance are still to be evaluated -- that evaluators are to "avoid turning important theoretical problems into crude utilitarian ones" (p. 28). He continues: "Many genuine problems do not have obvious practical value, but they nonetheless evoke interest and provide insight into student abilities."

A third principle of authentic assessment concerns the relationship between curriculum, assessment and instruction. It is a stated goal of authentic assessment that assessment reflect instruction as closely as possible (Baron, 1990; Cooper & Brown 1992, p. 41; Mitchell 1989,

p. 18; Wiggins, 1989a) and several writers emphasize that assessment plays an integral part in determining curriculum and setting instruction. Baron (1990) argues that due to increased accountability of teachers for their students' progress, tests have shaped the kind of instruction given to students. The effects of using multiple-choice tests, she observes, have been devastating.

> Where multiple-choice tests are used, they foster instruction that is broad rather than deep in scope, fragmented rather than holistic in form, and convergent rather than divergent in nature. Students are encouraged to memorize information, generally in the form in which they are exposed to it. They are not encouraged to take risks in their thinking. There is only one right answer to multiple-choice questions and the machines that score them are not equipped to read any comments that students might want to write about alternative interpretations, ambiguous items, or subtle or unusual ways to approach the questions. (p. 129)

Wiggins (1989a) stresses the importance that examinations have not just in monitoring student performance according to standards, but in actually setting those standards. He observes that tests are "central to instruction" and that tests and final exams "inevitably cast their shadows on all prior work" (p. 704). It is unavoidable that, given the importance placed on test results, they will influence curriculum and instruction. It is essential that curriculum, instruction, and evaluation each serve our aims in education. There must be complete harmony among

these three elements and, as Wiggins (1989a) asserts, assessments should be so closely tied to educational aims that teachers can comfortably "teach to the test" (p. 704). Evaluations should assess all the essential elements of the content and not simply that which is easily counted or observed (Wiggins 1992, p. 27). Assessment tasks should be an accurate reflection of curriculum and instruction not only because it is fairer to the students and presents a more accurate picture of their abilities, but because assessment inevitably shapes instruction and sways curriculum due to the high stakes involved.

A fourth principle undergirding authentic assessment relates to the purposes of evaluation. It often seems as if exams have become a static end-point of instruction, but central to the notion of authentic assessment is the idea that the results of such an evaluation are to be used to assist the teacher and student in mapping out a plan for continued achievement. This principle reinforces the strong connection between instruction and assessment previously mentioned.

While examinations have frequently been used simply to rank students, authenticity demands that assessment support learning with diagnostic, formative, and summative components. While comparisons among students may still be

undertaken, the examination is not only the end-point of a segment in a student's education, but a bench-mark in his continuing educational life. The test then not only discloses student capabilities to the assessor, but reinforces the actual challenges and standards of the subject matter for the student.

4.3 AUTHENTIC ASSESSMENT CRITERIA

Based on these principles, a number of criteria have been suggested which guide and determine the authenticity of assessment tools. While writers differ in subtle ways on the criteria for authenticity, the following seem to be represented in most conceptions.

1. An authentic assessment is one where the evaluation task is rooted in real-life contexts. The task is active rather than passive in nature and the conditions surrounding the assessment are as close to those normally found in the classroom as possible. Ideally, classroom conditions are themselves modeled after those found in real life situations. Duplicating problems found outside the classroom, tasks are open-ended and loosely structured, requiring the test-takers to define the problem and construct solutions rather than simply selecting them from packaged alternatives. Optimally, the tasks allow for diverse solutions and a variety of

strategies as students construct responses rather than make selections from among sets of possible answers.

2. An authentic assessment is one where the process or the reasons for taking certain actions are valued at least as much as the product of the students' efforts. The teacher's role is to watch the student identify, wrestle with, gather evidence for, find possible solutions to, and test the solutions to the problems given in the assessment. This necessitates that the teacher work closely with the student, not taking the role of final arbiter or judge when the assignment is completed, but working alongside the student and making observations for the duration of the process. Wiggins (1992) writes that "Evaluation is most accurate and equitable when it entails human judgment and dialogue, so that the person tested can ask for clarification of questions and explain his or her answers" (p. 704) thereby allowing the assessor clearer insight into the reasons and rationale behind the choices the student makes.

3. An authentic assessment is one which continues over time and allows for sustained effort and improvement rather than one which simply measures the students' capabilities at a specified point. In contrast to most testing formats, students should be allowed, and are

often encouraged, to return to previous work samples
with new insights they have gathered. The tester's
role therefore changes from one of technician working
under the restrictions put in place by the test
designer, to one of participant, constantly involved in
the student's work. As Mitchell points out, "the
entire process depends on the seasoned, generous
discrimination of teachers" (cited in Wiggins, 1989,
p.19).

4. Authentic assessment is committed to a holistic view of
evaluation where tasks examine major concepts or "big
ideas" rather than small bits of isolated information.
Wiggins (1992) points out that missing in many
evaluation tools is the demand for judgment on the part
of the student.

> Performance is not just doing simplistic tasks
> that cue us for the desired bit of knowledge.
> It entails "putting it all together" with good
> judgment; good judgment cannot be tested through
> isolated, pat drills. (Wiggins, 1992, p. 28)

5. Authentic assessment promotes self-evaluation as
necessary not only to assist in gaining accurate
information about the student, but as an integral
element in the student's learning process. The
evaluation process has traditionally been largely
removed from the experience of the student. Not only
do students rarely take part in evaluating their own

work, they frequently have little idea on what basis they have been assessed and therefore are unaware of what it is that they have to do in order to improve. Wiggins (1989a) writes:

> The true test of ability 'is to perform consistently well tasks whose criteria for success are known and valued. By contrast, questions on standardized tests are usually kept "secure," hidden from students and teachers, and they thus contradict the most basic conditions required for learning. (p. 706)

Authentic assessment encourages students to work side-by side with the teacher in evaluating their efforts. Not only does the assessor take responsibility for making sure that the criteria and standards for assessment are clearly known ahead of time, she often will include the student in setting those same criteria and standards. The goal is to make the student an active participant in the evaluation process, so students are encouraged to monitor their progress and continually make adjustments as their work nears completion -- a characteristic of how professionals carry out their work.

6. Authentic tasks are frequently collaborative in nature, because so much of what we want people to do requires varying degrees of group participation. The student and teacher collaborate on identifying criteria and standards of achievement, but the tasks themselves are marked by collaboration among students. Where

traditional assessment devices are marked by students working in relative isolation, authentic assessments are often characterized by communication and an atmosphere of cooperation. Group discussions in which the task is viewed from a number of perspectives are encouraged and seen as integral parts of both the instructional and evaluative process.

7. Authentic assessment is characterized by the variety of devices and tasks used to collect information about student achievement, and thus there is no one type of activity which can be pointed to as being typically authentic. The number of activities available for use as assessment tools is limited only by the number of real-world tasks associated with the field in question.

4.4 CATEGORIES OF ALTERNATIVE ASSESSMENT

Alternative assessment measures considered to be authentic vary tremendously, but generally fall into one of three categories: naturalistic assessment, performance assessment, and portfolio assessment. While such alternative forms of assessment are often authentic, they are not necessarily so, and their authenticity must still be judged on the basis of the criteria established earlier in this chapter.

4.41 Naturalistic Assessment

Naturalistic assessment is an informal, non-standardized approach where information on a social situation is collected with a minimum of disturbance to the subjects (Norris, 1986, p. 141).   Thus the teacher acting as participant-observer collects information about students while performing normal classroom duties. Information-gathering strategies include anecdotal records about actions and interactions of students, student-teacher conferences, and more systematic observations which might document the incidence of certain behaviours (Bailin et al., 1993, p. 33).

What distinguishes naturalistic assessment from other approaches is its nonintrusive format; what distinguishes it from the everyday interactions a teacher has with students when monitoring their participation and level of understanding, is the systematic way in which the instructor looks for and records certain qualities and characteristics in the student (Case, 1992, p. 17).

The advantage to this type of assessment is that since the student does not have a heightened sense of being evaluated, performance is likely to be more representative of what it would be in real-life contexts.   In other words, motivation to perform well or poorly would be determined more by interaction with the task and the other

students than by a desire to score well on any particular assessment. This approach, then, is particularly useful in assessing the attitudes and dispositions of the student toward particular types of tasks.

A disadvantage of this approach is 'that the teacher has reduced control over the direction of the assessment. The instructor may be seeking evidence of the student's ability to generate reasons to support a particular view, but for any number of reasons, the student may not actually take the opportunity to do so during the course of the class. Increased intervention on the part of the teacher may increase the focus of the assessment, but this moves it away from the fundamental characteristics of naturalistic assessment.

4.42 Performance Assessment

Performance assessment is a more structured approach to student evaluation where the performance of a feat or the creation of a product is the basis of assessment (Bailin et al, 1993, p. 32). The tasks are more complex than those normally associated with traditional assessment instruments, and are rich in contextual detail. The aim is to make the assignment as meaningful and as close as possible to the types of tasks normally faced in the real world. Wiggins (1992) sees rich contextual detail as the key and explains:

> A context is rich if it supports multiple
> approaches, styles, and solutions and requires
> good judgments in achieving an effective result.
> One must please a real audience, make a design
> actually work, or achieve an aesthetic effect
> that causes pride or dismay in the result.
> The test may be a contrivance, but it mustn't
> feel like one. (p. 27)

The effectiveness of performance assessment can be measured by the degree to which the participants can forget that the task is an assessment instrument. It is successful to the degree that the participants accept the premises, constraints, and "feel" of the challenge (Wiggins, 1992, p. 28) and work to meet the standards imposed by the activity itself.

For this reason, any constraints which exist should be necessary to the task at hand rather than tied to the demands normally associated with school tests. Typical constraints might concern (a) time limitations, (b) access to reference materials, (c) access to other people such as peers, experts, and test designers or judges, and (d) prior knowledge of the tasks and how they will be judged (Wiggins, 1992, p. 30). Traditional tests require tighter constraints on access to reference material, for example. The recall nature of most history exams demands that students be denied the use of any documents while writing an exam, but this is clearly far from the manner in which real historians actually do their work. A performance

task in history would place limits on the students similar to those imposed on actual historians.

Examples of performance tasks available to teachers are equal in number to the types of tasks people perform in their every-day lives. Debates, mock trials, concerts, and athletic events are all activities which can form the basis of an authentic performance assessment. The production of a school newspaper or an art display, and the construction of a bookcase or a school museum can each be used to assess student performance in a particular field.

What makes performance assessment particularly attractive is that it puts students face to face with the kinds of challenges they would be expected to meet outside the classroom. Some school departments have long recognized that students' abilities are not measured best by isolated drills, but by performances. Athletic departments, for example, measure success by their own type of performance -- the game. Music departments ultimately do not measure achievement in the classroom, but in the auditorium. Drama departments recognize that acting does not just involve the ability to memorize, or create facial expressions, or use body language, or project one's voice. It also involves a special kind of judgment which, when

judiciously applied, helps an actor "put it all together"
in a performance.

This approach to assessment recognizes that dentists,
voters, writers, mothers, and gardeners are also expected
to put it all together and structures assessment
instruments which require this of students.

4.43 Portfolio Assessment

The portfolio approach to assessment is patterned after
the types of collections which artists, writers, and
photographers compile as representative of their best
work. It basically involves the assessment of a number of
samples of student work, collected over a period of time
representing as thoroughly as is necessary the full scope
of work done by the student.

The portfolio could contain samples of student work,
teacher's observational notes, student self-evaluations,
and collaborative (between student and teacher) progress
notes (Valencia, 1990, p. 339). Items might include
reading responses and reading logs, selected daily work,
pieces of work at various stages of completion, tests, and
audio or video tapes.

Fisher (1992) sees portfolio assessment as a complex,
multi-faceted tool flexible enough to be adapted to a
variety of factors in the classroom. He has identified

three variables which form a framework for portfolio assessment: purpose/effect, context, and frame. Among the purposes (intended) and effects (unintended) of this type of assessment, he includes self-evaluation by the student, a focus on formative rather than summative assessment, reflection, ease of reporting, and de-centring the classroom.

Considerations such as the age, abilities, previous experiences, and attitudes of the students are all components of the assessment's context. How and by whom the portfolio is to be assessed is another component of context, as is the school, the community, and the classroom itself. The context affects not only the structure or form of the portfolio, but has to be taken into consideration when determining the purpose.

The frame is essentially the rules governing the creation of the portfolio including the extent to which the parameters themselves are established by the teacher or the student. Questions regarding selection and number of included materials, organization of the portfolio, and method of assessment are all part of the frame. Fisher says that frame can be viewed as a continuum with a loose frame (great degree of student control) at one end and a tight frame (great degree of teacher control) at the other.

The choices which can be made regarding each of these three variables give teachers a wide range of options in creating a portfolio suitable for their particular classroom. What makes the portfolio particularly effective as an instrument of evaluation is the degree of student involvement it allows. It views assessment as an integral part of instruction and encourages student participation in setting standards, monitoring progress, diagnosing strengths and weaknesses, and evaluating achievement according to relevant criteria.

4.5 CAUTIONS REGARDING AUTHENTIC ASSESSMENT

Despite the enthusiasm with which advocates embrace its use, authentic assessment brings with it several difficulties. The first is in the name itself. "Authentic" is not a neutral or value-free word, but a normative concept encompassing a set of assessment tasks with common characteristics and goals. There is danger, however, that all non-traditional or alternative assessment tasks be considered authentic, and that performance tasks, for example, automatically be used to replace other more traditional forms. "Authentic" implies a level of trustworthiness which cannot and should not automatically be assumed. As Bateson (1992,) points out:

> being labelled authentic does not carry with it
> immunity from careful examination. All
> measurement techniques, including those labeled

> as authentic, should undergo rigorous
> verification and scrutiny for validity and
> reliability considerations. (p. 5)

A second danger associated with the name is that labeling certain types of tests authentic implies that the remainder of tests are somehow inauthentic and therefore not genuine, trustworthy, or reliable (Bateson, 1992, p. 6; Peat, 1992, p. 52). This has caused the pendulum to swing away from nonauthentic evaluations to the point where much that is useful in those types of assessments is being discarded out of hand. As Bateson points out, there has to be a melding of the positive elements of authentic measurement with proven measurement applications.

4.6 SUMMARY

While different authors have treated the notion of authentic assessment in slightly different ways, certain salient features were found to have gained general acceptance. Assessments are termed authentic if they are designed to be less artificial and more closely patterned after the challenges faced in real-life contexts. Ideally, challenges are patterned after the tasks experts perform in their particular fields.

Several principles govern authentic assessments, namely that tasks should be (a) embedded in real-world contexts, (b) meaningful to students, (c) an integral part of

instruction, and (d) diagnostic, formative, and summative
in nature. From these principles a set of criteria for
authenticity are extracted. While there is no definite
line dividing authentic from inauthentic tasks,
assessments are authentic to the degree that these
criteria are followed.

Three categories of authentic assessments were identified,
namely, naturalistic assessment, portfolio assessment, and
performance assessment. The naturalistic approach was
found to be particularly effective in picking up on
critical thinking dispositions since through use of this
method students are not as aware that they are being
observed and therefore do not feel compelled to act out of
character. Performance assessments are not new to music
and physical education departments in schools, and give
students challenges which pattern themselves closely after
real-life tasks. Contextually rich, performance
assessments are successful to the degree that the feel of
the task allows the student to forget that it is an
assessment device. Portfolio assessments are patterned
after the types of collections of artists, writers, and
photographers. They generally contain various samples of
student work and evaluations of various types. Portfolios
were found to vary according to purpose/effect, context,
and frame giving the teacher numerous options for
designing a portfolio appropriate to particular students.

Despite enthusiasm on the part of authentic assessment advocates, cautions about uncritical use of the term were expressed.

5. AUTHENTIC ASSESSMENT AND CRITICAL THINKING

The assessment of students' thinking is currently being undertaken using measures which are inadequate when held up against the rich conception of critical thinking adopted in this thesis. While multiple-choice tests, essay assignments, and the text-book assignments reviewed in Chapter 3 have their place in critical thinking assessment, they fail, on their own, to take into account essential features of the concept.

Chapter 4 examined an alternative approach to assessment which holds much promise in broadening the array of assessment tasks at the disposal of the classroom teacher. In this final chapter, the potential of authentic assessment in evaluating critical thinking is more closely examined. In doing this, the three categories of assessment tasks outlined in Chapter 4 are evaluated. Naturalistic assessment, portfolio assessment, and performance assessment are each examined in terms of their effectiveness in evaluating student critical thinking. Naturalistic assessments are found to be effective in assessing critical thinking in unstructured or informal classroom settings and are judged useful in evaluating critical thinking dispositions. The naturalistic method of assessment also provides the teacher with information regarding the progress students make over time.

Portfolio assessments provide valuable information about the progress students make in their reasoning, but also are the focal point of critical thought in themselves. Students can be involved in establishing criteria and standards for the portfolio itself and will be active in evaluating the items they decide to include. The portfolio is useful as a point of discussion for meetings with students and parents, once again giving students the opportunity of assessing the thinking demonstrated by their work.

Performance assessment refers to student evaluation based on the completion of specially set, complex tasks (Case, 1992). They are patterned after the kinds of tasks normally performed by experts in any particular field or the kinds of challenges faced by people in their every-day lives. Performance tasks are unlike the neat, tidy kinds of problems often found in textbooks or standardized tests in that they bring together a number of factors from a range of subject areas. Students are expected to make judgments about various types of issues in the course of resolving the challenge. Two authentic performance challenges are examined in the chapter and features of authenticity are identified in each. A student response to the latter challenge is posited and the response is analyzed for its critical content.

5.1 NATURALISTIC ASSESSMENT

Naturalistic assessment has been characterized as a purposeful, systematic, documented evaluation of students as they go about the normal activities of the classroom. It is far less intrusive than other forms of evaluation since the teacher's role primarily becomes one of observer/recorder. This type of assessment can take place in a broad range of classroom environments, and its effectiveness lies largely in the fact that the students' behaviour is not unduly altered by the assessment. Under these conditions, student responses to the activities of the classroom match more closely their responses to similar challenges in contexts outside of school.

The demand for reasons is an integral part of critical thinking, and the classroom provides numerous opportunities for teachers to assess the reasoning of the students. Students working on a group project, for example, will undoubtedly generate conflicting views on how to proceed. By observing the way in which opinions are defended, the teacher can assess the quality of reasoning done by the participants and keep anecdotal records for use in student-teacher and parent-teacher conferences. Individuals who tend to provide adequate justification for their views or actions when working with their peers in the classroom are more likely to do the

same with peers in other settings. By the same token, students who passionately defend irrational views in loosely controlled settings at school aren't likely to become more rational when they enter their own home or place of work.

Naturalistic observations not only give valuable insight into students' ability to formulate sound arguments, they provide information about students' willingness to think critically. They are perhaps most effective in assessing the dispositional aspect of critical thinking. Class discussions, both formal and informal, are ideal times to assess a student's willingness to be open-minded -- to entertain opposing points of view and to adjust his views when the evidence warrants it. Controversial topics test students' tendency to withhold judgment and their commitment to bring good reasons to bear on an issue. This type of assessment also clearly identifies those students who participate in activities only when they sense that they are being evaluated. A student may construct a sound argument when involved in a formal debate, yet unless he also shows a willingness to involve himself in the exchange of ideas in the normal operation of the classroom, one can question whether his ability to reason will be used in his daily activities as a citizen.

Dispositional qualities can effectively be recorded by using checklists which identify key traits and attributes. A teacher might have a record sheet for each student, for example, and make notations either during the class or immediately thereafter. What is key to this method, however, is that the assessment take place over time. One-shot naturalistic assessments are no more authentic than one-shot multiple choice exams, and to judge a student's ability to reason based on one class discussion, or to assess an individual's willingness to entertain opposing viewpoints based on one observation is both unfair and unreliable.

This method of assessment is not only useful in the classroom, it is ideally suited for settings beyond the schoolgrounds. Student interactions with peers, teachers, and professionals on field trips, for example, provide sound information about students' critical thinking in a variety of contexts. I had the opportunity of traveling to California by bus with a high school band and choir this year, and gained new respect for a girl whose friend had an allergic reaction to something while sight-seeing along the San Francisco waterfront. Though relatively quiet and unassuming in class, she took charge of the situation, got help for her friend, and reasoning correctly that the band leader and I were more likely to be eating than shopping, found us holed up in a seafood

restaurant. My assessment of her ability and willingness to think critically in-action has changed.

The strength of this form of assessment is also its weakness. The fact that students have free reign to become as involved in critical activity as they wish improves the validity of the assessment, while at the same time reducing its focus. We might all wish to find out how students react to a particular crisis, but not everyone in the class will be in the right place at the right time, and we are unlikely to find someone willing to go into allergic shock in order to help us. While dispositional qualities and a degree of critical thinking ability can effectively be evaluated by this method, more focused evaluations must be made using other instruments.

5.2 PORTFOLIO ASSESSMENT

The portfolio approach is more structured and often more teacher-directed than naturalistic forms of assessment, but like naturalistic methods, gives valuable information about students' critical thinking. The direction or focus of the portfolio is clearly laid out ahead of time. In a portfolio where the criteria are set by the teacher (a tighter frame), the student could be instructed, for example, to include a piece of written work which contained a sustained argument about a particular issue.

The student might decide to include an essay which could then be used to assess the student's ability to justify a particular point of view. The student could also be asked to include a written critique of an article she'd read, an evaluation of a political speech or a party platform, a chart highlighting key differences between political ideologies, an evaluation of a classmate's oral presentation, and a self-evaluation of her term project. Each of the items would meet criteria which would themselves match course objectives. Upon completion, the portfolio provides a cross-sectional view of the student's quality of thinking in a variety of tasks. Errors in judgment on one assignment are balanced by sound reasoning in others, giving the teacher a multi-dimensional view of student achievement. Unlike one-shot evaluations, this method provides a much fuller account of student thinking ability.

Student weaknesses in areas of reasoning are more easily identified in portfolios than in naturalistic assessments since the criteria can be set to make similar requests in different contexts. A student who fails to distinguish fact from value claims, for example, would make this type of error on several assignments and make identification of the problem that much easier. This again assists teachers in distinguishing between cases of poor thinking which are

out of character, and those which are representative of
the student's everyday performance.

In addition to providing a broad view of student
achievement in different types of assignments, portfolios
can produce an accurate assessment of student work over
time.   Essential to monitoring progress toward long-term
goals, this feature of the portfolio can be utilized in a
variety of ways.   One method would be to give the same
type of assignment at different times of the year, and
then to compare the results.   Students might be asked at
three points in a semester to respond to newspaper
articles exhibiting faulty reasoning and to include each
of the three in their portfolio.   If the articles were
carefully selected, the assignments could provide an
excellent assessment of the students' improving ability to
critique poorly structured arguments.   Assignments could
also increase in difficulty over the course of the term,
and the resulting portfolio would again provide evidence
of the student's thinking progress.   Used in this way,
portfolios are seen not as final judgments, but as
checkpoints on the way to a larger goal.

Thus far I have examined portfolios and their role in
providing collections or samples of student products.
Portfolios have another function, however, in that they
themselves can become the object of student critical

thought. Portfolios with a loose frame can afford
students the opportunity of producing the list of items to
be included in the assessment, for example. Within
general guidelines outlining the number and variety of
assignments, (eight assignments, four from each semester,
not more than four essays) students can be instructed to
select samples of what they consider to be their best
work, and include a set of criteria by which they have
made their selections. The thinking therefore does not
end once the individual assignments have been completed,
but continues as students determine the best criteria to
use and decide which samples of their work fit those
criteria. The accompanying list of criteria can also be
quite revealing in that students will have to justify
their selections on the basis of their conception of "good
thinking".

The portfolio can also be used as a focal point for a
student/teacher conference. Students can reflect on work
done at various points of the term and can themselves
single out areas of progress as well as those requiring
improvement. The student in this way retains a sense of
ownership over her work, and takes responsibility for
evaluating her own thinking -- an important component of
Paul's strong sense of critical thinking.

The effectiveness of the portfolio as a means of assessing student thinking is limited primarily by the quality of its contents. It can be used to assess as wide a range of critical thinking competencies as the assignments it contains, but it also provides a structure for assessing particular strengths and weaknesses, for establishing overall reasoning, for monitoring progress over time, and for fostering reflection and self evaluation.

5.3 PERFORMANCE ASSESSMENT

Designers of critical thinking performance tasks strive to create challenges which force students to imitate as closely as possible the work and thinking of experts. Thus, performance tasks are characterized by their close match to the kinds of performances and products produced by experts in a particular field. What characterizes such tasks is the fact that, unlike the short answer textbook assignments reviewed in chapter 3, these tasks integrate a range of competencies into one critical challenge. They allow educators to watch students collect evidence, construct arguments, and take action as they tackle questions which are ambiguous rather than neat and tidy (Wiggins, 1989a). With authentic performance tasks, the reasoning element is built into the assignment.

A high school history teacher, wishing to assess her students' critical thinking in a history context might, for example, wish to reinforce the idea that historical events are always perceived and recorded from a particular perspective. She could identify three perspectives[1] of the War of 1812 -- the Canadian, the British, and the American -- and have the class come up with three different newspapers representing the three different perspectives. To ensure that students view events from each of the perspectives, individual students would be asked to contribute articles on select events to each of the three papers, ensuring that they saw the event from each of the points of view. They would be expected to (a) locate sources which support each perspective, (b) present facts in such a way as to promote particular perspectives, and (c) avoid obvious distortions of widely accepted facts.

One factor which makes this activity a critical challenge is that it places the students in the position of having to view events and make judgments from different points of view. Students are forced to make judgments about historical claims -- which claims would have come out of the British camp, for example, or which ones would have been disputed by the Upper Canadians. They will reach conclusions about who would have been considered a reliable witness and whose testimony would have been

ignored. They will also have to come to a determination about the role of the newspaper in deciding what is "news" under certain conditions. Judgments will be made about which version of an event should be reported, which stories should receive front page coverage, and which ones should be buried in the back. Headlines will be written, pictures will be produced, and decisions will be made about article length and type size based on judgments about their effectiveness in presenting a particular perspective on the event.

There is no clear line between tasks which are authentic and those which are not, so rather than making a definite judgment about the authenticity of a task, it is more useful to identify features which are included to increase the authenticity of the project. This particular task is made more authentic by giving editorial control of the papers to the students. They can, for example, decide which particular community each paper will represent, and can include stories with a local flavour. The additional research this requires gives them insight into the values of each community, as well as increasing the depth of their judgments about how the different communities would have perceived events of the war. The teacher is involved throughout the project by reinforcing standards, maintaining focus, and introducing questions which move students into areas they have not yet considered. This

involvement also gives the teacher an excellent opportunity to make observations of students' thinking in a non-threatening, collaborative environment.

The long-term, multi-faceted nature of such an assignment increases its authenticity by forcing students to use judgment in putting it all together. Unforeseen difficulties associated with locating information, working with classmates, and organizing the paper itself broaden the range of judgments the students make. It gives them the opportunity to formulate an idea, evaluate the result, and make changes as the project nears completion. They evaluate their own work and the work of others; they defend their own views and learn to accommodate the views of their classmates in an environment where the standards of good thinking are challenged and reinforced by their peers and their teacher.

Providing a real audience for such tasks increases their authenticity. While producing the papers for their classmates and the teacher might provide sufficient motivation for some students, having the newspapers printed in larger quantities and distributed around the school, or sold as part of a larger Canadian history museum display adds another dimension to the assignment. Finished products can also be placed in the regular library collection to increase readership and leave

students with the feeling that the work they are doing is of importance to a wider range of people.

Authentic tasks begin to blur the boundary between instruction and evaluation, and increasing the authenticity of the task makes it an even more effective tool for both teaching and evaluating critical thinking. While it is possible to have students interpret events from different perspectives in an essay assignment, placing the assignment in a newspaper context adds interest to the task and calls for judgments less likely to be demanded by an essay. Practical details such as article length, audience, and page layout muddy the task enough that new critical judgments will be required of the writer. The task requires students to demonstrate both the generative and the evaluative dimensions of critical thinking, and also gives the teacher first-hand information about students' critical thinking dispositions -- their willingness to act on the basis of reasons, their consideration of differing points of view, and their willingness to take or change positions when reasons or evidence demand it.

As was discussed in Chapter 2, critical thinking is judged according to various relevant standards. Critical thinking is not just the lock-step application of a series of mental moves, nor is it one's arrival at a

predetermined correct answer. An individual's thinking is critical to the degree that her judgments are supported by reasons which meet the criteria of good thinking in a particular context. Attempts to find neutral or general contexts for test questions are not often helpful since what counts as good thinking varies from one context to another. For this reason, it is vital that students' thinking is assessed in varied contexts which themselves are thought to be important. Tasks need to be embedded in contexts which are considered to be of educational value. Unfortunately, not all students care about what educators care about, and therefore the challenge in creating authentic challenges is to design tasks which test thinking in educationally valuable areas while including contextual elements which capture students' imagination. This may not always be possible. It is, for example, quite likely that some students care little about either the War of 1812 or newspapers, and are as difficult to motivate on this assignment as they would when asked to write an essay. This does not negate the usefulness of such a project, but it does point out the importance of using a variety of measures when assessing student thinking. The student's capacity to think well will not be tapped by every assignment, but increasing the variety of challenges will increase the likelihood that each individual will at some point be engaged by his work.

Most high school students are introduced to a variety of
forms of persuasive writing at some point.   While many
will   see   the   political   application   of   propaganda
techniques to be of little consequence in their lives,
most will recognize these tactics' when placed in an
advertising context. As an assessment of their ability to
identify   and   think   through   persuasive   techniques   in
advertising,   the   class   might   be   given   the   task   of
comparing the competing claims of four different stereo
outlets.   To add authenticity to the task, students would
assume   that   they   were   in   the   market   for   new   sound
equipment and work under the following guidelines:

a) they would be given a hypothetical budget;

b) they would be restricted to purchasing sound
equipment, but could purchase any combination of
components they wished;

c) they would be given a reasonable amount of time
within which to make their choice;

d) they would be required to go to the stores and
talk to the salespeople as if actually making a
purchase;

e) they   would   be   encouraged   to   go   beyond
price-comparing in making their selections;

f) they would compare their findings with the claims
of the advertisements;

f) their findings would be published in the school newspaper.

The assignment would be evaluated on the basis of the depth of the investigation and the soundness of the reasoning. Questions which a teacher might ask in evaluating such an assignment would include:

1. Did the student take important factors into consideration when evaluating advertising claims and making a 'purchase'?

2. Did she depend on reliable sources when making her judgments?

3. Was she reasonable in determining the relative importance of different factors or did she uncritically give all factors equal weight?

4. Were her methods of investigating the advertising claims adequate and uniform from store to store, or were there inconsistencies in her strategies?

5. Did her inquiry appear to be balanced and fair, or was their evidence that she favored one store from the outset?

Student reactions to such an assignment would vary, but a critically thinking student might respond in the following manner. An assessment of the response might look at

evidence of critical deliberations, dispositions, and strategies.

1. The student collects advertisements from various newspapers and flyers ensuring that she has information on each of 'the stores she is investigating.

(Disposition: Attempts to be well informed.)

2. She makes initial comparisons between the advertisements and begins to organize the claims made by the different stores. For example, some stores may focus on their low prices while others may highlight their after-sales service.

(Critical deliberation: begins systematic analysis of different arguments. Strategy: breaks complex issue into organized parts.)

3. She speaks to different school-mates who have recently purchased sound equipment. One student makes mention of having gone to one of the stores to purchase a sale item but having left with a more expensive component after finding out that the sale item was an older model. She decides to speak to the manager to find out if most sale items are less desirable models or if her friend's experience was unusual.

(Disposition: Open-mindedness - withholds judgment. Critical deliberation: recognizes that one instance provides insufficient evidence on which to base conclusion; determines credibility on the basis of responsibility and expertise.)

4. She decides to price-shop by phone, but after contacting each of the stores, she realizes that comparing advertised prices is virtually impossible since many advertised products are "house" models not sold by the stores' competitors. She decides to speak to a friend of her father's who is a sound engineer. After their meeting, she compiles a list of essential features she is going to look for when purchasing a stereo.

(Disposition: changes a position when evidence demands it; focuses on relevant information. Critical deliberation: recognizes that independent variables must remain constant in order for valid comparisons to be made; recognizes that vested interests make store employees less reliable about certain judgments than an impartial expert)

5. After speaking to sales people at each of the stores, she realizes that differences in price can result from factors other than differences in equipment. She then compiles a list of other

contributing factors, and after thinking over the advice offered her at the stores, arranges them in order of importance.

(Disposition: strives to be impartial and well-informed. Critical deliberation: recognizes that a conclusion based on one factor is not enough; recognizes that judgments need to be based on information from all concerned parties when impartial experts are not available; formulates questions which clarify and challenge; makes judgments on the basis of sound reasons.)

6. She takes the notes and lists which she has compiled and compares them with the advertisements she's collected. She realizes that the advertisements focus on factors which make the stores look good, but often fail to mention mitigating factors like limited quantities, old stock, or restrictive warranties which make the 'specials' less special. She also realizes that some of the advertisements imply certain things without really making them explicit.

(Disposition: pays close attention to detail. Critical deliberation: analyzes the evidence and recognizes contradictory claims; recognizes and identifies techniques of persuasion; makes critical judgments about actions and intentions)

7. Using her hierarchy of factors as a guide, she
   types out her report making sure that her
   evaluations and recommendations are well
   supported.   Realizing that she failed to get
   information about warranties from one of the
   stores, she qualifies her conclusions and mentions
   the oversight in her report.

(Disposition: commitment to basing conclusions on
sufficient evidence; commitment to intellectual
honesty. Critical deliberation: balancing factors
of varying kinds against one another while making
a critical judgment; deciding on a course of
action based on good reasons.)

While different students will react in different ways to
such a task, the foregoing analysis provides an indication
of the kind of critical response such an activity might
generate. Recognition of persuasive techniques such as
emphasizing or exaggerating favourable characteristics,
downplaying or ignoring unfavourable features, using
persuasive language, and making vague promises has
application in a number of areas, and is of considerable
educational value. Investigating such techniques in a
meaningful context, directed at a real audience, within
nonarbitrary limits is at the root of this assignment's
authenticity and its relevance to students. Many high
school students are faced with making decisions about such

purchases, and this type of task assesses the students' ability to critically evaluate certain written claims while weighing the relative values of different types of criteria -- price, quality, service, maintenance, reputation, etc. -- in a context that is meaningful. Assignments such as this one may ultimately be contrived, but they needn't feel like it. Students can take a number of approaches to completing the task, but the assessment of their work is based on the quality of the investigation and the judgments made along the way rather than on arriving at a "correct" answer.

This type of assignment also addresses the generative aspect of critical thinking. Having students generate criteria by which to judge advertising claims is clearly quite a different matter from selecting relevant ones from a list, for example, and the act of production mirrors more closely what people must do in real life. The judgments made in these less well-defined contexts are more difficult, but the weighing of various factors is a critical challenge which effectively assesses critical thinking. Authentic tasks provide opportunities for students to become vitally involved in important issues, but they also allow students to remain on the periphery of such activity. It is precisely this flexibility which gives authentic assessments their effectiveness in measuring the dispositional aspect of students' thinking.

Their willingness to take risks, their tendency toward involvement, and their drive to bring evidence to bear on their own and others' beliefs are assessed effectively in the atmosphere characteristic of authentic tasks.

5.4 SUMMARY AND CONCLUSIONS: AUTHENTIC ASSESSMENT AND CRITICAL THINKING

While authentic assessment is not a critical thinking panacea, it can play an integral part in assessing student thinking. Whereas student knowledge and understanding of specific aspects, terms, and strategies can effectively be tested using objective measures such as multiple choice tests, the complex nature of critical thinking demands that additional, more comprehensive evaluation measures are needed. While student reasoning can adequately be assessed using the essay format the types of judgments-in-action made by individuals on a daily basis are not addressed by that format. And while textbook programs which promote critical thinking as an educational goal provide tasks within educationally sound contexts, many of the activities don't provide an adequate fit to the ill-structured and complex issues most people face every day. Making assessment tasks more authentic addresses these concerns. An assessment array which includes a combination of naturalistic, portfolio and performance measures will evaluate those features of

critical thinking not adequately measured by other methods.

Those who adopt more authentic measures for classroom use must, however, use caution. For one thing, both the design and ultimate grading of student work in such assessments require a substantial amount of expertise and judgment on the part of the teacher. The use of authentic assessments in the evaluation of student thinking not only requires that teachers are knowledgeable about critical thinking, it demands intensive staff development to ensure that teachers have a good understanding of the principles and criteria of authentic assessment. Unlike more objective tests, authentic assessments measure the critical thinking of the teacher, and judgments about student thinking are limited by the judgment of the person doing the grading.

Time constraints will also severely test the resolve of teachers wishing to increase the authenticity of assessment tasks. Both the planning and the completion of such tasks is more time-consuming, threatening those areas of the curriculum where breadth rather than depth drives instructional practice. Designing a newspaper takes far more time than writing an essay, and analysing the practices of stereo outlets isn't done as quickly as memorizing a list of persuasive techniques, so the

benefits of assessing authentically will have to be weighed against the losses in coverage.

It should also be remembered that authenticity is only one of many educational values. A realistic and possibly the most efficient approach to purchasing a stereo might be to get a knowledgable friend or relative to do it for you, yet this method would fail to meet certain important educational goals. Having the student work through the problems herself and placing constraints on her to ensure that educational aims are met will in many cases override the benefits of strict authenticity.

A final comment centres on the nature of authenticity. As has been mentioned previously, there is no definite line between authentic and inauthentic measures. The authenticity of an assessment is a matter of degree, and has a great deal to do with the purpose of the evaluation. Case (1992) observes that under the right circumstances, any form of evaluation can be an authentic assessment. He points out that "a timed multiple-choice test of general knowledge may be the most authentic measure of the ability to succeed at 'Jeopardy' or at some other quiz show" (p. 22). Conversely, the most richly conceived, elaborate assessment instrument would be inauthentic if applied to the wrong purpose. This thesis, while providing justification for the use of authentic assessments in

evaluating critical thinking, is not suggesting that other methods be discarded out of hand. Authentic assessments need to be added to teachers' repertoires of assessment instruments to meet more effectively the demands of critical thinking evaluation. The adequate assessment of students' ability to reason well ultimately depends not upon the use of one particular assessment device, but upon the teacher's use of good judgment in selecting from a variety of evaluation measures those which best suit her educational purposes.

<u>NOTES</u>

CHAPTER TWO

1. Critical thinking is more than just the evaluation of
   arguments that happen to come our way; it includes both
   the inclination and the ability to search out
   considerations that are relevant to an issue. The
   ultimate objective in critical thinking is not to grade
   another's argument, but to determine whether to believe
   or do what that person would have us believe or do.
   This is a task that sometimes begins only after we have
   appraised the argument we've been given.

2. McPeck claims that as such it is impossible to
   distinguish critical thinking from "intelligence".

3. "Distinguishing relevant from irrelevant information"
   is one of the critical skills listed in Beyer (1988).

4. These examples are taken from Ennis's list of critical
   thinking abilities (1985, p. 46).

5. Paul (1985) questions whether blindly memorized true
   belief can even be considered knowledge at all (p. 38).

CHAPTER 3

1. See Moore & Parker (1986), "Exercises", for example.

2. Level X is an easier test, primarily designed for use
   in junior and senior high school and first year college
   while Level Z is geared for university level students
   and adults.

CHAPTER 5

1. This assignment creates a dilemma which the teacher and
   students would have to resolve. The Native perspective
   is missing, but not out of oversight. Native people
   did not communicate through newspapers and including a
   paper which presented the Native perspective in the
   assignment would detract from the authenticity of the
   task. Simply ignoring that perspective would reinforce
   what many history textbooks have done over the years,
   so the class would have to decide how best to present
   the Native view while keeping the feel of the task
   authentic.

BIBLIOGRAPHY

Bailin, S., Case, R., Coombs, J., & Daniels, L. (1993). A Conception of Critical Thinking for Curriculum, Instruction, and Assessment. Victoria BC: British Columbia Ministry of Education.

Baron, J.B. (1990). "Performance Assessment: Blurring the Edges among Assessment, Curriculum, and Instruction." In A.B. Champagne, B.E. Lovitts, & B.J. Calinger (Eds.), Assessment in the service of instruction. Washington, DC: American Association for the Advancement of Science.

Baron, J.B., & Sternberg, R.J. (Eds.). (1987). Teaching Thinking Skills: Theory and Practice. New York: Freeman.

Barrow, R. (1987). "Skill Talk." Journal of Philosophy of Education, 21(2), 187-195.

Barrow, R. (1990). Understanding Skills: Thinking, Feeling, and Caring. London, ON: Althouse Press.

Barrow, R., & Milburn, G. (1990). A Critical Dictionary of Educational Concepts (2nd ed.). New York: Teachers College Press.

Bartlett, B., Craig, R., & Sass, G. (1989). Towards Tomorrow: Canada in a Changing World - Government. Toronto: Harcourt Brace Jovanovich.

Bateson, D. (1992). "Adding Authenticity to Sound Measurement Practice." Research Forum, 9, 5-8.

Beers, B.F. (1985). Patterns of Civilization (Vol. 1&2). Scarborough, ON: Prentice-Hall.

Beyer, B.K. (1984a). "Improving Thinking Skills - Defining the Problem." Phi Delta Kappan, 65(6), 486-490.

Beyer, B. K. (1984b). "Improving Thinking Skills - Practical Approaches." Phi Delta Kappan, 65(8), 556-560.

Beyer, B. K. (1985a). "Critical Thinking: What is it?" Social Education, 49(4), 270-276.

Beyer, B. K. (1985b). "Critical Thinking Revisited." Social Education, 49(4), 268-269.

Beyer, B. K. (1985c). "Teaching Critical Thinking: A Direct Approach." Social Education, 49(4), 297-303.

Black, M. (1952). Critical Thinking. Englewood Cliffs, NJ: Prentice-Hall.

Brandt, R. S. (1992). "On Performance Assessment: A Conversation with Grant Wiggins". Educational Leadership, 49(8), 35-37.

Brell, C.D. (1990). "Critical Thinking as Transfer: The Reconstructive Integration of Otherwise Discrete Interpretations of Experience." Educational Theory, 40(1), 53-68.

Case, R. (1992). "On the Need to Assess Authentically." Holistic Education Review, 5(4), 14-23.

Chipman, S.F., Segal, J.W., & Glaser, R. (Eds.). (1985). Thinking and Learning Skills (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Cooper, W. & Brown, B.J. (1992). "Using Portfolios to Empower Student Writers." English Journal, 81(2), 40-45.

Craig, R., & Perry, T. (1989). Towards Tomorrow - Canada in a Changing World: Government (Teacher's Guide). Toronto: Harcourt Brace Jovanovich.

Cresswell, J. (1989a). Towards Tomorrow - Canada in a Changing World: Geography (Teacher's Guide). Toronto: Harcourt Brace Jovanovich.

Cresswell, J. (1989b). Towards Tomorrow - Canada in a Changing World: History (Teacher's Guide). Toronto: Harcourt Brace Jovanovich.

De Bono, E. (1984). "Critical Thinking Is Not Enough." Educational Leadership, 42(8), 16-17.

Diez, M.E. and Moon, C.J. (1992). "What do we Want Students to Know?... and Other Important Questions". Educational Leadership, 49(8), 38-41.

Dunlop, S. (1987). Towards Tomorrow: Canada in a Changing World - Geography. Toronto: Harcourt Brace Jovanovich.

Ennis, R.H. (1962). "A Concept of Critical Thinking." Harvard Educational Review, 32(1), 81-111.

Ennis, R.H. (1980). "A Conception of Rational Thinking."
     In J.R. Coombs (Ed.), Philosophy of Education 1979:
     Proceedings of the Thirty Fifth Annual Meeting of the
     Philosophy of Education Society (pp. 3-30), Normal IL:
     The Philosophy of Education Society.

Ennis, R.H. (1981). "Rational Thinking and Educational
     Practice. In J.F.Soltis (Ed.), Philosophy and
     Education (pp. 143-183), Chicago: National Society for
     the Study of Education.

Ennis, R.H. (1985). "A logical Basis for Measuring
     Critical Thinking Skills." Educational Leadership,
     43(2), 44-48.

Ennis, R.H. (1987). "A Taxonomy of Critical Thinking
     Dispositions and Abilities." In J.B.Baron &
     R.J.Sternberg (Eds.), Teaching Thinking Skills: Theory
     and Practice, (pp. 9-26), New York: W.H. Freeman.

Ennis, R.H. (1989). "Critical Thinking and Subject
     Specificity: Clarification and Needed Research",
     Educational Researcher, 18(3), 4-10.

Ennis, R.H. (1990). "The Extent to Which Critical Thinking
     is Subject-Specific: Further Clarification."
     Educational Researcher, 19(4), 13-16.

Ennis, R.H. (1992). Directions for Scoring Cornell
     Critical Thinking Test, Level WJX (Experimental
     Version). Unpublished draft.

Ennis, R.H. & Millman, J. (1971). Cornell Critical
     Thinking Test, Level X. Urbana, IL: Critical Thinking
     Project, University of Illinois.

Ennis, R.H. & Millman, J. (1993). Cornell Critical
     Thinking Test, Level WJX (Experimental Edition #11).
     Unpublished.

Ennis, R.H. & Weir, E. (1985). The Ennis-Weir Critical
     Thinking Essay Test. Pacific Grove, CA: Midwest
     Publications.

Fisher, D. (1992). "Portfolio Assessment: A Framework for
     a Range of Options." Research Forum, 9, 9-12.

Griffiths, M. (1987). "The Teaching of Skills and the
     Skills of Teaching: A Reply to Robin Barrow." Journal
     of Philosophy of Education, 21(2), 203-214.

Hawes, K. (1990). "Understanding Critical Thinking." In V.A. Howard (Ed.) <u>Varieties of Thinking: Essays From Harvard's Philosophy of Education Research Center</u>, New York: Routledge.

Hebert, E. (1992). "Portfolios Invite Reflection -- From Students and Staff." <u>Educational Leadership</u>, 49(8), 58-61.

Herman, J. (1992). "What Research Tells Us About Good Assessment." <u>Educational Leadership</u>, 49(8), 74-77.

Hirst, P.H., & Peters, R.S. (1970). <u>The Logic of Education</u>. London: Routledge & Kegan Paul.

Johnson, R.H. & Blair, J.A. (1983). <u>Logical Self Defense</u>. Toronto: McGraw Hill Ryerson.

Johnson, T.D. (1992). "A Rational Approach to Evaluation." <u>Research Forum</u>, 9, 23-26.

Maeroff, G.I. (1991). "Assessing Authentic Assessment." <u>Phi Delta Kappan</u>, 73(4), 273-281.

Martinez, M.E. & Lipson, J.I. (1989). "Assessment for Learning." <u>Educational Leadership</u>, 46(7), 73-75.

McPeck, J.E. (1981). <u>Critical Thinking and Education</u>. New York: St. Martin's.

McPeck, J.E. (1984). "Stalking Bests, but Swatting Flies: The Teaching of Critical Thinking." <u>Canadian Journal of Education</u>, 9(1), 28-44.

McPeck, J.E. (1985). "Critical Thinking and the 'Trivial Pursuit' Theory of Knowledge." <u>Teaching Philosophy</u>, 8(4), 295-308.

McPeck, J.E. (1986a). "Response to H. Siegel." In D.A. Nyberg (Ed.), <u>Philosophy of Education 1985: Proceedings of the Forty First Annual Meeting of the Philosophy of Education Society</u> (pp. 73-78), Normal IL: The Philosophy of Education Society.

McPeck, J.E. (1986b). "Response to S. Norris." In D.A. Nyberg (Ed.), <u>Philosophy of Education 1985: Proceedings of the Forty First Annual Meeting of the Philosophy of Education Society</u> (pp. 89-96), Normal IL: The Philosophy of Education Society.

McPeck, J.E. (1990). "Critical Thinking and Subject Specificity: A Reply to Ennis." Educational Researcher, 19(4), 10-12.

Mitchell, R. (1989). "A Sampler of Authentic Assessment: What It Is and What It Looks Like." Paper prepared for Beyond the Bubble, Curriculum/Assessment Alignment Conferences, Sacramento, CA, October.

Moore, B.N., & Parker, R. (1986). Critical Thinking: Evaluating Claims and Arguments in Everyday Life. Palo Alto, CA: Mayfield.

Morton, D. (1988). Towards Tomorrow: Canada in a Changing World - History. Toronto: Harcourt Brace Jovanovich.

Norris, S.P. (1985a). "The Choice of Standard Conditions in Defining Critical Thinking Competence." Educational Theory, 35(1), 97-107.

Norris, S.P. (1985b). "Synthesis of Research on Critical Thinking." Educational Leadership, 42(8), 40-45.

Norris, S.P. (1986). "Evaluating Critical Thinking Ability." History and Social Science Teacher, 21(3), 135-146.

Norris, S.P. (1988) "Research Needed on Critical Thinking." Canadian Journal of Education, 13(1), 125-135.

Norris, S.P. (1989). "Can We Test Validly for Critical Thinking?" Educational Researcher, 18(9), 21-26.

Norris, S.P., & Ennis, R.H. (1989). Evaluating Critical Thinking. Pacific Grove, CA: Midwest Publications.

Nuttall, D.L. (1992). "Performance Assessment: The Message from England." Educational Leadership, 49(8), 54-57.

O'Nell, J. (1992). "Putting Performance Assessment to the Test." Educational Leadership, 49(8), 14-19.

Passmore, J. (1972). "On Teaching to be Critical". In R.F. Dearden, P.H. Hirst, and R.S. Peters (Eds.), Education and the Development of Reason. London: Routledge & Kegan Paul.

Paul, R.W. (1984). "Critical Thinking: Fundamental to Education for a Free Society." Educational Leadership, 42(1), 4-14.

Paul, R.W. (1985a). "McPeck's Mistakes." Informal Logic, 7(1), 35-43.

Paul, R.W. (1985b). "Bloom's Taxonomy and Critical Thinking Instruction." Educational Leadership, 42(8), 36-39.

Paul, R.W. (1987). "Dialogical Thinking: Critical Thought Essential to the Acquisition of Rational Knowledge and Passions". In J.B. Baron & R.J. Sternberg (Eds.), Teaching Thinking Skills: Theory and Practice, New York: Freeman.

Paul, R.W. (1992). "The Logic of Creative and Critical Thinking." Paper presented at the Twelfth International Conference on Critical Thinking and Educational Reform, Sonoma State University, August, 1992.

Paul, R.W. & Nosich, G.M. (1991). A Proposal for the National Assessment of Higher-Order Thinking. Rohnert Park, CA: Sonoma State University, Center for Critical Thinking and Moral Critique.

Petrie, H.G. (1986). "Testing for Critical Thinking." In D.A. Nyberg (Ed.), Philosophy of Education 1985: Proceedings of the Forty First Annual Meeting of the Philosophy of Education Society (pp. 3-19), Normal IL: The Philosophy of Education Society.

Raths, L.E., Wassermann, S., Jonas, A., & Rothstein, A. (1986). Teaching for Thinking: Theory, Strategies, and Activities for the Classroom. New York: Teachers College Press.

Redding, N. (1992). "Assessing the Big Outcomes." Educational Leadership, 49(8), 49-53.

Reithaug, D. (1992). "Naturalistic Assessment: Capturing Success in Reading." Research Forum, 9, 46-50.

Romanish, B. (1986). "Critical Thinking and the Curriculum: A Critique." Educational Forum, 51(1), 45-56.

Selman, M. (1989). Critical Thinking, Rationality and Social Practices. Unpublished doctoral dissertation, University of British Columbia.

Siegel, H. (1980). "Critical Thinking as an Educational Ideal." Educational Forum, 45(1), 7-23.

Siegel, H. (1986). "McPeck, Informal Logic, and the Nature of Critical Thinking." In D.A. Nyberg (Ed.), Philosophy of Education 1985: Proceedings of the Forty First Annual Meeting of the Philosophy of Education Society (pp. 61-72), Normal IL: The Philosophy of Education Society.

Siegel, H. (1988). Educating Reason: Rationality, Critical Thinking, and Education. London: Routledge.

Sternberg, R.J. (1985). "Teaching Critical Thinking, Part 1: Are We Making Critical Mistakes?" Phi Delta Kappan, 67(2), 194-198.

Sternberg, R.J., & Kastoor, B. (1986). "Synthesis of Research on the Effectiveness of Intellectual Skills Programs: Snake-Oil Remedies or Miracle Cures?" Educational Leadership, 44(2), 60-67.

Valencia, S. (1990). "A Portfolio Approach to Classroom Reading Assessment: The Whys Whats and Hows." The Reading Teacher, January, 338-340.

Watson, G. & Glaser, E.M. (1980). Watson-Glaser Critical Thinking Appraisal, Form A. San Antonio TX: The Psychological Corporation.

Wiggins, G. (1989a). "A True Test: Toward More Authentic and Equitable Assessment." Phi Delta Kappan, 70(9), 703-713.

Wiggins, G. (1989b). "Teaching to the (Authentic) Test." Educational Leadership, 51(4), 41-47.

Wiggins, G. (1992). "Creating Tests Worth Taking." Educational Leadership, 49(8), 26-33.

Wolf, D.P. (1989). "Portfolio Assessment: Sampling Student Work." Educational Leadership, 46(7), 35-39.

Wolf, D.P., LeMahieu, P.G., & Eresh, J. (1992). "Good Measure: Assessment as a Tool for Educational Reform." Educational Leadership, 49(8), 8-13.

Young, L.E. (1992). "Critical Thinking Skills: Definitions, Implications for Implimentation." NASSP Bulletin, December, 1992, 47-54.