The Evaluation of Evaluation:
A Critical Perspective on Management Aspects
of Teacher Evaluation in Ontario’s
School Systems

by

Arnold Goodman

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Department of Theory and Policy Studies in Education
Ontario Institute for Studies in Education of the
University of Toronto

© Copyright by Arnold Goodman (2001)
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-59087-9
ABSTRACT

The Evaluation of Evaluation:
A Critical Perspective on Management Aspects
of Teacher Evaluation in Ontario’s
School Systems

Doctor of Philosophy (2001)

Arnold Goodman

Department of Theory and Policy Studies in Education
University of Toronto

As illustrated in the literature review, the professional literature in both education and management is surfeited with methods of evaluation, both hypothetical and established, in addition to criticisms and meta-criticisms of those methods. As far as is generally known, little has been done to correct or improve teacher evaluation from a management or systems perspective since the majority of evaluations is still done by principal observation. This position is thoroughly outlined in the section on the literature review.

The present study, as described in the methodology section, through probing and assertive interview techniques rooted in strategically-selected, literature-based provocation-hypotheses, attempts to investigate that situation. Those segments of the study, the data and the analysis, are, without apology, designed to expose those defensive attitudes that seem to have been inculcated upon both the teachers and principals. As evidence in the thesis illustrates, overall those attitudes are seen to be present as part of a general culture of
compliance. Not only has this environment of acquiescence in that culture acted to reinforce the status quo, but it also may have given rise to a complementary, even more specific culture of mythology heavily focused around the whole topic of evaluation.

Those portions of the study, the data and the analysis, reveal serious flaws in management's planning of evaluation. The discussion segment immediately following seeks to highlight serious deficiencies in management, which could possibly be corrected within most modern management systems. The focus on these flaws is outlined with respect to the current literature and extrapolated further into implications, recommendations and suggestions for future studies.

Additionally, a systemic gap in efficiency seems to exist when other more established professions and the teaching profession are contrasted. It is the judgement of this study that the practices still prodigiously defended by principals and sometimes by teachers have already been abandoned by other professions. These other university trained professions, notably engineering, law and medicine, have recognized the pernicious effect of avoiding serious, consequential evaluation.

It is in the last segment, however, speculation on implications and recommendations, that the outline for future studies are detailed. There, suggestions are offered that describe possible solutions. Those proposed projects could be instrumental in restructuring the culture and the environment of teacher evaluation. It is hoped that this
thesis will at least provide some impetus to begin the initial process that might help establish a solid link between teacher evaluation and teacher effectiveness.
ACKNOWLEDGEMENTS

I wish to acknowledge the support, perseverance and patience of my thesis committee: Drs. Richard Townsend, Susan Padro and Professor Marvin Zuker. Without the assistance of these people, I would not have been able to complete my thesis. Although not on my formal thesis committee, I would like to give special honourable mention to a very kind, benevolent and dedicated person, Miriam Blumenfeld, whose unselfish devotion in the form of hours of typing for no financial or other personal gain was equally responsible for this project’s completion.

In addition, I would like to recognize here the special moral support and assistance from Dr. John Brock, a true friend and a real mensch, whose humour and sardonic perspective kept me alive, emotionally and spiritually. Equally helpful was Michèle Schmidt, who was always there for me, and whom I have had the privilege of watching develop as a person and as an intellectual.

Furthermore, I would like to thank the following people: the OISE/UT Library staff for their dedication and work ethic; Jane Goodlet, a departmental secretary, for her assistance; Marion Morgan, the departmental chargé d’affaires and general manager, who generously offered to type the last version of this thesis; Rabbi Reuven Silver for teaching me to appreciate and respect learning in general; Robert White for helping me to edit; Gary Muskat for his moral support; and Jeanie Stewart, whose expertise, patience, diligence and devotion to duty far exceeded her official duties. She clearly exemplifies the difference between mere efficiency and high-level dedication and proficiency. It was truly a pleasure learning from and working with her. Although I am sure there are others, those mentioned above seemed the most significant at the time.
# Table of Contents

ABSTRACT .......................................................................................................................... ii

ACKNOWLEDGEMENTS ..................................................................................................... v

SECTION I

CHAPTER 1 .......................................................................................................................... 1

INTRODUCTION .................................................................................................................. 1

On Teacher Evaluation in Ontario: An Overview ................................................................. 1

Description, Rationale, Problem Statement and
Research Questions of the Study ......................................................................................... 8

Description of the Study ..................................................................................................... 8

Rationale of the Study ......................................................................................................... 10

Overall Research Question and Sub-Questions ................................................................. 11

Project Question ................................................................................................................ 11

Project Sub-Questions ....................................................................................................... 11

SIGNIFICANCE, LIMITATIONS AND DELIMITATIONS OF THE STUDY ........... 12

Factors that Determine the Significance and Delimitations of a Study:

Opportunity for Consequential Results ............................................................................. 12

Problems Claimed with the Administration of Teacher Evaluation .............................. 13

Limitations of the Study .................................................................................................... 15

CHAPTER 2 .......................................................................................................................... 17

METHODOLOGY STRUCTURE AND PHILOSOPHY --
AN ANALYSIS OF THE CONCEPTUAL FRAMEWORK AND THE
EVIDENCE FOR THE EPistemological INFRASTRUCTURE OF THE
STUDY ................................................................................................................................. 17

Frameworks of the General Methodology and their Limitations ...................................... 17

Bias as a Component of Philosophical Systems .................................................................. 17

The Inherent Subjectivity of Culture and Argumentation ................................................. 18

The Inevitability of Value Judgements in the Premise of Philosophical Systems ............. 19

vi
# SECTION II

## CHAPTER 3

### LITERATURE REVIEW

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Itinerary of Literature Review</td>
<td>72</td>
</tr>
<tr>
<td>On the General Process of Evaluation</td>
<td>74</td>
</tr>
<tr>
<td>Criteria of Evaluation</td>
<td>75</td>
</tr>
<tr>
<td>Variation in the Paradigms of Evaluation</td>
<td>75</td>
</tr>
<tr>
<td>Models of Evaluation</td>
<td>76</td>
</tr>
<tr>
<td>Systems Analysis Model</td>
<td>79</td>
</tr>
<tr>
<td>Behavioral Objectives Model</td>
<td>80</td>
</tr>
<tr>
<td>Decision Making Model</td>
<td>80</td>
</tr>
<tr>
<td>Goal-Free Model</td>
<td>81</td>
</tr>
<tr>
<td>Art Criticism Model</td>
<td>82</td>
</tr>
<tr>
<td>Accreditation Model</td>
<td>82</td>
</tr>
<tr>
<td>Adversarial Model</td>
<td>83</td>
</tr>
<tr>
<td>Transaction Model</td>
<td>84</td>
</tr>
<tr>
<td>The Future of Evaluation</td>
<td>85</td>
</tr>
<tr>
<td>Trends in Evaluation</td>
<td>85</td>
</tr>
<tr>
<td>Polemical Status of Evaluation Models</td>
<td>87</td>
</tr>
<tr>
<td>Problems Peculiar to Models of Evaluation</td>
<td>88</td>
</tr>
<tr>
<td>Status of Evaluation: Discipline vs. Science</td>
<td>89</td>
</tr>
<tr>
<td>Benefits to Educational Administration from Advances in Evaluation Theory</td>
<td>90</td>
</tr>
<tr>
<td>Evolution of Management Philosophy</td>
<td>90</td>
</tr>
<tr>
<td>Causes of the Conceptual Revolution</td>
<td>91</td>
</tr>
<tr>
<td>Problems in Early Twentieth-Century Management</td>
<td>92</td>
</tr>
<tr>
<td>Mechanistic Philosophy of Early Twentieth-Century Management</td>
<td>92</td>
</tr>
<tr>
<td>Hybrid Philosophy of Structural Analysis</td>
<td>93</td>
</tr>
<tr>
<td>Management in Other Cultures: The Re-emergence of Human Relations Factors</td>
<td>94</td>
</tr>
<tr>
<td>Lean Production: An Introduction</td>
<td>95</td>
</tr>
<tr>
<td>Total Quality Management: Introduction and Principles</td>
<td>97</td>
</tr>
<tr>
<td>Common Theme in Philosophies of Management: The Need to Excel</td>
<td>98</td>
</tr>
<tr>
<td>Teacher Evaluation: History, Methods, Problems and Guidelines</td>
<td>99</td>
</tr>
<tr>
<td>History and Development of Teacher Evaluation</td>
<td>99</td>
</tr>
<tr>
<td>Evaluation: Methods, Problems and Guidelines</td>
<td>111</td>
</tr>
<tr>
<td>Contemporary Problems in Teacher Evaluation Systems</td>
<td>111</td>
</tr>
</tbody>
</table>
Problems of Evaluation in Teaching .......................... 113
Systemic Problems in Teacher Evaluation Systems ................. 113
Weaknesses in Teacher Evaluation ................................ 115
Management Problems in Teacher Evaluation ...................... 116
Reasons for Weaknesses in Teacher Evaluation ................. 117
Systems of Teacher Evaluation and Problems Inherent in Their Development ........................................ 118
The Range of Problems in Teacher Evaluation ...................... 119
Inherent Flaws in the Systems of Evaluation ...................... 120
Conflict and Controversy:
  Objective Data in Teacher Evaluation Systems ...................... 122
The Merit Pay Controversy ........................................ 123
General Problems Inherent with Evaluation Systems ........ 124
A Conceptual Illustration of the General Principles of Effective Evaluation ........................................ 128
Legal Aspects of Evaluation ....................................... 138
Qualified Privilege: A Definition .................................. 139
Criteria of Malice .................................................. 140
The Legal Component of Evaluation ............................... 141
Influences of the College of Teachers:
  An Introduction .................................................. 141
Summary and Results ................................................ 145

SECTION III

CHAPTER 4 .................................................................. 147

THE INTERVIEWS: DATA ............................................. 147
Overview ............................................................... 147
The Best Aspects of Teacher Evaluation ............................... 147
  Teachers’ Views ................................................... 147
  Principals’ Views ................................................. 149
The Worst Aspects of Teacher Evaluation ............................ 150
  Teachers’ Views ................................................... 150
  Principals’ Views ................................................. 151
Influence of the Collective Agreement:
  Teachers’ Views ................................................... 151
  Principals’ Views ................................................. 152
Influences of College of Teachers .................................... 154
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teachers’ Views</td>
<td>154</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>155</td>
</tr>
<tr>
<td>Objective Data: Using Students’ Standardized Exams</td>
<td>156</td>
</tr>
<tr>
<td>as Evaluation Criteria for Teachers</td>
<td></td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>156</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>158</td>
</tr>
<tr>
<td>Formative and Summative Aspects of Teacher Evaluation</td>
<td>160</td>
</tr>
<tr>
<td>Introduction</td>
<td>160</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>160</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>162</td>
</tr>
<tr>
<td>Appropriate and Adaptive Aspects of Evaluation</td>
<td>162</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>162</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>163</td>
</tr>
<tr>
<td>Validity of Observations in Evaluations</td>
<td>164</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>164</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>166</td>
</tr>
<tr>
<td>Adequate Number of Visitations for Fair Evaluation</td>
<td>167</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>167</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>167</td>
</tr>
<tr>
<td>The Status of the Follow-Up Interview in the Evaluation Process</td>
<td>168</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>168</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>168</td>
</tr>
<tr>
<td>Evaluation &amp; Professional Status: The Promotion System</td>
<td>169</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>169</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>170</td>
</tr>
<tr>
<td>The Discriminatory Powers of the Evaluation System</td>
<td>170</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>170</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>171</td>
</tr>
<tr>
<td>The Cultural Compatibility of the Education System</td>
<td>172</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>172</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>173</td>
</tr>
<tr>
<td>Preview and Clarity of Evaluation Criteria and Their Inherent Problems</td>
<td>174</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>174</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>175</td>
</tr>
<tr>
<td>Teachers’ Reactions to Derogatory Evaluations</td>
<td>176</td>
</tr>
<tr>
<td>Teachers’ Views</td>
<td>176</td>
</tr>
<tr>
<td>Principals’ Views</td>
<td>177</td>
</tr>
</tbody>
</table>
Problems with Objectivity of Evaluation Criteria ........................................ 178
  Teachers' Views ........................................ 178
  Principals' Views ........................................ 179
Protection of Rights within Due Process .................................................. 180
  Teachers' Views ........................................ 180
  Principals' Views ........................................ 181
Self-Directed Criteria Selection ............................................................. 182
  Teachers' Views ........................................ 182
  Principals' Views ........................................ 183
The Principal-Teacher Relationship in Evaluative Situations ....................... 184
  Teachers' Views ........................................ 184
  Principals' Views ........................................ 185
Professional Training of Evaluators ......................................................... 186
  Teachers' Views ........................................ 186
  Principals' Views ........................................ 187
General Attitude of Teachers Toward the Evaluation Process ....................... 189
  Teachers' Views ........................................ 189
  Principals' Views ........................................ 190
Successful Evaluation and the Nature of Its Design ................................ 191
  Teachers' Views ........................................ 191
  Principals' Views ........................................ 191
Effects of Evaluation on Teaching Quality .............................................. 192
  Teachers' Views ........................................ 192
  Principals' Views ........................................ 193
The Reluctance to Dismiss Inadequate Teaching Staff ................................ 194
  Teachers' Views ........................................ 194
  Principals' Views ........................................ 195
Retrospective ......................................................................................... 197

SECTION IV

CHAPTER 5 .................................................................................................. 199

ANALYSIS AND LOGICAL CRITICISM OF THE DATA ............................. 199
  Conceptual Framework: ................................................................. 199
  The Technique Used for Analysis and Logical Criticism of the Data .......... 199
  The Best Aspects of Teacher Evaluation ........................................ 200
    Analysis of Teachers' Views ....................................................... 200
    Analysis of Principals' Views ..................................................... 203
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preview and Clarity of Evaluation Criteria and Their Inherent Problems</td>
<td>251</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>251</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>252</td>
</tr>
<tr>
<td>Teachers’ Reactions to Derogatory Evaluations</td>
<td>254</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>254</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>255</td>
</tr>
<tr>
<td>Problems with Objectivity of Evaluation Criteria</td>
<td>257</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>257</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>258</td>
</tr>
<tr>
<td>Protection of Rights within Due Process</td>
<td>259</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>259</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>261</td>
</tr>
<tr>
<td>Self-Directed Criteria Selection</td>
<td>262</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>262</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>263</td>
</tr>
<tr>
<td>The Principal-Teacher Relationship in Evaluative Situations</td>
<td>265</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>265</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>266</td>
</tr>
<tr>
<td>Professional Training of Evaluators</td>
<td>266</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>266</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>267</td>
</tr>
<tr>
<td>General Attitude of Teachers Toward the Evaluation Process</td>
<td>269</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>269</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>270</td>
</tr>
<tr>
<td>Successful Evaluation and the Nature of Its Design</td>
<td>271</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>271</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>272</td>
</tr>
<tr>
<td>Effects of Evaluation on Teaching Quality</td>
<td>274</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>274</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>275</td>
</tr>
<tr>
<td>The Reluctance to Dismiss Inadequate Teaching Staff</td>
<td>277</td>
</tr>
<tr>
<td>Analysis of Teachers’ Views</td>
<td>277</td>
</tr>
<tr>
<td>Analysis of Principals’ Views</td>
<td>279</td>
</tr>
<tr>
<td>Summary</td>
<td>282</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Perspectives on Evaluation</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Flow Chart of General Methodology</td>
<td>51</td>
</tr>
<tr>
<td>Figure 3A.</td>
<td>Demographic Chart (Principals)</td>
<td>53</td>
</tr>
<tr>
<td>Figure 3B.</td>
<td>Demographic Chart (Teachers)</td>
<td>54</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Conceptual Framework of Methodology</td>
<td>71</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Criteria of Modern Management v. Philosophies of Modern Management and Teacher Evaluation (Data Based)</td>
<td>307</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Flow Chart of Project Stages</td>
<td>343</td>
</tr>
</tbody>
</table>
SECTION I

CHAPTER 1
INTRODUCTION AND OUTLINE OF THE STUDY

On Teacher Evaluation in Ontario: An Overview

S.11 (3) In addition to the duties under the Act and those assigned by the board, the principal of a school shall, except where the principal has arranged otherwise under subsection 26(3);...

(f) provide for the supervision of and the conducting of any school activity authorized by the board;

(g) where performance appraisals of members of the teaching staff are required under a collective agreement or a policy of the board, despite anything to the contrary in such collective agreements or board policy, conduct performance appraisals of members of the teaching staff; (Excerpt from Regulation 298. Revised Regulations of Ontario 1990 V. 3, p.393)

As can be seen from the above brief pericope quoted from the Revised Regulations of the *Ontario Education Act*, evaluation is an option left to the discretion of the board or the teachers’ unions and only then, under that aegis, does it become the responsibility of the principals — it is not a requirement made mandatory at any level of government or school bureaucracy by the *Ontario Education Act*. Although other portions of the *Education Act* allocate the responsibility of evaluation to the principal under certain specific conditions, for example, as a response to a specific demand from a parent, the general rule remains the same. Unless certain exceptional conditions arise, which demand the principal evaluate, there is no absolute requirement under the *Education Act* for anyone to regularly evaluate a teacher in order for that teacher to maintain their teaching
qualifications in the province of Ontario. Furthermore, in one survey, the following conclusion was reached: "A significant number of Ontario school systems (about 27%) have no written policies on teacher evaluation" (Hickcox, 1988, p. 2). Based on this study, the researchers ascertained that many systems of evaluation used in Ontario would seem to be too arbitrary to be meaningful, if indeed they exist at all.

But the problem does not just remain uniquely associated with the standard evaluation systems formally administered within Ontario school boards. Even Lawton’s study (as cited in Hickcox, 1988) on evaluation systems in Ontario schools emphasizes this point. Not only does he conclude that the standard evaluation systems are generally ineffective, Lawton also states that the few evaluation processes that are exemplary are so unique in concept or application that they essentially dichotomize the Ontario systems into those that are excellent because of this exceptionality and the many that are poor, which probably refers to the standard evaluation format of principal observation. Emphasizing these deficiencies and making that study’s conclusions even more credible, is the population that Lawton surveys to collect his data, that is, exclusively teachers practicing in Ontario’s education system, who are, for the most part, well informed in these matters (Elbaz, 1981; Irwin, 1989). Therefore, this redoubtable data, reflective of their dissatisfaction, should not be dismissed lightly.

Furthermore, referring to the Royal Commission on Education (a commission appointed by the provincial government of Ontario in 1993 to probe general concerns
within the educational system), it certainly could be said that it did a fine job of bringing to light many contentious issues about education. However it was never designed to focus on the evaluation of teachers uniquely, as a management problem distinct from other issues of educational administration. Within its mandate, then, it did more than an adequate job of educating the public on the current problems and controversies in education.

Taking that into account, however, even the Royal Commission on Learning (ROCL), presented what seemed like only ambiguous suggestions when referring to the steps that might be necessary for effective teacher evaluation. This Commission (Caplan et al., Vol.1, pp. 77-81; Vol.111, pp. 36-38; Vol.1V, pp.136-144) in fact, had little new to offer in the area of teacher evaluation. Thus, while some might claim that the Commission’s tone fairly reflected a hotly contested domain, others could find a certain disingenuous quality in its remarks on teacher evaluation. Some of the Commission’s suggestions included comments such as the following:

When effective teaching is seen as involving complex professional judgements that are based on broad knowledge and skill related to content, teaching strategies and children, assessing performance becomes more difficult. (Caplan et al., Vol.11, p.37)

Some principals would argue that in-depth knowledge of the subject is not necessary, but many teachers would disagree. (Caplan et al., Vol.11, p.37)

The issues involved [in accountability], in fact, are quite complex and if people are serious about introducing accountability...response must be equal to problem (Caplan et al., Vol. 1V, p. 137)

[Teachers] are committed to strong, human values and create classroom climates in which such values provide the foundation for students. (Caplan et al., Vol.1, p. 78.)
These quotations from the RCOL could be perceived as, essentially, "motherhood statements," that is, cliche-ridden and nebulous enough to mean almost anything and unlikely to provoke any action or reaction from either those advocating the status quo or those rejecting it. Consequently, they offer little that is not already known.

My study here aims to highlight both deficiencies and strengths within teacher evaluation in Ontario and discover, when applicable, what allows these evaluation processes to continue within that educational system, in spite of any ineffectiveness. This study, therefore, will naturally concentrate on topics unique to evaluation and the management of evaluation. These include such issues as the validity of the data collected for evaluation purposes, the extent of the formative and summative components, the accuracy of the observations, the adequacy of the number of visitations, the connection of the evaluation process to a promotion system, the professionalism of teachers and suitability of the training evaluators receive.

It will, however, also focus on general management concerns. Thus, such issues as inter-hierarchical communication, professionalism, consistency between practices and policy, cultural compatibility, internal and external client satisfaction and fairness of the system from an employee and management perspective -- essentially management concerns responsible for the deficiencies in teacher evaluation practice -- will be the interspersed throughout my study (see diagram on page 6, Figure 1., Perspectives On Evaluation). The problems encountered in evaluation are a quasi-dialectical problem. To isolate the situation
of teachers evaluation as purely a management problem would ignore the technical problems associated with evaluation generally and to focus just on management and evaluation would tend to disregard the unique problems associated with management of evaluation specifically regarding teachers. Thus, on the basis of the above arguments, it was decided to concentrate on three general themes: management, evaluation and the management of evaluation. The diagram Figure 1. Perspectives on Evaluation (page 6) clearly demonstrates this solution and illustrates the areas which overlap between the three categories.

Throughout the study these concepts highlighted in Figure 1. are discussed without specific references to how they fit in to the general framework of the parameters of our discussion regarding the terminology specific to it, although the discussion itself tries by its substance to indicate that. When the discussion itself is not clear enough, I refer the reader back to this diagram which is there to maintain the integrity of the discussion and keep the reader focused within the proper subsets of the framework as illustrated in the diagram.
Figure 1.

PERSPECTIVES ON EVALUATION:

Management of Evaluation vs. Evaluation

MANAGEMENT
Meets Needs of Organization
Interhierarchal Communication
Professionalism
Satisfies Requirements of Clients, Employees And Organization
Culturally Compatible with Organization
Technically Appropriate Situation
Consultative Approach
Standardizable
Consistency Between Practice and Policy
Internal and External Client Satisfaction
Fairness of the System from an Employee and Management Perspective

EVALUATION
Reliable
Technically Capable
Valid

Accountability
Processes of Evaluation
Teacher Effectiveness
Format (formative, summative components)
Outcomes
Criteria of Evaluation Clearly Explained
Competency of Evaluators

MANAGEMENT OF EVALUATION
As is evident from this study’s parameters (as outlined throughout the thesis and in the title itself), the immediate concerns of this study are the teacher evaluation system in Ontario. Besides accessibility, this locale and its demographic, that is, a sample restricted to Ontario, was chosen because of its typicality, the likelihood to generalize to other sites in the United States and Canada, the lack of critical studies done on teacher evaluation in Ontario and the high level of education and sophistication of both teachers and principals (as demonstrated by Ontario’s Principal’s Qualification Program (PQP)), which gives a higher probability of getting answers with legitimate insight into the evaluation system.

These categories, concerns and orientations, then, will focus the thesis on the main issues of management and evaluation within the realm of teacher evaluation.

While undoubtedly there are differences between Ontario and other Canadian and American sites, Toronto influences the rest of Ontario, not just through the hierarchy of its governmental structures but through its cultural influence. Thus a cogent argument can be made that the general culture in the United States and Canada is established by its major urban centres, which in turn affects the rest of the country. Thus, Toronto, when taking a broad-based sample that includes urban and rural centres, should be a good example of the general cultural and educational environment in Ontario and the rest of Canada and the United States.
Description, Rationale, Problem Statement and Research Questions of the Study

Description of the Study

This study, in trying to clarify some of those issues raised by Lawton (as cited in Hickcox, 1988), compares the views of principals and teachers, based on data from interviews on teacher evaluation. Essentially, this project sought to discover the following within the range of its data, based on interviews with a small sample of educators (teachers and principals): (1) the perceived state of teacher evaluation, (2) how this sample viewed the elements that influence, enhance or restrict effective evaluation and (3) aspects of teacher evaluation relative to the general management of evaluation and (to a lesser extent) modern philosophies of management.

As explained in the rationale below, this study will only briefly outline the available types of teacher evaluation (portfolios, principal observation, form evaluation, modified observation formats and so forth). The study will, however, focus critically on the functional and dysfunctional aspects of evaluation, concentrating uniquely on principal observation, by far the most common form of evaluation, and those management dimensions that deter and encourage evaluation’s efficient conduct, by referring to the literature and the data, as gleaned from the interviews. The evidence for principal observation being the most common form will be discussed later on pages 12-13 in more detail. But I concede that it is based largely on American sources.
The issues to be investigated centre around teachers' and principals' views concerning their satisfaction with the evaluation system, the factors that restrict or enhance teacher evaluation, how the principals' and teachers' viewpoints differ and how the present structure of teacher evaluation can be improved. It is meant to explore, using the outlined guidelines on effective evaluation and modern management theory as applied to problems of evaluation from the literature review (pages 124-128), the most and least effective aspects of teacher evaluation systems, the general state of teacher evaluation, and improvements to teacher evaluation compatible with general guidelines and viewpoints of modern management theory. Through this process, it is hoped that a gradual stripping away of the layers of mythology surrounding the ideology of evaluation will occur.

Since this is the conceptual outline of my study, it will concentrate on the interviews as the primary source of evidence. This does not mean to imply all other sources will be excluded. Occasionally other sources will be brought in to support or cast doubt on the opinions or claims of the interviews. As will be discussed, in addition to the primary source, logic will be used as an adjunct resource as well as certain personal experiences, the professional literature and some document analysis. My project, as just described, and as will be later outlined in more detail in the methodology section, is conceptual in the sense that it focuses almost uniquely on the concept of principal observation. It is only empirical in the sense that it focuses on educators' outlooks to establish the details of that concept. Therefore the literature, which is overwhelming American, is perfectly valid for
Ontario, because the vast majority of it focuses on the conceptual frameworks associated with the methodology of principal observation, the aftermath and the validity of the results of that process. As such, it is not concentrating on any specific document of evaluation but rather on the general paradigm of principal observation, which has been shown by the extant literature to be almost unique as the methodology of choice. The intent here is not only discovering the systemic problems that allow these ineffective practices to continue their existence within management systems but on exploring the possible systemic organizational and culture intricacies, not yet brought to light fully in the literature, especially for the manner in which they affect teacher evaluation in Ontario. In essence, then, we are investigating the concept of principal observation regardless of the specific nuances offered by each board within that paradigm or the exceptional case where principal observation is not used.

**Rationale of the Study**

A considerable amount of information is already available in the literature analyzing the merit and quality of contemporary teacher evaluation systems without offering any firm conclusion on which is the "best" evaluation system. Were this study, then, merely a factual investigation of present systems of evaluation, it would essentially be a summary or meta-analysis of existing studies. Instead, what this study will illustrate are those aspects of teacher evaluations in Ontario that are problematic from a systemic and management
perspective and why these conditions may persist. Possible corrective measures available that would allow teacher evaluation to fulfil its formative, summative and accountability purposes and the potential for the evaluations to harmonize with modern management philosophies will be offered in a speculative Appendix A. It will also serve to illustrate aspects of the organizational culture, union input and collective bargaining that influence or could influence the climate for the development of effective evaluation systems.

**Overall Research Question and Sub-Questions**

**Project Question**

With regard to the group of teachers and principals being studied, what is the state of teacher evaluation that the group identifies in the Ontario public school system; what are the elements which enhance or restrict effective evaluation; and what appears to be evaluation’s compatibility with the general guidelines on evaluation and selected modern philosophies of management?

**Project Sub-Questions**

(1) What is the nature of teacher evaluation in the 12 schools being considered, as determined in data collected from the 12 interviews with principals and teachers regarding teacher evaluation’s fidelity to its intended objectives, standards of efficiency and values?
(2) What are the views of this particular group of teachers and principals on the factors that restrict or enhance effective teacher evaluation?

(3) Do these particular principals and teachers reveal any important data about the conduct of teacher evaluation with respect to the accepted guidelines of general principles of good evaluation?

(4) Do the data reveal a compatibility between the administration of evaluation and the theories of modern management?

(5) Are there any important differences between these views of teachers and principals?

(6) After the analysis and synthesis of the data of this exploratory study, what suggestions can the researcher offer to help advance the present structure of teacher evaluation towards a more effective and responsive system?

SIGNIFICANCE, LIMITATIONS AND DELIMITATIONS OF THE STUDY

Factors that Determine the Significance and Delimitations of a Study:

Opportunity for Consequential Results

It would seem that, to be worthwhile, a project must offer a reasonable chance of producing consequential results that advance the study of issues of importance in that field. While the opportunity of generalizing beyond the immediate sample of the study is limited, the importance of the study can still be appreciated. In addition, to the above limitations,
a delimitation of the study is its assumption that the primary or major methodology of the vast majority of teacher evaluations is principal observation. This is largely based on American information to which the situation is similar if not identical (Coker & Medley 1987; Kowalski, 1978; Lower 1987). This researcher, as best as can be determined, could find no statistics on how prevalent that situation (that is, principal observation) is in Ontario. It should be noted, however, that the assumption that principal observation was the dominant form of teacher evaluation was the basis for almost all of the interview questions, and I was not disabused of that assumption by any of my interviewees by having them question the basis for that assumption.

Problems Claimed with the Administration of Teacher Evaluation

A common claim made by school officials (supported by some of the literature) in defense of their teacher evaluation policies is the allegation of a fundamental difference between business and education (English & Hill 1994; Sergiovanni, 1996). Two fields that, incidentally happen to operate so similarly in many areas but seem to differ radically with regards to personnel evaluation. In effect, administrators in education have been claiming for some time that much of what teachers produce is intangible and therefore it renders impossible any attempt to objectively evaluate teaching in the same way businesses evaluate their employees. School administrators apparently believe that “good education and the behaviours conducive to it are inherently difficult to measure in an objective,
quantitative formal manner” (Chubb & Moe, 1990, p.36). As a result, this argument has been used by administrators to dismiss complaints about the purported laxness or inconsequential attitude seen in much of the school evaluations in the system. This rationalization has recently evolved into criticism used against those demanding rigorous evaluation standards by accusing them of “bureaucratization of personnel, [which] tends to ensure that public schools will lack the proper mix and balance of talents on which effective education inherently depends” (Chubb & Moe, 1990, p. 50).

This is essentially the major contribution of this thesis: to increase the understanding of the underlying myths that limit the practices of effective teacher evaluation in Ontario. It, in effect, points toward some explanation for the endurance of those myths and why they initially penetrated the educational system and were perpetuated throughout the educational system.

Myth here is employed in the standard vernacular usage, meaning any fictitious story, unscientific account or belief (Webster,1996). This definition stands out in contradistinction to the one used by Schein in Organizations and Leadership (1992) and to the definition used by cultural historians and scholars of literary criticism such as Northrope Frye and Joseph Campbell. There, myth is defined as a right-brain symbolic representation of an actual historical event, presented figuratively, however, in order to present it in a less threatening way. It is only by accepting the first definition based on
common usage that the paradigm of the thesis maintains any efficacy as a tool to facilitate a process of starting to demythologize the institution of teacher evaluation in Ontario.

**Limitations of the Study**

Generally, when selecting experimentally valid sites in a qualitative study, several significant criteria are necessary. In qualitative research, the samples are often small; consequently, the sites must be highly typical and therefore representative of the population.

For the purposes of this study, the aforementioned caveats need not strictly apply; neither the sites nor the population have to be highly typical of any other site or population, for the most part. In effect, just selecting a sample of teachers and principals from the publicly-funded school population should be reasonably sufficient to establish, at least from within a qualitative paradigm, what the problems, concerns and benefits of teacher evaluation are. What we are attempting to explore is the methods of extant evaluation within the various school organizations and the attitudes and reasons for their existence, as put forward by the teachers and principals in my sample. At this stage, I am not attempting to generalize my findings in order to render judgement whether or not any specific public school is typical of other public schools in that board or whether a certain separate school is typical of other separate schools in that board. Rather, the impetus of this study is to compare teacher evaluation from two different perspectives in education,
those of the principal and of the teacher and to proceed towards a possible understanding of evaluation as a result of those viewpoints. These perspectives, both the principals’ and teachers’, must be able to contribute, in the final analysis, to the gathering of information that can answer the questions or at least clarify the issues raised by this thesis about the factors that restrict or enhance evaluation; the fidelity of each evaluation system to its own objectives, standards of efficiency, and values; the compliance with the general principles of good evaluation; the compatibility with theories of modern management; the significant differences between principals and teachers; and finally, the ability to suggest recommendations for future study into more effective and responsive teacher evaluation systems. Outside of the limitations just listed above, and depending on the variables just discussed, this study can offer contributions of data towards the resolution of these problems. That said, it should be acknowledged that in trying to determine educators’ views toward evaluation systems with the goal of making improvements that may be achievable in management systems this researcher is often ambitious. In some readers’ views, then, the scope of this study may be too ambitious. Nonetheless, as an exploratory study that may improve upon the present outlook towards management of evaluation systems, the present work here may prove helpful.
CHAPTER 2
METHODOLOGY STRUCTURE AND PHILOSOPHY --
AN ANALYSIS OF THE CONCEPTUAL FRAMEWORK AND THE EVIDENCE
FOR THE EPISTEMOLOGICAL INFRASTRUCTURE OF THE STUDY

Frameworks of the General Methodology and their Limitations

In these sections on methodology, I have formulated a framework conceptualizing the basic ideas of the study and defined them for the purpose of this study. Yet a conceptual framework in isolation without the prerequisite philosophical framework is not sufficient. It is the philosophical infrastructure that underlies any particular conceptual framework that ultimately determines the conceptual framework’s viability to function within the parameters of the social-scientific paradigms and rules for evidence and to establish reliable results. Once I have conceded the necessity of this process, a close scrutiny of the study’s particular philosophical perspective through its paradigms seems to be required. In the following paragraphs, I will discuss certain philosophical issues inherent within any system of argumentation such as subjectivity, cultural bias, development of first principles, value judgements, relativism and the strengths of the interpretivist position.

Bias as a Component of Philosophical Systems

It is important to acknowledge here the limitations due to the inherent bias of any philosophical system in addition to the general limitations already discussed. This
limitation, inherent in philosophical systems, reflects one’s own cultural-hermeneutic filters. This flaw cannot be overstated. In fact, lack of vigilance in this area can result in a system that adopts a relative socio-cultural viewpoint as an absolute one. Such a perspective may seem inherently justified even to the point of claiming self-evidence as a justification. Nevertheless, such specious confirmations should not be misinterpreted as an affirmation of the system. A far less tendentious conclusion would be the realization that such “irrefutable” evidence almost invariably rests on one’s own biases and is often merely a tautological process. Instead, one should adopt an approach that allows circumspection against this illusion of complacency. One’s own inculcated values must never be able to claim an inherent connection to any absolute truth. When this does occur, it is usually indicative of a chauvinism that could prove deleterious to the quintessential veracity of the entire philosophical system.

The Inherent Subjectivity of Culture and Argumentation

If the influences of a tendentious cultural or intellectual approach do occur, they can make serious inroads into the veracity of any philosophical system. Any further attempt to extrapolate the basic principles of the system through concept reification or corollary development without recognizing that systemic error might compound the flaw of subjectivity by constructing a futile structure of either obfuscation or entrenchment. Rather than a clearly developed argument, then, a camouflage of subjectivity results, occluding
the basis of the argument’s origin, its first principles. What remains is a system infested with systemic fallacies based on heuristic conclusions. Ultimately it is to the detriment, if not the destruction, of the entire algorithmic development.

**The Inevitability of Value Judgements in the Premise of Philosophical Systems**

Even if these precautions, so inherent to traditional critical reasoning, are adopted, it is unlikely that the gravamen of the argument will ever be reduced completely to first principles in the sense of being absolutely culturally independent and assumption free. The most that can be expected to emerge is a coherent set of values that can be eventually expressed as a general theory. In the final analysis, any hypothesis to be tested must be weighed against practice and compared or contrasted to the alternative paradigm of any critics who claim they have a superior approach. Ultimately a realization must develop that the critic is only capable of investigating a finite number of possibilities even though an infinite number of possibilities exist. Therefore one can never claim infallibility or simply “consider everything” as an alternative algorithm (Anderson, 1990).

**Conceptualization of the Study’s Parameters and a Discussion of Its Weaknesses**

From a conceptual perspective, teacher evaluation can be viewed as any mechanism used for determining the quality, effectiveness and proficiency of the teacher. This
definition can be used whether the assessment is applied to individuals, teachers or members of a team or department. For the purposes of conceptualization within this study, evaluation will be seen as the comparison of certain measurable or observable practices of teachers in their professional capacities during classroom time to some set or agreed-upon standard by the evaluator or by a consensual agreement between teacher and evaluator. This follows Scriven (1991) who, in general terms, defines evaluation as the determination, in as objective a manner as possible, of the merit, worth and value of things.

Teaching effectiveness, defined here more narrowly than in vernacular usage, is perceived as the successful transmission of academic knowledge by the instructor during class time or anything done that succeeds in helping a student reach educational objectives (Medley, 1992).

Supporting his judgement within the conceptual framework of a duties-based approach, Scriven (1988, p. 320) takes a similar view that states the plausibility “of the main aim of teaching to be maximizing valuable learning, relative to the potentialities of each student and within the usual constraints of time and other resources…. “ Although this conceptualization may be criticized as narrow in scope, in the judgement of this researcher, based on personal experience and an overview of the literature, the major function of the teacher is transmission of academic knowledge and reaching the pre-determined educational objectives. Nevertheless I will concede a wider range of equally valid alternate
criteria of teacher effectiveness might exist, such as instilling values, counseling or developing work or study habits. These, however, will not fall within the primary range of the proficiency criteria to be investigated in this study, as they are not included in my principal definition of teaching efficacy. In addition, supervision, when used in this thesis, in a technical sense, will be taken to mean supervision of teachers and will be defined as any board-approved mentoring or evaluative relationship between a teacher and any other board employee.

**Relativistic Nature of the Conceptual Framework and Its Restrictions**

Despite using these choices just described as the conceptual perspective, I must again reiterate that the conceptualization process and even a subsequent meticulous and rigorous algorithm does nothing to strengthen the argument against potential refutations from different perspectives. This is true because even irrefutable logic is based on some pre-conception. If a challenge from a different conceptual framework is advanced, it could possibly be used to refute a seemingly impregnable argument. Again, this explanation should illustrate the relativistic nature of any conceptual formulation and any philosophical infrastructure.

It might seem, then, that all philosophical systems being tainted by this flaw are invalid. Nonetheless, since any refutation based on a philosophical system is equally biased, a refutation would merely be able to supply another frame of reference. The chief
criterion of a valid philosophical system is its consistency and lack of circularity in the flow of its argumentation, not its irrefutability. Consequently, competing philosophical perspectives in isolation cannot be used for the purposes of a definitive refutation. Each varying outlook, however, is dependent both on its particular paradigm and on its internal validity as a philosophical structure. The ultimate test of its validity would, consequently, have to be the practical success it can deliver in its respective field of endeavor.

Therefore, despite certain reservations previously discussed, any of the above caveats should not be taken as a definitive refutation of a potential system that might be developed. Instead, it should help develop within us, the researcher and reader, an awareness of the weaknesses inherent in any philosophical or interpretive hermeneutical system. In fact, then, to properly construct a conceptual framework of an evaluation system, one must construct the basic outlines that make evaluation of teaching possible. For the purpose of this study, these concepts include the following: (1) general principles of teacher evaluation and (2) the management principles of efficacy and supervision.

In deference to the parameters of these caveats, it should be noted that this study is largely based on a research approach using a qualitative methodology to discover answers to selected research questions. The qualitative approach is neither inferior nor superior to the quantitative approach; it is, however, different. Perhaps that approach can best be described as a non-mathematical, non-statistical description of a phenomenon. It exists in contradistinction to, but not in opposition to, a quantitative approach which uses statistical
indicators or other numerical data to summarize its results. The qualitative approach, for the most part, prefers to use prose, narrative, dialogue or other forms of verbal and visual documentation to summarize its results (Anderson, 1990). This is not a compromise or an attenuated form of evidence but a conscious choice made by qualitative researchers to use appropriate methodology to better understand a particular perspective of the world (Bell, 1999).

Awareness and open disclosure of the potential for any bias that could weaken the study’s validity should help maintain a balanced scholarly perspective within appropriate parameters. That this study uses diverse evidence from independent sources such as our own data base of principals and teachers, the professional literature and logical analysis, which is open to the reader’s scrutiny, should also serve apotropically (that is, specifically against researcher bias) and help establish the data’s veracity. In that vein, I believe I should declare any possible background influences that might be interpreted as bias.

My academic background is in religious studies and psychology, in addition to training as a Talmudic scholar at a yeshiva seminary and my experience in the public educational system. At one point it becomes fairly clear that the influences of that background weigh heavily on the perspectives I adopt within the study. In addition to two sources taken directly from the Babylonian Talmud, throughout the study there is a strong disapproval of the lack of professional respect accorded to teachers, a reaction that I believe also reflects the influence of a Talmudic and rabbinic background and attitude.
That this is a cultural bias I openly acknowledge here but which I do not necessarily regard as a flaw. Within these limitations, then, the degree of objectivity necessary for the study should be relatively well protected.

Regarding my choice of interview questions, some of the issues emerged from the literature and, therefore, were used directly in the questions. Additionally, I would be less than forthright if I did not reiterate that, besides the literature, my personal experiences also influenced my choice of questions and the position I took in some of the analyses. From my experience in the school system and because of my background I should forewarn the reader that I was appalled both at the lack of general professionalism of the teaching staff and administrative personnel over that and other facets of the systems’ failings. I served as a high school teacher for six years in Peel, North York and York Region, largely as a supply teacher and largely short term. The subjects I taught were science and history, but being a supply teacher, I taught everything. The evaluation of supply teachers was totally capricious, more often than not depending on the mood of the principal. Interestingly enough, despite its almost total arbitrary nature, it was more consequential than evaluation for regular full-time teachers and several supply-teacher colleagues lost their jobs because of a principal’s whim. However, when talking to the full-time teachers their opinions were almost identical to the teachers who made up my data sample in this study. That is, in my judgement, their opinions were angry and resentful at the lack of professionalism in their evaluation process.
Nevertheless, despite this general overview, there were teachers who said the evaluations were done by caring professionals who spent the time to allow them to go through a valuable formative process that did improve their teaching. I make no attempt to deny that these experiences just recounted have influenced my attitudes, especially while engaged in the analysis of data. Although this is not my main instrument for collection of data, in an ancillary capacity, this folklore source is a legitimate source of information to help interpret the data from my primary source. In this connection, as Boulding has described it, folklore consists of the impressions, opinions, information and data derived from common observations and insights I adopt from the normal exchanges and vicissitudes of life as I experience it in any situation (Boulding, 1966).

Subjectivity and Qualitative Research

It should be noted, however, that the "taint" of subjectivity derisively leveled at qualitative research originates largely from a positivist perspective. Yet even if the premises of that positivist argument are accepted, it should be acknowledged that the hard sciences themselves are not free of subjectivity. Because those sciences base their conclusions on an observation-generalization model, scientific "truth" has been thwarted by objective reality that demonstrates how the elimination of variability is impossible. In deference to this, the latter half of the twentieth century has generally seen science downgrade its discoveries from "laws" to theories (Hughes, 1990, pp. 55-59).
Social Reality, Subjectivity and the Interpretivist Viewpoint

Alternatively, the interpretivist viewpoint claims that reality generally, and social realities specifically, as they relate to meaning, are inextricably tied to the hermeneutic aspects, personal perspectives and the social interactions inherent within language (Hughes, 1990). In this view, social reality is more closely linked to interpersonal relations and ethnological practices as interpreted by the language used to describe them. Then, if this view is accepted, the rigor of positivism can act rigidly in an unresponsive detrimental manner by stripping contextual implications from the social reality being investigated. Thus, from a nomothetic viewpoint, positivist methods might construct, rather than investigate or discover, a social reality (Hughes, 1990). When one realizes that there not only are “multiple” or even “antithetical realities” in any specific social situation but also that social reality is largely perception-based, subjectivity as a flaw in a qualitative research approach becomes largely, if not completely, abrogated. This should tend to vindicate qualitative research and establish it as a valid research technique.

The Justification of Interviewing as a Valid Qualitative Methodology

The Interview as a Qualitative Research Tool

As the primary methodological structure of my project, it is important to discuss the validity of interviewing as a research tool for data collection. I will now attempt to lay out
the principles underlying my use of this method in some considerable detail, owing -- as it will soon be apparent -- to the uniquely aggressive nature of my approach.

Even as qualitative research tools, interviews have their own unique problems. Besides the general problems indigenous to all qualitative research, the problem of response veracity can be so crucial with this methodology that it can ultimately call the data into question. This problem can be partially mitigated by guarantees of discretion which allow people to feel less inhibited or uncomfortable about an answer which they feel might reflect badly on them (Dixon et al., 1987) because the interviewees know their privacy will be respected categorically. Other possible strategies to maintain veracity popular with psychometric tests like the Minnesota Multiphasic Personality Inventory (MMPI), include inserting questions that are either obviously true or false (McKeachie & Doyle, 1970). If they are not answered "correctly," then that entire survey, interview or set of data is thrown out. This process for raising the accuracy level of the data is based on the understanding that under the aforementioned conditions, a high probability exists that the testees or interviewees are saying what they want us to believe. As a result, the data are discarded due to the high probability of inaccuracies.

Criticisms of the Interview Methodology

Even when the criticisms of the interview methodology are specified in detail, it seems that these "drawbacks" are largely supported from the perspective of a positivist
paradigm. Nevertheless, these criticisms bear repeating if for no other reason than to heighten the general awareness of their existence, regardless of their paradigmatic orientation. These flaws could include the following: absence of control groups, difficulty of generalizability, problems with replication of results and establishing of causality and so forth (Runyan, 1982). The complaint is that, in essence, this methodology seems to offer only anecdotal evidence.

Rebuttal to Criticisms of the Interview Methodology

Because of the above problems, it must be reiterated that, generally, interpretive social sciences are more interested in internal rather than external states. As such, it requires a research approach that investigates the circumstances and reasons of a phenomenon rather than one that records large numbers of results to pinpoint a causal factor. Yet acknowledging the validity of positivist criticism, in principle, does not necessarily mean the criticism is either ubiquitous or valid when directed towards qualitative research. An interview does not have to be conceptualized as merely a solitary data point which requires repetition. Extending what Yin (1981, p. 61) states, conceptually, in this respect (with regard to case studies), to include the interview, an interviewer might also be seen as “an analytic unit on a par with a whole experiment.” This viewpoint has even more credibility since interviews are not just constructed to be used as a random series of questions. Properly formulated, the questions should represent
the product of a well-developed theoretical instrument which the interviews or surveys are used to test. This is well within the range of the criteria outlined previously for a hypothetico-deductive model of inquiry (Hughes, 1990). The consensus established there agree that the hypothetico-deductive model is epistemologically valid even within the parameters of a positivist paradigm. Furthermore, accuracy increases because of verification techniques like triangulation (Denzin, 1978) or interviewing a paradigmatically identical person at the same time (Hakim, 1987). The interview methodology, then, although not without problems, can be used in an intellectually rigorous manner. Certainly properly done, from within a social scientific viewpoint, this method provides acceptable evidence with little risk of that evidence being challenged on methodological grounds.

**Cross-examination as an Hermeneutical Principle**

A more specific example of an interview technique necessary to accomplish the types of revelations previously referred to is cross-examination. The process of cross-examination is multifaceted and includes many principles of logic discussed in this section. Those principles combined with the focus of the interviewer will determine the success or failure of the cross-examination and ultimately the project itself. Within the duration of the cross-examination, any prevarication or convenient fiction should be revealed if the process is conducted efficiently. Essentially, the unacceptable argument is the one that
cannot stand up to the scrutiny of the cross-examination and is therefore rejected; the more accepted version is its complement which can survive. Of course, this does not mean to infer that the converse, that is, the accepted version is true -- just that it offers a more probable interpretation than the refuted version. It is the responsibility of the interviewer to use the cross-examination with skilled integrity to expedite this technique and to achieve the resulting dichotomy as described above.

Psychological Substructure of the Methodology: Provocation

The dominant, if unorthodox, psychological principle which pervades the ideological structure of our methodology, as a substructure, is, in a word, provocation. While in itself it is not part of the commonly accepted interview methodology, the concept of provocation is integral to the interview process in this study because it provides the psychological substructure and establishes the parameters within which that methodology exists. It is a consistent theme throughout the methodology and is supported by the infrastructure of other methodological sub-components. The aim of provocation is to spur the respondent and (at a later period) the reader into taking a stand on the subjects at hand. The premise of this project's hermeneutic is that the act of provocation can goad readers and respondents to reconsider what they know or think they know.

It was felt by this researcher, and undoubtedly this was a judgement, (this is illustrated later, albeit indirectly) that if normative, non-aggressive techniques were used,
no progress would be made. This means to suggest that had I not used this particular or a similar methodology, instead of breaking new ground, an essential step for getting data that leads to the quintessential problems and solutions, I would have probably just received the same repetition of alleged facts that the Ministry of Education and the general educational establishment support.

While this is an unusual methodology, it is not totally without a prior instance. The basic concept of provoking to initiate a response, in a scholarly context, is found in an article in Technology Review in 1977 (Vogely, 1977). Similarly the Adverserial Model of Program evaluation (page 82) uses a similar, although not identical approach to our data method. Finally, the whole concept of provoking to stimulate the other person to produce their best “parry” to your best “thrust” has a basis in the great Talmudic academies of Sura and Pumpaditha in Babylon (3rd and 4th centuries C.E.), where the rabbis, the scholastics and legal debaters, used this technique to arrive at a decisive answer. As a result, the rabbis in those debates were called “shield bearers” (Babylonian Talmud, Tractate Berakoth 27b).

Somewhat innovative, and perhaps even contrary to the accepted pattern of most investigative interviewing styles, this confrontational, even belligerent, approach is considered necessary, here, in order to challenge the perceived complacency of the status quo. These conventional ideas, often accepted at face value by the teachers, principals and even some academics in the field of educational administration, require a confrontational
attitude to expose the weaknesses of the arguments defending that position. By means of directed analysis focused on the interviews, any logical fallacies and untenable arguments (because they seem to oppose too much of the literature, seem self serving, contradict their own position under different circumstances or appear to refute the common knowledge about the school system), that have been used to support the evaluation system are exposed. (Respondents could have terminated the interview at any point in the provocation, but none did.)

However, even conventional "passive" interviewing is not nearly as docile as imagined. Even there the interviewer must be an operative, participating listener and constantly be cognizant of the relationship between the interview and the main body of the research and even be prepared to modify the tone, substance and direction of the interview if necessary, in order to maintain that equilibrium (Hammersly & Atkinson, 1983). Even in those non-directive interviews, a certain amount of "steering" is deemed necessary in order to probe answers, switch focus and advance to more in-depth, penetrating questions for the purpose of eliciting more accurate information and more profound insights as opposed to those which are merely superficial or have been culturally or bureaucratically ingrained (Campbell, 1993). The advantages gained by these interviewing techniques are the greater complexity of questioning (Baily, 1987) and the greater general depth of discovery (Borg & Gall, 1983). The interviews as conducted here are perhaps, then, just one step beyond the accepted tradition of interviewing.
The Purpose of the Psychological Substructure

What this logical analysis is intended to do is to illustrate, despite the attempts of the participants, a more critical analysis of the interviewees’ perceptions of the nature of the present evaluation system. Indeed, this methodology is prepared to criticize the very attitudes professed by principals and teachers which, depending on how the exchange develops, may sometimes be the most devastating indictment against the prevalent optimism among certain professionals in education that remains about the current system of teacher evaluation in Ontario.

Components, Devices and Methodologies of Provocation

Some of these devices of provocation include analogy and logic, (that is reducing the argument to absurdity), precedents and parallel cases. In the following I provide examples and rationales for some of these categories that are included in the interviews during cross-examination and subsequently in the methodology of the data analysis. Here, for sake of clarity, I have chosen to immerse the readers directly into the specifics of my topic using actual illustrations from the data. This was deemed better, even at the cost of what might seem like repetition of that data, than proceeding along a circuitous and inefficient methodological route by giving examples alien to my topic and external to the data at hand. Definitions of these devices follow immediately.
Analogy, as used here, is the comparison of similar cases in order to raise the question, specific to evaluation, why evaluation processes are different in the respective cases despite the cases being so similar. When the analogy is especially strong, it can be used to refute specific responses that were introduced to serve as exculpating circumstances for the lack of rigor in teacher evaluation.

Precedents, for the purpose of my study, refer to the citation of other styles of teacher evaluation practiced successfully, in other school structures within a similar context and their degree of success at accomplishing formative or summative goals. If the literature finds a certain protocol of evaluation effective in a different situation, the argument that claims the same technique would be a categorical failure loses some of its strength.

Parallel cases, as defined here, involve comparing different teachers' views of the same scenario and contrasting those views among the teachers themselves but especially with those views of the principals. It involves comparing and contrasting the different interpretations of the same situations and using logic and judgement to discern, on the balance of probability, where the more correct version lies. With these methods, one may not necessarily arrive at a definitive answer. Rather, the goal is more general: to reveal the weaknesses or strengths in the arguments and accepted conclusions about how evaluation should be conducted and to illustrate the positive and negative aspects of communication between the different levels of management.
The Hermeneutic of Provocation: Provocation as a Hermeneutical Principle

Overall, then, this substructure is directed at breaking down any rhetorical evasions of the interviewees. This is a common problem in many organizations and therefore it was assumed that it would be even more extreme in highly bureaucratized organizations like the school system. As Becker (1970, p.128) states:

Officials usually have to lie. That is a gross way of putting it, but not inaccurate. Officials must lie because things are seldom what they ought to be . . .

Since they are supposed to, officials develop ways of both denying the failure of the institution to perform as it should and explaining those failures which cannot be hidden. An account of an institution’s operation from the point of view of subordinates therefore casts doubt on the official line and may possibly expose it as a lie.

Yet even Becker may not go far enough here in explaining the problems of investigating general inefficiency and production problems in an organization. Even more poignant is William Whyte’s analysis in his study of corporate culture, The Organization Man (1956), which shifts the focus on to the worker. There Whyte refuses to exonerate the workers and suggests that they may be just as guilty of complicity. He describes how the institutional chauvinism inculcated upon company employees ultimately pervades their weltanschauung:

But in searching for that elusive middle road, we have gone very far afield, and in our attention to making organization work, we have come close to deifying it. We are describing its defects as virtues and denying that there is -- or should be -- a conflict between the individual and organization. It is worse for the individual. (Whyte, 1956, p. 13)
Yet despite these two noted academics’ description of the official version given to the general public as a lie, I, here, have judiciously avoided dichotomizing the version revealed in the interviews, after scrutiny and analysis, as the truth and the version offered by the interviewees, both management and labor, as the lie.

Continuing with this argument, but focusing more specifically on the traditional academic disciplines, many of the famous controversies in history and anthropology have been fought over the legitimacy of sources, documents or other evidential discoveries. In history, such documents as the Minnesota Runes of 1898 (still controversial), the Donation of Constantine (refuted by Lorenzo Valla), the poems of Ossian (refuted) and the Protocols of the Elders of Zion (refuted and proven to have been anti-semitic propaganda forged by agents of the czar in 1905), are examples of controversial data either still under scrutiny or already refuted (Gustavson, 1955). Similarly, this type of scrutiny is essential when historians use interviews to collect data. As Edmund S. Morgan, Emeritus Sterling Professor of History at Yale University, has said, “Historians commonly have to cull facts from testimony not necessarily designed to disclose them” (Morgan, 1999, p. 30).

Alternatively, not questioning data in anthropology, for example, the Piltdown Man, resulted in one of the most embarrassing scandals within the scientific community in this century, and one from which physical anthropology has still not fully recovered. Furthermore, critical ethnography goes beyond even these parameters. Following Marx’s dictum, these ethnographers try to counter his complaint against impractical scholarship:
"The philosophers have only interpreted the world, in various ways; the point, however, is to change it." (Marx, 1978, p.145). Reformulating the definition of research within the accepted parameters of social science, critical ethnography refers to a discipline that is procedurally value-laden, that is, it is judgemental and allows intensive, investigative scrutiny to challenge previously accepted "truths," and any hidden agendas, biases or assumptions that might limit or destroy social, legal or economic advances. Critical ethnography demands a professionalism that challenges policy and confronts that which is accepted as the norm, with the intent to institute reform. Critical ethnographers endeavour to enlighten by arousing social consciousness in order to catalyze social change and ultimately to implement social justice (Thomas, 1993). Although this thesis does not actually represent an authentic display of critical ethnography, it is in that vein that the researcher has decided to further question his data, beyond the "in situ" interview stage, later, in the section on analysis.

Originally, among the ancients, rhetoric was the study of a technique of persuasion designed to be used on the hoi polloi, the common unlettered masses, impatient to get to an answer without putting the logical pattern of the argument under too much investigation for either consistency or cogency. When appropriately used, it was an effective methodology for the development of specious, often emotional, arguments with the aim of directing people to the desired result at any cost (Olbrechts & Perelman, 1957).
Because of this intrinsic property defining rhetoric, it was considered by some cultures as inherently flawed, perhaps even dangerous and worthy of being banned. In fact, the Babylonian Talmud (Babylonian Talmud, Tractate Menachoth, p. 64b), whose supporters saw themselves as a bulwark against many aspects of Hellenistic culture, decried the practice of rhetoric in the strongest terms. The repulsion of this feeling was so strong that it was extended so far as to eventually become manifest by the production of a malediction. This contained an imprecation in the form of a divine curse on anyone who educates children using a Hellenistic curriculum, that is, rhetoric.

**Provocation as a Substructure of Rhetoric**

Here, in this project, no authority exists that would allow anyone to promulgate any form of anathema against those who would resort to the use of a rhetorical style of argument to defend their position and (as will be shown in later chapters) some of the respondents, unfortunately, seem to use rhetoric in a defensive manner. Instead my method, although somewhat more benign, is to expose, what appears to be, the deceit. This is to be done by applying the counter-methodologies just discussed, that is, this study’s own hermeneutical principles, its own rhetoric, to reveal the rhetoric of denial often practiced by the interview’s respondents.
The Parameters and Language of Inquiry

Therefore, in this study, the conduct of inquiry proceeds from the logic of the interviewees themselves to define what is acceptable as an accurate description of the facts. By attempting to establish what is, in effect, rationalization, or a self-serving motive, by pointing to the possible sources of their motivation or by outlining their logical fallacies, in effect, I can eventually discern which statements are, (as used in the legal context) on the balance of probability, false or contradictory and which of their statements are viable. That, for the most part, is the modus operandi I have adopted. Perhaps it should be made clear here, then, that refuting an interviewee’s statement does not mean the opposite is true. Instead the most it could imply is that other alternatives are more viable or should be tested to see if they are correct.

But besides the core methodology, there also exists a crucial sub-system necessary to expose any fallacious or surreptitious argument. That sub-system employed here is the unique language used to remove the protection of the cliché and initiate dialogue between interviewer and respondent and to present those results as data. It is clearly stated here, and without apology, that the language used is blunt and contentious.

By adopting this aggressive stance, it can be argued in refutation that these terms could be construed as merely the biases of the author. Yet that stance is said to provide a methodology not to support any bias but to refute, in as neutral a manner as possible, any untenable arguments. Essentially this is an ancillary methodology, a corollary of the main
methodology of interviewing, a system of argumentation that prevents the researcher conducting the project interviews from taking the interviewees viewpoints at face value. Ultimately, this methodology facilitates my primary methodology of provocation and my understanding of evaluation in general by introducing such terms as organizational culture, politically motivated behavior, denial, fascist communications’ structure and institutional brainwashing. Even, perhaps, at the risk of offending the various stakeholders and participants, these terms, used in the interviews but even more so in the analysis, will be defined in the glossary or as they are introduced into the thesis.

This is not to suggest that there are no weakness when such conceptual paradigms used. Undoubtedly this approach can backfire. Respondents can perceive themselves as under attack and become defensive, even to the point of not revealing what they would have under normative interviewing conditions. Additionally, because of the nature of the provocation methodology, to some extent, a more narrow focus is undertaken which concentrates on some issues to the detriment and even exclusion of others. This, obviously, is instrumental in dismissing certain issues as unimportant, which, to some degree, increases any bias within the methodological system.

Central Themes as Contextual Thematic Structures

It is also important to note that because of its irregular methodology, a certain unique structure permeates the project. This structure works to centre its arguments around
several major themes whose validity keeps emerging but from different sources of evidence from within various perspectives. As a result, a certain redundancy might seem to exist. In fact, this "redundancy" is a reinforcement of some of the central themes of the project. Their recurrence serves to illustrate how ubiquitous the themes are as they emerge from the analyses of several distinct types of data produced from assorted question contents and styles. This format is pivotal to the structure of provocation and the development of this project. What follows in the next section is the presentation of these themes that I focused on, albeit structured as the provocations.

The provocations-hypotheses are issues which evolved both from my experiences as a teacher, where I were confronted by several themes because they seemed anomalous to expected professional behavior, and independently from the educational literature. Sometimes certain aspects of evaluation seemed to fulfil no purpose other than allowing members of the organization to claim that, in fact, these aspects existed as a legitimate and bona fide functional process for the improvement of the institution. However, as already stated, this does not imply that the issues used as provocations are not literature-based. The majority of the issue-based provocations are also rooted or implied in the professional literature, as outlined in a survey article by Haughey, Howard and Marshall (1996) from which many of the questions were developed. These provocations, then, are not themes that emerged from the data but are confrontational, controversial issues based both on personal observations and the literature that are used in the interviews to confront certain
myths about teacher evaluation. In fact, these provocations were directed at the respondents to produce data. They were deliberately designed to create a confrontational atmosphere so as to elicit the teachers’ and principals’ real views of these vital issues.

It is for the reasons presented above, by the nuances that inform this project’s unique methodology, it was decided that investigating any specific board document would be unnecessary. The reasons for the particular document chosen would be almost arbitrary. Furthermore, the conceptual, as opposed to the empirical approach (as described on pages 12-13), used here, in establishing the methodology to be studied, would seem to vitiate any advantage or benefits to be gained from studying subtle differences between systems whose main methodological orientation, principal observation, is identical. This is true despite using an empirical approach through interviewing. It is clear, however, from the entire nature of the study that those interviews are based on the conceptual assumption, based largely on literature from the United States (Lower, 1987), that principal observation is the main, that is, the most common, method of evaluation. I, as researcher, was not disabused of that assumption, the basis and underlying assumption for almost all my questions, by any of the participants -- principals or teachers. Regarding other methods of teacher evaluation used, for example, portfolios or peer assessment, those do not really come within the main purview of our study. Rather, they lie within the range of exceptionality.
Having discussed the provenance and motivation for the methodology, I can now delineate the process of the methodologies, the devices used to confront any avoidance or prevarication in the interviews (pages 33-34) followed by the specific provocations used (pages 43-47).

The Process of the Methodologies in Interviews and Analysis

Besides the outline of a philosophical framework dealing specifically with methodology, a further explanation is called for, largely because of the approach taken here, one that does not fully accord with the normative conventions of social science that the reader may be expecting (and which I am not fully adopting). As has already been explained (pages 30-33 and 34-38), a unique, more aggressive approach to interviewing was adopted in order to advance this study beyond any dogma that might have been encountered in dialogue with the interview subjects. Examples follow of these aforementioned methodologies included under the general rubric of cross-examination and logic.

Analogy, Precedents, Logic and Parallel Cases

Analogy I

Analogy as a specific device of provocation, comparing two similar but not identical concepts, might be used if a position seemed entrenched. At that point, a different aspect
of the issue could be shown and the position could be forced to be re-evaluated. This can be depicted by the following two examples:

*Interviewer (using logic to expose the initial position):* Some of the problems raised by teaching to the test were that if it's a very good test, a very well constructed test that covers the material well, there is nothing wrong with teaching to the test as long as the teacher hasn't seen the test before and can't see the test.

*Respondent:* But how can you guarantee such uniformity across the province so that what I'm teaching in grade 9 or grade 12 English fits so closely with what someone in Thunder Bay is teaching in grade 9 English or grade 12.

*(The analogy of a precedent is introduced in the following paragraph, comparing the old “Departmental” system to the proposed system using standardized exams.)*

*Interviewer:* First of all, in the days of Departmental, at the certain hour of a certain day everyone was teaching the same thing, but they wouldn't do it that way now. However, it doesn't matter what you're teaching. If you are teaching Shakespeare or Norman Mailer, they are going to test you on vocabulary and structure and the quality of English. They are going to test your grammar, or they are going to test you on your ability to write things like that. Those rules are standard.

**Analogy II**

*(A second analogy follows:)*

*Interviewer:* In the States where they do use these scores to evaluate teachers in some areas, they take a baseline of the kids . . . and see how much that is raised by the teacher's efforts. How would you feel about that?

*Respondent:* I don't like that approach to teaching at all. Because then my motivation as a teacher is this kid becomes a number to me. What I'm looking for is this kid to score a certain level on a six-point scale. There are ways that a teacher can manipulate scores.

*Interviewer:* Not if they are graded by someone other than the teacher?

*Respondent:* Yes you can. Absolutely, you can. I've delivered grade nine tests. I know. I saw kids who could barely read and write at a high school level who got level three on a six point scale. I know that without my assistance they would not have had [those] scores [that they had].

An analysis of the protocol of the methodology used here is as follows: the interviewer is being confronted by an answer, “teaching to the test” that is essentially a
cliché. He then confronts the teacher with logic, in the form of the consequences of her thoughts. What exactly is her complaint against teaching to a well-rounded, comprehensive test that covers the whole curriculum? Confronted with this, the teacher’s language becomes defensive and she takes an opposite position. That position is actually almost directly in opposition to what she had been asserting previously. She now states that the tests cannot be guaranteed to be standardized, as opposed to before, when she claimed that they were so standardized and so rigidly formatted that everyone would be teaching to them. Then the interviewer forces the issue by introducing the analogy of a precedent which compares the situation to the era of provincially standardized exams. When the topic switches to using these exams as criteria to evaluate teachers, the respondent protests that there are too many external factors. In response, the interviewer uses an analogy to illustrate how districts in the United States use base-line testing to compensate for external factors. However, the interviewer after hearing the next response, which seems ad hoc and defensive, realizes that the respondent’s position is too intransigent to pursue further and decides to move to the next question.

Precedent

Many of the principals vehemently claimed that it is impossible to judge teachers based on their students’ scores. To refute this, actual cases were presented from New York where student scores are used, and (especially) Tennessee, which uses the Tennessee
Value-Added Assessment System (TVAAS) (Sanders & Horn, 1994, 1998) where studies have shown that no significant correlation exists between economic factors and student performance, were presented. The presentation of these instances of student scores being used to evaluate seemed to refute the arguments of both teachers and principals.

Similarly, when claims were made that teachers cannot be dismissed because of special contracts but skilled industrial workers can, precedents were shown that proved industrial workers were subject to union restrictions and just-cause requirements almost identical to teachers. In Ontario, the vast majority of (even) non-union workers after a 90-day trial period are subject to a just-cause restriction on firing (page 280-281).

**Logic: Reducing the Argument to Absurdity**

This device, which extrapolates the principle advanced by the interviewees until its logical flaws become evident, was usually reserved for the analysis section (my Chapter 5). As this technique was applied to counteract inveterate positions, it was felt, in the judgement of this researcher, to be too incendiary to use explicitly in the actual interview. It was not done because the counter-arguments against the position of the principals and teachers are too weak to expose to the scrutiny of the respondents’ logic or to avoid any opposing arguments. Quite the opposite, restricting this method largely to the analysis section was done because these opposing arguments, presented in the analysis, tend to render the respondents’ arguments so ridiculous that they were saved for the analysis so
as not to embarrass the respondents or cause them to quit the interview. Therefore, they usually appear in the sequestered area of the analysis section.

A prime example is the analysis (page 279) of a principal’s response that uses an excuse of alleged professionalism to avoid dismissing a teaching employee because, being in the teaching profession, that principal felt responsible to demonstrate to an incompetent instructor how to teach. There, and on the following page (page 280), this researcher first comments on the irony of the principal thinking he was running a faculty of education as opposed to a professionally-managed school. As the refutation continues, it reduces the argument to absurdity by going so far as to show how ludicrous the situation might become if someone were to extend that logic.

A similar method was used to refute the claim of teaching to the test because the test is too narrow. It was shown there that if the principals and some teachers claim was extrapolated it would essentially restrict or even prohibit using tests to evaluate students as well (page 352).

Parallel Cases

Parallel cases are the comparison of diametrically opposed views or issues usually between principals and teachers on identical issues raised by the interview questions. An apparent inconsistency in the data arose within the perception of evaluation that seemed to exist in the form of a distinct dichotomy between the views of the principals and the
teachers. This did not occur only once but with some degree of consistency within the following categories: evaluation and the College of Teachers, pages 154-156; formative and summative evaluation, pages 160-162; validity of observations in evaluations, pages 164-167; the status of the follow-up interview, pages 168-169; discriminatory power of the evaluation system pages, 170-172 and so forth. Discovering a protocol for presenting this situation was initially problematic. Essentially, one could confront neither the teacher nor the principal. If members of either group were confronted, they could easily, in all probability, claim that either those teachers or those principals were "not from our school or our board because that type of discrepancy would not exist at our institutions."

Furthermore, it should be remembered that the principals and teachers in this study were from different schools and different boards. In addition, this insight, concerning the existing dichotomous patterns between teachers and principals, only developed as a bona fide theme after the analysis. Nevertheless, despite the limitation, this phenomenon will be commented on as an emerging pattern at several places in the analysis as it occurs. Building on this discrepancy, the researcher uses the frequent appearance of this anomaly to suggest in Appendix C, the section on implications and recommendations in the subsection titled, "Communication," that something of a pattern occurs in the form of a lack of communication between management and front-line workers. Of course, as with other inferences in this study, this suggestion was not intended to establish, hint at or even intimate a firm conclusion that establishes a general principle. That suggestion was used
as an implication that would have to be confirmed by a more broad-ranging and statistically-valid study. However, in order to highlight this apparent discrepancy, the views of teachers and principals were only compared on the same cases in the same questions. This contrasting device of parallel cases, which could only be used in a limited capacity when a blatant dichotomy occurred, was preferable to analogy. Analogy was seen as less exact because, in the final analysis, the two analogous cases were never perfectly identical nor was the comparison strictly made only to teachers or to an educational environment.

**General Methodology**

**Methods and Issues of Selection and Balance**

This study consists of an investigation into the present teacher evaluation systems used in Ontario publicly-funded school systems. Six teachers and six principals chosen randomly from graduate students at The Ontario Institute for Studies in Education of the University of Toronto (OISE/UT) were interviewed. The interview questions evolved from the original project questions, sub-questions, provocation-hypotheses, and finally formed recommendations and implications that are all provided for public scrutiny (see Appendix B and the diagram, Figure 2., Flow Chart of General Methodology, on page 51). This methodology of selection will be described in more detail later (pages 49-62). These questions employed in the interview are used to discover what the principals and teachers
believe are the normative practices of teacher evaluation in their schools, what their opinions are on the effectiveness of these methods, how the evaluation complements the management system and how school evaluation and management compare in accomplishing tasks with management processes in other organizations.
Figure 2.

FLOW CHART OF GENERAL METHODOLOGY

<table>
<thead>
<tr>
<th>Project Question</th>
<th>Sub Questions</th>
<th>Provocation – Hypothesis</th>
<th>Development of Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>With regard to the group of principals and teachers being studied, what is the state of teacher evaluation in the Ontario public school system; what are the elements which enhance or restrict effective evaluation; and what is evaluation's compatibility with the general guidelines on education and selected modern philosophies of management?</td>
<td>What is the nature of teacher evaluation (in the schools being considered) as determined in data collected from the interviews of principals and teachers regarding evaluations' fidelity to their objectives, standards of efficiency and values?</td>
<td>Indifference of Management</td>
<td>Interview Questions (See Appendix A)</td>
</tr>
<tr>
<td></td>
<td>Do the data reveal compatibility between the administration of evaluation and the theories of modern management?</td>
<td>Absences of Incentives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What are the views of teachers and principals on the factors that restrict or enhance evaluation?</td>
<td>Dearth of Objective Data</td>
<td>1, 10, 13, 15</td>
</tr>
<tr>
<td></td>
<td>Do principals and teachers reveal any important data about the conduct of teacher evaluation with respect to the accepted guidelines of the general principles of good evaluation?</td>
<td>Unprofessional Behaviour</td>
<td>9, 14, 17, 18</td>
</tr>
<tr>
<td></td>
<td>Are there any important differences between these views of teachers and principals?</td>
<td>Institutional Weaknesses</td>
<td></td>
</tr>
<tr>
<td></td>
<td>After the analysis and synthesis of the data of this exploratory study, what suggestions can the researcher offer to help advance the present structure towards a more effective and responsive teacher evaluation system?</td>
<td>All of the Above</td>
<td>4, 21, 24</td>
</tr>
</tbody>
</table>
The interviews took place in July and August of 1996 in my office on the 7th floor of the OISE building in the morning and early afternoons. In an effort to explore the difference between administrators and faculty, it was necessary to select and interview both teachers and principals from the secondary and primary panels; from rural and urban settings; (urban including suburban also) and from the separate and public boards. The demographic breakdown was taken from the following sample: two urban Catholic schools, seven urban public schools, two rural public schools and one rural Catholic school. Categorized according to gender, there were eleven women and one man. For a more detailed analysis of the demographic information pertaining to the sample see the following charts (Figure 3A., and Figure 3B.; pages 53-54).
**Figure 3A.**

DEMOGRAPHICS ON PRINCIPALS IN THE STUDY

<table>
<thead>
<tr>
<th>PRINCIPAL</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF EVALUATIONS</td>
<td>30</td>
<td>20</td>
<td>35</td>
<td>35</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>LOCALE</td>
<td>Rural (public)</td>
<td>Rural (public)</td>
<td>Urban (separate)</td>
<td>Urban (public)</td>
<td>Rural (separate)</td>
<td>Urban (public)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>Elementary</td>
<td>Elementary</td>
<td>Secondary</td>
<td>Elementary</td>
<td>Secondary</td>
</tr>
<tr>
<td>YEARS OF SERVICE</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>YEARS OF (EVALUATION)</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>EXPERIENCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRAINING</td>
<td>In-service at School Board</td>
<td>P.D. Days, but only indirectly</td>
<td>1 Day In-service at Board</td>
<td>1 Day In-service at Board</td>
<td>1 Day In-service at Board</td>
<td>1 Day In-service at Board</td>
</tr>
<tr>
<td>TYPE OF EVALUATION PROCESS</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation, but other methods also</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
</tr>
<tr>
<td>NUMBER OF VISITS</td>
<td>2 - 3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2 - 3</td>
</tr>
<tr>
<td>CRITERION FOR HIRING</td>
<td>Partially</td>
<td>No</td>
<td>Usually not</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Figure 3B.

DEMOGRAPHICS ON TEACHERS IN THE STUDY

<table>
<thead>
<tr>
<th>TEACHER</th>
<th>#1</th>
<th>#2</th>
<th>#3</th>
<th>#4</th>
<th>#5</th>
<th>#6</th>
</tr>
</thead>
<tbody>
<tr>
<td>NUMBER OF EVALUATIONS</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>LOCALE</td>
<td>Urban (separate)</td>
<td>Urban (public)</td>
<td>Urban (public)</td>
<td>Urban (public)</td>
<td>Urban (public)</td>
<td>Urban (public)</td>
</tr>
<tr>
<td></td>
<td>Secondary</td>
<td>Elementary</td>
<td>Secondary</td>
<td>Elementary</td>
<td>Secondary</td>
<td>Elementary</td>
</tr>
<tr>
<td>YEARS OF SERVICE</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>YEARS OF (EVALUATION) EXPERIENCE</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>TYPE OF EVALUATION PROCESS</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
<td>Principal Observation</td>
</tr>
<tr>
<td>NUMBER OF VISITS</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
Because of the sensitive nature of the material, specifically, the tough questioning and the unusual perspective of evaluation as a systemic management problem, it was important for the interviewer to maintain his independence and freedom from any type of institutional or hierarchal intimidation. For that reason, the interviews were conducted at the OISE/UT facilities and not at schools or boards of education.

In addition to the exploratory aspects of the study, it was also, at least partially, due to the above caveats and restrictions, considered important to maintain a balance between urban, rural, separate and public educators, with regards to both administrators and teachers. Furthermore, so as not to encounter and possibly escalate any oppositional attitudes that my patterns of questioning might arouse, all interviewees were informally pre-screened. It was felt that this would help balance the responses and avoid any unstable or overly-emotional or hyper-sensitive or unusually compliant people who might not feel comfortable being challenged in a direct, even confrontational manner. However, it should also be noted that a more quantitative approach adjunct to the qualitative aspect of interviewing, which could have applied a statistical analysis to the results, was not possible because randomness could not be achieved due to a perceived certain lack of openness to the type of scrutiny necessary for this study in the school systems and this researcher's own lack of professional connections at the higher levels of the education bureaucracies.

Furthermore, the qualitative approach was felt to be the best methodological paradigm to obtain opinion or other data about internal states. This has already been
discussed (pages 17-26) and, as has been stated clearly throughout the description of this study’s infrastructure, the primary conceptual framework is a comparison of the opinions of the principals and teachers about evaluation. For that framework, a quantitative approach is almost uniquely suited. It is acknowledged that the selection was limited. While every effort was made to achieve a balance between urban and rural, public and separate, elementary and high school in addition to other categories, this was not always possible, due to the restricted sample. One manifestation of that delimitation was the lack of balance in the sample between men and women. This was an unfortunate but inevitable delimitation of the sample selection process restricted as it was, by the number and diversity of administrators at OISE/UT, where women administrators and teachers greatly outnumbered men in the summer semester of 1996. But it also must be noted that this is a qualitative study that only intended to open possible avenues for future research and not be used in any conclusive way. Ultimately, of course, the choice was limited to principals and teachers attending OISE/UT, a population which I estimated was not too restrictive, as teachers from a wide variety of backgrounds enroll in upgrading courses at OISE/UT for career advancement in their fields.

The direct logistics of the selection procedure was fairly straightforward. The subjects were contacted informally first. Permission was granted by various professors to enter their classrooms and offer their students an opportunity to participate in this study. Informal meetings were arranged between this interviewer and the approximately twenty
educator-participant volunteers to explain the procedure and forewarn against any unsuitable subjects as already mentioned (page 55). After six principals and six teachers were selected based on their background, availability, practice in Ontario, emotional fortitude and experience with evaluation, the selection procedure was closed. All teacher participants had to have experienced an evaluation process, and all the principal participants had to have personally conducted an evaluation in order to be accepted. After the selection was made, the questions, which had previously been composed, were pilot tested on several participants, teachers and principals who were part of the screening process but not part of the official study or that study's selection process. This was necessary to complete the preparation process and ensure that all the questions were clear, comprehensible, unambiguous and not misleading, in addition to being fair-minded and not skewed toward giving a pre-determined response.

The problems with such an approach are obvious because it includes only a certain small sample and is totally removed from neutral sampling methods. There is a clear narrowness of representation. However, while this is a delimitation, it should be remembered that this is a qualitative thesis only meant to illustrate some possible patterns to guide future more rigorous studies.

The interviews ranged in length from just under one hour to an hour and a half. Occasionally, if the interviewer felt the question was misunderstood, the question was
explained more carefully. If the response was the same, it was assumed that the question was understood and the interviewer felt comfortable enough to begin probing the response.

To repeat an earlier caveat, the traditionally accepted protocols of interviewing were not practiced here. It was felt by this interviewer that if the usual pattern of academic interviewing was followed, "institutional disinformation" would be regurgitated. To avoid this practice, which essentially would have canceled most of the study's potential gains and insights, a more aggressive, even confrontational, cross-examination approach to interviewing was adopted. In essence, the interviewer tried to prevent the participants from "getting away" with anything, so he constantly confronted them with the implications of their statements. This practice was judged by the researcher as enhancing, rather than detracting, from the validity of the study, its fidelity in defining what it claims to define and its participants' contributions. Although this is not the intended primary contribution to the field of educational research, an ancillary contribution could be this adversarial approach to interviewing, which, if adopted by other researchers, might eventually bring a more tough-minded approach to interviewing when it is used to challenge beliefs. As mentioned previously, there are also some obvious weaknesses to this approach. The aggressiveness on the part of the researcher may tend to become overly coercive and intimidate the interviewees into agreeing to a position that echoes that of the interviewer. Alternatively, an aggressive approach by the interviewer might cause the interviewee to resist and establish more firmly a perspective until it becomes an entrenched position.
Also, more focus was placed on the provocative as opposed to the less controversial but, often, equally important issues.

Any decisions reached about the evaluation system from the interviews were based on the data and the analysis of that data, that is, the analysis of the answers to the questions used in the interviews. Because of the nature of the study, instead of conclusions, the results were recorded in the form of observations and implications in which possibilities for improving teacher evaluation systems beyond their present condition were considered. These insights, of course, were tentative due to the qualitative nature of my study and the speculative make-up of some of these points which go beyond the strict text of the data and presume to suggest ways to solve the problems the data appear to have presented.

The purpose of data collection, though, is not to treat them, necessarily, as an irrefutable truth. But, much as the historian, anthropologist or archeologist does with their data, this study’s data are analyzed and its veracity often questioned. The main purpose of this work is to discover the attitudes of teachers and principals towards teacher evaluation and the processes of their performance evaluations, but, after that determination is made, it does not preclude discerning the viability of those data. Consequently the study will then seek to ascertain, in light of the literature and logical analysis, if these evaluation methodologies and attitudes are detrimental or beneficial to teacher effectiveness and if they maintain logical consistency.
The direction of the study proceeds from the general to the specific. This guideline means to suggest that the study started with the respondents’ general impressions about teacher evaluation in Ontario education and its surrounding cultural legal-contractual environment. The study then moved into more specific questions pertaining to the individual schools and experiences of the teachers and principals. Finally, the study focused on the reconstruction and potential scope of the evaluation under ideal conditions, in addition to the interviewees’ level of satisfaction with the whole contemporary process of teacher evaluation in Ontario.

The rationale behind this sequence is primarily psychological. The respondents were first asked to voice their general impressions about the system in a heuristic fashion and then given the opportunity to voice reasons for those feelings, while still concentrating on the general procedures of the evaluation process. This was designed to be the least stressful part of the interview in an effort to enhance the respondents’ sense of security. Once accomplished, the researcher proceeded to question the respondents and challenge their views specifically and in depth to see how closely those feelings could be backed up with detailed facts and how serious and inveterate those feelings were. Finally, the study dealt with a few pervasive problems that were highlighted throughout the interviews to see if this affected their support for the system. At the same time, attention in the interviews and analysis focused on some general suggestions by the respondents for repairing any failings they found extant in the system in order to give those respondents a chance to test their
suggestions. This last stage, wherein the respondents “solve” the problems in evaluation that they mentioned, gives them the opportunity (after revealing some of the discrepancies in the system) to attempt to consolidate their knowledge into their own consistent system of solutions and to expose their new framework of evaluation to the scrutiny of the interviewer. Thus, the respondents were encouraged to first state their general views, then their specific views, which were more closely scrutinized and confronted and then examined for consistency, and finally, to integrate their initial views into a system of evaluation by constructing a methodology that would work better than the one presently employed for evaluation. Thus in a “psychological dialectic,” these interviews essentially work on a conceptual framework of thesis, antithesis and synthesis.

This pattern is already adumbrated in the first question. There I immediately contrast teachers’ views with those of principals on the best and worst aspects of evaluation within the framework just mentioned. First I present and then analyze teachers’ views, thereafter doing the same for the principals. As this pattern progresses, I see that the material in these interviews covers a wide range of concerns pertinent to evaluation, such as content and philosophy, processes, purpose (accountability, formative or summative purposes), its objectivity or subjectivity and so forth. Eventually, attention is given to the best and worst aspects of the evaluation process as well as to rectifiable and not-so-rectifiable defects in this process. My analyses throughout the dissertation are offered on a topic-by-topic basis, in the section on analysis, on each interview topic.
This pattern of case-by-case analysis was chosen because the topics were so diverse that grouping the individual topics under general headings for purposes of analysis was not deemed feasible. Furthermore, the focus of the analysis is not so much directed to the topics at hand, although this certainly is a redoubtable section of the data, as to the contrast between principals and teachers, which are better highlighted by categorizing the responses under their individual issues, for both principals and teachers. Essentially, until all the analyses on the data had been completed and trends or patterns emerged, this type of categorization would not have been possible. Periodically I use my findings to argue a point bearing upon evaluation. Occasionally too -- and the reader should note this well -- I repeat certain themes, but only as new data reinforce these themes, as a new aspect of those same themes arises from those data. This pattern of apparent redundancy does arise from various portions of my interviews, but that type of back-tracking occurs only because these themes were considered so central to the major thesis of this project that the "same" questions were re-directed from different perspectives.

The following "provocations" are my hypotheses, which represent the major themes from which the questions on evaluation used in the interviews were developed. However, it is within the psychological substructure of the interview process, as the rhetoric of provocation is used to confront the interviewees, where the potential of these provocations fully germinates. As explained earlier (page 23-24, page 41), these hypotheses emerged from my previous experience as a teacher, as well as the literature. To clarify, they do not
necessarily arise uniquely from the professional literature on education, but they do, nevertheless, represent what, in my judgement, were systemic problems. Despite that, though these hypotheses are legitimate and must be proven or refuted, they are also used to extract the opinions, beliefs and rationalizations of the teachers and principals; these hypotheses have proven effective at eliciting strong opinions that often seem to have previously lain dormant.

In addition, this study is equally, if not more so, dependent on conceptual definitions. As such, I have made every attempt to work those definitions and their sources into the text of the study. Where this was not possible, I have supplemented those definitions with an appendix of definitions at the end of the study titled, Appendix C - Glossary. Finally, as a concession at the end of a lengthy explanation, if the reader cannot concede the viability of this type of adversarial methodology, together with its tentative dimensions, it is hoped that the reader will admit that at least the possibility of exploring new insights into educators' views of the strengths and weaknesses of teacher evaluation does exist and might eventually prove useful.

The Hypotheses of Provocation: Issues and Confrontations

Provocation-Hypothesis 1: Unprofessional Behavior

The first hypothesis asserts that the 12 educators’ responses may point towards the educational system in Ontario being (a) inimical to fostering a professional culture, and
therefore (b) their views raise questions about the teacher evaluation being conducted professionally.

My argument is that throughout the system there may be unprofessional characteristics in the relationship between the employee and managerial levels and between the employees (teachers) themselves -- such as the inability to take criticism, previewing of evaluation criteria, paternalistic unions, the inability to dismiss any but the most blatantly incompetent teachers and so forth -- that prevent successful evaluation. A lack of professionalism is seen as pervading much of the educational system both from within the ranks of teachers and management. This condition manifests itself in the way teachers conduct themselves and the way management supervises the teachers. A main effect of this absence of professionalism, however, takes place during evaluation, whereby management does not conduct the evaluation with professional concern and teachers do not conduct themselves with professional vigilance in relation to management, colleagues and students.

Underlying this entire perspective is the expectation of failure, that is, that many teachers will “burn out” after a few years, but continue their duties anyway so as not to lose their livelihoods and pensions. To some extent, in some places, this has become an acceptable mode of organizational operation. As a result, a form of collegial denial exists, another aspect of the syndrome of unprofessionalism, in which the full-time teachers “cover” for each other and sanctimoniously declare each other “excellent professionals” and proud to be working in such an educational system. Privately, however, many make
statements that unequivocally announce their contempt for the entire system ("I have only eight more years to my pension") and make even more severe, inflammatory comments than some of the provocations used by this interviewer.

**Provocation-Hypothesis 2: Institutional Weaknesses**

This hypothesis states that the current institutional structure in Ontario does not provide (a) enough rigorous planning in its evaluation systems for (b) any substantial improvement to occur within it. While some evaluators in schools claim to have elaborate planning and conduct serious evaluations, even they cannot indicate anything especially consequential about their system. An argument could be made that in essence nothing occurs as a result of any of their evaluations. Even in those cases where the evaluations are carefully planned out, a general lethargic attitude is taken towards the whole process. Thus, in the teachers' view, the evaluations are seen as arbitrary which, in turn, reinforces the idea that the evaluations are basically "meaningless."

**Provocation-Hypothesis 3: Indifference of Management**

The argument of this hypothesis is that the current indifference to evaluation within the educational system hierarchy, from the director down to the principal, encourages the inefficient use of any evaluation system regardless of the inherent merit of that system.
In the final analysis, it seems arguable that many educators in the system might be apathetic about evaluation despite written policy statements and testimonial-style declarations upholding the necessity of evaluation. When actions are examined, there seems to be no professional realization about just how important the evaluation of teachers, if done correctly, could be. Indeed, nonchalance is seen as pervasive throughout the system, ranging from the top level of management down to the teachers. As a result, it appears that evaluation is basically treated, more or less, with open approval or at least compliance, but with almost covert contempt by all levels of the organization.

As before, all these factors -- inconsequentiality, absence of incentives, lack of professionalism and a contemptuous attitude -- work synergistically to produce a highly-inefficient system, that is, one that produces nothing of consequence to further its goals of either a mandatory evaluation (resulting in the dismissal of the incompetent or rewarding of the excellent teachers summative), or the rewarding and improving of good and average teachers (formative), or of guaranteeing a certain standard of professionalism to the community (public accountability) as defined by that system’s own parameters. In one sense, a vicious cycle is created that, in effect, destroys the efficacy of evaluation.

**Provocation-Hypothesis 4: Absence of Incentives**

This hypothesis assumes that due to the ineffectiveness of the present evaluation program operating in Ontario's publicly funded school system that teaching, teacher
professionalism and teacher motivation within the system will continue to deteriorate until evaluation improves very substantially.

An argument can be made that teaching is viewed by many people within and without the profession as a "dead-end" job. That is, traditionally there has been no real opportunity for bureaucratic advancement within a teaching hierarchy except by salary raises, which come inevitably on a grid system, based on years of experience after the acquisition of certain basic academic credentials but not on excellence. A pervasive attitude seems to be that the prospects may not be spectacular, but at least the job lasts forever. This attitude, ostensibly so common to public service jobs where no incentives exist, creates resentment and ultimately a contagious apathy which eventually spreads among the good teachers who are not rewarded. What comes home to these exemplary teachers is that, in the absence of tangible reward, no one is really measuring excellence and any attempt to do so is farcical.

In contrast to this attitude, most businesses have different pay scales correlated to their professionals' accomplishments. Compare this situation to the school system where a cogent argument can be made that complacency and mediocrity are entrenched and excellence is rarely recognized. Although it would be difficult to prove that apathy could have already filtered down onto the teachers, at least some of them seem to indicate an attitude that somewhat suggests trying harder serves no purpose because no one much cares.
Provocation-Hypothesis 5: Dearth of Objective Data

The basis for this hypothesis concludes that higher levels of excellence in teaching will be accomplished when attention is paid to objective data. The most likely format for this objective data will be the grades of students on standardized exams.

As the situation now exists, even the mention of the words "objective data" can be anathema. Admittedly, even the relatively objective data, in this situation of teacher evaluation, are not perfect. Nevertheless, there must be a method developed to integrate such results as student test scores on standardized exams into the evaluation formula.

As a result of this attitude that rejects any use of objective data, almost the entire system, at least on occasion, has tolerated mediocrity. It should seem reasonable by now that a system that pays little or no attention to objective evidence, in itself acts as an incentive to apathy.

This of course does not mean to imply that there is an inherent weakness in any interpretive or non-objective paradigm. Indeed, as discussed earlier, (pages 17-19) a certain amount of subjectivity is always present, and it should not be considered a flaw. In fact, subjectivity may be a necessary condition to obtain certain information within certain paradigms. Further to this point is the particular situation here, where our own paradigm, itself, is a subjective one. Our main argument here, then, is not against subjectivity, per se, but rather against those excuses offered by our interviewees that have proven untenable in their defense against using relatively objective data. This relatively
objective data, that is, student scores on standardized exams, in the judgement of this researcher, is more likely to be accepted in court and is more viable to be used as legitimate support for a more consequential form of evaluation if it were to be contested.

The process of denial takes on many different forms, but the common thread throughout the web of denial is the following: there are too many things that teachers do besides the transmission of information which cannot be quantified; therefore, objective data cannot be used because they are misleading. Their claim then is that quantification, is essentially deceptive because it only accounts for a narrow scope of the function of the teacher. However, if this were really true, those “other things,” which some students do well in and others do not, would be reflected in the students’ grades. Conveniently, this excuse allows educators to function without any hard data allowing those teachers who go about “transmitting information” most efficiently to remain unrewarded. It is the underlying assumption of this provocation that a rigorous educational system is not possible without some source of hard data with which teachers can be measured.

**Summary of the Conceptual Framework of the Methodological Structure and Philosophy**

Thus, as in any philosophical system, the philosophy and psychology of this methodology is not free from subjectivity nor immune from bias -- but independently of that ubiquitous weakness it operates within the accepted parameters of a philosophical
system. Essentially, it proceeds logically along a defined algorithm and is free of internal contradictions. Its uniqueness lies in its particular methodology, choice of hermeneutics, psychological substructure and style of interview (provocation). It is these elements which form the conceptual framework and the infrastructure of my methodology (see diagram, Figure 4., Conceptual Framework of Methodology, page 71). There, in that diagram, the algorithm of the methodology is presented in a conceptual framework. Besides outlining the main stages of progress in the methodology, the diagram also presents the sources of input in forming the interview questions, analysis and scrutiny of the data. Essentially, it is through this diagram that the flow of the methodology, its logic and purpose within the boundaries of a conceptual framework, is seen. This illustration is crucial to an understanding of the progress of the thesis, because, as explained earlier (pages 33-37) the specific devices of this methodology will manifest themselves throughout the process of the interview and analysis, which follow in Chapters 4 and 5.
Figure 4.
CONCEPTUAL FRAMEWORK OF METHODOLOGY

1. Project Questions
   - Sub Questions
     - Issues of Contention in Teacher Evaluation
       - Literature & Experience Based
     - Provocative Issues: The Hypothesis of Provocation
       - Research Questions
         - Principles of Teacher Evaluation
           - Literature Based
         - Literature Based
         - Interview Questions
           - Data
             - Analysis or Scrutiny of the Data
               - The Hermeneutics of Provocation: Components, Devices & Methodologies of the Rhetoric of Provocation
                 - Original (Thesis Based)
                 - Scrutiny
                   - Components, Devices and Methodologies of Provocation
                   - Discussion
                     - Analysis
                       - Components, Devices and Methodologies of Provocation
                         - Results, Implications and Recommendations
SECTION II

CHAPTER 3
LITERATURE REVIEW

Itinerary of Literature Review

After illustrating the social science paradigms upon which the emerging discipline of evaluation is based and the criticisms of these models, this review will attempt to trace the antecedents and subsequent evolution of modern management. Following that outline, an effort will be made to conceptualize evaluation in all its aspects as an integral component of management theory with a special concentration on the management of evaluation in education. Within that concentration, sources are cited that highlight specific technical problems within contemporary teacher evaluation. This section also suggests how the application of general guidelines common to all successful systems of evaluation can help in developing a system appropriate for teacher evaluation. Finally, a brief discussion of the Ontario College of Teachers and the legal ramifications of any evaluation for management involved in the evaluation process will conclude the literature review. In addition to the above mentioned topics, this literature review will present a broad range of the research and reflection on evaluation and other related subjects seen as important to the main theme. This account also critiques and, on the basis of the weight, strength and general pervasive opinion of the literature, even forms some initial and strictly provisional, prima facie results about selected evaluation systems and philosophies cited. Beyond that,
some legal considerations bearing upon the subject are also introduced. This literature review then is fairly wide-ranging. It perhaps should be mentioned here that a delimitation exists. This delimitation comes as a result of a distinct lack of Canadian professional literature on teacher evaluation in reliable sources such as peer-reviewed journals. As a result, most of the sources are American. Therefore the convention adopted will be that unless otherwise referred to, the sources will be assumed to be American.

While, in a study such as mine, trying to establish the conditions of teacher evaluation in Ontario and the problems therein and suggest possible solutions, this range of literature focusing on the American situation could be a serious problem. It should be noted clearly here, therefore, that while the study itself is based on data gathered from interviews, which is empirical, the literature review is used as an adjunct to that empirical data and does not seek to refute or establish that data, *per se*, concerning its allegations about the system in Ontario. What the literature review does endeavour to accomplish, largely, is to build on the assumption of the study, not refuted or even contested by any of the participants that principal observation is the dominant form of evaluation. It then seeks to critique that conceptual reality, a basic assumption of the study, while giving some details of the other methods employed to evaluate teachers, making comparisons to program evaluation and discussing teacher evaluation in light of the accepted principles of evaluation and various management philosophies in general.
Thus, while the study itself is empirical, the literature review focuses on the conceptual, that is, the concept of principal observation as a methodology of teacher evaluation, which conceptually is largely the same in the United States and Canada (that is, Ontario).

On the General Process of Evaluation

Perhaps crucial to the entire undertaking of this project is the subject of evaluation itself. As an emerging discipline, evaluation must be able to assure accuracy in discernment more so than a random-choice selection based on certain criteria or data gathered could guarantee. Certainly, evaluation encompasses a data-gathering process, but evaluation is also much more. In order for evaluation to provide useful and valid information for the clients, the evaluator has to know how to get to the bottom line and that cannot be done by the data reduction procedures of descriptive statistics (alone). It has to be done in the particular way that evaluation reduces facts and values to evaluative conclusions. (Scriven, 1991, p. 8).

Specifically, with regards to education and its unique evaluation problems, evaluation could be more accurately defined as

the formal determination of the quality, effectiveness or value of a program, product, process, objective or curriculum. Evaluation uses inquiry and judgement including [the following]:
(1) determining standards for judging quality and deciding whether these standards should be relative or absolute
(2) collecting relevant information and
(3) applying the standards to determine quality (Worthen & Sanders, 1987, p. 22).
Criteria of Evaluation

Admittedly, at cursory view, evaluation might seem little more than a biased selection following some data accumulation. There is, however, an evolving and detailed method emerging that is setting rules in selecting valid criteria; that process goes beyond mere data-gathering. Among the qualifications for criteria selection are the following: removing and identifying inconsistencies in established organizational evaluation criteria, guaranteeing that all important factors have been listed, supplying evidence to support chosen standards and having a method of determining and measuring the extent to which these values are present in the evaluand. Succinctly stated, evaluation is the process of using the aforementioned parameters for determining “the merit worth and value of things” (Scriven, 1991, p. 1) in as objective as possible a manner.

Variation in the Paradigms of Evaluation

While a general agreement might exist on the above outline as a definition of evaluation, disagreement becomes more explicit over the specific methodologies employed in formulating practical evaluation processes. As such, variance can be wide among the models, that is, among the philosophical principles of evaluation, their paradigms that influence their evaluation practices. These, to a great extent, define the methodologies used in evaluation and ultimately the focus of evaluation itself.


Models of Evaluation

The major models influencing the practices of program evaluation can be delineated in the following manner: systems analysis, behavioral objectives, decision making, goal free evaluation, art criticism, accreditation, adversary-based evaluation and the transaction model (House, 1978). While these models were originally constructed to be used for program evaluation, they could also be more generally applicable to developing systems of teacher evaluation. Some of these models such as the Decision Making Model (Rossi, 1983), which directs all raw data to the program managers; the Goal Free Model (Scriven, 1976) where the evaluators are intentionally oblivious to the program goals; and the Accreditation Model, which involves comparing the program(s) being studied to a handbook of standards, a self-study of the program, a site visited by a team of external assessors and direct inspection of program facilities and interviews with staff, clients and students (Madaus et al., 1987) might be difficult to apply to teacher evaluation. Others such as the Behavioural Objective Model (Madaus, 1987; Tyler, 1942), the Transaction Model (Shaddish, 1991), the Adversarial Model (Popham & Dale, 1977) and the Art Criticism Model (Eisner, 1976) will be referred to again for their application and adaptation potential to teacher evaluation. Nevertheless, because of their general importance to the topic of evaluation and their seminal role in the evolving evaluation field, all the models will be outlined here.
It perhaps should also be made clear, here, why it is important to make use of these models. While some of these models are adaptable, as I have already pointed out, they are not explicitly teacher evaluation models. Nevertheless, this project, focusing on the management of teacher evaluation, would be remiss in not mentioning a pivotal article in the field, in which many seminal evaluation models, indigenous and foreign to teacher evaluation, are based. Furthermore, although this thesis deals with evaluation, it should be stressed that it is not primarily of evaluation, but on the management of evaluation and evaluation as a component of that system. As such, these models are necessary to show how managers might effectively construct evaluation systems using valid models that do not compromise the integrity of the evaluation process, per se, while at the same time noticing and removing any systemic management problems in the evaluation process.

During the last thirty years, over 50 types of evaluation models were constructed. These evaluation models included ones developed by Stake (1973), Worthen and Sanders (1973), Popham (1973, 1975), House (1978), Stufflebeam and Webster (1980) and Worthen and Saunders (1987) (as cited in Worthen & Van Dusen, 1994). House’s classification system was chosen for discussion because it goes beyond the superficial characteristics of the paradigms and explores their quintessential teleology and underlying conceptual frameworks. But despite this depth of analysis, models remain only the first step in developing an evaluation system. Similar to other social science models, evaluation
models can only represent an infrastructure from which an empirical system can develop through experimentation and subsequent field-testing.

Nevertheless, in the judgement of this researcher, House’s classification system should prove especially conducive to the development of hybrid systems necessary for efficient evaluation. This is due to that system’s clarity, allowing the differences within and between the evaluation models to become manifest. This synergy, the hybridization of the models, is now being realized as an important aspect of good evaluation systems. As House has noted, “we’re relying more on eclectic approaches. [Models merely represent] a kind of grammar of evaluation….Once you become more fluent you don’t have to use that structure any longer” (House, 1991 as cited in Worthen & Van Dusen, 1994).

Yet despite this belief, a reluctance to use the models as coefficients in an eclectic approach toward evaluation still persists. The problem is perhaps best illustrated by the unique focus of the models developed for evaluation. While evaluation includes program evaluation, personnel evaluation, product evaluation, student evaluation and organizational evaluation, very little model development has occurred outside of the program evaluation area. It is therefore essential to mention these paradigms that give insight into the etiology of evaluation, despite their unique focus on program evaluation, so that their potential for teacher evaluation can be illustrated. To some extent, this aspect is elaborated on later in the chapter on Discussion of Data and Analysis (Worthen & Van Dusen, 1994).
Systems Analysis Model

In the system-analysis approach championed by Rivlin (1971), information is gathered by means of psychometric tests, then statistically-valid results of the entire sample are presented as quantified data after being processed by standard statistical techniques for these types of data. These values are essentially correlational coefficients of pre-designated factors that were determined crucial to the program. Within these conditions, Rivlin has also set up two basic rules for evaluation: "(1) Single measures of social service [or education] should be avoided; (2) Performance measures must reflect the difficulty of the problem (absolute measures are to be avoided)" (Rivlin, 1971, p. 142). Furthermore, Rivlin, a strong advocate of this approach, like her colleagues Tyler (1942) (who used objective-based evaluation) and Popham (1977) (who used statistical measures almost exclusively), assumes a fairly mechanistic approach. The approach seems to accept the assumption that the paradigm used for natural science would hold true for evaluation models in general (House, 1993). That model furthermore asserts that valid results are obtainable only with "the innovation [of planned, rigorous experimentation] which should be tried in enough places to establish its capacity to make a difference in the conditions [under the scrutiny] of [the] scientific experiments being conducted, [that is, the conditions used to validate natural science] should be realized as [closely] as possible" (Rivlin, 1971, p. 91). It should be noted, however, that this claim is still controversial and as yet not a proven assertion.
Behavioral Objectives Model

The behavioral objectives approach was initiated by Ralph Tyler in the 1940s (Tyler, 1942) and later enhanced by Popham (1977). In this approach, outcomes, in the form of pre-determined objectives, “explicitly described [as] what it was hoped learners could do after instruction” (Popham, 1977, p. 45), were compared with actual outcomes based on observable operationalized behaviours. These outcomes, which often included student test scores and other tangible goals of educational programs, were frequently used as data with which to evaluate programs (Madaus et al., 1987).

Decision Making Model

The decision making model transposes the focus of evaluation from the public to the executive branch of the organization and the information necessary for evaluation shifts from a recommendation format to raw data. In this model, almost all the information is channeled to the executives or managers depending on what information their decisions require. It is at this level where discussions are made to accept, enhance or discontinue the program. This model, then, would consider those decisions as the culmination of the evaluation process. Proponents of this approach feel that it is especially suited for large-scale government policy decisions because “our society and particularly its decision makers are no longer sure about what they are doing and are skeptical that theirs is an obviously correct and effective set of ways to accomplish a given social end” (Rossi, 1983, p.22).
Obviously, this approach creates an almost total isolation from the public and has been termed elitist, non-collaborative or even anti-democratic as a result.

**Goal-Free Model**

Scriven (1976) sees the chief impediment to effective evaluation as the inherent bias of the evaluator. This, as he mentions in his later writings, is even more profoundly problematic in the standard forms of teacher evaluation which uses the observation as it is presently conducted (Cangelosi, 1991, Scriven, 1980). In his “goal-free” model, explicit pronouncements of any bias are a requirement of the evaluator as a form of indigenous protection of the integrity of the evaluation (Shadish et al., 1991). As such, one of the chief requirements of the goal-free model, once the evaluator’s biases are clearly articulated, is the evaluator’s absolute ignorance of the program’s goals. With this condition in force, “you don’t need to know what the goals of the program are” (Scriven, 1979). Thus the evaluator should be able to avoid any hierarchical pressure to obfuscate or ameliorate the actual agenda in order to deliver a decision favourable to upper-level management (House, 1978). The use of this model makes “the observer-evaluator struggle hard to find any and all effects without prejudice since his or her reputation is on the line” (Scriven, 1976, p.137).
**Art Criticism Model**

The art criticism model espoused by E.W. Eisner (1976) has been used primarily with educational programs. Because of the retraining of the evaluator's perception that takes place in this process, the evaluator as art critic theoretically learns to discriminate between effective and ineffective practices and policies. "The function of the critic is to illuminate, to enable others to experience what they may have missed" (Eisner, 1979, p. 15). This transformation occurs through a tripartite development process: (1) description (portrayal of observed methods and qualities); (2) interpretation (significance of employed methods, qualities, protocols and so forth); and (3) evaluation (goals achieved by organization's extant policies) (Eisner, 1976).

**Accreditation Model**

The accreditation process has been used with considerable success chiefly in the United States for almost three-quarters of a century. Primarily employed for determining which institutions of higher learning merit degree granting status, the process is restricted in relation to institutions of higher learning in Canada where these institutions are publicly funded and government-approved. For the most part, accreditation in Canada remains limited to specific professional designations granting their imprimatur upon programs within university departments and their graduates by colleges or other similar organizations.
The model itself is similar in many ways to the art criticism model. Nevertheless, Scriven lists several distinctive, if not unique, characteristics among which the following are included: the use of a handbook of standards for comparative purposes, a self-study of the institution resulting in a preliminary evaluation report; a site visit by a team of external assessors who evaluate the self-study; direct inspection of facilities and interviews with staff, clients and students; a penultimate report which recommends for or against accreditation; a panel review at which the right to appeal is sometimes granted; and a final report with a decision (Madaus et al., 1987).

**Adversarial Model**

A more confrontational, perhaps even polemical approach, is the adversarial model. This model uses a forensic paradigm based on the adversarial system espoused by British common law in which two sides present opposing views of various organizational policies. Those policies backed by the strongest arguments are designated for the core of the program. Some of this approach’s sharp tone of disputation appears in this work, without the parry and thrust of competing advocates in a courthouse.

Objections against the effectiveness of this model are similar to those raised against our legal system. These objections claim, with some justification, that the strongest arguments do not necessarily represent the strongest policies as much as they represent the most astute debater. Furthermore, it is doubtful that all policy-oriented issues can be
formulated in congruence with an advocate-adversary design in such a way that will offer useful information. Nor is there any way of preventing an administrator from rendering a decision based on prejudice or bias, despite obviously convincing arguments in opposition to that decision. Any administrator can always claim to have found the alternative more persuasive — which just happens to be the way that administrator would have decided independently (Popham & Carlson, 1977). A possible method of circumventing that detriment would be for the administrator to appoint an independent council to make those decisions. Then, however, the evaluation may seem to be too external and exclusive and even may devalue the program in the eyes of the public by making it look as if its managers are unable to come to an internal settlement. Nevertheless, this method should not be dismissed without further study (Brown, Levine, et al., 1978).

**Transaction Model**

The transaction model maintains that evaluation should be based on a consensus of all stakeholders, both internal and external. This model, acknowledging the inherent subjectivity of using observation as the sole epistemological criterion, maintains that an intrinsic connection exists between the observer (evaluator) and participant (evaluand). As such, this model concludes that observation can never be an objective or dispassionate methodology on which to base policies. This paradigm then conceptualizes the evaluator
role as a supervisory one, deregulating the flow of information. Primarily, it guarantees that all information is accessible to all parties (House, 1978). The evaluator, in this Rogerian-oriented model, "does not provide solutions but provides stakeholders with better understandings of their program so [they] can design their own solutions to their own problem" (Shadish et al., 1991, p.174). As Cornbach et al. have stated: "It is the reader’s task to ‘evaluate’ in the literal sense of the concept, and the evaluator’s task to provide the reader with the information which he may wish to take into account in forming his judgement" (Ambron, Cronbach, et al., 1980, p.154).

The Future of Evaluation

Trends in Evaluation

Although a consensus exists among scholars in the field of evaluation that, in effect, a hybrid approach would be the most practical, a definite division still exists between the quantitative and qualitative approaches (House, 1990). The reason might be the imbalance that still exists in terms of precision, the degree of scrutiny and quality of critical examination between quantitative and qualitative evaluation -- in quantitative evaluation's favour. Inherently linked to this problem is the imprecision in qualitative evaluation and the related problem of forming value judgements and deriving them accurately from data. In addition, the difficulties associated with generalizing from qualitative data and combining fairly the opinions of different participants also represent a formidable challenge
to establishing credibility and equivalency between quantitative and qualitative evaluation (House, 1994). Perhaps accentuating this gap are the advances of the last decade primarily geared toward quantitative evaluation. Such innovations as powerful statistical packages (for example, SPSS, SAS, STATA, SYSTAT) geared to home computers, open access to data sets, e-mail and the Internet (although these last two have uses in qualitative evaluation as well), hierarchal linear modeling and meta-analyses have greatly advanced the cause of quantitative evaluation (Rossi, 1997).

Nevertheless, despite this demarcation, a relatively ecumenical period seems to have recently developed and the demarcation does not anymore, in and of itself, represent a crucial problem. This does not mean to imply that there are no flaws in present day evaluation. In fact, in the 1990s, the following weakness seemed the most pervasive:

(1) technical inadequacy;

(2) lack of sufficient field testing;

(3) insufficient meta-evaluations;

(4) inattention to political dimensions that influence the acceptance and outcome of the evaluation process and results;

(5) inappropriate formats of evaluation due to chauvinism, habit or a false sense of loyalty to a mentor; and

(6) lack of clarity in outlining the criteria of evaluation.
The above can probably be eliminated by taking appropriate precautions in the design of the evaluation programs. Several other problems exist, however, and probably will remain as permanent deficiencies, due to the intrinsic properties of the process of the evaluation. These systemic, intrinsic flaws include the following:

(1) conflicting interests of stakeholders leading to the selective release of information;
(2) bias selection in the choice of variables;
(3) difficulties with potential for standardization and generalization of results;
(4) the obsolescence of the evaluation study by the time the results are produced; and
(5) the inability of the evaluation to do more than identify problems as opposed to produce solutions (Worthen & Van Dusen, 1994).

Polemical Status of Evaluation Models

In summation, although there seems to exist, in a broad sense, a philosophical truce, nevertheless, an almost polemical disagreement among the specific proponents of these various models of evaluation also survives, despite the general consensus on certain major issues mentioned above. Consequently, it cannot be said that any real agreement exists demarcating the paradigmatic variables necessary to establish a generally conceded objectivity, accuracy or epistemological veracity necessary to accomplish an effective,
relevant and significant evaluation model. While this controversy is undoubtedly crucial to evaluation, it is not unique to evaluation. As students of the history of social science know, controversial and apparently seminal issues are still hotly debated within the precincts of social science. To some extent, these remain serious and controversial issues that need to eventually be resolved within the broader spectrum of social science.

**Problems Peculiar to Models of Evaluation**

Aside from these general problems endemic to the social sciences, as well as evaluation, certain methodological problems based on evaluation paradigms for model construction are peculiar to evaluation as an epistemological process. These problems are intricately connected to the theoretical infrastructure on which the models of evaluation are based. One such problem is the degree to which a significant portion of these models are based on organizational and management theory and, to some extent, legal theory (Alkin & Ellet, 1990). If evaluation is to be considered seriously as a science, it must develop a more discriminatory selection process for choosing its epistemological foundation. The disciplines it presently uses for this infrastructure have not yet, themselves, reached the level of rigor usually associated with the scientific disciplines. Consequently, a valid question could be raised about the legitimacy of the level of accuracy one can come to expect from evaluation procedures regardless of what particular model is adopted. These are extremely serious issues intimately involved with the credibility of the entire field of
evaluation and measurement as it is presently constructed and will ultimately determine the level of their acceptance in the scholarly community.

**Status of Evaluation: Discipline vs. Science**

Nevertheless, despite some seriously defining flaws, evaluation can, for the most part, be considered a field moving into intellectual independence, progressing toward the ultimate goal of achieving the status of a science. As of yet, no model has emerged allowing evaluation to claim that distinction. Yet the trend in the literature for some time has been concentrating on the differences and even the advantages and disadvantages of the various models. This trend has generally benefitted the field by refining certain models and forcing them to evolve into more viable sophisticated forms while exposing some as weak and others as extraneous or invalid (Alkin & Ellet, 1994). This situation would seem to indicate not only how far the field has progressed but, paradoxically, how much more it needs to develop in order to mature into a bona fide, reliable process. This progression seems to be part of a normal dialectical evolution that emerging sciences or theories of science undergo. Perhaps, ultimately, a definitive model will be developed that will meet all the requirements of a science. If such a model eventually does develop, society’s social, scientific, political, medical and educational institutions could stand to gain immeasurably. The improvements that could result might provide some of the most significant social and scholarly benefits in the history of applied social science.
Benefits to Educational Administration from Advances in Evaluation Theory

Perhaps even more significant is the potential this avenue of research could provide the entire field of educational administration. As more and better models of management are developed, flaws within the educational organizations become more salient and the necessity for reorganization becomes imperative. In that shift, approaches to evaluation for the systems of management and the organizations themselves become almost crucial for productivity, compatibility and the prevention of costly errors and repeated mismatches between organizations and systems. As these models of evaluation become increasingly more sensitive, they provide the infrastructure to improve educational administration even further. This is clearly in congruence with most of the modern philosophies of management, as is the need for constant improvement. Evaluation, then, becomes not only timely but almost essential to discriminate the functional from the dysfunctional. The key to this process seems to reside within the confines of evaluation theory and should prove valuable to administrators of educational organizations as an important skill for managing their organizations efficiently and effectively.

Evolution of Management Philosophy

Initially, then, it must be realized that evaluation is not an independent, isolated process. If evaluation is to have merit as a component of a management philosophy, it must be integrated and made compatible within the general principles of that philosophy, just
as management philosophy itself must be compatible with the organization's culture (Musella, 1988). During its formative years, management theory seemed to largely ignore this caveat, resulting in the adoption of a largely mechanistic model that based personnel decisions, for the most part, on time and efficiency studies (Taylor, 1911). As a result, management philosophy received the reputation of being exploitive, controlling, narrow-minded and antagonistic toward its employees. As will be shown in the following sections of this chapter, this disappointing beginning has, on the whole, been overcome. The inchoate dogmas of scientific management gradually have given way to management stratagems of greater complexity that are more concerned with the human component inherent within its organizations (Wren, 1972).

**Causes of the Conceptual Revolution**

This conceptual revolution of sorts, however, developed slowly and was not the result of any subliminal altruism. This revolution was probably due to an inexorable pragmatism. More likely, the technological gains that caused unprecedented growth of the industrial complex in the late nineteenth and early twentieth centuries were responsible. As powerful industrialists emerged, they began seeking effective management techniques. Essentially, what they sought was a network of supervision capable of using technical expertise to efficiently direct bureaucratic institutions controlling huge amounts of capital and personnel (Perrow, 1970). Their management philosophy was inefficient and therefore
cumbersome, detrimental and expensive. Consequently, out of necessity, these business leaders began looking towards new philosophies of management (Braverman, 1974).

Problems in Early Twentieth-Century Management

A major problem that had resulted from early twentieth-century management techniques was process compartmentalization. For the most part, this meant workers were no longer personally involved in each stage of production. As a result, the worker stopped feeling either any gratificational or ethical, communal or psychological responsibility for the product's final status. Furthermore, it became virtually impossible for the assembly-line worker to check product quality any more as he or she knew little about the total process (Sallis, 1993).

Mechanistic Philosophy of Early Twentieth-Century Management

Yet even in the developmental stages of management sciences, there seemed to be tacit acknowledgment of the weaknesses inherent in those incipient systems of management. Even then, ab initio, those systems organized along the lines of scientific management were criticized as overemphasizing the components as opposed to the totality; the major flaw within these systems seemed to be the absence of a gestalt perspective. In time, variations of scientific management developed which refined certain aspects of this problem. Pioneers such as H.L. Gantt (as cited in Ling, 1965) and F.B. Gilbreth (as cited
in Gouldner, 1954) were instrumental in starting to move administrative philosophy toward a more holistic approach. Eventually, management became influenced by such disciplines as organizational behavior, industrial psychology and personnel management. Subsequently more humanistic schools of management philosophy would develop. Such scholars of management as Mary Parker Follet (as quoted in Metcalf and Urwick, 1941), Chester Barnard (1968) and Herbert A. Simon (1965) would heavily influence the shift in perspective to an even more humanistic approach.

**Hybrid Philosophy of Structural Analysis**

Despite these developments, scientific management did not completely disappear. Apparently certain advantages in that system and certain disadvantages in the human relations system seemed to co-operate synergistically, prolonging its viability. Eventually, however, Structural Analysis, a hybrid philosophy with tenets adopted from both systems, took hold. Now organization-technology-environment incongruence was seen as the core problem in industrial systems (Guillen, 1994). Because the proponents of structural analysis believed that component-oriented work was the most efficient system, structural analysis developed in many ways to become almost as strongly rooted in functionalist philosophy as Taylor’s (1911) Scientific Management philosophy. However, structural analysis tacitly acknowledged the advantages available for awareness of the emotional and social components within an organization. This attitude nevertheless may have encouraged
a more detached and unemotional management philosophy. Ultimately though, structural analysis seems to have prefigured the modern decentralization movement with its emphasis on appropriate, structural-organizational-environmental compatibility. Yet to its detriment, the philosophy inherent in structural analysis seemed to de-emphasize the psychological and personal concerns of the individual that were to prove crucial in the next stage of management evolution (Guillen, 1994).

Management in Other Cultures: The Re-emergence of Human Relations Factors

While advances were made in western management techniques, they were not commensurate with advances in some other cultures. This apparently was true despite huge investments in research and development in western countries. The Americans, especially, began to resent their consistently low performance when they compared their production figures to those of the Japanese (Womack, 1990). As markets began to develop a more international focus, remaining competitive with the Japanese gradually became crucial to the American economy. A consensus began to form in American management literature urging greater cognizance of human factors within a technocratic system. Ideas emphasizing pride, achievement, quality control and customer satisfaction — qualities alien to the North American sociological and management community’s emphasis on individuality and the workplace’s independence from the social or cultural environment — began to play major roles in formulating management philosophies (Alston, 1986;
Drucker, 1971). As a result of the above problems, and the motivation instituted by new demands for higher quality in a more selective world, radical changes began to occur. Ideas that had been formerly ridiculed in America as too "touchy feely" or too communal, juvenile or affectatious, but already adopted in Japan with much success, were now meeting with authentic enthusiasm in North America for the first time. Such leaders in the field as Joseph Juran, W. Edwards Deming, Philip B. Gosby and Armand Fagenbaum, previously labeled as "crackpots," were now being hailed as visionaries and "management gurus" (Warrin, 1991). As a result, two major new schools of thought became manifest, Lean Production and Total Quality Management (T.Q.M.) (Gabor, 1990; Guillen, 1994 Womack, 1990).

**Lean Production: An Introduction**

Lean production is a system of management that combines insights from the three seminal models of management, (that is, scientific management, structural analysis and human relations) and might be conceptualized as an early form of Total Quality Management. From scientific management, it adopted the "one best way" ideology emphasizing time and motion studies in addition to streamlined accounting processes. Because of the influence of the human relations school of management, teamwork, motivation, feedback and worker satisfaction all formed important components of the Lean Production System. Furthermore, because of the influence of Structural Analysis (Guillen,
1994), decentralization and environmental analysis were also important elements of Lean Production (Guillen, 1994).

Essentially, Lean Production regards buffers (extra supplies, utility workers, surplus space and so forth) as extraneous and expensive camouflage that tie up capital needlessly (Krafcik, 1988); such buffers are seen to obscure crucial problems in the organization's infrastructure, as well as impeding group cohesiveness. As a result of buffer reduction or elimination, workers must be versatile, resourceful and highly adaptive, though not highly specialized if that thwarts their skill flexibility. Workers, theoretically, then, should, in the course of this process, develop a more productive and sincere attitude largely because they realize their dedication to the company is reciprocated. Consequently, the Lean Production System demands that its organizations guarantee a high degree of employment security, compensation based on performance and open communications with all levels of management, in addition to investment in professional development (Lean Construction, 1999; Womack, 1990; Zayko & Hancock, 1998).

This philosophy has an expectation of more productive results from any enterprise incorporating this approach into its practice. Especially noteworthy is the attention given to employee satisfaction, which is thought to result in significantly fewer worker absences. Presumably, management philosophy is instrumental in focusing on excellence in production and customer satisfaction. Lean Production is a system of management that emphasizes teamwork at the expense of alternates, multi-skilled training, minimal
hierarchical authority, responsibility for the final product and strong informal co-operation across lines of authority (Womak, 1990). Total Quality Management (TQM) is in harmony with this outlook but operates with a more pervasive, systems-oriented approach.

**Total Quality Management: Introduction and Principles**

The basic philosophy of TQM demands that management adopt a holistic systems approach for its organization. Essentially the organization must agree to relinquish bureaucratic compartmentalization, that is, departments operating in isolation of each other and the entire organization. TQM advocates believe that when rigid compartmentalization occurs, it is usually to the detriment of the organization despite any ephemeral advantage to an individual department (Deming, 1993). A total quality system, however, trains its employees and managers in statistical control, production management and job analysis. Employee participation becomes essential to such a system in order to obtain accurate data and various strategies to increase it are essential to the success of the TQM system. These strategies can include any of the following: employee attitude surveys, interdepartmental team work, feedback from employees and consumers, as well as transformational leadership techniques. These devices have been used successfully to make employees and managers more motivated, co-operative and responsive to the demands of the organization (Guillen, 1994), attributes which are also essential for the proper evaluation of personnel. Once these ideas are implemented correctly, a TQM management system could be said to
exist. As a result, an organization under TQM auspices is regarded as becoming totally dedicated to enhancing the quality of its psycho-social climate, product and goals. Ideally the organization would begin to run at peak efficiency in all of the areas significant to achieving its goals. It is these systems of management, emphasizing cohesion and employee excellence and, presumably, also conducive to successful evaluation systems, which have, until recently, been seen as lacking in most western bureaucracies.

**Common Theme in Philosophies of Management: The Need to Excel**

Almost synonymous with these contemporary philosophies of management is a passion for excellence in the quality of production. Excellence, roughly identical with an ideology of constant improvement, can be defined as a continuous refining process involving the quality of the manufacturing procedures, the employees applying those procedures and consequently the resultant product. A corollary, almost as important, is the customer satisfaction component. Yet despite a consensus in the literature of modern management philosophy’s esteem of these ideals, (even expressed as essential components in any organization’s management protocol), administrators in educational institutions have largely ignored their significance (Frase & Streshly, 1994; Pohland & Wood, 1983). Furthermore, while TQM has entered educational literature and thus the realm of educational administration theory, if not practice, theoretical references to Lean Production, in that body of literature, are rare.
Teacher Evaluation: History, Methods, Problems and Guidelines

History and Development of Teacher Evaluation

Yet, despite the agreement on quality as a component of most systems of management, professional evaluation within educational administration has developed very slowly. To illustrate this a brief history of evaluation will be shown based largely on American sources. It can be assumed that while there are strong similarities to Canada in general and Ontario in particular, there are also significant differences. This is an admitted, but unfortunate, deficiency necessary because of the lack of many sources for the historical development of teacher evaluation in Canada and specifically focusing on Ontario.

Chronologically, teacher evaluation in the United States began, though in an inchoate and narrow format, even before the industrial revolution. At that time, with the founding of the first public schools in seventeenth and eighteenth century colonial America, evaluation was used sporadically and primarily as a means of dismissing teachers for blatant incompetence (Vold, 1985). That type of ineptitude was usually defined only by serious breaches in either professional responsibilities or in the strict socially-imposed moral code of that era, made even stricter for teachers. Later that period, however, some latitude appeared in the evaluation process, which came to be used as a means for establishing a teacher's credibility for increased remuneration (Clark, 1993; Peterson, 1982).
With the thrust of the industrial revolution at the beginning of the nineteenth century and its emphasis on the regimentation of production subsections, under those influences, a more specialized approach to education gradually caught on. Concomitant with this narrow focus, however, somewhat paradoxically, an incipient public demand for more competent, qualified, professional teachers surfaced. Despite these demands, teacher quality did not improve much and, overall, a relatively low quality of teacher professionalism prevailed during that period. The rule in New England at that time seemed to be that parents should not interfere (Bettencourt, 1982). Even as late as 1920, this virulent trend towards teacher incompetence was so pronounced that instructors were generally portrayed as “the misfits [who] seem to have been [numerous enough to be] so conspicuous that they set off an unflattering image of the teaching profession” (Hofstadter, 1963, p. 314).

To halt this deleterious image, teacher evaluation was suggested as a solution, in order to maintain, at the very least, a basic, minimum professional standard for teachers. Consequently, because of this emphasis on minimalism, evaluation, as a redoubtable force within formative, summative or accountable educational management, did not materialize. In fact, although in contrast to evaluation practices of the last century, a more professional attitude prevailed at the beginning of the twentieth century; still, only three fundamental criteria were used for evaluation:
(a) unquestioned obedience to authority, as defined by the observance of autocratic rules;
(b) the enforcement of certain "standards of education" as decreed by a council of lay advisors; and
(c) a summative evaluation process that assumed responsibility for the dismissal of grossly incompetent teachers (Peterson, 1982).

Interestingly enough, despite the persistence of these restrictive and regressive ideas, by the second half of the nineteenth century, the first teachers' unions started to appear in North America. Despite the imbalance in the power they faced when confronting contemporaneous educational establishments, these unions began to demand equitable and professionally-planned evaluation procedures and limitations on those procedures in the form of clear, specific criteria that would determine teachers' possibilities for continuation, dismissal or advancement. However, like their counterparts in industry, these teachers' unions, in both Canada and the United States, were only in an inchoate stage of development and in no position, therefore, to enforce any of their demands (Clark, 1993; Johnson, 1968; Stamp, 1982;).

Causing that situation to deteriorate further, the subdivided forms of micro-focused supervision, then in vogue, meant that management was inimical, if not antagonistic, or blatantly hostile to the goals and values of those incipient teachers’ unions. Unfortunately, this tension existed not due to any ethical or quality-based issue but merely, the general
principles of a rigid hierarchal management structure of the educational bureaucracies. And this persisted despite the potential of those unions to offer valuable contributions to the quality of education, if indeed a synergistic relationship had developed. Consequently, local communities, counties or states largely determined the criteria, parameters and philosophy of any sanctioned evaluation procedure within the public school system. To aggravate matters further, little if any consultation with teacher representatives took place, official teaching bodies or any benign unofficial group extraneous to those incipient unions (Clark, 1993).

For the most part, then, this autocratic brand of teacher supervision continued, with relatively little variation, until the beginning of the twentieth century. In effect, the evaluation then extant was generally undemocratic, hierarchal and both dismissive and condescending towards students’ rights and teachers’ professionalism. Nevertheless, management continued to rationalize this approach based on its opinion (which, to be fair, was not entirely inaccurate) of the low caliber of teacher available (Peterson, 1982).

Yet prior to the twentieth century, any professional evaluation was heuristically based, and only tentative prospects existed for any potential success as an evaluative tool. Legitimate scholarly research of teacher evaluation can only be dated from the works in 1905 of J.L. Meriam: he demonstrated the low correlation between grades obtained in teacher-training programs, scholarship in those programs and actual teaching ability or between practice teaching and professional teaching ability. Only a nebulous correlation
occurred between experience and teaching efficiency, if one excludes the first year of teaching (Peterson, 1982). This was the case in Ontario too, where inspectors were almost exclusively concerned with whether or not the teachers were enforcing a sequence of autocratic "top down" rules at the expense of any real beneficial supervision, either formative or summative (Houston & Prentice, 1988; Prentice & Theobald, 1991). Besides that, these inspectors only other real concern was whether or not the schedule regulating the exact amount of time for each subject and duty was performed to the letter (Danylewcz & Prentice, 1986). In addition, there was also strong discouragement for any innovative or even independent thinking by either superintendents or teachers (Stamp, 1982). This period of teacher evaluation stressed strict uniformity and unquestioned obedience (Salisbury, 1912).

While labour and management began to become involved increasingly in conflict, various management strategies came under the influence of the factory-inspired values of Frederick Taylor. As a result, before the second decade of the twentieth century, production and task analysis became the dominant forms of industrial supervision. This philosophy transferred to education and permeated educational management to the point where such apparently innovative educators as Bagley and Cubberly were among Taylorism's strongest supporters. Their advocacy of that truncated, structured management outlook went so far as to concede the necessity of unquestioned authority to a chain of command, dominated by better-educated, upper-echelon bureaucrats as a management tool.
for successful education evaluation (Callahan, 1962; Gitlin & Smyth, 1989; Spring, 1986; Wood & Pohland, 1983;). Thus, the majority of contemporary mainstream thought about school management remained stagnant in a mechanistic mode, caught up in the ideas of the early part of that century and the latter part of the previous century.

Despite that tendency, even then, there were already exceptions to such fierce regimentation. Such critics as Sara H. Fahey and Ava L. Parrot in the 1920s (as cited in Peterson, 1982), both frontline schoolteachers, opposed the prevailing management systems of evaluation on the grounds that they were patronizing, superficial and detrimental to the professional status of teachers. Their criticisms claimed that the rigidity and bureaucratic nature of those contemporary management systems isolated teaching from its dynamic in-situ context, thereby failing to conceptualize teaching as a vital gestalt process (Peterson, 1982). Even more outspoken in their opposition were critics like Edward Elliot (Herbst, 1996) who insisted, indignantly, that the public school systems had the right to be exempt from the pernicious effects of the bureaucracies that had already corrupted the various levels of government. Yet despite these voices of dissent, the rising strength of unions and teacher resentment of that system of management and evaluation, the cautionary outlook of the teaching profession coupled with the innate conservatism of the era, thwarted any concerted or sustained effort towards meaningful change. Therefore, the status quo defined by a mechanistic efficiency-oriented evaluation system continued as the accepted norm (Herbst, 1996).
In the period ranging from 1920 to 1940, the evaluation systems of the previous two decades came under considerable scrutiny, resulting in criticisms of their unscientific nature. Even the anticipated positivist rigor usually associated with the minimalist, efficiency-centred approach as a major benefit was noticeably absent from the evaluation repertoires. A breakthrough of sorts, however, occurred when Rugg published an influential study in 1922, *Self-Improvement of Teachers Through Self-Rating* which concluded that the contemporary rating systems were essentially useless (as cited in Peterson, 1982).

That study, along with a changing industrial climate, established the existence of a management vacuum. As a result, scientifically-designed “objective” teacher evaluation methods were being suggested. These evaluation methods included the following: pupil’s evaluations of teachers (backed up by such studies as Bryan’s *Pupil Rating of Secondary School Teachers* (1937) and, Knight’s *Qualities Related to Success in Elementary School Teaching* (1922)); measurement of pupil progress over a given period; initial interviews of candidates before entering teacher training to assess potential for success; and the identification of personality characteristics and other factors relating to success in the profession (cited in Peterson, 1982).

Although those studies represented some highly innovative thinking on supervision and evaluation, most of the actual changes that distinguish evaluation procedures from those of the previous era focus on psychometrically-designed rating scales by such pioneers
in the field of evaluation as Knight (1922), Monroe and Clark (1924) and Barr (1929), (all as cited in Peterson, 1982). Yet in this search for a "philosopher’s stone" of educational evaluation, these researchers may have overlooked some of those strategies previously mentioned that were more innovative but less paradigmatically positivist. As a result, those scholars came to employ various pseudo-psychometric instruments which were, in fact, crude and only possessed limited potential for accuracy, validity or reliability (Peterson, 1982). Although these practices influenced Ontario, this type of reliance on pre-packaged psychometric tests never reached the epic proportions that it did in the United States (Gillet, 1969).

In the following twenty-year period between 1940 and 1960, the philosophical basis for evaluation, probably influenced by humanistic trends in other fields of management, shifted, in general, toward a more inclusive model. This paradigm shift caused a democratic revolution of sorts, in that, for the first time, all staff members were consulted in the evaluation at all levels of its formation and implementation. Also, for perhaps the first time, evaluators began looking at other aspects of teaching besides actual classroom exchanges. These concepts included the following: leadership potential, peer relationships, dedication and professional growth. This trend towards democratic management gradually filtered through to the general public. Consequently the public stakeholders became involved in criticizing the evaluation policies and practices designed to achieve excellence in the schools (Peterson, 1982).
Furthermore, by the early 1950s, more men began entering the teaching field. With them came their male-oriented concept of organization and expectations of prestige and political power. This resulted in stronger unions, higher salaries and the conceptualization of teaching in the public perception as a bona fide profession. By 1957, however, with the spectre of the Soviet space craft Sputnik “threatening” the not-yet-in-space United States, the years of complacency seemed ready to be replaced by a new and vociferous demand for excellence. Overall, there seemed to be a degree among the general public that the Soviet Union was producing better and more technically- and scientifically- qualified teachers. Eventually it was felt that the Soviet Union might even be able to win the cold war, an idea not taken seriously before 1957. Consequently the unions with their growing powers, coupled with the lay movement for democracy in education and educational management, acted synergistically to produce a professional and public outcry for a more professional teacher evaluation format (Clark, 1993).

However, because of this greater democracy within the professional ranks among teachers, principals and supervisors, certain integral aspects of the educational system itself also began to be challenged from within that profession. Questions confronting the basis for selection of curriculum content and even the professional duties of teachers were starting to crystallize among the public and professional ranks. As these challenges began to influence choices in curriculum and the venue of professional responsibilities, they also began to influence policy formulation, especially regarding evaluation. Because of the
nature of these challenges from within the ranks of educators, which emphasized a greater professional deference for teachers and increased democratization, the concept of essential teacher responsibilities seemed more nebulous than in the past. Subsequently, a form of educational anarchy began to develop where equity under the guise of humanitarianism was coupled with a tacit acceptance of incompetence (Peterson, 1982). In Ontario, too, teacher evaluation shifted from the rigid Taylorist or even the psychometric model to a more human-relations oriented approach. In that paradigm, a more concerted effort was made to gain the respect, co-operation and feedback of the teachers (Fleming, 1972).

In that climate, then, a brief period of acquiescent acceptance occurred. However, by the end of the first lustrum of the 1960s the demand for excellence and rigorous evaluation, which was to last until the 1970s, grew with almost inexorable strength. Again, it would be naive to attribute that phenomenon entirely to increased public assertiveness and awareness. Undoubtedly, real deficiencies existed in the educational system that were serious enough to arouse that awareness. By this time, scholars of education such as Redfern, Heald, Moore, Biddle and Washington had also entered the fray. A consensus among researchers seemed to emerge that some sort of evaluation of teacher effectiveness was necessary to re-establish standards.

Although a certain consensus about the need for improved evaluation had materialized, beneath it all lay fierce divisions. Washington’s (1971) study cited nine distinct goals of evaluation in contrast to G. B. Redfern’s three objectives, though neither
enumeration is consensual among most researchers. However, as Heald and Moore (1968) observed, almost all scholars agree that teacher improvement is a major goal of evaluation.

During this twenty-year period, 1960-1980, then, several rating scales intended for use in evaluation of varying effectiveness, foci and strengths were crafted, in an attempt to improve the quality of teaching from different perspectives. The following are examples: The Torgeson Scale (rates 18 characteristics usually associated with teaching); The Purdue Rating Scale for Instructors (rating scale in which pupils are asked to rate teachers); and The Bryan-Yntema Rating Scale (another scale used by students to evaluate their teachers but focuses on the secondary school level). Nevertheless, as early as 1969, scholars such as Redfern cast serious doubts on the whole performance-goal-observation approach to evaluation because of the difficulty in defining or describing in any clear operationalized format the essential and determining elements which describe effective teaching (cited in Peterson, 1982). In this regard, Biddle (1964) has also stated that despite thousands of studies dealing with teacher characteristics, behaviour and goals, few apodictic "facts" have been established, no universally accepted rating or psychometric model has been developed nor has any methodology been discovered that guarantees some process to ascertain teacher adequacy.

Therefore, despite innovations within evaluative systems, the basic paradigm of observation and transcription remained the most widely used. Despite the longevity of its tenure, however, there was a resounding dissatisfaction with that process, as it existed,
from both administrators and teachers -- with the stronger feelings of dissatisfaction coming from teachers. If any important innovations were to occur that would revolutionize evaluation, they would have to happen in the next twenty year period (Peterson, 1982).

Following a brief respite and a period of seeming public satisfaction or at least acquiescence to the prevailing system of evaluation and education, a crisis call to arms occurred. In 1983, a report was issued by the National Commission on Excellence in Education called *A Nation at Risk: The Imperative for Education Reform* that decried the present American system of education and evaluation and challenged the public to seek excellence in public education. As a result of that report and others, most notably one by the National Education Association titled, *Schools for the 70s and Beyond: A Call to Action*, innovations began to appear (Clark, 1993). In fact, within two years, 33 states had implemented at least five of the 40 suggestions, and half of those states had taken on ten or more of the suggestions (Shinn & Van Der Silk, 1985,1988).

The shock and harshness of that report seemed to release a torrent of suppressed criticism. Suddenly, but not without some justification, evaluation was being labeled as a potential saviour, but one that thus far had been a misused as a useful resource. Held up to the harshest criticism, perhaps because it is the oldest and still most widely used, was the practice of observing the behaviour of teachers in situ, that is, while teaching students during an actual classroom session. This often repeated claim of invalidity is based on the error known as the "halo effect," the tendency in impression-oriented and evaluative
situations to give the same teacher a consistent rating on virtually every item in the scale. Concerning the evaluation of teachers, this particular pernicious attribute manifests itself by forming a general impression of the teacher, usually based on previously formed informal observations. Furthermore, that type of untrained observation provided little, if any, useful information with the potential to increase teacher effectiveness (Cangelosi, 1991). Consequently, this inefficiency often resulted in just repeating different aspects of that previous impression, regardless of how the teacher actually performed in the specific rating category. In essence then, the observer, usually the principal, would give the teacher the evaluation that was felt to be what the teacher deserved any way. This regimen, for all intents and purposes, vitiates the evaluation procedure as a process with any real fidelity to a legitimate protocol (Cooper, 1981).

**Evaluation: Methods, Problems and Guidelines**

**Contemporary Problems in Teacher Evaluation Systems**

The most common form of contemporary evaluation, is classroom observation, (usually principal observation) and can be subdivided into five distinct variations: ecological (the recording of every event during the observation period), ethnographic (the recording of significant events as determined by the observer), in-class rating scale (recording of pre-determined behaviours on a Likert or similar scale), structured observation (observer's attention is focused on several pre-determined behaviours) and
informal (impressions based on chance or casual classroom observations) (Cangelosi, 1991).

In addition, since most of these observations take place after the teacher has seen the evaluation instrument, a test preparation mentality may result where teachers produce a specific *ad hoc* behaviour which may, in fact, rarely occur in the teacher’s actual professional teaching practice (Bridges, 1986). So far as is known, no research-based answer yet exists to resolve what the probability of that type of affectation has of occurring during an evaluation-based observation. However, just from a logical, heuristic perspective, it would appear that a teacher under pressure would give examiners what they want to see regardless of how that performance relates to teachers’ usual practices (Medley, 1992; Scriven, 1980). In addition, such validity and accuracy-enhancing strategies as team observation, familiarity with the subject (Scriven, 1987), a dramatic increase in the number of observations or different criteria for different teaching situations (Stodolsky, 1984) are still seldom used (Stodolsky, 1990).

In that light, the severe argument can be made that, “modern” teacher evaluation in its most common form has, conceptually, changed relatively little in the last 50 years. Ignoring blatant flaws in systemic design and recommendations from educational research, that field appears rife with poorly designed evaluation instruments, untrained evaluators, that is, principals, blatant bias and other methodological weaknesses. Furthermore the evaluation instruments do not even meet the normally accepted social science standards of
validity generally demanded of psychometric rating scales or tests (Cangelosi, 1984; Peterson, 1983; Quirk, Witten & Weinberg, 1973; Soar, Medley & Cocker, 1983).

Problems of Evaluation in Teaching

Systemic Problems in Teacher Evaluation Systems

This section on problems of teacher evaluation will begin by outlining some of the alleged aspects of ineffectiveness present in contemporary evaluation systems due to poorly designed evaluation forms, untrained evaluators and other methodological weaknesses. A clear indication of this situation, that is, the lack of progress in the areas of educational management and evaluation, is the 1979 Gallup Poll data declaring the number one suggestion for improving U.S. schools to be the improvement of teacher quality, of which a significant prerequisite should be teacher evaluation (Fraser & Streshly, 1994). There is little indication that the situation has changed since then. This statement seems reasonably certain since not measuring or monitoring excellence almost precludes the possibility of ascertaining, with any degree of objectivity, whether it is being attained or not. As further evidence of this conjecture, a 1986 Gallup Poll showed that 85 per cent of its respondents felt that prospective teachers should be required to pass a state exam in their teaching subjects (Fowler, 1988) -- another example of public dissatisfaction with the proficiency of teacher professionalism and the concomitant demand for evaluation and quality control. Despite this demand, evaluation of teachers, in almost any contemporary form, arguably
remains, perhaps the weakest point in the entire educational management system. This is evidenced by the existence, in the majority of scenarios, of the same general paradigm for selection of the components of teacher evaluation as was used at the turn of the century. Essentially, the concept of teacher competence accepted then is still considered accurate (Pohland & Wood, 1983). This result could hardly be called a customer-oriented modern management approach, essentially, it was not accurate then, and it does not appear to be so now.

Nor is it encouraging when one realizes just how thoroughly the general feeling of frustration and dissatisfaction with teacher evaluation permeates the literature. The extent of these feelings has reached the point where teacher evaluation is viewed as either having so little merit as to be useless, essentially considered a waste of time and money or at the very least a seriously flawed process or one with significant procedural errors (Anctil, 1991; Boyer, 1983; Darling-Hammond, 1990; Duke & Stiggins, 1986; Frase & Streshly, 1994; Goodlad, 1983; Haertel, 1987; House, 1989; Housego, 1989; Musella, 1988; Newton & Braithwaite, 1988; Peterson, 1983, 1986; Pohland & Wood, 1983; Popham, 1987; Poston & Manatt, 1993; Schwab, 1990; Scriven, 1995; Scriven, 1980; Smyth & Gitlin, 1989; Soar, et al., 1983; Stiggins, 1986; Wiggins, 1989; Wise, et al., 1985; Wise & Darling-Hammond, 1984;).

Even the standard forms of teacher evaluation used in schools are considered so inaccurate for predictive or assessment purposes that researchers have used their general
formats and methods of application as an argument for supporting teachers' reluctance to be evaluated at all. These same authors claim the source of teacher reluctance to accept evaluations is largely due to evaluation being "subjective, unreliable, open to bias, closed to public scrutiny and based on irrelevancies" (Soar et al., 1983, p.246). According to those researchers, their resistance does not seem to come from any fear that teachers have of being exposed as incompetent.

**Weaknesses in Teacher Evaluation**

Problems with teacher evaluation are seen as so widespread that they nearly outweigh any other systemic criticism in educational administration. When one adds to the above data information about principals conducting the vast majority of evaluation with no absolute requirement to evaluate or significant special training to prepare them (Housego, 1989), it should come as little surprise that their evaluations have often been challenged as unprofessional. To reintroduce a point, principals' evaluations were even viewed as detrimental when it became known that, in general, they were based on inconsistent classroom visitations and provided little if any constructive criticism or guidance for teachers (McGrail, 1982; Sergiovanni & Starratt, 1992). Thomas (1984, p.32), commenting on this lamentable situation, proclaims that "while it is not feasible for all districts to have learning specialists, principals [nevertheless] need extensive training to properly evaluate." As Thomas correctly implies, there is far more to evaluation than
marking off a checklist. The preponderance of actual opinion, then, is that, in its present form, teacher evaluation has become just that, a perfunctory routine that exists as an organizational requirement, serving only political or legal purposes thereby rendering it useless (Housego, 1989). As Scriven (1980) comments in agreement with the above views, the present most common form of teacher evaluation based on untrained and unrehearsed observation by a principal is a disgrace.

Management Problems in Teacher Evaluation

As alluded to earlier, evaluation generally exists in two formats, formative and summative. Formative evaluation refers to evaluation that is used primarily for feedback so that the employees can improve themselves without fear of repercussion. Summative evaluation is used to determine rank and to assess the ultimate value of the employee up to and including dismissal. While summative evaluation is not inherently antagonistic to modern management systems -- since reward, promotion and dismissal are often necessary -- some version of formative evaluation is essential as a major component. Furthermore, the principle espoused by modern management stresses a corporate culture conducive to the recognition of enhanced performance and excellence. This environment can include a salary structure, the encouragement of a more professional environment, career ladders, job enrichment and practical, valued professional development (Freiberg, 1985). Yet outside of the Tennessee Career Ladder program, which offers pay incentives largely
based on unbiased criteria, there appear to be few entrenched, incentive-based programs of any kind in public educational organizations in North America (French, 1985). That situation, as far as this researcher has been able to determine, has not changed radically since 1985. This suggests, *prima facie*, that an anachronistic management approach might be prevalent, even today, within these educational organizations.

**Reasons for Weaknesses in Teacher Evaluation**

The reasons for the existence of laxness are more than likely variegated. However, in all probability, a significant share of the responsibility lies rooted in the culture of the schools, their management philosophy and their bureaucratic infrastructures. Commenting on the consequences of these factors, Mintzberg (1979) has noted the rigorous consequential approach business takes towards evaluation. This approach seems directly influenced by the methodology of evaluation itself within business organizations, which is product-oriented. In contradistinction, educational evaluations are focused on process (Levin, 1979; Mintzberg, 1979). Two of the most important factors responsible for maintaining teacher incompetence are the nearly unbreakable teacher contracts and the problems inherent in evaluating classroom teaching (Bridges, 1986). Bridges’ confirmation of the importance of those two criteria affirms that a major problem with developing consequential teacher evaluation lies at the management level. If product-oriented evaluation were to be used and acknowledged in the contract as valid, a major obstacle
preventing the dismissal of incompetent teachers would no longer exist. However, probably because these caveats have been ignored, ludicrous situations occur, like the one in California where in 1984 it cost $166,715.70 to dismiss a teacher, almost seven times that teacher's annual salary of $26,000 (Bridges, 1986; Van Sciver, 1990). Much of that cost was expended on court costs (which would have been much higher if the teacher had won). Unfortunately, decisions to dismiss are almost inevitably challenged, so costs in that range are not exceptional.

Not only is teacher evaluation inefficient, it is often arbitrary (and as mentioned previously) bureaucratic, non-professional and based on unsuccessful formative models (with respect to likelihood of employee improvement) (Bridges, 1986). Furthermore, these just mentioned flawed and common characteristics offer some explanation why contemporary teacher evaluations usually concentrate only on the superficial aspects of the teaching process (Wise & Darling-Hammond, 1985).

**Systems of Teacher Evaluation and Problems Inherent in Their Development**

Specific methods of evaluation are often difficult to rate, compare or even categorize. These methods are diverse and usually operate under various assumptions, perspectives and biases. These methods can include classroom observation, peer evaluation, teacher self-evaluation, student evaluation of teachers, the use of students' standardized test scores, teacher portfolios, psychometric tests, classroom simulations or
a combination of any of the above. While this is not a complete or exhaustive list of systems, these categories alone should be able to indicate just how many systems or system permutations of teacher evaluation might be encompassed within these aforementioned categories. Yet few of these innovations are used on a pervasive basis. What I will attempt to show in this section is that despite this narrow approach, by following certain guidelines and certain management philosophies, much of the criticism presented here might be avoided. Discussed below are general problems in evaluation systems such as process evaluation, subjectivity, arbitrariness, lack of employee consensus and lack of clear guidelines.

The Range of Problems in Teacher Evaluation

As previously argued by citing a number of sources, a general and legitimate dissatisfaction is manifest with teacher evaluation. Yet, while dissatisfaction is widespread, if not ubiquitous, the intensity of disapproval is not uniform. Even within the extant systems, a quality gradient can be said to exist. For example, of the systems regularly used to evaluate teachers, the paper and pencil psychometric, knowledge-based style of standardized tests are the most severely criticized (Haertel, 1987). Seeing as these systems have almost none of the core parameters (see pages 124–127) of efficient effective evaluation systems, it would be inaccurate to stigmatize all the evaluation systems presently in use as equally inefficient.
Inherent Flaws in the Systems of Evaluation

It seems, however, that almost all these systems fail, at least to some degree, because of non-compliance with the "Levin Rule" (Levin, 1979) that criticizes process-based evaluation systems due to the inconsequentiality of process-based criteria without connecting them to a product result. Essentially, a logical circuit, fueled by philosophical argumentation, develops because of the bias inherent in the particular values of the system's authors. (Even the much vaunted System for Teaching and Learning Assessment Review (STAR) evaluation protocol described in more detail on page 128, which prides itself on objectivity, is also based heavily on phenomenon observation, which may or may not ultimately relate to teacher effectiveness.) This complaint concerning emphasis on observation is a valid one because, to reiterate an earlier vital point, observation itself, by definition, is a highly idiosyncratic activity. This format becomes crucially detrimental when untrained observers, that is, principals, who have been shown in many studies to have approximately zero correlation between their observations through rating scales and any aspect of effective teaching (especially with student gains in information) (Bridges, 1986; Brookover, 1945; Cangleosi, 1991; Coker & Medley, 1987; Hellfritsch, 1945; Hill, 1921; Jayne, 1945; Jones, 1946; La Duke, 1945; Lins, 1946; Medley & Mitzel, 1959), record behaviors that are likely altered by the mere presence of an observer and almost invariably processed through inherently value-laden, culturally-oriented systems. While not all the authors have conducted independent studies, for example, Cangelosi, Leese and
Bridges have not, they all support the idea that principal observation, as it is presently practiced, is an unreliable method of teacher evaluation, based on their insights, practical experience or reading of the literature. In these systems, what is often recorded and adjudicated is teaching style. This is highly unreliable because what one evaluator calls "obstreperous and disorderly," another might call "redolent with unrestrained enthusiasm for learning." Thus, the position taken here is that systems that are value-laden, process-based or observation-based, especially for untrained observers, and much of principal observation as presently practiced in Ontario may contain all three problems, are inherently flawed. In effect they become oxymoronic systems claiming internal objectivity but are, in fact, based within their own system on inherently biased parameters external to the monitored situations. As a result, these systems often fail to produce tangible results that can be applied to teacher improvement. Because too much of human activity is dependent on underlying motivation and intention, evaluation of this type, based on observation, almost always becomes a non-impartial undertaking. This is in addition to human error, human weakness, prejudice, dishonesty and psychological motivation (conscious and subconscious), all factors involved in and affecting accuracy in observation.
Conflict and Controversy: 
Objective Data in Teacher Evaluation Systems

Acknowledging the above, moreover, does not mean that the objectivity of a system automatically repudiates any claim against the system’s veracity. It does mean, however, that certain major flaws such as the use of arbitrary criteria and process orientation are, usually, at least partially, corrected. Illustrative of this is the fierce controversy surrounding assessment strategies that use the relatively objective, product-oriented criteria of standardized student grades for evaluation of teacher capabilities and competence, all routines that much of the literature rejects on other bases.

The standard arguments, from which most others derive, presented in the literature against using student performance to gauge teachers’ effectiveness include the following:

(1) the extraneous socio-economic factors which affect the teacher and learner (Milman, 1981);

(2) the variation in student grades in different subjects may be due to different innate student abilities, independent of any professional teaching skill (Ross, 1984);

(3) the narrow range of knowledge contained in a standardized test (Darling-Hammond, 1984);

(4) the serious problems that prevent one from obtaining a statistically valid result (Medley, 1973); and
(5) the teacher teaching to the test, in effect, nullifying students' grades as reflecting their true capabilities and knowledge base (Page, 1972).

The Merit Pay Controversy

While there are arguments against merit pay (Hanushek, 1981; Johnson, 1984; Newton & Braithwaite, 1988; Peterson, 1986; Rosenholtz & Smylie, 1985), possibly it can, nevertheless, be an important component, in an evaluation system, as an incentive for excellence. In support of this possibility, about 63 per cent of teachers polled in one nationwide American survey agreed that at least some part of their compensation should be contingent on performance (Rist, 1983). Some have claimed that it is possible to find the relatively unbiased criteria necessary to establish a merit pay system (Haertel, 1986; Shulman, 1987; White, 1983). Overall then, career ladders, as used in Tennessee, could be tried in some districts (Ficklin & McCormick, 1983). Functioning as an adjunct to reward exemplary evaluations, a system of merit pay might result in a more effective teaching force because officials would be identifying, rewarding and providing an impetus for excellence, whereas now they equate, on some level, all these, that is, incompetence, mediocrity and excellence, by distinguishing neither from the other in the form of graduated compensation. Although some form of career ladder and bonus incentives are strong elements in most systems of modern management (see Appendix C), systematically
and efficiently-collected data on classrooms under experimentally-valid conditions would be needed to prove either side of the argument.

**General Problems Inherent with Evaluation Systems**

Admittedly, therefore, the formation of a consistent evaluation system is difficult regardless of the particular management philosophy. Besides the problems specific to contemporary teacher evaluation just discussed, by definition, indigenous flaws exist in any appraisal process, as can be seen in the following summary:

I. While summative evaluation seeks to reward, diminish or perhaps maintain the employee's status, data are required for its administration. Those data mostly emanate from the employee. It is not always to the employee’s benefit to reveal accurate data (Meyer, Kay & French, 1965; Mohrman, 1984; Porter, Lawler & Hackman, 1975). Consequently, then, collegiality forms an essential element of any evaluation system (Sergiovanni & Staratt, 1983). Systems that use summative evaluation are then intrinsically weakened by the inevitable dichotomy that this conflict inculcates upon them. Processes of evaluation, like many teacher evaluation systems, that use one scenario for both the formative and summative aspects of evaluation just exacerbate this problem (Popham, 1987; Scriven, 1987).
II. Evaluation knowledge and opinion of prior performance, which may be biased, erroneous or irrelevant to the present situation, tend to bias formal evaluations (Huber, Neale & Northcraft, 1987), as in the halo effect. (Nisbett & Wilson, 1977; Rudner, 1992). This is especially salient in light of revelations forthcoming in this project’s data as revealed by principals who may rely for a major part of their evaluation on “informal visitations” to supplement the inadequate number of formal visitations. This would then enhance the influence of the halo effect and skew the veracity of the results even further away from any possibility of objectivity, validity or reliability (Cooper, 1981).

III. Factors such as age of rater make a significant difference (Ferris, et al., 1985), although not gender or race (Pulakos, White & Oppler, 1989). Regardless, though, there are other extraneous factors which might tend to bias evaluation results.

IV. Although individual evaluation is necessary for many systems of evaluation, perhaps even TQM, some scholars, for example, W.E. Deming himself, are against individual, as opposed to departmental, evaluation because of the potential for interdepartmental conflict (Deming, 1986). Regardless, then, this is a serious drawback in all evaluation system used on organizations requiring teamwork. Therefore, it is difficult to reward excellence on the
basis of evaluation for fear of inciting rivalries and jealousies among workers. This problem of rivalry and resentment has been a formidable problem when trying to evaluate the importance of merit pay within an evaluation system.

V. Summative evaluations and their concomitant promotions, terminations or pay increases can be and have been successfully challenged by the courts. This reinforces the need for significant, relevant and unbiased objective criteria in order to be able to demonstrate fairness and relevance conclusively in a legal challenge to an organization's decisions. This demonstration of evidence is becoming increasingly difficult, increasingly contentious and increasingly necessary. As discussed later, writing an evaluation about somebody, even if under the aegis of formal authorization, could result in a violation of the slander laws if "reckless disregard for the truth" were to be established. To avoid any possibility of that, one must ascertain that the evaluations are as fair as possible and in accordance with those legal limitations just mentioned. Thus, even ignoring weaknesses in evaluation systems' methodologies, inherent problems external to the evaluation system itself exist that might tend to dilute the effectiveness of any assessment process, although some of these problems probably can be compensated for.
VI. The method of feedback of the evaluation data to employees is hotly debated, controversial and even polemical. Apparently, language itself is a significant factor when presenting feedback. In fact, reporting an employee as *above average* or *outstanding* has led to improved performance, but reporting an employee as *satisfactory, or not significantly different from average*, has led to resentment, negative attitudes and decreased company dedication (Pearce & Porter, 1986). Who presents the feedback is apparently also pertinent. Feedback from a supervisor properly presented increases performance. However, if the feedback is from one’s self or fellow employees, it seems not to lead to increased performance (Becker & Klimoski, 1986). This dimension could prove valuable for determining evaluation methodology in Ontario schools. Because of that variable, some of the recent innovations involving portfolios or peer evaluation may backfire entirely as they fail to produce the anticipated improvements in teacher effectiveness.
A Conceptual Illustration of the General Principles of Effective Evaluation

Nevertheless, there does seem to be somewhat of a consensus as to what certain core principles of any good evaluation system should be from a management perspective. The literature suggests that those basic parameters should be composed of the following:

1. A clear declaration of what the evaluation criteria and their processes are;
2. The adoption of a consultative approach with the evaluated staff members so they understand why certain criteria were chosen; why they are important to the organization; and why they are important to their own professional development;
3. The formative nature of the evaluation is clarified to both employees (teachers in this case) and supervisors, but it must remain distinctly independent from its complimentary summative component;
4. A balance is maintained between local or site-based autonomy and central hierarchal control;
5. A culture of evaluation that is compatible with the organizational one;
6. Assured, clearly recognized and acknowledged competence of evaluators; and
7. Consistency between evaluation practice and policy within the parameters of due process (Cohen & March, 1974; Daresh & Playko, 1995; Deal & Celotti, 1980; Musella, 1988; Richards, 1984; Weick, 1982;).
Choosing these principles is not just a matter of adopting a consensus of the literature. These principles, besides being independent of any specific process, are compatible with and reflect the basic elements common to modern management philosophies. Such qualities as constant improvement, consensual authority and mutual credibility are ideas inherent in these principles. Beyond that, these principles themselves do not directly represent any particular evaluation system either in education or any other field. In following these principles, however, one can select from among the systems of evaluation, ultimately selecting the effective from the ineffective systems. It is believed that this approach can be used successfully to adjudicate the evaluation methodologies of the school systems being studied in this project. It is these criteria that should be used as the conceptual framework to adjudicate between extant evaluation systems in the school systems studied.

As a result of the aforementioned criticisms and criteria, new approaches to evaluation, not as strongly based on the deficit model or improved by certain compensatory features, are currently being introduced. In recent years, some effort has been undertaken to improve the observation rating process mode of evaluation. Some of the most innovative are the Criteria for Effective Teaching (CET) protocol and the previously mentioned System for Teaching and Learning Assessment Review (STAR). According to Ellet (1994), STAR represents the cutting-edge of classroom-based observation-rating assessments, which focus on finding, eliciting and evaluating teaching in the upper level of skills, not
just the basic minimum competencies. It was also found effective for discriminating between the journeyman and expert teachers. Unlike most of the former protocols of evaluation based on these models, assessors must complete a comprehensive seven-day training program. Any assessments are made independently by each member of an assessment team composed of the school principal, an outside assessor and a peer teacher. The total STAR classroom-based assessment observation component consists of six visitations -- three in the fall and three in the spring -- that is, two separate independent assessments, all conducted by each assessor. The initial results on the use of this method suggests that trained and experienced educators certified in assessment systems like STAR are qualitatively different than the more commonly accepted principal observation methods.

Other attempts at rectifying a flawed system include teacher portfolios, student ratings and peer assessment. Although none of these systems have overtaken the popularity of the observation-scoring format (usually overseen by principals) and estimated to define 88 per cent of the evaluations in the United States (Kowalski, 1978; Lower 1987), some of these methods are appearing with increasing frequency in North America, in general, as well as specifically throughout the educational systems in Ontario.

Because students are in a unique position, in the sense that they are consistently observing teacher behavior, they seem to be in the ideal situation to perform evaluation. Initially, then, this form of evaluation would seem appropriate. Yet despite this positive
outlook, there are problems. Some of the questions raised about the validity of this type of adjudication focus on the “natural” student bias. Are teachers, perhaps, not just giving students with good marks a chance to reciprocate the “favour” to their teachers regardless of the quality of teaching? Conversely, students who do poorly might be motivated to seek revenge by giving their teacher a poor assessment. If students below a certain grade-received level are excluded, then isn’t the assessment not being skewed away from any serious criticism toward a positive evaluation? Even despite these potential flaws, the balance of scholarly opinion remains favourable towards this type of assessment and adds an important component to the spectrum of available teacher assessment tools (Centra, 1972).

Teacher portfolios, an evidence file of teachers’ professional practices in carrying out their responsibilities, are also becoming more popular. The professional portfolio could include taped lessons, lesson plans, developed instructional material or evidence of student achievement (Bird, 1990; Wolf, 1991 & 1996). Or as Lee Shulman more generally defines portfolios, they are

> the structured documentary history of a carefully selected set of coached or mentored accomplishments substantiated by samples of student work and fully realized through reflective writing, deliberation and serious conversation (Shulman, 1992; as cited in Wolf, 1994 p.115)

Regardless, it is a paradigm used for formative and summative evaluation of teachers.

There are some clear advantages to the portfolio paradigm. It can be used in a single format for all teachers or more flexibly, adjusting to the needs and circumstances of
individual teachers, depending on its use as a summative or formative tool. For purposes of evaluating teachers other than traditional classroom teachers, portfolios can also be adjusted accordingly in order to make their substantive content relevant to the assessment situation requirement to rate the performances of different kinds of teachers.

The entire practice of using portfolios in evaluation (especially formative professional development) began in 1983 as a demonstration of the proper criteria expected for beginning teaching (Terry & Eade, 1983). In one reckoning (Dietz, 1995), there are three types of portfolios: the Presentation Portfolio -- a collection, résumé, or album representing an individual's accomplishments, learnings strengths and expertise; the Working Portfolio -- an illustration of the compliance to and accomplishment of any pre-determined professional standards or competencies; and finally the Learning Portfolio -- a collection of knowledge, experiences and feelings that provide a perspective from which teachers can focus their progress, vicissitudes and flow of ideas, as through a progression on the way to their final goal. Each teaching portfolio should theoretically cover four facets of the evaluation process:

(1) Purpose - the reasons for evaluation of teaching;

(2) Focus - the aspects of teaching being covered;

(3) Process - the decisions involved in organizing and illustrating the chosen materials in order to demonstrate the evaluation's focus; and
(4) Outcomes - the choices of a concrete conclusion, description and organization that successfully portray teaching experiences.

In these cases, the teacher chooses another teacher, a portfolio partner, to discuss and criticize their plans, usually meeting for about 30 minutes every other week (Dietz, 1995).

But it would be a mistake to equate a teacher portfolio with a "state-of-the-art" scrapbook. A portfolio is a collection of information that illustrates the ordered, gradual progression of professional development of the teacher’s career. Merely collecting random “snapshots” of that career defeats its purpose and essentially removes the collected material from the category of a bona fide portfolio. Instead, a professionally constructed portfolio should begin with certain background information such as a resumé and the particular teacher’s educational philosophy. The core of the portfolio should always include some or all of the following: an overview of the unit goals; standards and overall instructional plans; lists of the resources connected to those plans; two consecutive lesson plans; a videotape of a teaching exercise (probably based on those lesson plans); pre- and post-standardized tests where applicable; student work samples and evaluation of those samples, especially illustrating the value of teacher feedback; a summary of the teacher integrating the different parts of the portfolio; the reason for their choices and the strengths and weaknesses of those goals; lesson plans; and evaluation methods. In addition to student and peer evaluations, the portfolio should also include informal evaluative statements and comments, letters from employers about the teacher’s abilities and letters from teachers
of the next grade or level, that is, high school or university teachers. Finally, the résumé should include a summary of ancillary professional activities, professional conferences, committee work, positions of enhanced responsibility and commitment to minorities, relevant academic courses and professional development seminars, in addition to new methods tried as a result of those courses or conferences and letters of recommendation (Wolf, 1996).

Nevertheless, criticism of teacher portfolios suggests that the very flexibility which is a necessary element of teacher portfolios is also one of its weaknesses. Although this attribute empowers teachers by providing them with the authority to exercise professional discretion, it also pushes portfolios into the realm of subjectivity where pre-determined criteria necessary for precision may ultimately be lacking. A corollary of this characteristic is the duration of time commitment. Since its parameters are so vague, the time expended can extend well beyond any net gain. In fact, teachers have claimed that portfolio preparation took such depth of commitment that during the period of organization the quality of their teaching decreased. Another possible weakness is in portfolio assessment. As the situation now stands, the principal usually evaluates these portfolios, but if that principal is not familiar with either the logistics, the demographics or the subject matter and its validity, the judgement rendered by that principal can be seriously questioned (Fisher, 1994).
The next area of evaluation being developed is peer evaluation and its sub-categories of clinical supervision and self-evaluation. The first mention of peer-supervision as a viable method of evaluation appear in Manaat and Stowe's (1984) article on the use of master teachers in evaluation. This article was subsequently followed and enhanced by Hunter (1988). In this paradigm, which is usually employed just formatively, a master teacher observes and subsequently critiques the classroom practices of less expert teachers. However, the concept of "master teacher" is itself elusive to define operationally, as a discerning epithet.

In addition, although most currently employed methods of peer evaluation use either the paper and pencil format or some form of administrative observation, neither method has proven sufficiently satisfactory to merit the status of a definitive approach. As a result, improved methods have been suggested. Sparks and Lipka (1991) have proposed a nomination process with appropriate active roles for students, teachers and administrators with an 11-characteristic scale ranking, developed by Azumi and Lerman (1987). Interestingly enough, a study by Sparks and Lipka (1991) has shown that a choice demarcation between teachers and administrators in one category and students in another exists, based on the selection of master teachers. In that study, students chose "craftsmen" criteria (that is, categories such as the following: uses strategies to meet the needs of all students, monitors student progress, gives feedback and provides reinforcement) while principals (administrators) and teachers chose "commander" criteria (that is, classifications
such as the following: subject competence, organization and discipline). Nevertheless, focusing on the personality characteristics of master teachers, Sparks and Lipka (1992) revealed no significant differences between either of the three group choices, for master teacher, teachers, students or administrators. All the same, in that later study (1992), master teachers were described consistently as warm-hearted, socially-outgoing, attentive and generous in personal relations and maintaining interpersonal contacts. Furthermore, the master teacher is said to be hard to fool, has a high drive level and respects traditional ideas while simultaneously possessing sensitivity and warmth. Evidently the teacher herein described provides a warm stable atmosphere within a structured environment of coherence and authority. This result based on characteristics and their previous findings based on teaching technique provide a detailed profile of the master teacher.

A further innovation in the available methods of evaluation is self evaluation, a specific form of peer review is based on the concepts that (1) the teachers know themselves in the professional capacity status of a teacher better than any "external" supervisor or colleague and (2) as professionals are capable of sufficient emotional and intellectual detachment to give a fair impartial assessment. As Haertel and Wheeler define it (1993, p. 131), self-evaluation is the "process of judging one's own performance for the purpose of self-improvement." According to Barber (1990), the process of self-evaluation should invoke seven steps: (1) understanding the underlying philosophy of self-evaluation, (2) using audio and video tape, (3) identifying essential teaching behaviours, (4) identifying
verbal cues, (5) identifying non-verbal cues, (6) learning how to evaluate the instructional behaviours under study and (7) using evaluation forms. One example of evaluation forms is the Teaching Goals Inventory in which the teachers rate each teaching goal’s personal value to them as a professional, in order to assess its priority within their teaching value systems (Angelo & Cross, 1993). Another example is the Teacher Evaluation Rating Scales (TeachERS). This form includes separate supervisor and teacher forms, whose scores are subtracted to highlight any discrepancy. This score theoretically should help the teacher in a self-improvement pathway (Wheeler, 1993).

In summary, then, each of these systems mentioned has its own unique set of weaknesses. CET and STAR are based on an observational format that may focus the teacher’s behaviour toward a certain standard of behaviour, one not necessarily practiced as customary behaviour. And despite the improvements it offers over other observation systems, there is the well-established policy declaring that numerous studies of teaching methods have shown us that there is no one standard methodology that either works with every teacher or always produces the best result. Furthermore there are no definitive teacher competencies that are observable (Darling-Hammond, Wise & Pease, 1983) or even a consensus on what categorically constitutes effective teaching (Dunkin & Biddle, 1974; Good & Brophy, 1987; Johnson, 1980; Shulman, 1987; Turner & Clift, 1988).

Those weaknesses, in addition to the restricted amount of observation periods, lack of training by observers, prior knowledge of criteria, and so forth, might call into question
the accuracy of data obtained by observation. The problems with student assessment of teachers have already been discussed. For example, possibilities exist for abuse of the system due to disgruntled students, grateful students or apathetic students. In portfolios, the teacher often suggests areas for improvement that are not strictly monitored for compliance -- there is no guarantee that a good portfolio means increased student learning. So too does peer evaluation have its own indigenous problems such as master teacher selection, the master teachers' methods of adjudication and how applications of the results can be employed without creating animosity. And, of course, the use of student scores to evaluate teachers is, as mentioned previously, controversial. Thus, as has been suggested several times, some combination of the discussed methods should be considered.

**Legal Aspects of Evaluation**

Consequently, it seems a consensus forms that evaluation involves the transmission of both positive and negative information. It is not as obvious though, except in the case of self-evaluation, when these evaluations cross the line between critical and defamatory communication. Furthermore, because the evaluator is usually a manager, supervisor or other person in administrative authority, the mitigating factor of qualified privilege often becomes relevant. As this is the case, it becomes very unclear which evaluations can be challenged in a Canadian court of law for constituting either libel or slander and which fall within the legitimate domain of the manager. It is the purview of this section to outline the
basic laws of defamation and how they affect evaluation in terms of slander and libel so that the reader can see how the law itself can limit or assist evaluation procedures.

Slander and libel (oral and written defamation respectively) have been clearly defined by Clement Gately (1981) as any written or communicated word or act "which may tend to lower the plaintiff in the estimation of right thinking members of society generally or to cut him off from society or to expose him to hatred, contempt or ridicule" (p.6); it is this definition I will use in the following exegesis.

**Qualified Privilege: A Definition**

The main problem confronting administrators, however, is the challenge that threatens their provisional immunity of qualified privilege. This privilege is said to exist when a legal, social or moral duty, or a common interest is present between the receiver and the perpetrator of the communication (*McGuigan v. Davidson* 1985). In such cases of qualified privilege, malice, as always, has to be proven. An important difference, though, is that in these cases of qualified privilege the defendant, usually a manager or supervisor, is given the benefit of an assumption against malice, a necessary and sufficient cause to establish defamation. Thus it is incumbent upon the plaintiff to provide evidence of malice.

At this point it seems that precedent allows a plaintiff to establish malice only if the defendant can be shown not to have had an honest belief in what was said or written or if
the privileged occasion was used for an improper purpose (*Korach v. Moore* 10R, [Jan. 2, 1991] at 281; Rogers, 1979, p. 239).

**Criteria of Malice**

To establish malice, one does not necessarily have to prove that the defendant knowingly made derogatory remarks that they knew were untrue. All the plaintiff has to show is that the defendant was reckless. Recklessness, under Canadian law, means indifference to the truth or falsity of what was said (*Horrock v. Lowe* [1975] A.C. 135, [1974] 1 All E.R. 622 [1974] 1 W.L.R. 282 H.L. at 153 A.C.). What complicates the task of the plaintiff, though, is that recklessness is not to be equated with carelessness, lack of judgement or irrationality (in thought).

The matter becomes even more vague when trying to establish improper purpose as a criterion for malice and therefore defamation. Currently, while the Ontario Court of Appeal quoted theoretical criteria for malice that included “vindictiveness, dislike, revenge or improper motive” (*Korach v. Moore* [Jan. 2, 1991] 1 O.R. [3d] at 282), the courts have not been as forthcoming with a reified operational definition of equal clarity. However, the courts do seem to strongly indicate that the establishment of proof of the defendant’s statement going “beyond the subject matter giving rise to interest” (*McGuigan v. Davidson* [1985] 58 V.B.R. [2d] at 123) is sufficient to cancel the status of qualified privilege.
The Legal Component of Evaluation

It appears, then, that evaluation by administrators must be seen in a legal and situational context. It is no longer sufficient for successful, relevant evaluation to operate solely by the rules of social science. To be safe, efficient and effective, evaluation must always be conscious of legal ramifications. Thus, because of the implications inherent in these rights, precautions should be taken to avoid incurring the prohibitive cost of a civil defamation suit. This knowledge must become part of all educational executives’ administrative portfolio, if they are to function effectively and enhance the goals of their respective organizations.

Influences of the College of Teachers: An Introduction

Perhaps complicating the entire evaluation situation even further is the recent formation of the Ontario College of Teachers (OCT). In September 1996, building upon ideas from a recent report of a Royal Commission and on previous advocacies developed during the preceding fifteen years, Bill 31 was passed in the Ontario Provincial Legislature guaranteeing the establishment of a College of Teachers with an initial budget of $17,000,000.00. Among other mandates, this College of Teachers had the authority to regulate the profession of teaching and to govern its members, to accredit education programs for teachers, to oversee certification, to discipline and to establish and enforce
professional standards. These powers might then eventually extend to include the entire jurisdiction of teacher evaluation (The Ontario College of Teachers Act, 1996, c. 12 as amended).

Acknowledging this enactment before any discussion on the teachers’ views on the Ontario College of Teachers proceeds, it is important to offer a preamble. Despite the apparent diversity of opinion on the College, in my data, there seems to be an underlying anti-College of Teachers viewpoint, which might be due to the proselytizing against the College by the unions and even by some of the principals (though not noticeably in the group of principals represented in this sample). To some extent, these educators may have vested interests against this new OCT. That is, teacher unions and principal groups may see themselves as losers in any ultimate power struggle against a government institution, which might either directly or indirectly usurp some of their power (from the unions) or duties (from principals), especially in the matter of evaluation. In addition to these possibilities, the teachers’ unions are on record as protesting their members’ lack of majority representation on the College’s governing board. This attitude persists despite the majority of members (17 out of 31) being from the education field (Lewington, Globe and Mail, June 2, 1997). Furthermore, the College of Teachers was just evolving contemporaneously with this study’s interviews and, in this incipient state, some principals may have been circulating contradictory information about the College’s provenance and scope.
In May 2000, though, the OCT had begun taking a more active role in advising the government about the controversial issues such as teacher testing and professional standards. Practical experiences, such as the College’s indignation at not being informed prior to the government’s apparent decisions concerning mandatory testing, may seem to belie this relationship claim of consort and advisor. The OCT, nevertheless, has suggested an amendment to Regulation 184 of the Education Act that would demand teachers in Ontario take a language proficiency test. This suggestion is likely to be adopted.

Summarizing all the above in their own words, as of June 2000, the College seems to concede the following duties as their major responsibilities:

The College of Teachers is a self-regulating body created by the Ontario Legislature to regulate the teaching profession in the public interest. It sets ethical and professional standards for teachers, establishes teaching qualifications and accredits education program providers and programs. The College also receives and investigates complaints of teacher misconduct or incompetence made by school boards and the public... [Thus] the responsibility for the removal of a teaching certificate has moved from the Ministry of Education and Training to the profession—[that is, the Ontario College of Teachers].

(Retrieved June 18, 2000 from the World Wide Web: http://www.oct.on.ca)

In summation then, for the most part, teachers in this study’s sample seemed ambivalent and ill-informed even within the parameters just mentioned, almost as if the College of Teachers was a policy decision beyond their range of expertise and control. To some extent, of course, teachers were right: after many years of hovering between differing viewpoints of policy deliberations, the idea of the College was just beginning to gain political legitimacy outside of professional teaching circles. In general, however,
teachers had not widely argued that the College was an assured way to put Ontario in the front rank of reforms that members of the public seemed to want for their schools. However, the educators’ anti-College attitude undoubtedly influenced their answers; this should be taken into consideration in the adjudication of those responses as accurate information about the OCT. Educators’ attitudes, although not important from the perspective of their power to approve or disapprove, are substantial because of the necessity for their co-operation if this professional accrediting body is to succeed in having an influence on the profession. Essentially, the more support the College elicits from the rank and file, the less influence the teachers’ union leadership will have in opposing the College’s formation.

Finally, it would perhaps be advisable to insert a further brief caveat, early on, about my own data. Generally, even the principals were not fully briefed on the exact nature, purpose and structure of this emerging institution, at least partly because of the inchoate stage of the College’s development at the time of my interviews. It was the perception of this researcher, based on general attitude, that the teachers, as compared to the principals, were even more poorly informed. The recency and uncertainty of the College should consequently be taken into account when assessing the veracity, credibility or sophistication of the views that are discussed later in the thesis. Ultimately only time will be able to inform us as to the exact nature of the OCT and the role the OCT will play in teacher evaluation and the broader field of educational administration.
Summary and Results

It seems, therefore, that because of the culture that has become dominant in schools, the assorted problems associated with evaluation are often intensified. For example, student evaluation of teachers, which has the potential to be accurate and helpful, is uniformly and inexorably opposed by most teachers' unions (Marsh, 1977; Wilkinson, 1989) with what might be characterized as almost irrational fervor. Similarly, while other researchers encourage the inclusion of the results of objective student performance in a teacher's evaluative process (Ross, 1984), teacher organizations are, again, strongly opposed to involving student results in any stage of the evaluative process (Soar & Soar, 1979).

In essence then, it is difficult to form any solid conclusions about appraisal systems' construction except for the controversial nature inherent in their application. Neither is the literature definitive in this respect, failing as it does to adopt any clear choice of one methodology of appraisal over another. Some of the problems become even more egregious when focusing on teacher evaluation. Since many systems evaluate teachers based on the style of their teaching, style becomes basic to this discussion. Yet the findings are that outside of some very basic behaviors, no firm consensus exists in the literature on what constitutes good teaching practice as opposed to bad teaching practice. This point has already been well made (Levin, 1979; Scriven, 1994) in a trenchant criticism of the present process-oriented evaluation systems. Levin, as other researchers already mentioned,
believes that few effective methods of evaluation presently exist in education. A major reason for this may be the focus on process evaluation.

In summary, then, this segment has been concerned with the relationships between the themes of my study on evaluation and relevant issues as described in the literature on educational administration. It seeks a balance and understanding between these various topics and a sense of coherence in the opinions and conclusions offered therein. As such, this segment has discussed those administrative topics and their relationship to this project. As delineated in the following section, it is these topics incorporated within that context that define many aspects of the project’s major themes.
SECTION III

CHAPTER 4
THE INTERVIEWS: DATA

Overview

This study deals, from a qualitative perspective, with the teachers’ and principals’ perceptions of certain teacher evaluation issues. These issues include content and purpose of evaluation, adaptability (of format), perception of methodology, consequentiality, organization’s cultural compatibility, formative aspects, relationships (between evaluator and evaluatee) and client satisfaction. Based on the literature and the hypotheses-provocations, these issues were seen as the most likely to synchronize with my specific methodology and the most relevant to either support or refute any claim of a well-managed evaluation system.

The Best Aspects of Teacher Evaluation

Teachers’ Views

The teachers, interviewed to ascertain their views on the best aspects of teacher evaluation, collectively saw the following as the most positive attributes of the evaluation processes: getting practical advice, having their professional growth stimulated, establishing standards within the profession, receiving feedback, gaining reinforcement and increasing teacher efficiency.
Certain teachers suggested that the personal advice and comments of the administrators in charge of their evaluations and the ensuing post-interview discussions were important because they provided the greatest potential for improvement. But, this opinion, although expressed by four teachers, was not unanimous. One of the teachers valued the specificity of the comments, both positive and negative (accolades and criticisms), in both the pre-interview and the post-interview sessions with the administrator: "For example, principals who would comment on the way I would concentrate materials to help with my math program . . . ." Another teacher had strong opinions that emphasized her administrator's encouragement for the initiative she displayed in independently setting her own goals: "The whole process of setting goals for yourself, and [then] having the evaluation done was like a pre-conference followed by a classroom observation [format] and then [you had] your post-conference [which involved] sharing the goals . . . ." was a main factor in making evaluation a personal positive experience. This teacher explained that for her there was a form of autonomy implicit in this particular process of evaluation which reinforced her sense of professionalism, a sense, she said, that teachers often seemed to struggle over.

Others may have seen the core process of evaluation as serving to restate obvious facts about general teaching methodology or style and failing to concentrate on their own unique teaching methodology to any significant extent; these teachers nevertheless regarded certain aspects of the evaluation experience as having some positive attributes. They still
thought it was “nice” to get an official positive evaluation, even though the recognition was regarded somewhat dismissively as ephemeral, (largely) pro-forma, inconsequential and, invariably, positive. As one of the teachers elaborated, “There wasn’t follow-up, there wasn’t professional development offered, so it was simply a pat on the back.”

Another teacher stated that, ”you want to know that you are on the right track, as a first-year teacher anyway.” This teacher explained that in her first year she was told that her classroom was “not welcoming enough.” As a result, the teacher explained, “by the time the next evaluation came around I had more welcoming displays on my bulletin board, the students’ work was up, [other] work that had been produced was up, as well as posters from different stores, teacher stores and public libraries.” All these, this teacher implied, enhanced the culture and general teaching environment of the classroom.

**Principals’ Views**

Overall, the principals claimed that the professional growth and the technical improvement in teaching, as a consequence of the evaluation process, were the most positive aspects of staff evaluation. One administrator stressed the collaborative nature of the evaluation systems at her school; evidently, the staff sits down together to decide what factors determine a good teacher and how these factors can be incorporated into an evaluation system. Another principal, in addition to the growth aspect of teaching, emphasized the summative dimension when backed by administrative follow-through. This
summative component was thought to be sufficiently powerful for removing teachers who were incompetent.

The Worst Aspects of Teacher Evaluation

Teachers’ Views

For two teachers, one of the worst aspects of teacher evaluation was their administrator’s “detached,” even indifferent, attitude. The teachers claimed that they often sensed an apathetic demeanor which, in some cases, extended to teachers writing their own evaluations; this disinterest was most obvious when principals showed up late (almost an hour late sometimes) and would not conduct the post-evaluation interviews or classroom demonstrations. This reinforced those teachers’ initial suspicions that the administrators’ criticisms were trivial or perfunctory. Four teachers felt frustrated at having to “stage” a performance in order to fulfil an alleged arbitrary, merely bureaucratic requirement. Furthermore, there was considerable tension whenever evaluators came into their classroom. This resentment existed alongside equally negative feelings towards non-consequential appraisals to evaluate the teachers based on situations that were far from reflective of their usual classroom. Four of the teachers in this study held that the entire process resembled a hypocritical ritual that neither they, the principals, nor the board administrators felt mattered at all or really wanted to undertake as a professional responsibility.
Principals’ Views

Half the principals perceived that the worst aspects of evaluation was the antagonism often found when confronting teachers with negative evaluation reports. This antagonism, contrary to the alleged professionalism expected from teachers during an evaluation process, was experienced by three of the principals in my study. These incidents involved “teacher attitudes of anger and resentment,” which often occurred “when somebody doesn’t agree with your evaluation … particularly if they had been teaching for ten years [or more], and had not [yet] philosophically bought into the idea that [all] instruction needs to be continually improved as they go along.” Other major negative aspects for the principals were lack of time, lack of documentation by former principals on incompetent teachers and the intense nervousness among teachers who were undergoing evaluations.

Influence of the Collective Agreement:

Teachers’ Views

As is generally known, the vast majority of permanent teachers in Ontario sign contracts of employment. These contracts are based on a negotiated collective agreement that outlines the rights and obligations of the teachers towards the board of education. The latitude determining the restrictive nature of those obligations and those rights ultimately determined by the settlement between the teachers’ unions and their employers (that is, the boards).
Two teachers categorically stated that the collective agreement does not inhibit an efficient evaluation process. A third teacher had the impression that the administrator, at least some of the time, was reluctant to write down any strongly-worded or highly-critical evaluations because of teacher-oriented union by-laws or because of formidable job protection clauses in the collective agreement achieved by the teachers’ union. As that teacher elaborated, “[If] there has been something that they’ve wanted me to work on . . . they’ve [principal or vice-principal] always told me face-to-face.” A fourth interviewee felt strongly that the collective agreement constrains any strict enforcement of evaluation judgements because “I would think that administrators [who themselves until recently were members of the same union as the teachers] are very aware of the collective agreement and the last thing they want is to get entangled in a [collective agreement] dispute.”

**Principals’ Views**

Only two principals thought that the collective agreement contributed to the evaluation process. One of this pair perceived the collective agreement as “beneficial to the improvement of teacher quality.” The other principal did not see it “as a bad thing because of the need for teachers to be protected.” Throughout the study, though, the bulk of the principals’ opinions on the collective agreement was negative. In this vein, one principal said that she would like to initiate or scrutinize more but couldn’t because of the contract’s legal constraints. Another principal suggested that the collective agreement
“[unduly] lengthens the process of evaluation” especially in cases where a dismissal is mandatory. When asked if this period should be reduced for the sake of process efficiency, the principal responded without hesitation, “Oh, for sure.” She then went on to explain:

It [evaluation] lengthens the process more than it needs to. It becomes very long and arduous .... I would say that it's more the level of documentation that has to happen, the involvement of the federation [that is, teachers' unions]. Sometimes lawyers have to be called in. And that would happen over a long period of time, perhaps two to three years.

Another principal conceded that the evaluation process could facilitate the dismissal of an incompetent teacher when “people are being more diligent” all the way along the chain of command. One principal, commenting about the effect of the collective agreement, said the following: “Basically a person would have to kill someone in the classroom before (being dismissed)... and even then I'm not too sure they would get their teaching certificate removed.” One of the other principals also stated his impression, that as a result of the collective agreement, “people in our (particular) board did feel that the evaluation process was not conducive to getting rid of (an) incompetent teacher.”

In summation then, four of the principals suggested that the collective agreement had a negative influence on teachers’ professional standards and prevented the evaluation process from having any power as a summative tool for the purposes of dismissal or demotion.
Influences of College of Teachers

Teachers' Views

It can be said that all teachers felt, either due to formal or informal meetings they had attended or through “the grapevine,” that the College of Teachers would take over all or a meaningful portion of the evaluation process. There is, however, a range of opinion concerning the value of this change. Three of the teachers suggested that it would be an improvement over the present system because the evaluation would become either more standardized or more professional. All three agreed that there would be more pressure to maintain established standards and perhaps even to improve professional qualifications. These teachers said that at present the system did not seem to be very stringent in these areas.

One teacher, while forecasting that the College of Teachers would eventually manage the profession’s evaluation programs, was neutral about any effect that change might have. Conceding that she was drawing upon the views of her peers, another teacher anticipated no significant change from the current system that she viewed as basically controlled by the principals and teachers’ unions. One teacher expressed views unequivocally opposing eventual transfer of evaluation to the College of Teachers. That teacher believed that evaluation would become too impersonal. Elements such as the context of evaluation, the background knowledge, unique to each teacher, and the communication style of that teacher, usually acquired through personal knowledge of that
teacher in action, would all be unavailable and therefore disregarded in a College-run system. Finally, one teacher, while acknowledging the College of Teachers would eventually supervise the profession’s evaluation programs, remained neutral about any effect that change might have.

Principals’ Views

Five out of six of the principals agreed that the formation and influence of the College of Teachers was positive. Four of those principals were also enthusiastic enough to express confidence in the College of Teachers’ ability to increase professionalism among teachers. The other principal who was positively disposed toward the College’s development declared that the College “would be positive because it calls for disciplining within your own profession.” Comments from the majority group of those who approved on grounds of professionalism included the following:

(1) “It will allow them [teachers] to reach the level of being designated as a professional organization….They [members of the College] are there to protect the profession.”

(2) “They will establish professional standards for evaluation.”

(3) “Teachers would benefit from rising to a level of professionalism where there is a demand for courses and such.”

Besides the belief that the College of Teachers will help teaching evolve into a bona fide profession, five principals were of the opinion that the evaluation of teachers itself would become more professional. The gamut of views on how this would occur ranged from speculation that the College would be providing courses for principals to a hypothesis
that suggested the College would actually be adopting the entire job of evaluating, thereby providing external professional evaluators — in essence relieving the principals of this duty. One principal strongly favored this realignment. Another principal, approvingly and firmly, believed that the College of Teachers would send in an evaluator who does the evaluation of the teachers in isolation, without the advice or consent of the principals or unions. Only one principal was categorically opposed to the College of Teachers. This principal claimed that, generally, there was not a lot of support from principals for the College idea: “Most of us feel that the evaluation process, the accountability that is already built in, is one that we work very hard at, and it seems to negate that by having this body overlooking us.”

**Objective Data: Using Students’ Standardized Exams as Evaluation Criteria for Teachers**

**Teachers’ Views**

When teachers in this study were asked about using the grades of students’ standardized content exams for evaluating teachers, much less diversity was apparent in their responses. Except for one teacher, all the rest were strongly against that type of teacher evaluation. One teacher insisted that these types of student test scores were an unfair evaluation of teacher skills:

> Perhaps they [teachers] spend a lot of time getting half the content done, and the class knows that half the content and they ran out of time, so those kinds are going to do badly on the provincial
exam, [especially if that exam is held before the school year's end]. But they [students] know one half very well instead of knowing the whole thing a little bit.

Another teacher asserted that test scores should not be used because of "the human aspect of teaching"; for her, testing would not capture all the developmental, social and psychological gains that a teacher cultivated. A different teacher stated that using standardized student scores to evaluate teachers would be wrong because of the following reasons: (1) teachers would expressly teach to the test; (2) tests would be based on prior knowledge rather than on knowledge of any one grade; (3) tests would never really be content based; (4) too many variables exist in teaching; and (5) tests would inherently be dishonestly graded. This same teacher further opposed the idea of using test scores because too many external socio-economic factors, beyond the control of the teacher, are involved that affect the grade — such as child nutrition, community violence, family dynamics and structure, community expectations, peer pressure, stability of the home, personal expectations and so forth. Still another teacher had serious reservations about the whole concept of standardized testing from the student perspective, although she thought that this approach, evaluation based on student-test scores, would not in itself cause serious professional problems for teachers. Finally, while one teacher was neutral as to the efficacy or equity of using student scores to evaluate teachers, she held, regardless of the official use that was assigned to the test, that whenever these type of tests were
administered to students, inevitably, such tests were exploited in some way for teacher evaluation.

**Principals' Views**

The principals' views on using standardized exams for the purpose of teacher evaluation were all negative. The one principal who offered some support, felt that standardized exams might be used for a "tiny" part of the evaluation, with the current evaluation process still making up the bulk of the system. Five principals, however, were opposed to any such endeavour. One of the reasons was that "parents would be comparing schools and that would be dangerous." This same principal felt that the use of student scores, for purposes of teacher evaluation, would be "ridiculous" and supported her allegation by claiming that "We want to teach kids to be critical thinkers, not test takers."

However, in the give and take of this study's interviewing, this last principal was eventually confronted with the argument that a well-constructed test does actually cover the material in a comprehensive manner and could be constructed to test critical thinking. The corollary of that argument also seems to expose the fallacies inherent in using the "teaching to the test" proposition as a potential weakness to the teaching process. As a result, at the end of her interview, she remarked that "teaching to the test may be appropriate, if the test has value."
Other principals, after probing and cross-examination, gave somewhat grudging, if not outright reluctant support for the idea of using standardized student exam scores to evaluate teachers. One principal felt that this approach could contribute some benefits to the process and another (as previously mentioned) felt that this type of evaluation may one day perform a very minor role. One principal was in agreement with using students' standardized test scores to evaluate teachers but only "if it was definitely a way to increase excellence." As with the other principals, she felt it necessary to add a restrictive clause to her endorsement. Four principals were more critical:

No, I think there are too many other factors.

I believe that standardized testing should not be used as a method to evaluate teacher competence... [because] it depends on the kids, the classroom, the area, the kids are living in, it depends upon the particular standards the school and the teachers might have.

I don't believe it [testing] tells you what a teacher really is. I believe we're there for the educational gain of the student...but I don't believe that's necessarily the qualifier of what makes a good teacher.

Interviewer: Do you think a paper and pencil test is a good way of measuring whether they learn something?

Respondent: It would undermine the very complexity of the nature of classroom teaching because one cannot judge a teacher that way. And the other thing is, do we want to be teaching to CAT (Canadian Achievement Test) tests and the basic skills test?
Formative and Summative Aspects of Teacher Evaluation

Introduction

This section of the interview concentrated on the different purposes of teacher evaluation, formative and summative. Formative evaluation is defined as a method for stimulating growth in employees, forming them into better employees or in some cases, better professionals. Summative evaluation is defined as a method used for purposes of accountability, rank or employee termination. These terms have been commonplace in the educational literature for over 30 years (see Appendix C - Glossary).

Teachers’ Views

When asked if their system of evaluation contained a formative component, all but one teacher agreed that it did. This dissenter stated that the entire process of evaluation served no real purpose in its present structure other than to satisfy the management’s requirement of having an evaluation: “I don’t think it’s [the evaluation] used to do either [that is, formative or summative]. I think it’s [the evaluation] done because it has to be done, and I don’t think there is much follow up to any of it.”

However, the opinions of the teachers on the formative components of their evaluation were not quite as homogenous as the immediately preceding paragraph might seem to suggest. One of the teachers felt that “principals tend to back away from making suggestions for growth;” and another declared that whether the formative option within
evaluation was exercised-operated independently of the style of the evaluation and therefore was likely to be present. Still another teacher related:

There was just a small [portion] at the end [of the evaluation form] where recommendations [were called for]. He [the principal] said two things that he would recommend that I do to improve my program and that was the only part of the evaluation that was formative.

The second part of the question, seeking to determine whether or not there seemed to be any contradiction between the formative and summative components of evaluation, found that four of the six teachers acknowledged a contradiction. Despite this overall agreement, again the homogeneity of the responses was not quite as pervasive as one might expect from their apparent congruence.

While two teachers saw no conflict, one saw no contradiction at all and one viewed the entire process (formative and summative) as a homogenous single continuum and merely routine, the rest of the teachers confirmed that a contradiction between the two philosophies, formative and summative, did exist, albeit along a continuum. One teacher considered their presence simultaneously as somewhat contradictory while a teacher in another school saw a strong contradiction. Two more teachers agreed that there was, indeed, a serious contradiction because of the dual, somewhat antagonistic, nature of the evaluation process.
Principals’ Views

Four of the principals claimed that no contradiction existed between the two components, summative and formative. But, while three of the principals expressed relative satisfaction with their evaluation’s formative component, describing it as strong or adequate, another three apparently believed that there was either no meaningful formative component or no formative component whatsoever. Those who stated explicitly that there was no formative component held that this was a serious failing in their evaluation systems.

Appropriate and Adaptive Aspects of Evaluation

Teachers’ Views

Nearly unanimous agreement emerged indicating that all classroom teachers, regardless of their experience, were evaluated by identical techniques. Very little diversity appears in the responses concerning this issue of flexibility in the evaluation format. In evidence of this, one teacher complained about the push for evaluative consistency, forcing a “one size fits all” evaluation format at her school. Apparently, a new principal and a new vice-principal were “going by the book” more faithfully than teachers at this school had ever experienced. According to this respondent, the push stemmed from the new administrator’s lack of experience and because a probationary teacher was experiencing
vast difficulties. Ostensibly, rules were being adhered to more scrupulously than in the past, so that everyone would be seen as treated equally.

When teachers were asked about the amount of flexibility available for evaluating personnel with duties other than regular classroom duties, the responses were more heterogeneous. Two respondents claimed that teachers with different duties, for example early child education teachers or guidance counselors, were evaluated differently with a methodology that apparently took into account their different roles. The rest of the teachers surveyed either did not know or stated that different types of teachers were not evaluated by different processes or in any way that would distinguish their evaluations from the evaluation format of the classroom teacher.

**Principals’ Views**

Four of the principals reported that the evaluation is conducted the same way for all teachers regardless of their role as regular classroom teachers, guidance counselors or resource teachers. One principal mentioned that the evaluation forms had an addenda section within which it would be possible to comment more appropriately about other types of professional staff who were not involved either in administration or classroom teaching. Two principals reported that at their schools the evaluation procedures for various kinds of teachers were different. One principal stated that “[the evaluations] are very specific to the individual’s role.” In acknowledging a different evaluation process for non-classroom
teachers, another principal stated that "with guidance counselors and child youth counselors, [we] don’t use the same format as the teacher...because of the fact that they are of a different union and that their job is basically different."

When commenting on the relationship between years of service and type of evaluation, four of the principals reported that no relationship existed. No distinctions were made between experienced and inexperienced teachers regarding evaluation. Two principals dissented from this view, one stating that "the outline process that is given to us is slightly different." Later that principal declared the evaluation criteria "could be different depending on the principal but the basic form is the same." The other principal offered no clear distinction between the two types of evaluation, those for experienced and inexperienced teachers.

**Validity of Observations in Evaluations**

**Teachers’ Views**

When asked whether or not the observations on which principals base their evaluations accurately reflect the classroom situations, most teachers had reservations. Specifically, four of the teachers responded with a categorical no, that is, that the observations were not representative of actual classroom situations. But other teachers were not as unequivocal. Of the two remaining teachers, one teacher made the following dissent: “they represent what goes on in the classroom for the most part except that the
teaching strategies that I’m using that day I have given more attention to, more thought to, than I would on a regular basis.” The other teacher, however, stated that the observations were accurate “only if they are coupled with visitations outside the evaluation period.”

The second part of the question deals with the opportunities available for the teachers to see their evaluation forms prior to formal evaluations and the effects that this viewing has. Five out of six of the teachers said they had an opportunity to see the evaluation forms beforehand. The one remaining teacher did state that she did not see the evaluation forms before the administrator appraised her nor was that the regular practice in her practice in her school: “in fact, I’ve had veteran teachers come to me (a probationary teacher) and ask me what happens, because they weren’t sure what the administrator would be looking for, and I simply told them what I did.” What the apparent effect this opportunity for teachers to preview evaluation criteria had on the validity of their evaluation generated more of a controversy. Four of the teachers agreed that viewing the criteria beforehand would affect the authenticity of their evaluations by either influencing the teachers to do something that they might not normally do or by otherwise giving the teachers some sort of informed advantage. Conversely, one teacher stated that previewing would help the teachers put the criteria being adjudicated in their program on a more regular basis.
Principals’ Views

Here, in contradistinction to the teachers, five out of six of the principals believed that what they evaluated did represent an authentic picture of what transpired in the classroom. The principals felt justified in claiming this level of accuracy because they frequently visited the classroom informally.

All the principals reported that the teachers did have the right and opportunity to see the evaluation content before their formal evaluations. Only one principal acknowledged the possibility for teacher deception because of that privilege. This principal agreed with the majority opinion of the teachers. She argued that since teachers were generally given the opportunity to see the evaluation criteria beforehand, the opportunity for the teachers to “incorporate” them into their teaching strategies just for the occasion of the principal’s visitation would be too tempting. The result, in this principal’s opinion, suggests a strong likelihood of an affected performance, one done only for “appearances,” and not one representative of usual classroom work. The other principals countered by claiming that multiple informal visitations probably allow for accurate pictures of the classroom situation. These principals further asserted that the teachers’ right to preview the evaluation forms was justifiable because they “need some sense of what the game is, due to it being a public document,” “the ability to show what we’re capable of” and this process’s capacity to raise the comfort level of the teacher. One principal, however, held that there were probably not enough visitations and the possibility did exist that the teachers could
fabricate teaching proficiency by substituting an expected script for their usual retinue of pedagogical routines.

Adequate Number of Visitations for Fair Evaluation

Teachers’ Views

When the teachers were asked for an estimate of an adequate number of visits, they were unanimous in condemning the present approach to visitations because of their short duration and low frequency. When further questioned on what would be better, five out of six of the teachers wanted an increased number of visitations or longer periods of observation. Two of the teachers were strongly enough opposed to the present system of visitations to actually want surprise visits. As one teacher stated, “I think if you have a good program, you should be ready for every day of your job.”

Principals’ Views

All the principals generally agreed with the teachers regarding the question of number of visitations. They also claimed that there were not enough occasions for formal observation for them to give fair and accurate evaluation. In fact, only one principal asserted that there were adequate visitations to conduct an effective evaluation. One of the principals, moreover, adamantly insisted that there were not enough formal visits to do an effective evaluation; this high school principal concluded that “anything less than five
formal visits is not enough.” She added that in her board there were no prescribed amount of visitations and the number of visits the evaluation was based on were “totally capricious.”

The Status of the Follow-Up Interview in the Evaluation Process

Teachers’ Views

When teachers were asked if a follow-up interview occurred after the evaluation, all of the teachers responded that there was some follow-up. The second part of the interview question asked for details of their follow-up interviews. Here, five out of six of the teachers expressed doubt about whether, in its present form, the follow-up interview would be an effective tool in furthering the goals and aims of evaluation. Comments like the following were common: “The principal does not usually take it very seriously;” “We discuss it generally for ten to fifteen minutes;” “Yes, [we have] a follow-up meeting maybe five or so minutes [and] there wasn’t [really] much;” and “[Basically] he handed it to me, I read it [and] I said I agreed or disagreed.”

Principals’ Views

All the principals concurred with the teachers that there was a post-evaluation interview. However, a clear, salient difference was evident in the description of these interviews. One principal stated that she would “sit with them [that is, teachers] for an
hour after class...and go through the whole lesson almost [not minute by minute necessarily] to make sure they saw exactly the kinds of things I saw in the classroom.”

Another principal said that the interview resulted in the teacher having to correct or improve certain aspects of their teaching. Yet another principal commented that the post-evaluation discussion sometimes continues to leave the teacher feeling that the report is too negative. This principal concluded that “an outside person coming in from the College of Teachers would be better for a school staff -- that in itself would provide some objectivity.”

**Evaluation & Professional Status: The Promotion System**

**Teachers’ Views**

When asked whether evaluation results were used to benefit the teachers’ professional status, most teachers agreed that the evaluations were not major components of the promotion systems in their schools. One teacher felt that the promotions to curriculum leader or department head were almost totally arbitrary, stating that “whoever is free [is selected], that is, whoever doesn’t have a classroom basically. [And] I mean someone that the principal feels they can count on, obviously.” Other comments concerning the reasons for promotions were even more emphatic when insisting that any advancement was independent of evaluation procedures. One teacher stated that promotions were “absolutely not [based on merit]: promotions are based on seniority and they [the
people promoted] are never demoted." Even those teachers who conceded some minor role for the formal evaluations in the promotion process seemed very tenuous about it and unable to articulate the nature or procedures outlining that role.

**Principals’ Views**

The principals’ opinions, in this case, seem to coincide fairly closely with those of the teachers’. Only one principal stated that evaluations were used for promotion and this was said guardedly: “Yes, to some extent; the principal, who has to write an assessment [of the applicant] relies on past evaluations.”

The rest of the principals were solidly in agreement that promotions were based on either experience, that is, seniority and accumulated courses and/or university credits. One principal went so far as to state that this promotion system, which for the most part ignores the evaluations, was the preferred one: “It’s very important, because you have one evaluator writing one report on a couple of situations they saw.”

**The Discriminatory Powers of the Evaluation System**

**Teachers’ Views**

Even though all of the teachers felt that although the vast majority of their colleagues in the classroom were above-average, nevertheless, as one teacher put it, “even probationary teachers are not dismissed for incompetence” as a result of their sub-standard
evaluations. Thus evaluation was thought to have had little connection with the degree of their colleagues’ competence. Another teacher, when asked if the evaluation system would expose the incompetent teacher, replied (somewhat indignantly):

To whom? Who reads these? I don’t know anyone who has read these. They go into a file folder and they’re kept in the principal’s office and I don’t know if they go anywhere after that at all.

In fact, only one teacher felt that the evaluation system did successfully discriminate between the competent and the incompetent teacher.

**Principals’ Views**

When the principals were asked the same question, they responded homogeneously, for the most part. Somewhat unexpectedly, one principal for the most part, implied that there was a stronger dichotomy among teachers’ abilities categorizing them as either very good or very bad. One principal even suggested that all the teachers at her school were “outstanding” and said that since “it is so hard to get into Teachers College right now that the people that you are getting really want to be there,” and that degree of selectivity would account for the excellence.

Where the opinions between teachers and principals diverge sharply is on the matter of the evaluation system’s efficiency at discriminating between good and incompetent teachers, particularly at weeding out the incompetent ones. Here, most principals believe that the system is effective. In fact, of the principals interviewed, only one conceded that
incompetent teachers were transferred rather than dismissed: “For that administrator of that school, it’s easier and [more] expeditious than the whole process of trying to determine if someone is incompetent. That’s just the nature of the beast as we have it right now.”

All the other principals’ opinions, however, stand in contradistinction to that one just mentioned. One principal was very specific in her opposition. She stated emphatically that “there is a whole review process. If they aren’t competent, they are put on some goals and if they don’t meet the goals and the time line, they are put on review and the superintendent is involved.”

The Cultural Compatibility of the Education System

Teachers’ Views

The teachers in this section of the study were asked for their opinion on the comparability between their schools’ evaluation system and culture. Here the prevalent answer, with only minor variation, was that the degree of compatibility is more dependent on who the administrator doing the evaluation is, than on the actual form or system of evaluation. Only two teachers gave unqualified affirmative responses to the questions concerning how compatible their evaluation processes were with their schools’ cultures. Two of the respondents replied in the negative. One of these teachers stated that “the board does not take culture into consideration as a big factor.”
Interestingly enough, one of the teachers that characterized her school's culture as compatible did not see this cultural correspondence as a positive attribute. In many ways, this responding teacher apparently had the impression that the culture that permeated the school and the classroom in particular was not contemporary enough. The positive aspect according to this respondent was that the schools in general and the evaluation system in particular were trying to evolve that culture into one more compatible with leading-edge teaching methodologies.

Principals' Views

By and large, the principals affirmed their evaluation system's congruence with their school's culture, largely in opposition to the teachers' views. Only one principal expressly stated that culture was not directly paid attention to, but also mentioned that this incongruence was mitigated because the evaluation process at that school was so generic that any cultural incompatibility influencing the evaluation process would be almost irrelevant. All the other principals seemed to be of the opinion that the evaluation was compatible with the particular culture of the school. All the same, one principal did agree with the teachers' view that cultural compatibility was more closely correlated to the evaluator's unique personality, that is, the insight, awareness and sensitivity of the individual administrator was the crucial factor in determining that congruence.
Preview and Clarity of Evaluation Criteria and Their Inherent Problems

Teachers' Views

Four of the teachers implied that the criteria for evaluation were explicitly outlined. However, two teachers dissented, stating that this was not the case. One of those teachers expressed her views on the issue of clarity, by stating that it was evident only some of the time, and even that depended, for the most part, more on the school where the evaluation was taking place than on any stringent policy demanding a certain standard of lucid comprehensibility.

All of the teachers were of the opinion that nothing needed to be deleted from the listed criteria for evaluation at their boards. One of the teachers declared, however, that the omission of co-curricular and extra-curricular activities in the evaluation of teachers was a drawback and the system should, therefore, be broadened to include some other areas that teachers work to foster improvement in their students. For example, independent remedial work should be introduced. In other words, the definition should be expanded to include extra-curricular (volunteer) teaching and other categories that contribute to student development, some of which have nothing to do with the classroom or academics.

Several teachers, independently, seemed to have reached a consensus, arguing for the addition of certain completely new criteria. These criteria were based on a philosophical construct that demanded a more comprehensive outlook for defining evaluation, such as using the contents of a portfolio as a criterion, as well as using
standards that recognize contributions to the school cultural community and the individual's dedication to professional development.

Some of the teachers advised against revealing any of the evaluation criteria prior to the actual evaluation because those revelations might unfairly slant the results in their favour. Essentially these teachers were of the opinion that "you act a script" or that the preview detracted any presumed spontaneity because "everyone knows" exactly what they would have to do in order to merely imitate the requirements of a "good" evaluation.

**Principals' Views**

Five of the principals believed that the criteria for evaluation were clear and the teachers had ample opportunity to see them before a formal evaluation. One of the principals, however, proposed that whether or not a teacher saw the criteria was dependent on the particular administrator's perspective on how beneficial that sharing would be for evaluation.

The principals also seemed to be comfortable with the present preview format and implied that spontaneity was not affected by teachers knowing the evaluation criteria in advance; they held that it would be unfair to "surprise" the teachers with an evaluation. One principal went so far as to suggest that it was his duty to present the criteria in advance to the teacher. Most of the principals were satisfied to the extent that they believed nothing needed to be added or removed from the evaluation criteria, in their opinion. Two
principals, however, dissented on that point. One stated that some criteria were unclear, that is, the guidelines were too general because of the checklist approach, and should be removed; similarly, the other suggested that “school involvement in extra-curricular activities should be added.”

Teachers’ Reactions to Derogatory Evaluations

Teachers’ Views

When the teachers were asked how they react to criticism of their teaching by evaluators, the answers seemed ambiguous. Five of the teachers implied that some teachers (but not they themselves) would become defensive. The one remaining teacher said that teachers would not resent the criticism of the evaluation report but, in general, would resent the process of being evaluated just because “somebody is there taking notes on your performance.” Interestingly enough, one teacher initially said that any teacher who is given criticism would “just blow it out the window” because nothing usually happens “unless the person says something to you formally about it afterwards.” However, on further probing, the same teacher said that “it depends if you are given constructive feedback or…if he says ‘You are really lousy at this…’ and then doesn’t offer to help you out or tell you how you can improve yourself.” Further probing into this issue was followed by that teacher mentioning a hypothetical scenario in which the following conditions were outlined: the principal was not constructive; suggested a need for serious improvement in a couple of
vital areas; added some derogatory ad hominem epithets referring to the teacher as lazy and incompetent; criticized that teacher’s pedagogical methodology saying she was using teaching methods from 20 years ago; and so forth. This teacher replied that she would probably become defensive and conceded that “I’m sure it might make things kind of uncomfortable.”

**Principals’ Views**

Among principals the responses were more heterogeneous. Three principals were of the view that the main factor affecting the attitude of teachers receiving criticism was the personality of the individual teacher. When probed for an elaboration, one principal responded by stating that “experienced teachers are generally more easygoing and more reflective on their own practice and therefore willing to sort of discuss how they might improve.”

The other three principals presented heterogeneous views. One principal held that the criticism was always taken in a positive frame of mind and thought that “most people are generally positive and wanting to grow...I haven’t had anybody ever be upset and say ‘This isn’t right.’” Another principal conceded that most teachers were categorically defensive. When this point was probed, the principal responded that he had no idea why: “I just find teachers, generally, somewhat insecure and a bit defensive.”
Problems with Objectivity of Evaluation Criteria

Teachers' Views

When asked whether or not the criteria used for evaluation were objective or subjective, two-thirds of the teachers replied that they were subjective. One teacher was particularly emphatic that the subjectivity inherently rendered the appraisals less effective. Consider the following dialogue between this interviewer and one respondent.

Respondent: I think more objective would be good. It would certainly help make it more equitable across all the teachers, because even in a high school the Math Department may be evaluated by vice principal “A” and the English Department is evaluated by vice principal “B.” They are going to be very different types of evaluation even in the same school.

Interviewer: And vice principal “A” may know nothing about Math?
Respondent: That’s right.

Those teachers who insisted, however, that the criteria were objective, seemed unable to supply any valid reason or example to substantiate that opinion. One teacher insisted that the criteria were objective but on further probing conceded that “the report itself is subjective” and that none of the criteria were totally objective. A different teacher, who incidentally was consistent in defending the status quo of the evaluation system, insisted that the criteria were objective but was not able to provide a reason why she felt this way.
Principals’ Views

The principals, once again, were almost totally at odds with the teachers in their defense of the evaluation system as it now exists. Five of the six principals viewed the criteria used in their present system as not subjective or at least as “not too subjective” to give valuable information about the teacher. The following citations are indicative of the principals’ defense of the status quo.

1. Interviewer: Have you ever felt that the evaluation system in your school has been too subjective to offer useful information?
   Respondent: No.
   Interviewer: Does your appraisal system use any form of objective criteria?
   Respondent: No.

2. Respondent: I don’t think the criteria are subjective. I think it lies in the hands of the evaluator...I’ve gotten reports back about my own teaching practice and they’ve been very complimentary, but I’m not so sure what they’ve seen.
   Interviewer: Does your appraisal system have any form of objective criteria?
   Respondent: No.

3. Respondent: I’ve never felt that (the evaluation system is too subjective)
   Interviewer: Does your appraisal system use any form of objective criteria?
   Respondent: No.
   Interviewer: Would you like to see that?
   Respondent: No.

Interestingly enough, the same principals who deny any serious problem with subjectivity in their evaluation criteria freely admit that they do not use objective criteria in evaluating teachers. Apparently, as the preceding interview segments demonstrate, the principals fail to recognize this duality as a serious contradiction. One principal conceded
the problem with criteria subjectivity, but she objected against using objective criteria like scores on standardized student tests in an evaluation of teachers.

Protection of Rights within Due Process

Teachers’ Views

When teachers were asked whether their rights were protected within a system of due process, all teachers responded affirmatively, stating that they regarded their rights as adequately protected. However, when the second part of the question was raised concerning the potential for systemic abuse, the opinions offered were more diverse. Two teachers discarded the question as basically irrelevant because, in their opinion, no administrator could become desperate enough to use an evaluation procedure to harass a teacher when there are many other better, more powerful, ways. One of those teachers, however, did not even “think evaluations are important enough to cause abuse.” Another teacher reacted similarly, by stating that abuse was highly unlikely because “it [that is, the evaluation] doesn’t mean anything, really.” A different teacher suggested that the administrators might abuse the system “because the administrator is in a more powerful position -- they are the evaluator.” Still another (in a departure from her usual views) also offered that the administrators were in a position to abuse the evaluation system. She felt that:
If you are a teacher, like the previous example I gave, who does not know that your federation [that is, union] is there to help you and you have an administrator there to crucify you, you'll get crucified.

In a position almost at antipodes with the two just cited, one of the other teachers in the study stated that any exploitation of the system would be from the teacher's position because “you can’t, in the final analysis, get rid of the incompetent teacher.”

**Principals' Views**

The principals, in complete agreement here with the teachers, assessed the system as scrupulously preserving the rights of the teachers by providing due process. Similar to the teachers, the principals, as administrators of the evaluation system, had varied responses to the question concerning the possibilities for abuse. Half the principals conceded that it could be possible for an administrator to use an evaluation to unfairly harass a teacher. However, one principal insisted that she was aware of only one incident where harassment actually had occurred: “In one case where I sort of suspected -- when I was a teacher -- that a principal went after a person.” In fact, several of the principals maintained that this harassment never takes place. Concerning this, one principal stated confidently that at her board that type of abuse would be impossible “because of the collaborative nature of it, at least with our collaborative nature with [in] our system.” An even stronger view was expressed by another principal who claimed any systemic failure that did result would be because “someone had not been doing that ‘paper trail’ [that is,
Moving on to a slightly different view, though, she next admitted that although abuse in the form of harassment could take place, it would be very difficult for an administrator to do.” This was probably because of the union that would defend the teacher strongly from any trumped-up accusations.

**Self-Directed Criteria Selection**

**Teachers’ Views**

In response to the question of whether or not teachers themselves should be involved in setting the goals or criteria for evaluation, all the teachers answered unanimously in the affirmative. Furthermore the teachers also agreed unanimously that this co-operative effort was of primary importance. Since it was expressed in so many different ways, however, unanimity might only describe these results in a very general way. This consensus, however, did exist on the role that teachers should have in helping to formulate the conditions for evaluation. All teachers in one form or another seemed in favour of a collaborative effort between teacher and evaluator (usually principal).

Examples of this type of feedback can be seen from the following reflections: “I think that in that [collaborative] way, change can come through teachers having more ownership of the process in setting goals for themselves.” Based on the requirements of the board, “the principal and the teacher should be able to discuss in which way they
would like to see that [teacher’s competence or excellence] demonstrated.” Another teacher thought that setting up a portfolio following a discussion of its basic criteria between the teacher and the principal could be an important step in the creation of a climate necessary for effective evaluation. It seems fairly clear, then, that all the teachers seem to want some role in setting up evaluation and feel it would be best served by working together with the principal in a co-operative effort to produce an evaluation following board guidelines with which they can both be satisfied.

**Principals’ Views**

The principals, too, admitted that evaluation should be a joint effort between the evaluator, usually the principal and the teacher. They seemed to agree, although not quite as cohesively as the teachers, that evaluation would be less effective in a non-collaborative context, and probably not capable of inducing professional growth. Similarly, all of the principals stated that this factor was of primary importance.

These ideas on the collaborative nature of evaluation suggest certain roles that the principals see the teachers as playing, especially in the incipient stages of the evaluation process. The following are declarative statements commenting on those roles:

*Principal No. 2* - Yes, hopefully, but not necessarily, I think it [evaluation] should be done jointly, but if I believe there is something that a teacher needs to work on and unfortunately the teacher may not agree, but I will give my criteria as to why I believe that that’s the area that they need to [work on]. So, hopefully, it’s joint but not necessarily.
Principal No. 4 - I find it [collaboration] works effectively if they suggest some areas where they would like to see some growth within themselves.

Principal No. 5 - And, of course, the union representatives sat on that committee as well, and he is in fact a representative of teacher [that is, the committee formulating the evaluation criteria].

In addition, as with the teachers, all the principals said they wanted a partnership with the teachers in the evaluation process and that this initial stage of the evaluation was necessary and of primary importance.

The Principal-Teacher Relationship in Evaluative Situations

Teachers’ Views

When the teachers were asked to further characterize the type of rapport between teachers and principals regarding the evaluation procedure, no strong consensus was identified. Two of the teachers said that it was arbitrary, that the type of relationship essentially depended on the already existing relationship with the principal. These teachers maintained that their experiences with their evaluator prior to evaluation were the most significant factors in determining their association later, within an evaluation process, regardless of the nature of the evaluation, positive or negative. Three teachers stated that the professional exchange between evaluator and evaluatee was beneficial under nearly all circumstances. One of these teachers who described the procedure as collaborative said the following:

I think there is an understanding that this is a bureaucratic hoop that we both have to jump through, so let’s get through it together and get on with things.
In contrast to those views just mentioned, one of the teachers in this study suggested that the usual feelings that existed between evaluator and evaluatee were antagonistic. This situation was due to the inherent nature of the relationship which is, for the most part, unbalanced, from the power perspective, in favour of the principal. Essentially, this teacher implied that the power component of the relationship, due to the control with which the board imbues the principal, was so overwhelmingly weighted toward the principal that it intrinsically defined the relationship as an anxiety-inducing one.

**Principals’ Views**

The opinions of the principals were found to be just as variegated as the teachers. One principal declared the process very open and another very positive while another principal went so far as to claim the following:

In the four years that I have been evaluating, I have not encountered anyone who really seemed to object to me going in [to the classroom]. I think a lot has to do with the way that you approach it.

Two principals said that the quality of the co-operation was dependent on the evaluator and on the type of relationship that existed prior to the evaluation. A different principal, surprisingly, for someone so supportive of the status quo, stated quite openly that the determining factor in acceptance by the teacher was whether or not the outcome was positive. As that principal said:
I guess it’s the outcome that is the determining factor, isn’t it? If you need an evaluation and one had to be done and it was negative, you would think the process would suck. So the outcome is the determining answer in that factor, I think.

Apparently then, overall, no consistent opinion emerged among the study’s principals illustrating the cause or nature of their relationship with the teachers when evaluating them.

**Professional Training of Evaluators**

**Teachers’ Views**

When asked about the professional training that principals underwent to qualify as evaluators of classroom instruction, teachers gave a wide variety of responses. Four of the teachers said they did not know and did not feel qualified to discuss that aspect of the evaluation system. One teacher said that she did not think that the principals received any specific training that would qualify them to act as professional evaluators. Another teacher was of a similar opinion but stated even more emphatically that whatever training they did receive as evaluators was “highly inefficient.” A different teacher conceded that the principal clearly did not seem to receive training, but she was not sure what training the principal did receive and, therefore, could not comment.

When asked what they would do to improve the evaluation system and specifically the principals’ training in evaluation, again some teachers questioned their own competence to respond. Those who did agree to reply, did not produce any consistent theme. One
teacher expressed the view that peer evaluations, which in effect leave the principal out of the evaluation system, would be better because then "a good teacher could come in and watch my class and help me develop my skills in that subject."

Another teacher, evidently, was exasperated enough to state that, although she could not suggest any definite new strategy, it would be better to "throw out what we're doing now and come up with a better system." To a degree, this opinion fortifies the belief that what exists now is just not working, at least according to this teacher.

The final teacher in the group interviewed stated categorically that principals should be trained through simulation exercises, which she said "they are not using now." This training would encourage principals "to come up with a more standardized way of evaluating or have a hundred cases of evaluations that you need to read and review."

**Principals' Views**

Surprisingly, the principals were generally even more critical of the training they received in evaluation than the teachers were. All were at least somewhat critical, some were defensive, and most had suggestions for what would be better.

One principal abruptly stated that there was "no such thing as a professional evaluator at a high school level." Another principal, when asked about the amount of training, responded that there "probably was not as much as there should be." Another principal concurred, stating that it was "minimal at best." A different principal was vague
about the value of the training but, when asked for specific details, conceded that "we didn’t actually take the document and practice on someone." The penultimate principal interviewed agreed, saying, “I don’t think they addressed it nearly enough in the sense of practical terms.” As that principal described it:

They brought someone in and they talked about it and so forth. What they did was mostly look at [theoretical] models of evaluation as opposed to really looking at how you [practically] approach the process with teachers and so forth....But I would have liked to have asked some of these questions you are asking.

The final principal admitted that, “In our board, to my knowledge, we have no formal inserviceing for administrators to conduct formal evaluations.”

No consensus, however, seemed to materialize among the principals that would sanction a better process for training evaluators in any clearly defined manner. Some principals, as was the case with the teachers, could not come up with anything concrete at all. One principal suggested doing some graduate courses in evaluation to supplement the government of Ontario’s Principal Qualification Program (PQP) requirement, although he found it an extremely valuable resource for general information about school administration. Supporting this view, another principal expressed support for a certificate in applied evaluation stating that it might prove valuable depending on how it was structured. With respect to the College of Teachers, only one principal came out strongly in its support as an important element in the training process of evaluators. This principal reasoned that the College would eventually imbue the profession with the rigorous innovations and professional standards that were desperately needed in the evaluation
system. She saw the results of a program run by the College as producing better trained and more aware evaluators.

General Attitude of Teachers Toward the Evaluation Process

Teachers’ Views

Teachers, when asked about their reluctance to be evaluated, all agreed that it was an unpleasant experience and something to be avoided if possible. One teacher, however, did make the distinction between probationary teachers who are held to a much stricter regimen and teachers with a permanent contract. This sense of unpleasantness was attributed to the avoidance behaviour which was said to be fairly prevalent among the permanent teachers. Another distinguished between the evaluation, which teachers generally dislike, and the results, that teachers are interested in seeing because of their potential summative effects in terms of promotions, transfers or just retaining their jobs. As that teacher reflected:

I think they try to avoid the evaluation but not the results ... The results, as a teacher, you surely want to know if you are interested in keeping your job or being promoted or just want to work at a different school. You want to be sure that whoever wrote your write-up wrote something positive or something extremely positive.

One teacher had an interesting view, suggesting that teachers don’t avoid their evaluations but “try and get them over with because they are a hassle.” Although this view, superficially, may seem in opposition to the view expressed by other teachers, the
basic underlying theme seems the same: evaluations generally disrupt and do not materially enhance the daily teaching and administrative routine of the school.

**Principals' Views**

The principals took a contradistinctive view, that is, one completely in opposition to what the teachers in this study believe. Four of the principals were of the view that the teachers have a positive attitude toward evaluation and they, as principals and evaluators, had not experienced any resistance to the evaluation on the part of the teachers. Two principals differed. One, who seemed a little uncertain, nevertheless volunteered that her “perception was that maybe they [the teachers] were trying not to do it; they don’t like it” [the evaluation process].

The other principal was more definite. This principal openly said that he “think[s] teachers think it’s a waste of time, especially for those who have been veterans.” Continuing with these thoughts, that principal said, that “there are some who don’t want to be evaluated because, and I go right back to the beginning, there seems to be some insecurity and defensiveness.”

Clearly, the majority of principals and teachers in this study do not agree on what teacher attitudes are toward evaluation.
Successful Evaluation and the Nature of Its Design

Teachers’ Views

In this section of the study, teachers and principals were asked how they would design an evaluation system if they had total authority to do so. Among the teachers a partial consensus seemed to develop. Four of the teachers, if given the choice, would choose to be evaluated by a portfolio system. In this type of system, teachers design their own goals, work toward what they are interested in improving and then provide documentation of their accomplishments, which are usually presented to a council of their own peers. Commenting on the reason for that choice, one teacher said:

I would much prefer to have an experienced member of my own department take a look at my teaching in math than have a vice principal who doesn’t know anything about calculus in my class, because how can that person tell if they [the students] don’t understand it [the math]. High schools are so subject and content specific now that I would need someone who understands calculus in my class.

The two dissenting teachers were not actually opposed to a portfolio system either. One teacher was satisfied with the one currently in use in her school. Another teacher supported any evaluation that shows growth over time, an approach that would probably not be inconsistent with a portfolio system of evaluation.

Principals’ Views

Although two of the principals were satisfied with their present evaluation system, half of the principals thought their system was inadequate because it granted the teachers
too much input. One of that group who was satisfied openly stated that the system was fine the way it was and that it shouldn’t be “necessary to involve teachers, except that I think we need more peer coaching, peer mentorship and the administration may be providing a little more guidance on what they deem to be excellence in instructional practice.”

In contrast to that view, another principal vocalized her general exasperation with the whole field of teacher evaluation: “[I] would like to see somebody do some looking at the different evaluation processes, and I would like someone to come up with some reasons and rationales [for the various evaluation systems].”

Expressing her frustration based on the inherent conflict of interest, a different principal suggested that school principals should be relieved from conducting teacher evaluation as any part of their professional duties. That principal observed that she “would support independent evaluation so that the administration of the school is not given the responsibility of evaluating teachers within that school. I think it provides more objectivity, too and would create, perhaps, a better evaluation.”

Effects of Evaluation on Teaching Quality

Teachers’ Views

A majority (four) of the teachers considered evaluation as having had no tangible effect on the quality of teaching in their schools. One teacher did demur, although she was ambivalent about the strength of that dissent. As far as this teacher’s personal teaching
history was concerned, evaluation “has impacted on my personal teaching because I have been given advice on how to improve,” but when further questioned by this researcher (“And you felt it was significant advice?”), she claimed that it was good for her only because at that early stage in her career “any advice is good.” This teacher, continuing with her comments, seemed obliged to voluntarily add that regarding post-probationary teachers, “I don’t think it really impacts on what they do.”

Another teacher, however, was more complimentary when commenting on the evaluation system’s effects. This teacher’s opinion was that the system often had a positive effect on teaching because “some teachers have left our school because of poor evaluations. A teacher has [even] left the profession as a result of a poor evaluation.”

Principals’ Views

The principals, as opposed to the teachers, agreed that evaluation has had a redoubtable impact on the quality of teaching in their schools. The principals’ opinions only differed as to the degree of contribution evaluation has made. Five of the principals were emphatic that their evaluations were beneficial and formative. One principal was implacable in asserting that evaluation had a major impact because, if not, “I would just be doing it for the sake of paper work, and I don’t believe in doing that.” Another principal justified her evaluation system “because the fact that we’ve been able to get rid
of teachers who teachers on staff have been saying for years and years should have been gotten rid of” shows the effectiveness of the evaluation system.

Only one principal criticized the evaluation system at all, and this seemed to be because her school had just switched from a comprehensive portfolio system that she felt was considerably more effective than the depersonalized checklist format that had been adopted.

The Reluctance to Dismiss Inadequate Teaching Staff

Teachers’ Views

All the teachers agreed that schools were more reluctant than businesses to dismiss their employees. This consensus should not be taken to mean, however, that their opinions were identical. When the reasons for their opinions were stated, quite deep divisions were revealed in their reasoning. Three of the teachers suggested that this phenomenon was due to the collective agreements’ restrictions on management withholding the authority necessary to dismiss incompetent teachers within a reasonable amount of time. Some of the teachers opined that the process was seen by managers as so cumbersome, restrictive and time-consuming that they were reluctant to initiate it at all. One teacher in this group commented, “Once you have a permanent contract, I have never known of a teacher who has been asked to leave the profession or [even] the school.”
Another teacher said that these problems existed because the principals were the evaluators and the school cultures seemed to demand a non-competitive atmosphere which restricted the more hegemonious aspects of the employee-employer relationship making the principals “too close with the people.” One of the teachers viewed the main problem not as the culture or the unions but as the inherent nature of the evaluation used within the teaching profession, and the way society defines the role of the teacher. As that teacher stated, “we’re very different from business and because of that we’ve only got subjective criteria and you can never fire someone [based] on subjective criteria.” Furthermore, she indicated that “because of the function of the school, we don’t have to show we are good.”

**Principals’ Views**

The principals’ views, when contrasted with the teachers’ views, were in agreement half of the time about the legitimacy of the problem of dismissing incompetent teachers. Three of the principals agreed with the teachers that schools were more reluctant to dismiss their teaching staff than businesses. A surprising discovery was the number of principals who were quite candid about the problem. One of these principal bluntly stated regarding dismissals, “There is a huge reluctance, and that concerns me because we have so many exemplary teachers [who could be brought in to replace the incompetent ones].” Another principal conceded that the alleged inordinate reluctance to dismiss was probably true but attributed this discrepancy to organizational and management problems, while a different
principal insisted "that’s absolutely the case, however, the nature of teaching means you never fire for incompetence because teaching is growth, teaching is [learning]." Interestingly enough, in this section of the study, no principal stepped forward and openly laid the blame on the teachers’ unions and their demands for such stringent requirements as the sufficient condition for retaining an incompetent teacher.

The other half of the principals’ group disagreed, arguing that the claim of reluctant dismissal was either nonsensical, foolish or libelous. For example, the first principal interviewed seemed, to this researcher, angry and fiercely defensive; she bristled as she shot back the following:

It’s ridiculous to compare business and to compare teaching. It’s not a valid comparison. Teachers are not comparable to the medical profession in terms of how they can and can’t discharge certification. I mean I can go in as an employer in a business and the only thing that ties me is the Labour Relations Act where I have to give ‘X’ number of days notice to somebody with not necessarily just cause in writing or anything.

However, a different principal, but still from the same group of those interviewed, based her feelings, that the charge of tolerating incompetence was false, on her experiences "with the Bank of Montreal, with Petro Canada and with a couple of the architectural firms we have, [who] deal with it exactly the same as we do.”

Also defending the status quo, but with some mild concession, one of the other principals commented that while "that might have been the case ten years ago, but definitely there has been a big change [toward more expeditious and efficient dismissals].”
Retrospective

This essentially represents the opinions of the principals and teachers in response to the interview protocol located in Appendix B. How valid these opinions are and whether they reflect a fair and honest judgement of the culture and organization of their school systems, how they compare to the professional literature and how they effect evaluation generally, will be discussed in the next section.

This section of the study, as presented thus far, has tried to represent the raw data. Thus, the results of the interviews were shown, for the most part, at face value, without scepticism and with the quotations substantiating their validity held to a minimum. Thus, any analysis that subsequently appears in these interviews was not presented in this section. In the next section, The Analysis, the data were briefly summarized and then the scrutiny of the interview was recorded in addition to any post-interview analysis. It is through these analyses, interview and post-interview, that the credibility of this study’s population, the teachers and principals, was verified or refuted. It is for this reason that many of the rhetorical devices (logic, reduction to absurdity, analogy, sarcasm, parallel cases, and so forth) are used exclusively in the analysis. For the same rationale, most of the more complex quotations were reserved for the section on analysis, where the subtlety of language was necessary to support the type of exegetical analysis employed to glean evidence from the interviews. It is on this basis that much of the more complex, intricate
dialogue was not cited in the data section but, rather, in the following section of the analysis.
At this point I should emphasize the break in the established methodological procedure of interviewing due to the unique problems endemic to the topic of this dissertation. As explained in some detail previously, in the section on methodology and philosophy, the interviewing done in this project goes somewhat beyond the regularly accepted protocol for interviewing. Yet even that more assertive format was found to be insufficient for the purposes of this project. Frequently, and admittedly this did involve the discretion of the interviewer, the position taken by the respondents, either teacher or principal, seemed to inexorably and inveterately prohibit further questioning due to the emanating tension of the situation. As such, the interviewer had at least two choices available in order to keep the information flowing: (1) to abandon further questioning and let the position stand or (2) to develop a section on logical and critical analysis and expose the arguments to further scrutiny. Of these two choices, the one most helpful to the purposes of this study seemed to be the second one. It is for that reason that I have chosen to continue applying the methodology of confrontation in this section on analysis.
Although some researchers will take issue with this approach, I am comfortable using it, and I am reasonably confident that it can be defended within the accepted norms of social-scientific research. Though this critical scrutiny of sources may not be the accepted regimen for gathering interview data, it is well within the normative parameters of other methodologies such as historical research, which scrutinize their sources, and document analysis used in many fields of study such as English, law or philosophy as well as history. It is an accepted norm in those realms for researchers to question their data in order to verify its legitimacy (see pages 36-37).

In addition, as outlined in the title, this section of the dissertation will concentrate on analysis. While the same topics as in the Data Section are discussed here, they are analyzed in depth to show subtle differences of opinion, strengths and weaknesses of the arguments and broad general themes that have developed through the course of the research. While some of this section may seem repetitive, what actually is occurring is a deeper, fuller, more analytical explanation of the previously presented raw results of the data.

The Best Aspects of Teacher Evaluation

Analysis of Teachers’ Views

Apparently, a general attitude existed among the teachers that any form of professional growth that resulted from their evaluations validated that process as beneficial.
Yet, on further examination, this view was not as monolithic as it seemed. A content analysis of these remarks revealed that the growth aspect which teachers mentioned appeared dependent, as far as could be discerned from a long and probing interview, on the specific teacher's orientation and that teacher's general philosophy. Therefore, some teachers suggested that it was the guidance and feedback that they received from their administrative evaluators that was crucial in making evaluation a worthwhile experience; the more specific the comments, the better. One teacher, however, went to some length in emphasizing the importance of having the autonomy to set professional goals as the essential elements that allowed her to feel more proficient in her role as a teacher. A different teacher asserted the importance of independent external reinforcement, one which essentially reinforced her skills as a competent teacher. Confirming this dynamic as a positive aspect, another teacher added that since this reassurance was the main enhancing aspect of the whole experience, the process could justifiably be declared fairly trivial and perhaps even "frivolous." This teacher went on to say that in her opinion, "The process makes a difference to me [only] insofar as it is cause for a little bit of anxiety."

This information is important because it highlights the teachers' professional concern with growth of most teachers in this study's population. It seems to suggest an alternative to one of the popular stereotypes of the teacher as a browbeaten dullard just "serving time" and waiting for retirement. Perhaps a seminal illustration of that stereotypic teacher, the one so burnt out that their sole goal is physical survival until retirement, can
be found in Mordechai Richler’s classic, *The Apprenticeship of Duddy Kravitz*. In that novel, Richler portrays that type of teacher’s despair by presenting us with the loss of hope and expectations of Mr. Macpherson, a once young and idealistic teacher, who long ago began to bitterly regret going into the teaching profession. Macpherson each year continued to regret that decision in an ever increasingly heart-breaking but pathetic manner:

> Long ago Mr. Macpherson had vowed never to strap a boy... That he no longer believed in not strapping was beside the point. As long as he refused to do it, Mr. Macpherson felt that he would always land safely. There would be no crack-up. He would survive. Another eight years he thought. Eight years more and he would retire. (Richler, 1959, p.15)

The data here, however, seem to highlight an alternate perspective of teachers here steadily honing their performances with the aim of improving them and wishing there were more sophisticated rigorous opportunities to do that. In fact, the strongest overall complaint about the system seems to suggest that these particular teachers in this study may be quite dissatisfied with the relatively unsubstantial degree of rigor that the evaluation system is presently prepared to offer. This gap, between what they expect and what they receive from the process, might also suggest that the problem with instituting more discerning evaluation not necessarily geared to the lowest common denominator, something that the literature has been demanding for years, may not lie completely with resistance from teachers.
These more stringent attitudes of teachers, as one respondent said, are perhaps reflective of the increased competition and the higher standards, in recent years, demanded for admission to faculties of education in general and in Ontario specifically. In the main, the teachers seem to present themselves as demanding high standards of the system and of themselves, and their criticisms and aversions seem to focus on their disappointment with the lack of quality and consequence in their schools’ evaluation systems and not on any desire to evade evaluation.

**Analysis of Principals’ Views**

By and large, it seems that principals in this study feel that the growth aspect is the main reason for the evaluation process. What this suggests, even *prima facie*, may be misleading. As will be discussed later, some of these same principals found their training as evaluators inadequate; they see themselves as entering into evaluation situations almost totally unprepared to conduct as proficient a professional evaluation as they would have liked had their preparation as evaluators been more thorough. This self-acknowledged lack of expertise gives rise to the question whether some of the evaluations are being handled in a non-professional manner or at least not being done at the highest professional level, and if so, can the process still provide the professional growth teachers require?

This researcher is skeptical of the overwhelming, almost unanimous support for professional growth by teachers and principals alike and, the contrast with reality, if the
professional literature on the subject is to be believed (Bridges, 1986; Coker & Medley, 1987; Freiberg, 1985; Housego, 1989; ). There, the apparent absence of growth as a result of evaluation conducted under the observation model with the principal as observer is largely seen as the rule and not as the exception.

Should this provocation have merit, that is, should this researcher be correct in interpreting the responses and subsequent dialogue in light of relevant contextual connotations, then these data might genuinely reflect the respondents’ views. If this is so, then the apparent dichotomy between the principals’ actual claims and practices as reported by some teachers and documented generally in the literature and even more clearly in the immediately following section may still be valid. Principals here may still be more interested in administrative, legal or even political expectations than in actually ensuring that the evaluative process is beneficial, rigorous or capable of inspiring teachers toward greater professional growth.

The Worst Aspects of Teacher Evaluation

Analysis of Teachers’ View

While these opinions represent the majority of teachers in the study, it cannot be assumed that the entire process of evaluation is meaningless to them, as the teachers themselves have found credible aspects to it. Nevertheless, a persistent and consistent tone in the teachers’ remarks seems to indicate a marked dissatisfaction with their
evaluation processes. Perhaps the most pervasive feeling throughout this section of the interview was an apparent sense of exasperation with their administrations’ perceived insincerity or fraudulent sincerity. Teachers seemed resentful of their having to display various unauthentic behaviors for an administrator’s benefit, particularly for an occasion which, in the final analysis, essentially has no real effect on their careers, public accountability or professional growth. In essence, then, at the risk of some hyperbole, it would in all likelihood be correct to summarize the teachers’ views by suggesting that the teachers seem to feel that they were only obligated to be evaluated to help the administrators fulfil their organization’s mandate for this requirement.

These inferences would appear to suggest that one of the major defects within this process, as they experienced it, at least from the teachers’ perspective — is how inconsequential the whole evaluation procedure seems to have become. Therefore, a perception of evaluation as merely a bureaucratic necessity, but not one beneficial for professional development, is steadily recurring within our data.

This attitude can be traced to the way one teacher described her frustration with the entire experience. Other teachers, she noticed, got their permanent contracts “even though their evaluations, throughout their probationary contract, were less than satisfactory.”

Still another teacher argued that the process was often almost a parody of what evaluation was meant to be. When, at the conclusion of an exchange, wherein the
respondent had been adamant with her dissatisfaction over the evaluation process, this teacher was asked to confirm whether or not she felt the evaluation might be something of a joke, she replied, “Yes, at times.” She then added that despite her opinion that administrators themselves don’t take it seriously, “they [nevertheless] built it up to be such a serious affair.”

*Interviewer:* Is there opportunity [then] for abuse of the system?

*Respondent:* Of the evaluation? I think it’s abused right now, in that it [evaluation] is something that is stated that [it] needs to be done . . . The whole evaluation procedure is just a formality and then that’s it -- it’s done. The evaluation is done; you never go back [to it].

The respondent here seems to be suggesting that an abuse exists because of the mechanical, ritualistic nature of the evaluation process. Essentially, she implied it is an abuse because it is used only as a bureaucratic mechanism that has no practical application. If this outlook is widespread, it may underscore the point made above: that the principals in the study appear more attentive to their administrative routines or micro-political agendas than they are to actually ensuring that the evaluative processes they administer are mutually beneficial, rigorous or capable of advancing teachers’ growth as professional educators.

An experienced teacher, recalling a personal experience, seems to support this view. On one occasion, she was called into the principal’s office to discuss her evaluation. After a few minutes, the principal suddenly realized that he was talking to the wrong teacher. Then, apparently as a face-saving gesture, the principal accused her
of trying to make a fool out of him. On the basis of considerably more such evidence of mismanagement, incompetence and apathy towards responsibility (allegedly including drunkenness and sexual harassment) reported by herself and other teachers, the board is said to have put the principal on review. Nevertheless, he was said to have remained as principal for almost seven years, but reportedly his behaviour did not improve.

In this connection, a different teacher was even more blunt: “A lot of people don’t find them [evaluations] very fruitful. And the feedback? It’s just another part of the bureaucracy.” Another teacher was equally terse and disparaging: “It’s a nuisance. It’s a hoop you know you have to jump through in your first couple of years of teaching and then every so many years, but it doesn’t have benefit to me as a teacher.”

A third teacher was just as convincing of the process’s futility in saying the following:

_Interviewer:_ Let me paraphrase. You’re saying that evaluation won’t really expose the incompetent?

_Respondent:_ To whom? Who reads these? I don’t know anyone who has read these.

_Interviewer:_ Well, they go somewhere.

_Respondent:_ They go into a file. I don’t know where they go after that. They go into a file folder and they’re kept in the principal’s office and I don’t know if they go anywhere after that at all.

_Interviewer:_ Are you saying that they are rather irrelevant in general?

_Respondent:_ All together, yes. Definitely.
Thus, for teachers to appreciate the process and to take it more seriously, it seems this fatalistic attitude would need to change.

**Analysis of Principals’ Views**

Many of the principals’ complaints, although seemingly different from the teachers’ dissatisfaction, seemed to nevertheless share some common ground regarding those grievances. The absence of sufficient time to document teachers properly, claimed by some of the principals, seems to correlate with reports cited earlier (Hickcox, 1988) and those throughout the literature.

Nevertheless, this researcher must admit some surprise in learning that so many of this study’s principals found the attitude of teachers toward any negative reports to be a major problem. This “problem” is perhaps even more surprising when one takes into consideration all the claims to the contrary and the emphasis lately being placed on teacher professionalism by unions, Ontario’s newly founded College of Teachers and even by teachers themselves who profess a strong and sincere desire for constructive criticism.

Despite this being a qualitative study based on a small population, in our judgement the aforementioned attitude of belligerence might suggest a basic difficulty of managerial training and professionalism for at least some portion of the teaching profession. This, of course, would have to be supported by other statistically valid
studies. There may be a nearly pervasive culture in the several different school boards represented in this research that acts to belie the professionalism that the teachers and administrators themselves claim to seek and which the administrators often already appear to possess. Several other characteristics of this possible deficiency, not yet alluded to here, may be hypothesized. The points immediately below may seem initially tangential, and not emanating from my data, but they do bear on one of this project's central hypotheses of provocation: the mismanagement of evaluation due to lack of professionalism. These points, then, are being brought in, as per the previously outlined technique of logical analysis, to support an implication in the data. In essence, then, these pieces of evidence in the following paragraphs should be conceptualized as elements of a possible solution to the problem just presented, drawn from my interpretation of implications within my data and the basis for possible future study.

One general and recurring theme I find in my data is the lack of professional respect accorded to teachers. Some anecdotal evidence, presented as folklore, to support that data, should make the reader aware of this. The fact remains that other than teaching, there is probably almost no other profession that does not assign separate offices to its professionals. If a very rough analogy were to be made to certain other professions, the classroom can be seen to have the same relationship to the teacher as the courtroom, the operating room and the university lecture hall to the lawyer, surgeon or professor, respectively. The symbolic importance of each of these other three
professionals having their own offices and telephones in addition to their more public 
professional forum should not be totally dismissed out of hand as negligible. Continuing 
this point, it should be noted that the contrast between teaching and other professions is 
even greater because department heads or curriculum leaders (in the K-8 schools) have, 
in general, not crossed this demarcation line between management (who may be treated 
more professionally than rank and file teachers) and labor (teachers who, although 
professionals in name, are generally not treated in as respectful or dignified a manner, 
by being assigned their own private workplaces). This is true despite their positions 
being quasi-managerial.

Several other conditions, which support our data here, further appear to 
exacerbate the entire structure of the working conditions for teachers and continuously 
contradict teachers’ professional judgements. The lack of authority in their own 
classroom, which manifests itself when teachers are forced to send an unruly student to 
the principal’s office (because teachers do not have the authority or power to handle the 
situation), is a fair example of the system inherently questioning a teacher’s 
professional judgement. Another illustration of this attitude might be found in the 
practice of assigning para-professional duties to teachers (such as hall supervision, 
lunchroom duties, book stacking and so forth). Evaluating teachers in front of their 
students further demeans their professional status. With these points, we are raising the
possibility that it may be difficult to act professionally if one is continuously treated unprofessionally by management.

A further undercutting of teachers' professional status might have its provenance in the substance of principal training. As will be elaborated on later, and has already been alluded to, significant deficiencies might exist in the formal instruction that principals receive to prepare them for conducting an evaluation. As one principal, speaking for her colleagues, said:

What they [instructors of teacher supervision in the Principal's Qualification Program] did was mostly looking at models of evaluation as opposed to really looking at how you approach the process with teachers and so forth. So it was more academic ....

Maybe you can only learn on the job. But I would have liked to have asked some of these questions you are asking... [Like] what do you do in those cases ... what happens if a teacher walks out or something?

[In our service programs] the whole concept of evaluation is mentioned and there are some things done about evaluation but it's not very specific or subject related .... I feel it needs to be improved.

Another principal allowed that there is probably not as much [specific training in evaluation] as there should be: “You need to go back and revise it.”

Influence of the Collective Agreement

Analysis of Teachers' Views

When examining the views of teachers, one should constantly bear in mind that the natural expectations are that their viewpoints would be supportive of any union
restrictions on derogatory evaluations. Ostensibly, any organizational structure that potentially would be instrumental in preserving a teacher's job or in shifting power away from management towards teachers would be expected to be given a favourable review by teachers in general. Yet half of the teachers in this study either felt ambivalent ([the contract] "doesn't really deter a good evaluation") or believed strongly that, because administrators have to be so careful about what they can and cannot do or say, the written evaluation often becomes vague to the point of being "meaningless."

Still, one must remember, the contracts that teachers sign with their boards have usually not been textually identical to their collective agreements. By talking with their union representatives for specific details, the possibility exists that teachers may merely catch a general impression about the collective agreement.

It is, therefore, probably reasonably safe to assume that, as teachers estimate about their teaching peers here, relatively few teachers have read their entire collective agreement in detail. Thus one teacher, for instance, believed that the question concerning the collective agreement was beyond her expertise. That teacher held the opinion that she simply was "not familiar enough with the collective agreement" and, therefore, suggested that my question might be too abstruse.
Analysis of Principals' Views

These negative views on the relationship between the collective agreements and the evaluation process, declaring contracts detrimental to the summative aspect of evaluation, might be used eventually, if proven in a statistically valid study, to support an impression that elements of the general public may have about the negotiating skill and position of the school boards. This opinion suggests that the boards, operating under the province's decentralized bargaining, may have been too lax in negotiating collective agreements with the various teachers' unions. It is even assumed by the majority (four) of the principals interviewed in this study that these collective agreements could serve to maintain the certification or tenure of an incompetent teacher. In any event, this evidently is the opinion of half the teachers and the majority of principals in this study. It should be stressed in this project, and this will be done as the rest of this study unfolds, that this impression was even confirmed by half of the teachers who also felt for the same reasons that their collective agreement was detrimental to effective evaluation.

Any positive facet sifted from the data of these interviews would seem to be derived from the professionalism still remaining in the teachers' weltanschauung. It was at least somewhat surprising to this researcher to find that half the teachers surveyed felt the collective agreement was detrimental to effective evaluation. If that is so, one wonders why boards of education have not built upon that consensus between principals
and teachers to pressure the unions in a more inexorable fashion to obtain concessions in the collective agreement and remove some of those impediments to effective or more effective evaluation. As quoted previously in the data section, one principal said, “In the province of Ontario I would think if you researched the frequency of dismissals, you would find very few people have ever been fired. In fact, that was brought up at a superintendent’s course I took.”

From the immediate previous section on the data and my analysis of principals’ and teachers’ views on the collective agreement, it seems there exists the following common theme among both these groups in the study: the collective agreement is restrictive and detrimental to good evaluation in general and the summative aspects specifically because it intrinsically bypasses certain fundamentals of the evaluation process. In essence, the interviewees perceive that the collective agreement strips the evaluation process of any power it has to effectively identify and deal with incompetence and ineffective teaching practices.

Arguably, then, the problems seem to be that any negative evaluations may have been ignored in (1) the granting of permanent contracts and (2) any dismissal hearing. Also, it may be inferred, that any low rating on an evaluation can be counteracted by the teacher’s corresponding right to register a complaint, which considerably detracts from the evaluation’s focus and considerably weakens its acceptance as a true description of the teacher’s ability.
As additional support to these views of the principals, there is data showing the difference between the harsher demands that teachers’ unions in Ontario put on the boards of education as opposed to unions in other professions (Downie, 1992). It then becomes particularly salient with respect to the limits set for management on dismissal requirements. I am also citing here some of these contractual restrictions directly from the various collective agreements between boards and federations:

(1) Consideration shall be given to visiting a variety of classes or teaching situations if so requested by the teacher (p.30 Collective Agreement of Toronto District School Board, 1998-2000).
(2) Where a teacher has an unsatisfactory Evaluation placed in the Teacher’s files under 3.4.5.0.0 the Teacher shall upon request have the right to an Evaluation by another evaluator. (P. 31, Collective Agreement, Toronto District School Board 1998-2000).
(3) The teacher shall have the right to respond in writing to the evaluation within five (5) school days of receipt of the evaluation. The teacher’s response shall be included in the teacher’s personnel file with the evaluations (p. 39, Collective Agreement, York Catholic District School Board 2000).
(4) At the request of either party, a meeting shall be held to discuss the evaluation. At such a meeting, the teacher is entitled to be accompanied by a representative of the Bargaining Unit (p. 35 Hamilton Wentworth School Board 1998-2000).

While some of these restrictions may be seen as necessary protection against abuses, they also tend to at least partially support the claims of the principals (and teachers) that the restrictive nature of the collective agreement stays the summative (and to some extent the formative) power of the evaluation. Such conditions as almost automatically having to visit a variety of classes, having the right to request alternative evaluations, being able to write a dismissive response and having the right to a representative present to discuss the evaluation might tend to make the evaluators very
cautious about how they administer a supervision session, what they wrote and what they say to the teachers.

This is how the principals and the majority of the teachers felt, and it is supported by the actual text of various collective agreements.

The Ontario College of Teachers (OCT)

Analysis of Teachers’ Views

The previous information from the data section, approving a role for the College of Teachers in Evaluation, raises the possibility of a demand among the teachers for more stringent standards in their evaluations. Yet even within this demand there seemed to exist, paradoxically, a subtle but formidable apprehension of what that request could mean.

Examples of this subtle sub-textual attitude are the following:

*Interviewer:* Do you think they [The College of Teachers] are going to be [involved in evaluation] more [at the level of] an advisory capacity [or in a more applied role]?

*Respondent:* I would hope that they would. But I assume since it is legislated they will have a little more say than that.

*Respondent:* I’ve heard it’s supposedly meant to be punitive -- that it’s a way for the public to get at teachers that they would like to get out....Unless you are really bad, they are going to make it positive because it’s in your personnel file.

*Respondent:* I would not like to see that [The College of Teachers have a more applied role] because then you are having outside evaluators who know nothing about you as a person, nothing about you as a teacher, have never seen your rapport with students [and] don’t know you.
To some degree, though, this is to be expected. After years of largely ineffective evaluations, teachers in this study appeared wary of their own organizations regarding this matter and, therefore, seemed increasingly anxious about more “consequential” evaluation because that evaluation might actually be punitive. This possibility, of course, raises the question of whether the teachers displayed a negative attitude with regard to evaluations because they feared it might expose them as academically weak or professionally incompetent.

It already has been well established through several sources that teachers are technically professionals by the accepted sociological definition. For instance, to buttress the definition that “teaching is professional work,” Weiss (as cited in Hoy & Miskel, 1996 p.328) asserts that “because the work teachers do is largely intellectual and requires the consistent application of judgement, [it] cannot be standardized or routinized and calls for prolonged preparation through education.” Therefore, by this definition it is professional.

Hoy and Miskel (1996), forming an abridgement of the literature on the criteria for professionals based on a wide variety of authors, have declared four criteria as essential in defining a professional as such: knowledge, regulation and control, ideology and association. Teachers have always had knowledge and ideology, and in Ontario, they have had association since the 1920s and the rights to collective bargaining since 1972. Presently the criterion of regulation and control is being fulfilled by the establishment of
the Ontario College of Teachers. This seems to leave little doubt then as to the professional status of teachers within the technical parameters of the definition.

Yet despite this technical interpretation, in this researcher's estimation of the data, as it pertain to information about the cultural environment in the schools, there seems to be an embedded administrative attitude, possibly held over from the days when teachers were largely uneducated by university standards, which regards teachers as low paid and untrained para-professionals who do not merit professional treatment. As mentioned earlier, this seems to manifest itself by denying teachers offices, adequate authority to do their job and insisting that they perform para-professional duties. This perspective also seems to exist as a dialectic within the teachers' own conceptual interpretations of what it means to be professional; that dialectic conceivably may influence their receptiveness to accept certain responsibilities.

A professional, like those in other specialties, then, when extrapolating beyond the literal definition into the connotation of professional, that is, professionalization, also means one should be treated like and act like a professional in all contexts of their duties, with respect to both the obligations and the privileges that this term implies (Herman & Herman, 1993). The concepts of ideology and service for the general good, often disregarding any personal gain and absorbing personal loss, are seen as directing the decisions of the professional and also defining their status as professionals. Some elements of that "professional concept" are already present in the oaths and allegiances of the other
more firmly established professions such as medicine or law, demonstrating the concept of public service as also indigenous to the concept and definition of professional.

It is chiefly this latter connotation that is used here as one of the criteria to determine the level of professionalism in the school culture and as a criticism of teachers for unprofessional behavior when they fail to fulfill that criterion. This idea, which de-emphasizes the other more promulgated but literal definition of professional, is actually an aspiration aimed for, in order to professionalize teaching and its affiliated organizations. It is strongly supported by the following criteria for teacher professionalism established by Darling-Hammond (1989):

(1) knowledge in the service of the satisfaction of client needs;
(2) concerns for the benefit of the client; and
(3) responsibility for correct standards, practices and ethics as assumed by the various organizations concerned with the criteria of excellence.

It seems to this researcher, on the balance of the logic and evidence presented so far, however, that despite their reservations, the teachers in this study appear to feel that teachers, as a general entity, may get a “better deal,” with respect to the potential to fulfill their professional mandate to the public, if a more professional, regulatory body like the College of Teachers assumes responsibility for all evaluations. Taken comprehensively, it seems that at least with regards to the question of the College of Teachers, the teachers in this study are, for the most part, trying to put forth a professional demeanor. Yet the
recurrence of this type of attitude that might be characterized as self-serving will be noted again, whenever it does recur, to show how it counteracts the alleged professional responsibilities and the claims and purported ideologies of teachers.

It should be acknowledged here that there is no unbiased way of gauging different shades of commitment or any solid comparisons of the levels of commitment of teachers as compared to doctors, lawyers or engineers or to definitively outline when an action legitimately represents a justifiable self interest and when it becomes self-serving. Nevertheless, in the judgement of this researcher, there are times when this line is crossed by teachers and principals here, who seem to be abandoning their professional responsibilities in favour of personal advantage and when this occurs as a systemic issue, seemingly as a result of ineffective or dysfunctional policies, it will be illustrated from the text of the interviews and left to the judgement of the reader whether the point is valid or not. Therefore, in order to make this point, regardless of the teachers' attitude here, and at the risk of skirting some dimensions, this is the aspect of the definition of professionals that I will focus on, based within the parameters of Darling-Hammond's definition immediately above: any member of a paid consortium who has technical knowledge, regulation and control, ideology and association as well as dedication towards fulfilling the goals of their mandate and responsibilities, regardless, within reasonable means, of the sacrifice or loss invoked and despite any conflicting personal reservations, opinions or ideologies.
Analysis of Principals’ Views

Perhaps one of the most interesting discoveries in this study was the strong and almost unanimous support that the College of Teachers received from the interviewed principals, where five out of the total six were supportive. Their reasons were generally centered on the expectation that the College of Teachers would help raise professionalism. This also co-substantiates some of this researcher’s scepticism about the extant level of professionalism presently in teacher evaluation. It seems that this concern with professionalism is what the principals were responding to when they said that a College would raise professional standards. Yet even within that attitude, one can start to detect implicit traces of a tacit, if not almost subliminal, begrudging of the inevitable loss of power that would result: the principals, after all, might no longer be in charge of the evaluation of their own staffs, nor would they be in control of the process and its consequences. This impression is fortified further by citing, in order of strength of feeling, the following portions from this study’s interviews of the principals:

Interviewer: Do you think it (College of Teachers) might take over evaluation completely?

Respondent: Never! [the implication is that this respondent does not want or would be against such a move].

They may [take over evaluation] and that scares me a bit because they are not all teachers in the College of Teachers.

No, [The College won’t take over evaluation]. It’s never been suggested in any of the literature they [the College] would do that [that is, evaluate]. The administrators will continue with that particular role...that will still remain in the hands of the administrators in the building because they are ultimately responsible.
I think the College of Teachers has been something that hasn’t had a lot of support with my colleagues. It seems to negate that [our work in evaluation] by having that body overlooking us.

The accountability that is already built-in is one that we work very hard at, and it seems to negate that by having this body overlooking us."

They [staff at the College] will probably end up training the evaluators, but these will be principals and vice principals who are not members of the College, and that seems to be an impediment to the College of Teachers [itself doing the] evaluating.

To be sure, these quotes do not demonstrate a stern begrudging sense of resentment among the six principals toward the College with respect to any jurisdiction it might exercise over teacher evaluation, but they do point towards these administrators’ wariness.

Objective Data:

Using Students’ Standardized Exams as a Criterion to Evaluate Teachers

Analysis of Teachers’ View

First, it should be noted that it is an accurate perception that much of the literature, (as detailed on page 121 of the Literature Review), supports the validity of the teachers’ opposition to the format of evaluation which advocates the use of students’ standardized exams to evaluate teachers (Berk, 1988; Darling-Hammond, 1984; Medley, 1973; Millman, 1981; Ross, 1984;). In the judgement of this researcher and despite the acceptance by a large segment of the literature, the teachers in this study (as most of the literature itself does) offer, as will be shown in Appendix A following, only weak arguments against such a results-oriented evaluation.
The point that should be taken here in this analysis is that only one teacher was willing to grant any credibility at all to using student scores as a criterion to evaluate teachers. The fact that they were so adamant indicates to this researcher and perhaps should to the reader as well, just how well the cultural values of the education system have been inculcated on the teachers.

**Analysis of Principals’ Views**

While it was really no surprise to this researcher that all the principals were, to varying degrees, against using standardized testing of students, that is, in the form of a student score or average to evaluate teachers (although one principal recanted partially) what was interesting was the intensity of feeling or even the vehemence with which the principals demonstrated that opinion. Even less surprising, however, were some of their supporting arguments. In Appendix A, I repeat some of the arguments, but there the arguments are made to principals’ themes. (The reader who is not interested in this quasi-recapitulation, with a principal’s focus, should omit that part of Appendix A.) For now suffice to say there is a fierce resistance against using student scores, in any format, as a criterion for teacher evaluation.
Formative and Summative Aspects of Teacher Evaluation

Analysis of Teachers' Views

Initially it would seem, for the most part, that in their entirety the teachers' views represent a very positive view of teacher evaluation. Subjected to further analysis, however, teachers' comments do not seem quite so approving. While some of the comments were a blanket approbation of the system's formative component, many of the other affirmations of that same component were actually quite limited in their approval or only based on limited experience. This guarded endorsement tinged with reluctance can be seen, for instance, in the following pair of comments: "The formative component of the evaluations that I've had has been the largest part" and "The one I've had, which is the only one I've seen, is very much descriptive of what actually gets done." Furthermore,

The evaluators will be in your room regularly as a teacher and if there is something that needs to be addressed, they will talk to you about it. If you have set a plan for professional growth with the aid of somebody or just on your own, the natural thing would be that you follow whatever steps you were going to follow to achieve whatever goal.

Even comments that seemed to wholeheartedly confirm the existence of formative aspects also added reservations in the form of stipulations like the following: "Principals tend to back away from [even] making suggestions for growth because of the unions and the [resulting] collective agreements, and the extent of the formative component," "[the amount of the formative component] depends on the style of the evaluation" and the formative dimension "was just a small [part] at the end where [there were] recommendations...."
These reflections may even suggest that only a minor overall role for the highly touted formative component presently exists for the study group, one that is only ancillary to the summative component, at least to some degree. In the case of the particular teachers in this study, there does sometimes seem to be too much done in the field of teacher evaluation just to comply with policy demands and much less done to actually fulfil a formative objective. What is presented here seems to be a stellar example of something that superficially appears to actually correct a problem and improve teaching, but when that patina of apparent functionality is removed, a different image seems to be manifest. For the most part, evaluation, as it presently exists for these twelve educators, might turn out to be something which has less practical formative value than initially thought.

This flaw seems to feed naturally into the next part of the same question, in the interview, whether the formative and summative components are antagonistic. Any artificiality in the formative component may be due to the infrastructure of the evaluation being summative in nature. If this is so, the contradiction becomes more blatant. This is perhaps why a majority of the teachers might have seen the duality inherently present in their evaluation process as antagonistic and not synergistic. As far as could be detected the teachers seemed to affirm that the predominant role of the summative component was to dichotomize the evaluation processes and even sometimes limit the scope of the formative sections. Therefore, they may have recognized this situation as a contradiction between the
two components and, for the most part, regarded it as, if not incompatible, at least incongruent.

Analysis of Principals' Views

As mentioned in the previous paragraph, some interesting discoveries can be discerned in the principals' answers. While the majority of views given by the principals negating any discrepancy between summative and formative evaluation seemed, in the judgement of this researcher, somewhat unlikely, some of the responses are intriguing enough to merit discussion.

One of the principals, commenting on the apparent contradiction in the nature of formative vs. summative evaluation, stated that “Both are probably critical because if there was no accountability aspect to evaluation [no possibility of negative consequences resulting], it probably would not be as important to the teachers.” As the following comments reveal, there seems to be general agreement on this point, although, two of the principals were more equivocal on this point:

I believe there are dual tracks there. They both have very different objectives. I think they can be done at parallel times. I don’t think they are mutually exclusive of one another....I think they both can be done for very different purposes.

The same as I explain to parents, teachers are not there to sabotage, and deliberately ruin your child’s life. That’s not why people get into teaching. And I think teachers have to believe that that’s not why I’m there as well -- I’m not here to go at them, but if they are doing something wrong, I would like to help them make it right.
Two of the other principals adopted a firmer stance by claiming not to see the contradiction between the two philosophies of evaluation, and they almost had to be interrogated to break down their intransigence. An example of that type of inveterate resistance is the following exchange:

*Respondent:* I don't see any contradiction [between summative and formative evaluation] at all.

*Interviewer:* The argument is that the teacher may not be as forthcoming if it's [that is, the evaluation] being used as accountability [in a summative manner].

*Respondent:* Oh sure. And there's a lot of that where teachers play the game just to get through the performance review.

*Interviewer:* On the surface of it, someone might wonder if there is any contradiction in the purposes to this process. How do you see it? Take your time in responding.

*Respondent:* Yes I guess when you explain it further I do see where that can be a contradiction in terms to some degree [vocal emphasis in conversation].

The point being made by the principals is revealing: it suggests that a much smaller percentage of the teachers would put in a full effort under purely formative conditions. That answer appears problematic. Firstly, it belies much of what this study’s principals and teachers have already said about dedicated professionals who want and demand growth opportunities. Secondly, two independent evaluations could co-exist, one mandatory, just concerned with incompetence and minimum standards of professionalism and one that voluntary, would have to be requested (perhaps available only on the recommendation of a principal, department head or professional organization) for career development, thus avoiding any ideological conflict and its concomitant problem. If the latter was requested
and completed satisfactorily, there could be, in this variant paradigm of evaluation, substantial remuneration or promotions. Although some might claim that a process so defined by reward might push this evaluation into a summative category, in the final analysis this researcher believes that the voluntary nature of the evaluation, coupled with the absence of any formal requirement of a public or archival record that could prove detrimental, would make the valuation more, if not completely, formative rather than summative and could, therefore, improve the evaluation process. While this is not even meant to be a definitive solution, as it is as yet untested, there are also other solutions, ones (including this one) that experienced administrators, principles or teachers may critique if they were presented to them properly.

Again, it seems that the underlying assumptions attributed to the evaluation system by the study’s teachers and principals are largely similar. Both seem to believe they are implicated in a system that has allowed so many bureaucratic strictures and obstacles to proliferate that the intended purpose of the process somehow seems lost.

Up to a point, the existence of so many strictures and bureaucratic obstacles in the systems they are familiar with has been verified by both teachers’ and principals’ repeated references to their presence. The relative length of time and complexity of the intricacies involved in the dismissal of a teacher as compared to the dismissal of a business or industry professional are highly incongruent in favour of business as the positive example,
and this is openly admitted by the professional people on both sides of this study. As one principal said:

[An incompetent teacher] would not be dismissed as quickly as somebody in industry. But I think there are steps in between. I think you can remove an incompetent teacher from the classroom yet still maintain the teacher and provide a year or six months additional training.

Another principal, a staunch defender of existing evaluation and educational systems, described the situation as being encumbered with “all kinds of different legislations that are tied to the teaching profession as well [and] that are not tied to business….The teaching profession is regulated by different Acts which constrain [us] in terms of how you [can] get rid of somebody.”

Contrary to what one might expect, many of the teachers felt just as strongly as the principals about the bureaucratic structures and the debilitating rules they enforce as necessary to dismiss someone legitimately as being too restrictive, time-wasting, expensive and responsible for inadvertently carrying teachers who were “dead weight” and who probably should have been dismissed long ago. Some of the teachers’ comments about collective agreements expressed these reservations by emphasizing, for instance, that administrators are “very aware of the collective agreement and the last thing they want is to get entangled in a dispute. So I would think that they [principals] are very careful in how they phrase anything on paper.”

*Interviewer:* And do you think that deters effective evaluation?  
*Respondent:* Yes I guess it does. Yes, I would say so.  
*Interviewer:* And why is this?
Respondent: I think the administrator simply knows how incredibly difficult it is if you try and label a teacher incompetent and you record things. The process is so long that it’s not worth the administrator’s time or effort.

Complementing these opinions, a different teacher asserted that since “teachers are [members of] a profession [that is, consider themselves responsible enough to correct their own mistakes and independently enforce high standards of practice and have even had this independence protected], I don’t know right now whether a vice-principal could make someone improve if they [that is, the principal or vice-principal] wanted to or not”; that judgement on the uncertainty of the evaluation’s impact seems to belie any effective summative consequence in our study. Emphasizing this point, a colleague was even more insistent that if any charge of abuse exists, it should clearly be laid against the unions without which “there wouldn’t be some teachers there now that are [teaching there].”

An even more intimidating deterrent to effective evaluation seems to be the apparent lack of judgment within the organization in not compensating for these restrictions by supplying the practical situational responses necessary for effective evaluation. The possibility even exists that the executives within the education system begin to believe their own publications about the widespread efficacy of the teaching force. This researcher realizes, however, that with that type of unequivocal comment the possibility exists for exposing this report to harsh criticism from within the field of education. Nevertheless, since this type of opposition to institutionalized incompetence is common to principals and teachers and the boards’ failure to exploit it is encountered repeatedly in our study, at least
tentatively this researcher might have the right to offer that as a possible hypothesis to be tested. Let us, however, leave the option open to change that opinion if the evidence merits it at a later date.

**Appropriate and Adaptive Aspects of Evaluation**

**Analysis of Teachers’ Views**

Since the opinions of the respondents were unanimous (and almost categorically so), this study can suggest, based on the responses of its subjects, that teacher evaluation, within the small data base of this study, is not dichotomized along experiential lines. Therefore, generally, a different evaluation process is not applied contingent on these teacher’s amounts of experience. Little verbal backing is given by the respondents to substantiate the importance teacher experience. If there is actually so little awareness of the importance of teacher experience, the entire process might seem to be flawed. Extrapolating from our data, therefore, this researcher wonders whether this dynamic seems to represent the type of systemic error that only a mechanical and unresponsive evaluation system would tolerate. In some ways, using one form of evaluation regardless of the experience of the teacher is tantamount to assuming that experience as an irrelevant factor, reducing everyone to the most primitive stage of professional development and despite the contradictory fact that, in Ontario, pay scale increases are based almost uniquely on experience. Thus, if experience were irrelevant, then the boards of education,
as the representative of the government educational enterprise, have been basing the pay differential on an irrelevant factor.

Although one teacher did mention, in an oblique manner, that higher standards were used for experienced teachers, she only claimed that there are "higher expectations," despite the evaluated items being the same on the agenda. Again, this researcher believes that this kind of statement, by itself, would appear to miss the point or at least our construction of the impetus of evaluation. To base the evaluation of a teacher with ten years of experience on a form that includes basic requirements like classroom control as a criterion may well be counterproductive. Although certain circumstances may occasionally work against the veteran teacher, the claim here is that classroom control is largely a function of experience with a class of students and presumably after ten years, skill in this domain should have almost certainly been adequately learned. To have important informative value, the evaluation should seek to discover more subtle and sophisticated measures of good teaching as experience accumulates and leave the basic categories of evaluation for the less experienced teachers. This is of course in addition to having an evaluation system that is adaptable enough to change along with a milieu or school culture that does not remain the same.

It was also somewhat unexpected to discover that, according to half of the teachers in this study, resource teachers and guidance teachers are being evaluated in the same manner as classroom teachers. This might suggest that these types of evaluation must be
using at least some irrelevant categories to evaluate certain very integral members of the staff. It seems much more reasonable to this researcher that an entirely different process of evaluation is required in those cases of staff not primarily involved with regular academic classroom duties. More research would have to be done, however, to prove or disprove this proposition.

Analysis of Principals' Views

A majority, four out of the six, of these principals surveyed claimed to evaluate all teaching personnel in the same manner despite the lack of any official classroom duties by some of them. Under those conditions, then, it would seem difficult, if not impossible, to evaluate efficiently that segment of the staff not involved in traditional classroom teaching because of the duties they undertake. Although one principal openly stated that it might be wise to design different forms of evaluation for these different roles, none of the other principals seem to be bothered by this apparent discrepancy. In the judgement of this researcher, this seemingly nonchalant attitude in this study reflects in many ways an inconsequential attitude toward evaluation here. In the final analysis, and in the event that such a cavalier attitude is found to characterize whole groups of teachers in whole groups of schools, it is not too much to suggest that it may be necessary, if evaluation is ever to be taken truly seriously, that the evaluators and those designing the evaluations may need to become more sensitive to role and to innovation. Essentially they need to adapt their
evaluation systems so that they can be usefully and diversely applied especially for purposes of growth.

This same logical concern should apply to the evaluation of classroom teachers themselves. Should there not exist a different format for experienced teachers? Since Ontario teachers are rewarded financially in proportion to their years of experience and education, it seems that a reasonable conclusion would be that they have gained something due to that experience. Therefore should basic teaching skills still be used as criteria in the evaluation process of experienced teachers? Therefore, in the view of this researcher, attributes related to experience like classroom control should be eliminated or, at the very least, should be observed casually, when evaluating experienced teachers, in order to concentrate on the more advanced, sophisticated aspects of teacher evaluation. Perhaps evaluation will become more accurate and therefore more valuable as a management tool if the time allotted for it were used more efficiently. Naturally, though, other studies with larger, more diverse populations would have to be used to establish the veracity of this principle beyond the narrow range of this exploratory study.

**Validity of Observations in Evaluations**

**Analysis of Teachers’ Views**

Four of the teachers in this study do not believe that the observations on which the evaluations are based are true reflections of what goes on in their classrooms. Even the two
who do, appear quite circumspect about the extent to which they feel it does successfully capture authentic teaching practices. Again, this prompts us to wonder, if this proves to be the case universally, whether this study could be instrumental in showing that the principals might be ignoring or (possibly) not even giving the teachers any effective voice in determining the methods and criteria of the evaluation process. The teachers, for the most part, agree that seeing the evaluation forms prior to the evaluation might distort the veracity of the evaluation to some degree by having the observations and subsequent evaluation influenced enough to be possibly based on staged situations.

Although it could be argued that seeing the criteria gives focus to the teacher, after hearing justifications for such viewpoints I am given pause. Nevertheless, this researcher sees that as a somewhat specious claim. Overall, the evaluation is not there to focus the attention of the teacher for the short period of time defined by the immediate act of the evaluation, but to see, on average, how the teacher is usually accustomed to teaching over the length of their entire tenure. In my view, there would be nothing wrong with the teacher, on their own accord, preparing a list of techniques as an aid, as long as this was done independently and reflected that teacher’s conventional practices. Otherwise, any external prompting or coaching probably biases the true nature of the evaluative teaching episode and tends to prejudice the observations and result.

The whole idea of prompting the teachers, besides being counter-productive to the evaluation process, seems extraneous to this researcher. After all, as a graduate
professional teacher, the teacher should know what constitutes good teaching; how a good evaluation would be defined; and how to perform at that level consistently. If this is correct, and no other variables intervene in particular situations, then viewing the criteria for evaluation prior to the actual evaluation may well be extraneous. I am not suggesting, however, that teacher evaluation, either in its formative or summative format, be limited or abolished. I am, however, stating that in order to have evaluation function effectively and accurately as an efficient, continuous process, it might have been prudent for teachers in this study to be examined in the performance of their duties in as natural and unaffected a habitat as possible. Otherwise there exists at least some possibility that the temptation, on the part of the teacher, will overcome them, and they will feign compliance momentarily, before reverting back to their regular routines.

Furthermore, if some of the teachers believe that they can “fake” their way through these evaluation sessions, as the data in this study suggests (pages 164-165), arguably then, there seems little chance that these teachers will consider their evaluations seriously enough to actually make any changes in their daily teaching strategies, even if they do so during the inspection. This type of affected behaviour reflects incidents that occurred repeatedly and were recounted in the professional literature (Bridges, 1986). The pattern repeated here is described in some of the earlier interviews, that is, where principals’ declarations and implications were being denied by the teachers. If this situation questioning the validity of the observations is just another aspect of that phenomenon, and this proves to be the
case in a broader more universal study then the literature on evaluation of the late 1980s and early 1990s, the same academic literature that condemns teacher evaluation as ineffective, might then still be valid. Certainly information such as that given above by the teachers seems to negate those types of favourable opinions repeatedly gathered from principals, claiming that the evaluation situation has improved radically. While any number of scenarios are possible, two of those are that the principals in this study may be apathetically repeating whatever the educational bureaucracy expects them to repeat or are in denial on this point.

**Analysis of Principals’ Views**

If these principals do make several other visitations before a formal visitation, the admittedly relatively shorter period of time they spend in the classroom on those occasions almost certainly does not compare to the formal visitation where they are supposed to be looking for certain specific qualities. Possibly, not many reports would reflect such unannounced visitations because they were only random, brief visitations as opposed to their official formal visitations. If what they say about the contribution of their combined (that is, formal and informal) observations proves to be accurate, their data could be misleading. As far as is known, these observations usually are not recorded in any formal way. Therefore, the chances of these informal impressions of teachers’ classroom methodologies being remembered accurately and then used properly in a manner that
successfully integrates them later with the formal visitation are (in the estimation of this researcher) slight. Furthermore, as mentioned earlier, these informal observations merely create a "general impression," which is detrimental and, essentially, accentuates the "halo effect."

Yet even more disturbing to this researcher, unfortunately, were, and I can use no other word, the rationalizations that the principals kept using to justify entrenched management policy. To be sure, they may have felt that they were obliged to put on a brave "front" for their boards, and that may have been done just in this study alone. It seems rare that careful thought or scrutiny of an accepted practice is employed by the principals alone. A typical example of this in this section is the principals’ favourable response to the question of why the teachers are allowed to see their evaluation forms prior to evaluation -- a widely accepted and approved practice (by the principals) but one criticized by the teachers -- and to this researcher one that is a highly illogical practice. While it is not possible to "get into the heads" of the study’s principals, careful thought or scrutiny of evaluations’ accepted practices do not seem to have been diligently espoused by the study’s principals.

Further rationalization for showing teachers the evaluation forms in advance of visitation included the following: "We want [them] to show what they’re capable of," "It makes them more comfortable with evaluation," "They still can use their personal teaching styles" and so forth. Yet the same principals in this study who fight inexorably against
evaluating teachers with the scores of students on standardized tests, which teachers do not see beforehand, because teachers then would be "teaching to the test," do not seem to mind their teachers teaching to an evaluation form which they do see beforehand. The comparison seems clear to this researcher. If "teaching to the test" limits the repertoire and scope of the subject matter and (according to the principals interviewed) teaches test-taking skills and not actual knowledge, then "teaching to an evaluation form" limits the scope of teacher testing, that is, evaluation, and teaching strategies and may only train the teachers in acting skills and compliance with a bureaucratic format. The principals remain adamant, however, that a clear distinction exists and the situations are not comparable.

Number of Visitations for Fair Evaluation

Analysis of Teachers' Views

Here, in this section of the interview, five out of six teachers openly stated that there are not enough visitations to get a fair and accurate assessment of the teacher and/or the teaching situation. Two of the teachers surveyed were in favour of unannounced visits because they felt this would give the evaluators a better understanding of what actually goes on in the classroom, and they anticipated deeper insight into their own professionalism and teaching methodology. Although this could be a risky proposition, instituting a general policy of unannounced visits could offer great rewards too. Nevertheless this researcher posits that there is little chance that this policy would receive
approval of the teacher federations because it would limit the control they now have over
the evaluations; an unlimited number of visits would also broaden the conditions under
which one could get dismissed, thus making the protection of their members more difficult.
In any event, it is refreshing to see an attitude among teachers which, if more widely
accepted, might help to change evaluation for the better.

**Analysis of Principals’ Views**

Teachers in this study come out “four-square” against accepting the present
system’s version of what an adequate number of visitations is, and they, with the
principals, are almost totally in agreement (only one principal felt that the two visitations
she made were adequate). This consequence leads us to puzzle about whether these
educators have pressed their boards or federations toward change. Again, the study here
keeps reinforcing the literature (which most principals fiercely claim is outdated) that
asserts evaluation is largely a mechanical process with little practical use, as it is now
employed.

The decision-makers at the management levels for these 12 educators may be
ignoring the viewpoints of other members of the workforce as well. Such blinders certainly
would not be a very strong sign of good management practices.
The Status of the Follow-Up Interview in the Evaluation Process

Analysis of Teachers’ Views

Here, to this researcher, it seems indicated that the principals in this study are not being adequately trained for the evaluation or are choosing to ignore that training or the possibility exists that the interviewees did not recall what they knew. As a case in point, after a fairly exhaustive series of probes during the interview in this section, the teachers questioned the integrity of the entire evaluation procedure, particularly concerning the veracity of the follow-up interview. Then again, if it is proven to be generally true that the principals are not being adequately trained or if the evaluation is just cursory, undertaken (it would appear) at least partly to satisfy organizational routines, aside from symbolically assuring the public that teachers are, at least to some extent, being monitored, one could ask what is the point of doing any evaluations at all?

In this instance, five of the six teachers stated that they see serious deficiencies in the follow-up interview. This conclusion follows the previous question, where most principals themselves believed that there were too few classroom visitations to conduct a fair evaluation. If this is generally so, administrators must be prepared to confront the possibility that the classroom observations are likely to be inaccurate and that the post-evaluation interviews do not seem to be conducted for any reason other than the requirements of board policy. Should the conditions described herein be pervasive, the likelihood of fair, accurate or consequential evaluation may not be strong.
Analysis of Principals’ Views

Considerable discrepancy seems to exist that differentiates among the two groups, principals and teachers, concerning the merit of the post-evaluation interview. To be blunt, most of the teachers’ attitudes (that is, five out of six), suggested that they viewed the post-evaluation interview as a useless, redundant exercise done more to comply with an organizational regulation and symbolic posturing than to provide anything of functional value for either teacher or principal. Essentially they seem to believe, for the most part, that the entire structure of the post-evaluation interview, because of the way it is presently employed, has been exaggerated when described as an effective evaluation component.

Most of the information distilled from the teachers’ opinions revealed that the post-evaluation interview took 5 to 15 minutes while at least one principal said that she met with the teacher for an hour. This can be illustrated by four of the teachers’ comments:

1. *Interviewer:* After the evaluation is over, do you sit down and talk about it? How long would it take?
   *Respondent:* 10 minutes, 15 minutes.

2. *Interviewer:* Is there a follow-up interview after evaluation?
   *Respondent:* Yes. I wouldn’t call it an interview but there is a follow-up meeting, maybe 5 minutes or so.

3. *Interviewer:* Is there a follow-up interview?
   *Respondent:* Yes, very brief. The first time I did [have such an interview]; the second time, I didn’t.

4. *Interviewer:* How long was the follow-up interview?
   *Respondent:* Probably 10 minutes, maybe.
   *Interviewer:* So the follow-up isn’t very extensive?
   *Respondent:* No. And not even positive [that is, of real value].
Yet some of the principals claimed that their interviews lasted for an hour. As this study’s teachers were most probably not evaluated by the interviewed principals, it must be acknowledged that both parties could be correct. Still, this researcher becomes increasingly skeptical when hearing that a principal remembers sitting down for a meeting with a school teacher to conduct a post-evaluation interview for as much as 60 minutes. In a typical suburban high school of 1500 to 2000 students, there would be about 40 to 50 teachers, which would easily take over two months just to do the post-evaluation (assuming that each teacher had been treated equally). In a job with so many external and internal demands, it is nearly incredulous that a principal would be able to spend that amount of their time just on the post-evaluation interview. This researcher is especially skeptical about these claims since, if the teachers are to be believed, little of real substance usually happens with these evaluations. If so, why would the principals spend so much time on them? Regardless, about half the principals suggested that the teachers were basically accurate, viz., the post-interview consists of a brief discussion about the content of the written evaluation report, and then the two parties sign it. Furthermore the principals and the teachers in this study seem to agree that almost none of the reports are negative. Thus it seems possible that one of the following conditions exist: (1) almost all the teachers in this study or those teachers whom the principals have ever dealt with are truly excellent (statistically unlikely), (2) the evaluators may be going through the process mostly for reasons of board requirements (perhaps to some extent), or (3) (more likely) the evaluation
is neither being done very professionally nor being taken very seriously. These and other possibilities warrant investigation.

**Evaluation and Professional Status: The Promotion System**

**Analysis of Teachers’ Views**

It seems reasonable to infer that this study is questioning, within its narrow range, whether the evaluations are serving the 12 educators in a very practical summative way, at least as regarding promotions. If for the most part, the teachers (and principals) in this study’s data agree that evaluations are not being taken seriously enough by the administrators and the evaluators to be a factor in their promotions, this implication may even demand a broader statistically-valid study to see if this situation is epidemic. If that subsequent broader study confirms this attitude and also establishes that the vast majority of these evaluations are positive regardless, again, it may well be asked, why is evaluation being conducted at all? Nevertheless, this exploratory study does tell us that the study’s six teachers and five out of six of the principals surveyed believed that evaluations are not used for purposes of promotion.

**Analysis of Principals’ Views**

Eight of the teachers and the principals in this study agree that the evaluations have little, if anything, to do with promotions. Indicative of this discrepancy is the
aforementioned remark from a principal opposing a link between evaluation and promotion “because you have one evaluator writing one report [based on] a couple of situations they saw.” Even if that is her real opinion, as mentioned previously, this is hardly a “ringing endorsement” of that evaluation system and certainly contrasts with the general outlook of principals, including her own previous stance on teacher evaluation certifying it as generally effective. Indeed, frequently, when it comes to discussing consequential services in the educational system, the evaluations, which are publicly defended by those who work professionally in educational management as a valuable and an integral part of the educational system, are privately dismissed as inconsequential, unreliable and cursory. This “confidential” opinion would seem to disqualify evaluations as criteria for decisions pertaining to professional advancement. That the teachers, who may not really in a position to know the intricate details of how promotions work, appear to sense this anyway, is even more disturbing. What could be more extraneous than a system of evaluations that both the evaluators and evaluatees tacitly know is, for the most part, insignificant regarding promotions and not used much for either accountability, summative or formative purposes, as illustrated by the data collected? Although it presently remains inconclusive how consistently these perceptions about evaluations exist throughout the rest of the Ontario school system, this qualitative study raises the possibility that the inconsequential nature of evaluation may exist commonly and, therefore, may be a phenomenon that merits further investigation.
If in fact, it turns out, in a series of systematic studies that promotions are independent of evaluations received and evaluations are weak, generally the process may be wasting time and money when those same resources could be put to more direct use in improving the educational system. Furthermore, these evaluations, if improved, could indeed become consequential and be used to promote teachers into higher-paying teaching positions or mentoring roles. In the latter, they could act as supervisors to the probationary or even less experienced teacher.

The Discriminatory Powers of the Evaluation System

Analysis of Teachers’ Views

It seems that a certain sense of cynicism may be creeping into these teachers’ attitudes towards evaluation in which they just do not believe in the integrity of the process any more. While they infer that the majority of their teaching peers are at least competent, they do not think the system, particularly the evaluation system, has anything to do with documenting, exposing, improving, rewarding or encouraging that level of teacher competence. Essentially, in the opinion of the teachers, competence or excellence, under their present educational organizations, is an independent entity that develops regardless of any formative or summative staff development that the system has to offer.

Despite the aforementioned, a minority of the teachers digressed from this view and seemed to find the evaluation system somewhat effective but only when used as a
foundation for building a case against probationary teachers who were incompetent. Nevertheless, the overriding feeling still persisted claiming that the evaluation system is highly ineffective, except perhaps for eliminating the very worst of the probationary teachers, but a few teachers even disagreed with that. Two of the teachers expressed doubts that it was the educational establishment, per se, that made the evaluations relatively unimportant but rather focused their blame on the unions which made it too dangerous, from a legal-political standpoint, to identify the incompetent teacher. This seems to synchronize with three teachers’ views that if the evaluation system has any effect it might be with the probationary teachers. As apprentices, these teachers, for the most part, do not have a permanent contract and are often, therefore, not fully covered by the aegis of the collective agreement, although they may receive some representation from the union in a dispute.

**Analysis of Principals’ Views**

There seems to be a pattern materializing which usually has the principals claiming that the system is effective and the teachers opposing that view. In this case, particularly, a distinct discrepancy arises between the two viewpoints. As just cited above, one teacher believed that no one of authority even reads the evaluations and further states that even probationary teachers are not refused full-time contracts regardless of evaluations.
Although I state this as a matter of judgment, this researcher tends to believe that the teachers’ report here generally offers a more credible position. This is not to suggest a categorical reprieve for any scepticism towards the teacher-based data. If the principals clearly display a certainty, supply better evidence or present better logical reasons within the framework of this study’s conceptual paradigm, then, undoubtedly their credibility would surpass that of the teachers’. However, when it is just a matter of straight, unsupported but contradictory claims between the two groups and their status as reliable witnesses depends almost solely on the alleged motive, then it seems prudent to grant the teachers’ testimony more strength. This seems to be confirmed by Becker (1971) and Whyte (1956) and their theories of hierarchy and its relation to indoctrination and independent thought within a bureaucracy despite the lack of direct statistical information, (see pages 35-36).

The Cultural Compatibility of the Education System

Analysis of Teachers’ Views

Over the range of responses, it seems that the teachers do not believe in the importance of cultural compatibility in the sense of a general congruency with the cultural mores of that particular school. Even one of the two responses that seem to affirm a certain cultural compatibility as an important factor is suspect. The first of those respondents stated that there already exists a cultural compatibility with the evaluation system, but out
of necessity, the evaluation system is working to change that culture to update that school's teaching methodologies. I find it difficult to believe that an evaluation system trying to change a culture is also simultaneously compatible with it.

Even if the ultimate result culminates in something quite positive, certainly during this period of flux in which the culture is undergoing a certain adjustment, the system could hardly be described as compatible with an ambient culture that seems to tolerate the older methodologies. One teacher suggested that the evaluation form was basic enough to adequately cover all cultural scenarios. I find this "one-size-fits-all" paradigm a doubtful proposition because a culture that is complementary to one system, for example, a college preparatory program, may be incompatible with another, for example, a vocational classroom. Although I have not gauged more than one educator's view in each of the 12 schools, this researcher tentatively infers, nevertheless, that, in the schools discussed, the possibility exists that those boards responsible do not consider cultural compatibility a very significant factor, or they have not yet decided to consider culture as a factor in their evaluation systems.

**Analysis of Principal's Views**

The teachers, for the most part, did not see cultural compatibility as a major factor intended in the design of their evaluation systems. By and large, they seemed to think that any cultural compatibility that resulted was incidental. Then again, on the whole, they
seem to believe that the evaluation system was inefficient and inconsequential, and their opinion on the effect of the organizational culture may just reflect this. The principals certainly seem to be repeatedly in direct opposition to the teachers’ views. One often wonders, however, if this is really the case or if the six principals interviewed have become indoctrinated with what they are “supposed” to believe.

So that this will not be taken as just the bias of one particular researcher, an analysis of some of the principals’ statements is in order. For example, one principal declared that the atmosphere is so collegial that “My staff evaluates me as well as my evaluation of them, and that’s done on a yearly basis.” If the principal in that quote is trying to convince this researcher that an equivalent and balanced relationship exists synergistically, such that the processes of the principal evaluating the teacher and the teacher evaluating the principal are of equal weight, then this researcher is not convinced that such a relationship does or could even possibly exist in reality. While this topic is tangential to the purview of this study, it seems fairly clear to this researcher that a healthy amount of scepticism should be exercised when adjudicating the veracity of any assertion claiming that a collateral and equal evaluation exists between teacher and principal oblivious to the power dynamic. In the judgement of this researcher, based on his general knowledge of the highly structured educational bureaucracy, it is unlikely because (1) that bureaucracy affords no mechanism for this synergy to occur and (2) almost by definition, due to the basic rigidity of the school bureaucracies, authorities there would not tolerate an exception of that pronounced
a magnitude, especially one that seems to negate the very authority of that hierarchy. While it is possible, there would have to be a formal mechanism for reciprocal evaluation or some type of ombudsman whereby a teacher could claim the right to evaluate the evaluator.

There also seems to exist a lack of understanding among the principals of the nature of such terminology as compatibility with the professional culture. One principal claimed the evaluation system's compatibility with the school's culture based on the following reasoning: "because...the kinds of things we advocate with our school goal package...is reflected in the document creating an environment that students feel comfortable in." This seems to suggest a misunderstanding of what professional culture is. Professional culture is not dictated just by following the goals advocated by the organization but also by encouraging a consensual atmosphere that the professionals (and other workers) develop together to their mutual satisfaction. A misunderstanding of this magnitude could lead to serious discrepancies with consequential results.

Preview and Clarity of Evaluation Criteria and Their Inherent Problems

Analysis of Teachers' Views

Only one of the teachers apparently believe that the evaluation, as presently structured, can be falsified. It would seem reasonable that in a quasi-summative situation where the subjects are highly motivated to do well, the possibility exists that these teachers
will show the evaluator whatever they want to see. Certainly as an alternative, an opportunity could be arranged for videotaping several sessions without giving the teacher advance knowledge of which sessions were being taped, in order to get a more accurate representation of the teacher’s abilities. If this option for more accurate data is available, then it may not make sense to tell the teacher when an evaluator will be visiting, particularly if the teacher has the opportunity to study the selected criteria for adjudication beforehand. Consequently, if this type of preview and advance warning is available to the teachers, based on our limited sample, a serious problem may exist in their evaluation processes.

Most of the other criteria mentioned as missing in present evaluation systems seem fairly straightforward. Those suggestions include recognition of academic courses, the readings teachers are doing on their own, outside activities (besides the above) that contribute to professional development and contributions to the school community. One wonders why those criteria have not already been incorporated into these teachers’ respective evaluation systems.

Analysis of Principals’ Views

It is difficult for this researcher to remain emotionally detached in the face of some of the anomalous responses certain of the principals gave. In support of that, this researcher refers to the following remark made by one of the principals stating that, “If
as leaders we hope to foster excellence in learning, then surely to goodness we should be able to indicate what makes excellence in instructional practice.” This is the reply given to the question of whether or not spontaneity is compromised if teachers previewed the criteria used for evaluation. To this researcher, and I hope I am not overstating, this is an almost incredulous response.

Basically it claims that, because educators know the criteria, they necessarily must reveal them to those being evaluated despite the feasibility of possibly encouraging playacting. But just because a physics teacher knows the questions to the exam, should that teacher necessarily give them to the students? Sometimes it is difficult not to point out the ironic nature inherent in some of the illogical and self-serving answers. None of the literature (or at least none that I was able to find) has supplied this researcher with a logically solid reason why evaluators would not get a more accurate result if spontaneous observations were used. Essentially, as I have been arguing, all the evaluation processes are already largely unscientific because they contain such previously mentioned flaws as too few observations, untrained observers, short duration of observation, unfamiliarity on the part of evaluator with subject material being taught, lack of scientific psychometric, statistical or qualitative evidence that the criteria being employed actually correlate to good teaching, too few visits and so forth. Certainly evaluators are only increasing the evaluation’s heuristic, coincidental properties and the likelihood of obtaining unreliable data if they allow teachers the opportunity to rehearse the criteria and perhaps pretend the
criteria set out on the evaluation sheet are the norms of practice regularly used in their classrooms. What motivation, other than pure altruism, would the teachers have to incorporate any new strategy into their repertoires and make it habitual if they have a list of the criteria readily available for them to study and then display on the day the evaluation takes place, of which they are also informed? I believe those six principals defending that system may not be thinking the matter through clearly enough to see the discrepancies inherent in it and the problems it might cause.

**Teachers’ Reactions to Derogatory Evaluations**

**Analysis of Teachers’ Views**

This part of the study is noteworthy if for no other reason than the manner in which the teachers answer. Four of the teachers said that they themselves would not be defensive, but other teachers probably would be defensive. Of course, one cannot argue from absence, but this researcher wonders how so many of the teachers can have such good insight into the foibles of other teachers who seem to only exist in theory but consistently exonerate themselves from the same faults. Perhaps if they ever did receive a derogatory evaluation, they would be equally defensive. But, as most teachers keep saying, they never have received and, most probably, never will receive such a review.

One teacher’s answer was especially puzzling. For instance, the first part of her answer is illogical when it assumes that a negative evaluation would not receive a
belligerent or conflicting response. It seems illogical to assume that a written negative evaluation would not, according to the criteria set up by this teacher, be reciprocated with some type of negative reaction by the evaluatee. After all, this teacher had already agreed that even the realization that "somebody is there [just] taking notes on your performance" in the classroom, regardless of whether those notes are critical, derogatory or complimentary would generate such a reaction. On further probing, however, this same teacher, when asked if an experienced teacher would resent being told her lessons were not well planned, responded that,

   Of course she would mind. Because she would probably say, 'Look, I've been doing this for ten years . . . I'm not going to sit down and write the same amount of detail as in my first year teaching.'

   To this researcher, it seems the teachers may be so bothered by the whole evaluation process that they just do not want to admit how upsetting it actually is. They seem to be feeling attacked and therefore overly zealous to protect themselves from admitting that they don't like the process. It also seems that they would be at least somewhat resentful about criticism of their teaching by non-teaching management personnel who may not have been in the classroom for a decade or more.

Analysis of Principals' Views

The principals, as opposed to the teachers, stated categorically, for the most part, that the criticism is and should be received in the spirit in which it was given, that is, as
essential, professional and practical information. One of the principals stated, however, that teachers are usually bitter about any criticism they are given. This last principal aside, the bulk of these principals interviewed had views that contrasted sharply with the teachers who believed that most of their colleagues would take criticism personally and in a defensive, bitter manner. (Again, those teachers in the study excluded themselves from this theoretically perceived majority group.)

The question again of why this dichotomy continues to exist again seems to confront us. These principals, who arguably should be empathic and caring leaders, seem to be consistently misinterpreting the mood of the teachers. If these principals, who claim to be interested in the professional growth of the teachers, actually want to instil a sense of professionalism within the teachers, it seems they should be more sensitive to their prevailing moods. A major element of successful management is communication skills and, inextricably tied to that skill, is awareness of the mood and culture of the members of that organization. Regardless of who is right or wrong in this case, it seems that this study would suggest that this type of skill may be weak or missing from certain principals, at least when required to facilitate staff evaluation. However, this interpretation can only be made for the administrators referred to in this study and even they may be aware of that failing but not able to act to contravene it as of yet.
Problems with Objectivity of Evaluation Criteria

Analysis of Teachers' Views

Apparently four of the teachers in this study were of the opinion that the criteria used in their evaluations were subjective criteria and that they would like to see more objective criteria used. While it would demonstrate bias on the part of this researcher to dismiss the two dissenting teachers, it may also be somewhat ingenuous to accept their views as equivalent in veracity to the majority since, under probing, neither of those dissenters would provide explanations or concrete examples to justify their viewpoint that objective evaluation criteria exists.

This phenomenon of holding tenaciously to untenable views because of organizational loyalty, expediency, or some other political motive keeps occurring among both teachers and principals, although it seems much stronger among principals. However, in the judgement of this researcher, throughout the course of the interviews, members of both groups sometimes tended to report, mechanically, beliefs with which they seem to have been imbued. This seemed indicated due to rigidity in body stance (cross arms, tight facial features, authoritative sententious tone among other features that usually indicate lack of willingness to communicate). In addition to these more subtle connotations there was also the repetition of cliches which, from their tone, seemed to end the discussion, although it usually raised more questions than it answered. They seemed unwilling or unable to relinquish these views even under scrutiny or cross-examination and even when
it becomes apparent that their views cannot be supported either logically, with any tangible examples or other evidence.

In this case, the problem seems to be even more inveterate, where even those who take the affirmative position, admitting the subjectivity of the present evaluation criteria, are still very reluctant to use one of the few product-oriented criterion, student test scores on standardized exams for teacher evaluation. This reluctance persisted despite these same teachers conceding that some form of this type of evaluation was bound to take place and may eventually prove to be an improvement over the present evaluation system.

Analysis of Principals' Views

There is a pronounced opposition to the use of scores of students on standardized tests for evaluating teachers. This opposition exists among the teachers and principals interviewed, but it seems more deeply entrenched among principals. Interestingly enough, one of the principals (who agreed that it might be appropriate to use students' standardized test scores to evaluate teachers when a specific problem arose) still would not even consider the possibility of using these scores on a regular basis for general evaluation of the teachers.

This researcher cannot come up with a strong argument that would prevent using a student's standardized content exam results for some portion of periodic teacher evaluation, at least for a trial period. The one principal who did agree that most
evaluations used criteria that were "somewhat subjective" was even more strenuously against using student scores to evaluate teachers. This principal was of the following opinion:

It isn't part of our board's evaluation for the teacher, because that's how the student did, that's not how the teacher did. It doesn't mean that the teacher was a poor teacher just because the scores were low this year.

This researcher is surprised at the way some principals can dichotomize the learning and teaching processes as almost two independent solitudes. With no connection, it seems to this researcher that these principals may be claiming a complete disjuncture between learning and teaching. Stated baldly, the logical implication of this assumption suggests an indifference to and an equivalence between good and bad teaching, because teaching might be irrelevant. While this interpretation may, initially, seem astounding, the aforementioned attitude of principals and teachers dichotomizing learning and teaching seems to logically demand this.

Protection of Rights within Due Process

Analysis of Teachers' Views

Initially, this researcher was impressed by the high level of integrity with which the teachers instilled into the interview dialogue. After all, they all considered their rights protected adequately. On further reflection, however, one could conclude that perhaps attributing their affirmations as indicative of a belief in the system's level of integrity could
be a misinterpretation. This affirmation of due process in the system was, in two cases at least, based on a belief in those evaluations' irrelevance. Others believed that their teachers' union was necessary to protect them. This belief implies that the due process aspects of the evaluation system itself, were it not meaningless, would be inadequate to offer a safe level of protection from harassment.

In general, then, after an analysis of the responses, it seems evident that four teachers either sensed a requirement for protection from abuse or concluded that they were inherently protected from any abuse stemming from their evaluations because they were irrelevant. Since, as we have argued, their evaluation situations were relatively meaningless, in both instances the teacher did, in fact, feel protected. Surprisingly, one teacher felt that if any abuse did appear in the system, it would be the teachers and not the administrators who would bear the responsibility for its existence, owing to the overwhelming restrictions that would have to be overcome if it became necessary to dismiss a teacher.

Conceivably a main factor causing the problem lies in the apparent security the teachers have by virtue of their union contracts and the lack of protection the administrators, and by implication the whole education system, have from incompetent teachers by virtue of the same union contract. In such a situation, what can an administrator do to dismiss an incompetent teacher after repeated poor evaluations if in fact, the unions, as the data here imply, maintain such restrictive conditions for dismissal?
Analysis of Principals’ Views

Contrary to what one would have expected, there is not a large discrepancy between teachers and principals over the issue of fairness. One unexpected discovery that this researcher did find was the inferences of two teachers that abuse of the system by administrators was impossible because the evaluations were inconsequential. While it is true that not much seems to result from the initial documentation of an incompetent teacher that a derogatory evaluation provided, it is nevertheless often an essential element in beginning the paper trail that can result in the review process and eventual dismissal. Some of the teachers know this, and all the administrators in this study seem to know it, but as one teacher mentioned, some of the teachers are naive about the potential harm the documentation process is capable of initiating and eventually accomplishing.

Yet in general, it seems that both principals and teachers agreed that if abuse is possible by administrators, it is unlikely, largely because of the strong union protection against unfair or unwarranted evaluations that could result in that type of harassment. Even more interesting is the consensus that developed between teachers and principals over the evaluation system’s ineffectiveness to dismiss even the most incompetent teachers. If this is so, it seems to represent a serious problem in the evaluation system’s summative component in particular. What could the purpose be for having an inconsequential evaluation system that is neither, according to this study, effective from the perspective of formative evaluation, summative evaluation or accountability?
On a global level, supporting this question is the American literature which claims that estimates of 5 per cent incompetency only produce a dismissal rate of less than 1 per cent (Bridges & Gumport, 1983; Fuhr, 1993; Lavely, Bergman & Follman, 1992). Bridges (1992) found, in the California school system, a fairly low dismissal rate of .6 per cent. However, he found that 70 per cent of that .6 per cent were temporary teachers, and probationary teachers accounted for the other 25 per cent. Thus, only .03 per cent of full-time tenured teachers are dismissed annually in California. When this is contrasted with the 5 per cent estimates of incompetency, the seriousness of the tolerance for inefficiency becomes even more blatant. While complete figures are not available in Ontario, of the 175,000 Ontario public and separate school teachers only 35 (all men) have been dismissed since 1997 it is now 2000. Of those, 32 have been terminated for sexual improprieties. Thus only .0017 per cent of the teacher population could have been terminated for technical incompetence, and there are no data available stating how many of those were full-time tenured teachers or if they were merely probationary teachers (Wente, Globe and Mail October 17, 2000).

**Self-Directed Criteria Selection**

**Analysis of Teachers’ Views**

A consensus appears here among the teachers that the collaborative approach to evaluation is not only necessary but of primary importance for them to achieve an effective
evaluation with which they are comfortable. The differences between the opinions of the teachers exist only regarding procedural methodologies, chiefly over which particular strategies to emphasize within that collaborative paradigm and how best to implement those methodologies. Some suggested a portfolio approach, where the teacher brought tangible evidence representing growth, was best; others proposed an open discussion of all the stakeholders would be better; and still others offered the opinion that teachers should have the dominant leadership role in a collaborative but not necessarily equally balanced relationship with the principal in this area. Regardless of the preferred means, it is reasonably safe to conclude, based on the available data, that the teachers surveyed wanted a strong say in the formulation of the evaluation protocol. Further research in future studies, more focused on this particular aspect of evaluation, however, might have to be undertaken to discover if this interpretation is valid over a larger statistically significant segment of the teacher population.

**Analysis of Principals' Views**

When the responses of the various principals were examined, they seemed approximately equivalent to those views expressed by the teachers. However, on further analysis, that conclusion may not be completely accurate. One principal, while not excluding the teachers as partners in this planning effort, nevertheless thought that the union representatives would be equivalent in viewpoint to what the individual teacher
would want, and presumably would therefore be adequate to represent the teachers’ concerns. This, for the most part, is plausible, but it is also possible that these viewpoints are not identical. The union negotiator, for instance, might be more concerned with what is good for the system and how that reflects on the union or the political position of the union. If the individual teacher’s criteria for evaluation are not conducive to that paradigm, even if the majority on staff agreed, the federation representative might not advocate it. Indeed, from a teacher’s perspective, it may sometimes be detrimental to his or her individual position to have the union representative there at the pre-conference. One principal, just previously cited, seems to believe in collaboration as long as she can convince the teacher to pick the criteria that she, as principal, thinks are important. That does not seem very collaborative; it seems more like “We collaborate as long as it’s my way.”

Principals in this study seem to feel they must appear to be collaborative because that is what is thought to be philosophically correct when, in fact, the principals are not really in favour of any genuine power-sharing. If this is so, this attitude might extend to other areas of this question and even other segments of this study where the principals appear to concede the importance of power sharing but are, in fact, not generally prepared to do so (see page 183, principals 2 and 5).
The Principal-Teacher Relationship in Evaluative Situations

Analysis of Teachers’ Views

Much to the surprise of this researcher, the tension between evaluator and evaluatee that one might have thought existed was only mentioned as a defining condition of the relationship by one of the teachers. Other teachers did not regard this sudden focus on their skills as highlighting the power aspect of a previously established relationship and was not so overwhelming as to redefine the relationship. Most of the teachers, however, stated that during and after evaluation, the relationship at that time was largely, if not completely, defined by the quality of the relationship prior to the evaluation taking place.

It seems, then, that these opinions may not be that far apart. Both opinions seem to concede that the administration is instrumental in setting the mood for the cultural subtext of evaluation in the organization. One teacher’s experiences, however, caused that teacher to see the power the administrator had as so overwhelming that tension became an inherent, if not a defining part of the relationship, during evaluation.

Perhaps the most interesting, if not amusing, comment came from a teacher who considered the evaluation a tacitly acknowledged but useless, ritualistic game, in which almost everyone involved in the process agreed that it was an unfortunate necessity. Therefore, both sides subtly agree to go through the steps in such a way as to cause the other side the least amount of annoyance. Future studies would be needed if the accuracy of those opinions are to be determined.
Analysis of Principals' Views

The principals were apparently not consistent in their views of teachers' attitudes toward evaluation either. This does not mean, however, that the principals' views paralleled those of the teachers in the study. In fact, in most ways, their views contrasted with the teachers' views of evaluation. Not one of the principals felt that the process was a useless ritual which, in itself, may have helped to define the relationship between the two participants, principal and teacher. Also, none of the principals were prepared to acknowledge the inherent power of their position and its potential for intimidation. To their credit, however, several of the principals recognized the importance of their previous relationship with their teachers as the defining factor in the evaluator-evaluatee association. Perhaps, then, the most important aspect of the principals' answers is how they apparently indicate a recurring pattern of discrepancy between teachers' and principals' views on the important issues inherent within teacher evaluation.

Professional Training of Evaluators

Analysis of Teachers' Views

There seems to be both a general lack of knowledge and a sincerely expressed disappointment concerning the level of training that principals receive to prepare them to conduct evaluations. Invariably one would wonder why four of the teachers claim they have no knowledge as to what kind of training the principals had undergone to qualify
them as evaluators. It seems fairly unequivocal that if evaluation was consequential to the teachers and critical reports were being written forcing teachers to change aspects of their teaching styles, those teachers and their unions would want to know what training the principals had received that precipitated that information and how valid they were, and the unions, which have the right to know, would then inform their members.

Surprisingly, not many of the teachers were optimistic enough to believe that a program in professional evaluation would help much. Only two teachers, however, came up with concrete suggestions as to what would be better. One suggestion might be acceptable as it was based on constructing interviews that were standardized and to which every teacher had to submit. This was suggested in addition to the principals being trained by reviewing case studies and analyzing them in a formal classroom situation. One of those suggestions, however, might not be acceptable to any board of education as it was based uniquely on peer review where a mentor teacher comes in and does a formative evaluation. This would likely be rejected as an infringement of management rights, despite any intrinsic merit that this type of review might have.

**Analysis of Principals’ Views**

In the opinion of this researcher, several of the principals seem to have adopted an intransigent stance over their level of training in evaluation. Instead of responding in a forthright manner, they responded with a torrent of jargon, which in the judgement of this
researcher, was just used to obfuscate the issue further rather than clarify. One principal stated the following:

Because in terms of what the Ministry of Education says on the *Education Act*, part of the training to become a principal means that you have to have different levels of education as well. The different levels of education right down to even taking principals’ courses deals with the evaluation of staff. So, yes, the training is there, it doesn’t have to be local.

My own summary depicts this exchange by the following description: “I see the words but seem unable to determine what they mean.” Certainly if this principal means that any professional course or workshop improves someone’s evaluation skills, he or she cannot be taken very seriously. Certainly there is nothing in that circumlocution that would convince most impartial listeners of the credibility of the evaluation based on the skill and training of the evaluator. In the judgement of this researcher, they are an incomprehensible farrago. Another principal thought a certificate in applied evaluation would be detrimental because she didn’t think it should become too structured: “I don’t think it should be of that particular nature. … [And] yes it’s important that it’s within the framework of the full role …”

Yet, despite this, several of the principals, as just cited, were quite straightforward about their lack of training and the consequent unprofessional quality of their own teacher evaluation procedures. Most of these principals fervently wanted the school board to provide stronger in-service training sessions in evaluation. These complaints seemed to echo the teachers’ previously cited objections about the lack of professional evaluation.
That is, other studies report that teachers consistently claim that evaluation training is egregiously poor. Teachers consistently complain that principals give indiscriminately high evaluation, lack evaluation training, attend evaluation sessions only for brief, infrequent periods and are generally too incompetent to evaluate teachers (French, Mallow & Rakow, 1988; Housego, 1989; Kauchak, Peterson & Driscoll, 1985; Turner, 1986).

General Attitude of Teachers Toward the Evaluation Process

Analysis of Teachers’ Views

To this researcher, the above teachers’ opinions seemingly represent a fairly strong indictment against the present system of evaluation. Apparently all the teachers surveyed believed that their respective evaluations are aggravating and sometimes even demeaning to their professional status. The reasons why, as discussed in the literature review, are still speculative, but in the judgment of this researcher, the attitudes of the six evaluators in this study, their admitted lack of training as evaluators and the largely inconsequential nature of their evaluations, as reported here, seem only to serve as an encouragement to the commonly held (teachers’) negative disposition toward evaluation.

Regardless of the reasons behind this aforementioned attitude of the teachers toward evaluation, it seems detrimental to effective formative or summative evaluation. Furthermore, as a result, any improvement in teaching that administrators would expect to occur from evaluation is probably stymied. That this negative attitude seems fairly
ecumenical may even emphasize the acute nature of the problem, that is, the serious faults extant in the evaluation system as it presently exists. This, however, could only be determined by another statistically valid study.

**Analysis of Principals’ Views**

The principals, as mentioned several times, often seem to be at odds with the teachers’ viewpoint. While one cannot definitively say how widespread this is outside the studied group here because of the small sample size and due to the teachers and principals being from different schools and boards, the literature generally also supports this proposition. The cause is certainly open to speculation, but in the estimation of this researcher, the culture of the school system seems to contribute to this problem. Arguably, this culture is so bureaucratically stratified that it creates an almost sub-cultural divisiveness that dichotomizes the atmosphere between upper and middle management on one side and labour on the other (Fernandez, 1993; Glasman, 1986; Mohrman, Cooke & Mohrman, 1978;). While this may sound extreme, in this study alone, the majority of principals felt confident that the majority of teachers were generally satisfied with their school’s evaluation systems. Yet in the same study, the teachers in general were, in fact, not satisfied, but highly dissatisfied, with their evaluation systems.

As a result, the principals often misconstrue the teachers’ positions. Again, as mentioned previously, we cannot posit this as a definite conclusion because the principals
and teachers are from different schools and different boards. In effect, they are discussing different segments of different populations. Since the split is so pronounced, however, it may indicate a real problem pervasive in the general population, as well. Such a split, if indeed it existed throughout a school and beyond just the 12 educators of our studied sample, it would seem to be highly damaging in a profession where communication is seen as an essential skill and held at a premium. Communication between teaching staff and management is at least equally important, seeing as part of management’s job is to be aware of the attitudes and opinions of their staff, especially regarding formative assessment situations. This study then may point to a serious management problem in educational administration and needs to be tested in a larger statistically valid study.

Successful Evaluation and the Nature of Its Design

Analysis of Teachers’ Views

What is perhaps the most significant part of this section is the majority opposition to the system presently being used. Four of the teachers who opted for the portfolio paradigm, were also dissatisfied with the system presently in use in their schools or boards. Even the teacher who was more general in her preferences was still opposed to the evaluation system presently followed at her school. Again, here is a fairly clear indication of dissatisfaction among the teachers. While it is not feasible within the purview of this study to determine definitively why, it again appears that no one seems to be especially
heeding the concerns of these particular teachers regarding evaluation. It seems true that some of the teachers were not clear as to what exactly was wrong with their system, and while they did not offer any innovative solutions with suggestions of how they would fix it, some of the teachers were quite clear. Specifically emphatic was the teacher quoted above, who was insightful enough in her outlining of a formal evaluation system of evaluation that, *prima facie*, it seemed better than anything that appears to be presently operating at the schools discussed, based on the information derived from these interviews.

In the judgement of this researcher, it seems, relying on the information presented within this study and on this study’s literature survey, that the evaluation system presented here is not perceived to be accomplishing its purpose efficiently from either the formative, the summative or the accountability aspect.

**Analysis of Principals’ Views**

As just mentioned, within this sample of principals, there is a strong core of conservative opinion firmly rooted in the status quo more or less satisfied with the present situation. This conservative opinion holds that, in contradistinction to the majority of teachers in this study, little if any changes are needed to improve the evaluation system. Certainly the principals did not see any need for changes that make the system more empowering from the teachers’ frame of reference. One principal, expressing her own
dissatisfaction with evaluation, pleaded for an analytical, applied research that would put an end to all the controversy and misinformation over which system to adopt. She further says:

We haven’t looked at exemplary practices ... but I don’t believe that anybody has done any research out there to say, ‘Here are exemplary practices that teachers feel are effective, principals feel are effective and these are the ones that really work, and these are the components that have to be within it to make it very effective ...’ I would like to know what are these exemplary practices and take a look at our system and make it better.

Unfortunately, that is just what scholars in the evaluation field have been trying to do by applying the emerging discipline of evaluation to educational administration for quite some time (Clark, 1993; Darling-Hammond, 1984; Levin, 1979; Millman, 1981; Mohrman, 1989), but with only moderate success. It would then seem, based on these data’s suggestions, that unless a consensus of what the essential elements of good teaching are emerges from the literature and is then able to operationalize them and train observers to look for them specifically, no such definitive system may be possible. However, even that would have a strong element of subjectivity, and as the evidence now suggests, the style that works for one teacher probably does not work for another. Thus, any general guidelines would have to be in the form of obligations, which would leave evaluation assessing a basic form of teaching. The alternative, which has been adopted in some districts in the U.S. (although vigorously opposed in others), is to use, at least as a component of teacher evaluation, product-oriented criteria like student scores on standardized exams. Yet, almost any suggestion to adopt such a system here is, in my
experience, continually met with fierce resistance by teachers’ unions and academics -- often, as I have argued above, supported only by the most feeble of excuses (Darling-Hammond, 1984; Florio, 1986; Soar & Soar, 1979).

Another principal suggested that all principals be relieved of their evaluation tasks. This researcher would have to agree with that view. Evidently principals receive little formal training to qualify them as evaluators in this highly technical field even according to the principals themselves (pages 186-187). Furthermore, as the same principal states, these evaluations rely on formats that are seldom based on rigorous scientific or psychometric data nor could one expect them to be under the present conditions. Again, in the estimation of this researcher, a change in policy reflecting those views might contribute to making teaching and teacher evaluation more congruent with other professional groups by initiating higher standards and results-based criteria of professional behaviour.

Effects of Evaluation on Teaching Quality

Analysis of Teachers’ Views

If the majority of the teachers interviewed, that is, four out of six, stated or implied that evaluation has had no real impact on their teaching, it might be time to experience a different evaluation system. Even the teachers who dissented were severely limited by the scope of their approval in the present evaluation system. The most unequivocal praise came
because the evaluation system was beneficial in expediting the removal of incompetent teachers. While it is certainly important to remove teachers who are incompetent, that summative function as very nearly the only positive aspect of the evaluation system does not speak very well for evaluation within the education system. Evaluation systems that exist uniquely for the purpose of instituting a dismissal are, at the very least, not living up to the fullest theoretical potential of the purpose of evaluation. Again, if a teacher is that incompetent, documentation outside of the formal evaluation process would seem more efficient, from an economic and even a bureaucratic perspective (Glasman 1986). This has also been partially confirmed by several principals in this study itself, who have stated that an entire “paper trail” would have to exist to dismiss a teacher, implying that single evaluation or even a series of evaluations would not be sufficient in and of itself although admittedly, some felt it would probably be a necessary first step. If that is taken into account, no teachers’ comments can really be said to be a very convincing endorsement for claiming the evaluation process presently in the schools studied is efficient.

Analysis of Principals’ Views

Again, I see a sharp dichotomy of opinion forming between what teachers assert about the evaluation system, and what principals appear to believe. Most of the teachers did not indicate that evaluation had any impact as either a noticeable or valuable contribution to the quality of teaching. One principal considered the evaluation process a
valuable tool because he was able to dismiss teachers who were incompetent. This statement seems to be misleading because an evaluation, no matter how unsatisfactory, would not be sufficient cause to dismiss a teacher but only to begin a review process. This use of assessment to dismiss teachers also highlights only the narrowest application of evaluation: the final step of a summative process.

To this researcher, these impressions seem to point to the problem in the present-day evaluation's failure to dichotomize between formative and summative evaluation. Certainly, by the time an administration is convinced that a teacher should be dismissed, an official evaluation is either extraneous or routine. What seems necessary is a documentation process which probably should be a totally different process from evaluation and may not even involve much formal classroom observation but certainly very little, if any, real evaluation.

This objection notwithstanding, this study's principals were not able to offer any explicit, concrete example of how evaluation works in their schools as a process that could improve teachers or motivate them to teach more effectively. Admittedly, though, it would be wrong to blame principals or teachers entirely for this gap in their knowledge about their evaluation system. Again, it must be reiterated, that until research clearly identifies exactly the teaching behaviours that cause students to learn, teachers cannot be expected to have faith in their evaluation results, or possibly change their teaching styles based on an evaluation that is itself, for the most part, arbitrary and perhaps even inaccurate.
The Reluctance to Dismiss Inadequate Teaching Staff

Analysis of Teachers’ Views

As a researcher, I had not expected that half the teachers surveyed would paradoxically be so highly dissatisfied with the formidable strength of the job security that their union was able to obtain for them because it tolerates incompetence. Although it is beyond the immediate purview of this study, it may prove interesting to see how universal this attitude is over a statistically valid sample. If, for instance, this outlook is found to be a common one, one might wonder, then, how these naturally beneficial attitudes get ignored by the incumbent scrutiny of the board’s negotiators who negotiate a counter-initiative and counter-productive contract, which works ultimately against their own best interests. Furthermore, it also seems conceivable that a significant percentage of the union membership might reject this type of security because it impinges on their professionalism, credibility and the quality of student education. Put another way, it would be helpful to know if it is the contract that is impeding efficient summative evaluation and subsequent dismissal, and if it is the boards of education that approve these restrictive union demands, it seems, then, that the balance of probability implies that the boards of education that these principals and teachers are affiliated with are approving these contracts despite the opposition of both teachers and the boards themselves.

Although one interviewee seemed to infer that the culture of the school system and its influence on what was acceptable as criticism was responsible, this point still has to be
confirmed. Furthermore, other professions, for example, the medical profession, which dismisses a certain number of doctors for technical professional incompetence every year, does not seem to have this problem in nearly as severe a manner, despite an equally collegial atmosphere. Perhaps though, that teachers' insights were correct, and the culture of her school might indeed breed a type of familiarity and tolerance of mediocrity that is counterproductive to the summative or even formative aspect of evaluation.

Also the possibility exists that, because it is generally accepted that only the most incompetent get fired, the dismissal of a teacher would not only result in his or her dismissal from that board or school but may result in the effective removal of that teacher from the profession for all intents and purposes. This may be a further restraining factor in initiating dismissals.

Another teacher again raised the problem of impartially evaluating something as multi-faceted and judgement-based as teaching. This researcher holds, as postulated several times before in different contexts relating to this issue, that this argument is specious at best. Again the question must be asked if doctors can, for some purposes, be evaluated on the basis of patients cured and lawyers on the basis of cases won, why can teachers not be evaluated on the basis of student results on standardized tests for at least part of their evaluation program? Why must principals continue to defend the use of arbitrary, unproven criteria and then complain that they can't dismiss teachers because they only have data that are too "soft" to be used as precise or consequential criteria for summative
evaluation? To this researcher it sometimes sounds like such educators need this excuse to protect themselves. In many ways this is a self-created, unsolvable problem that is typical of a psychologically neurotic dilemma. There is no solution because of certain premises but to abandon the premises would force acceptance of the solution which would leave everyone without the protection the premise provides. Perhaps that is why this researcher has been exposed to what he perceives as an inveterate, adamant resistance by teachers and administrators even at the intimation of conducting a study on the use of student scores on standardized exams as a component in teacher evaluations.

Analysis of Principals’ Views

One of the principals cited above insisted that since both they and the teachers were in the education profession it would be wrong and unprofessional to dismiss a teacher because, as educational administrators, they should be prepared to instruct an incompetent teacher until they have improved. While a cogent argument could be made for having a separate training facility at school to retrain teachers or demote them to assistant teachers while they work with other teachers until they improve, it seems almost indefensible to allow clearly incompetent teachers to continue to teach indefinitely, withholding bona fide professional teachers from the students and possibly damaging countless students in the process while they learn the necessary professional skills, just because the administrators are in the education profession.
Yet the principals, here, who disagreed with the teachers’ views were, in the judgement of this researcher, making that argument. The first principal was perhaps the most adamant about that position. She made, what appeared to be, several misleading statements. Firstly, this principal claimed that boards of education cannot decertify other teachers as easily as other professions weed out their incompetents. The point of the question, at least in my interpretation of it, was attitudinal, seeking to discover whether or not there was a more reluctant perspective to dismiss incompetent teaching employees as compared to employees in other industries (or professions); so, the question would still remain, “Do these other professions seek to expedite dismissal diligently while the teaching profession does not?” This was asked irrespective of any actual problematic aspects of the situation.

Secondly, the question was asked concerning why incompetent teachers, in practice, were kept on as paid teaching employees as opposed to an entirely different question, why teachers were not decertified, which that principal chose to answer. Therefore, the intent of the question was missed. Surely, this principal knows that any board of education can fire someone without decertifying them, which, indeed, is much harder, as the principal correctly claims.

Thirdly, that principal’s last point is factually incorrect. She claimed that business can dismiss any employee even without just cause while she, as a principal, does not have that option. However, in actual fact, it is quite difficult to dismiss any employee who has
been employed for a significant time regardless of the type of employment. This is because the costs of any wrongful dismissal suit could prove prohibitive, unless there was just cause, if this type of reckless dismissal was done on a regular basis. This seems to be confirmed by the next principal. She had previously said that the problems of dismissing an employee were the same for industry. Since this principal offered the Bank of Montreal and Petro Canada, two largely non-union institutions as her examples, this researcher personally doubts that their difficulties with dismissal were identical with those of boards of education who have to deal with unions, although they probably did share some similarities. Furthermore, although private enterprise firings are most likely not done as simply as the first principal suggested, people are, in fact, fired more quickly even in business-oriented and blue-collar union establishments than they are at union-represented boards of education (Downie, 1992).

Another principal thought that this pattern, indicating a reluctance to dismiss incompetent teachers, might have been the case ten years ago but was certainly not the case now. This researcher remains skeptical because, as this study's data show, all the teachers and half of the principals agreed that difficulties dismissing a teacher present a serious and immediate problem. In summation, then, while sometimes there is a certain forthrightness on the part of the principals, nevertheless, in their role as administrators, they do not seem to have fully come to terms with this problem. For the most part, even those principals who clearly acknowledge the problem seemed to lack an explanation when asked why
solutions to an apparently fairly straight-forward though critical problem have not been implemented or even developed to a rudimentary level.

**Summary**

While there are several positive aspects in the various evaluation systems, such as some potential for professional development, external reinforcement, and the acknowledgment of a moderate degree of professional autonomy, it is clear, if the reader can accept the data of this study, that 12 educators conducting teacher evaluations experienced serious systemic problems. While no system of evaluation is without flaws, are there strategies based on the literature and further developed in my analysis that might maximize the chances of effective evaluation in the formative, summative and accountability components?

The necessary precautions, according to the data of this study, have not been taken. As a result, a dissection of respondents' views suggest that their evaluation systems seem to have fallen into disarray with the expected benefits of professional evaluation occurring only very rarely. Lack of communication (between levels of management and between employees and management); lack of professional evaluation and evaluators; the absence of strong reasons to avoid using objective data, discrepancies in general school culture and professionalism; and the restrictive nature of the collective agreement would appear to be
just a few of the major problems that confront educational management in its performance of evaluation duties.
SECTION V

CHAPTER 6
RESULTS AND DISCUSSION OF DATA AND ANALYSIS: PROBLEMATIC THEMES IN TEACHER EVALUATION AND THEIR RELATION TO THE LITERATURE

Introduction

From an initial reading of our study, it might seem that there is almost nothing in the described evaluation systems that attributes any positive characteristics to the extant forms of evaluation; however, this is not what the study intends to declare. It would, in effect, be an egregious misreading of the interviews and analyses to remain with that impression as a conclusion. In fact, the contents of the first interview question and its corresponding answer belies any such misunderstanding of the study. However, like Lawton’s earlier research (as cited in Hickcox, 1988), these positive aspects are unique enough to be considered as incongruent and not indicative of any overall consistent trend of efficiency in the various commonly used evaluation systems. They are, for the most part, unique and even exceptional in the sense that they do not seem to be part of any concerted systemic effort to integrate evaluation into a more directed component of a modern management system. Consequently, these aspects are not accurate, overall descriptions of the extant systems of teacher evaluation.

Nevertheless, those positive aspects exist and are included among the following: (1) reinforcement of perceived correct teaching methodologies, (2) directed constructive
criticism, (3) opportunities (for teachers) to exercise their own professional judgement in an equal and collaborative relationship with their peers, (4) an affirmation of their professional status due to the conscious acknowledgement of standards inherent in efforts made to maintain effective practices within the teaching profession and (5) clear, unambiguous evaluation criteria. These attributes were present in the evaluation systems strongly enough to be confirmed by both teachers and principals but only nonconsensually as individuals when asked specifically for positive aspects of the evaluation system. The exception to this positive-aspect rule was number (5) above, which was confirmed by the majority of teachers and principals and therefore mentioned in the Results of the Data subsection listed below in 6).

That having been conceded, the gravamen of this thesis must also be recognized and dealt with. One of the major themes of this project is the existence of serious flaws within management of the teacher evaluation processes presently used in Ontario. For the most part, that can be shown by the results of the following data that seem indicated within the range of our sample:

**Results of the Data: A Systemic Description**

(1) Teachers and principals unanimously, to various degrees, are dissatisfied with the excessive protection afforded by the collective agreement to borderline and incompetent teachers.
(2) Teachers and principals (although they were recently teachers) often display profound differences in their attitudes toward the same issues. These include the following: the subjectivity of evaluation criteria, teacher attitudes toward evaluation, post evaluation interviews and the criteria used for evaluation.

(3) There is no clear demarcation between the formative and summative components of the evaluation processes and four of the principals did not recognize the necessity to make that distinction.

(4) Teachers and principals (eight) agreed that the evaluations only slightly influence promotions.

(5) Most teachers and principals (eight) agreed that all staff regardless of professional experience or duties were evaluated in the same manner.

(6) The majority (nine) of the teachers and principals believed that the criteria for evaluation were clearly outlined.

(7) Teachers and principals (eight) agreed that the criteria and the evaluation report are, to different degrees, subjective.

(8) All the principals and teachers who responded because of an assumption of adequate knowledge (eight) agreed to varying degrees that training for evaluation was inadequate.
(9) There was strong, almost unanimous opposition from principals and teachers
(ten) against using product-oriented data in the form of student scores on standardized
exams to evaluate teachers.

(10) There was strong, almost unanimous agreement among principals and teachers
(ten) that the number of visits allotted for observations in the evaluation process was
inadequate.

It seems, then, an emerging pattern of judgement within this thesis suggests that
most problems that can be discovered within the evaluation process are systemic ones.
Modern management philosophy would seem to extend this assumption in the sense that
it believes that systemic organizational problems nearly always originate at the upper levels
of management (Sallis, 1993). This principle is even more applicable in this situation here
because teachers have only limited power in formulating policy. What this implies is that
even if the problem was not the policy itself, which, if followed, would have produced
good results, but, for example, if the problem was the teachers’ inability to understand the
policy, it would still be management’s fault. Since the problem is systemic, then the fault
lies with management’s failure to provide clearer instruction.

Of course this perspective, and, indeed, the concomitant entire analysis of the
professional environment, as it effects evaluation, is based on an organizational model
heavily steeped in the business management organizational paradigm. While this may be
initially unattractive to some, that approach is well established and assumed to be workable
and constructive. This assumption has become so inveterate that “in essence,…the study of schools as organizations has come to mean the study of organizations [themselves]…” (Allison, 1983, p. 14). Consequently, in any standard book of educational administration, as the rule, not the exception, schools are not conceptualized as unique or isolated phenomena but merely as the construct generally known as an organizational bureaucracy, a social system or a loosely coupled system (Allison, 1983).

Lennington, seeming to advocate this paradigm even more strongly, has stated, regarding the structure and the organization of higher education that, “the Holy Grail of any human endeavour should be excellence, which is [only another way of describing] quality…. If we have developed a quality product for our students, then we will have produced a quality product in our students” (Lennington, 1996, p. 166).

This strong dependence on business models, then, has perhaps developed, as Allison has already suggested, generally, because of the “well developed images and models of organizations and their major properties which could be expected to provide some valuable insights into the nature of schools (Allison, 1983, p. 13). Specifically four main reasons can be given explaining why education seems to automatically orient itself towards business for its organizational paradigms:

(1) organizations, on fundamental issues, have the same problems to deal with

(2) theorists in education have developed very little systemic management theory of their own
(3) beliefs persist that businesses are well managed but public schools are not
(4) concerns which vociferously demand that all government-funded services use
resources as efficiently as businesses do (Bottery, 1992)

There are other obvious commonalities between the organizational structure of
schools and businesses:

1) the need to monitor situations and personnel;
2) the necessity for excellence;
3) the requirement to create a sense of team and community spirit (Bottery, 1992);
4) the primary motives and beneficiaries;
5) the methodologies of resource allocation;
6) the concern with quality control of products;
7) the measures and evaluation of professional personnel;
8) the awareness and control of organizational culture and professional environments
(Edmonstone, 1982)

As a result, the following terms and analogies, essentially business terminology,
have become widely accepted in education: consumer-led aspects, standardization of
product, human resources (employees), efficiency in finance, accountability, quality,
management by objectives (school-based management), job market orientation, hierarchy
of management and the relationship between performance and competition (Bottery, 1992).
While the notion of bureaucratic related ideas (human resources, standardization of product, hierarchy of management and so forth) may initially seem inimical to the structure of educational organizations, it should be recalled that Mintzberg himself, often opposed to applying certain aspects of the business model to education, sees no contradiction. He elaborates by describing the professional bureaucracy as a specific variation on the bureaucracy paradigm, developed from and oriented towards educational management structures (Mintzberg, 1979). Additional objections to the accuracy of the analogy can be weakened by noting those aspects of the human psyche ubiquitous in all organizations and therefore factors in the efficiency of all management systems. These human aspects have already become well established, important criteria and documented in detail by Drucker (1968) and Waterman (1982), both accepted authorities on management.

This is not to suggest that there are no alternative viewpoints that distinguish themselves in contradistinction to those mentioned above. Among those are included the following: Lotto & Clark (1986), Sergiovanni (1996) and English & Hill (1994). Most of these opinions come from a vigorous humanist perspective of management theory and see schools not as organizations but as unique entities socially organized along community lines. Furthermore, they do not seem to offer any detailed, concrete alternative paradigm with which to replace management business models and along which to determine an organizational protocol.
Pivotal to our view that asserts a system deficiency are the documented effects on evaluation, indicated by the evidence of the data of three general but pervasive and systemic problems, which can be extracted from the results, that recur consistently. These general categories are the following: (1) Lack of professional culture, evaluators and evaluations (objective data and incentives); (2) Poor communication; and (3) the Limiting nature of the collective agreement.

**Lack of Professional Culture For Evaluation and Evaluators**

As perhaps the most prevalent of these problems, I must list the lack of professionalism from both middle management (that is, the principals) and the front line workers (teachers), both of whom, according to the suggestions of our data, are instrumental in constructing the following “vicious cycle”: the more management refuses to treat teachers as professionals, the less likely the teachers are to act professionally, and the less they act professionally, the more management will refuse to treat them professionally. Despite this allegation, both principals’ and teachers’ organizations staunchly claim that they are professionals, have professional values and are operating within a professional culture. It is an implication of this thesis that the culture inherent within the school systems is unprofessional and fosters poor evaluation through a tolerance for mediocrity. Yet one of the chief signs of professionalism and even trade-unionism, from the times of the first guilds, which were formed to safeguard standards of quality, is
a high regard for professional and technical competence and a low tolerance of serious incompetence.

It seems that the entire study indicates a lack of standardization in its teacher evaluation systems in Ontario, in isolation to other aspects of that education system, which infers a lack of professional clarity on the evaluation issue itself. The corollary of that point implies that only an attenuated professional status presently exists for the entire evaluation process and, by implication, possibly teachers themselves. This criticism seems justified because professionalism, by its very nature, must provide an expertise that does not allow or tolerate arbitrary or experimental measures except at the initial stages of a process or under the most extreme or unusual circumstances. In any event, mediocrity should certainly not be a perennially accepted practice, especially in an issue as crucial to excellence of education as teacher evaluation. Such random variations as are presently accepted in teacher evaluation would almost certainly not exist if a truly professional concern permeated management's attitude toward the evaluation of teachers at a ministry level and a professional culture permeated the school system. That the literature review itself demonstrates evidence of this unprofessional and, indeed, inimical attitude toward the process of evaluation is again manifest in the desultory attitude the revised statutes of the Education Act itself espouses toward evaluation. While the Act itself relegates the duties of evaluation to the principals or to the board, in no way does it insist that any teacher evaluation at all has to occur, let alone delineate criteria for either a general
protocol or a specific process. In essence, it is left up to the local board’s discretion whether or not to have evaluation at all. This researcher cannot imagine any other serious enterprise, especially one with an operating budget in excess of thirteen billion dollars, (Ministry of Education, 1997) besides public education, take such a capricious, cavalier attitude toward the evaluation of its employees.

External evidence, assumptions and tangential implications are not necessary to illustrate the unprofessional culture that surrounds the evaluation system. It can be argued based on the data here, that a flawed culture may pervade the educational system that might in many ways bear the responsibility for the absence of professional evaluation. From these data, it might be deduced that the culture might be somewhat responsible for the evaluations being seen as a perfunctory, arbitrary and even pre-determined process. As these data suggest, so few teachers claim to have ever received poor evaluations and the standards for the evaluations are said to be so low that receiving a poor evaluation is considered insulting or even demeaning. Consequently, as the data have shown, teachers often seem to consider that form of criticism as a refutation of their professional status and not as an impetus to improve. Yet, even conceding this behavior and attitude as a major problem, the existence of professional discrepancies within teacher evaluation might not be seen as peculiar to these attitudes alone. The serious deficiencies of professionalism revealed from this study’s data in the form of apathy and incompetence also seem intrinsic to the evaluation systems themselves. Lack of professional training for evaluators in
evaluation, lack of consequentiality, arbitrary and indiscriminate application of evaluation criteria, low number of visitations, insignificant post-evaluation interviews, the preview of evaluation criteria, poor excuses supplied to avoid the use of objective data from standardized tests to evaluate teachers, a lack of clear demarcation between formative and summative aspects of evaluation, unprofessional attitudes of teachers towards derogatory evaluations, a general attitude of scepticism and even contempt by teachers toward their evaluations all could be interpreted as systemic weaknesses that reflect on the poor quality of the professionalism of the culture of evaluation. While some of these practices are defended by the teachers, most are conceded as professional deficiencies, but ones with which they feel at a loss for solutions with which to correct them.

Tainting the entire professional perspective of teacher evaluation, the one which distinguishes itself as perhaps the most serious cause of professional misconduct because of its pervasive influence, may be a lack of bona fide professional training for the principals who are required to perform the evaluations. The principals in this study themselves acknowledge that they have not received sufficient professional training as evaluators. Almost all the teachers interviewed, who felt competent to comment in this area, also agreed that the principals had inadequate training to conduct professional evaluations. According to the responses in this study, confirmed by Ken Leithwood who coordinates one of the principals’ programs in Ontario, the PQP (Principal’s Qualification Program) is usually the only training principals are required to undergo as evaluators.
However, one must be very cautious in extrapolating this confirmation because each principal’s program is not only independent but often unique in content and format. Furthermore, as mentioned in the Literature Review, based on studies going back to 1921, the correlation between teacher efficiency and principal observation (the most common form of teacher evaluation) is effectively zero, although this form of evaluation, principal observation, can be improved somewhat with training.

The reasons presented in the data, as given by six teachers and six principals from different settings, used to avoid the approach which uses objective data seem to further undermine the level of teacher professionalism because they fail to stand up under logical scrutiny, which, actually, renders them as weak, if not feeble, excuses. One of the main claims seems to be that there are external factors responsible for student learning beyond the teachers’ control. However, those are the same excuses that could be used by groups as diverse as salespeople and heart surgeons, who are nevertheless almost uniquely evaluated and rewarded on the basis of tangible results that are not under their direct or absolute control. In those domains outside education, those or similar excuses are no longer offered nor are they accepted. Furthermore, cardiac surgeons, whose failures would seem to be excusable on grounds of external factors, are able to be judged by their successes and failures and are openly compared to other cardiac surgeons once the “risk factors” of their patients are considered in the assessment of their skills (see page 353). Again, if the excuse of “external factors” was legitimate, here would certainly be the place
to apply it, but, in general, it is not accepted as an excuse viable enough to avoid rigorous
evaluation in the same way the data here tells us teachers and principals, as a means of
avoiding evaluation, accept it in their professional culture.

The other main claim offered in these data to avoid using this type of more objective
data is the matter of teaching to the test. This, as has been argued in the main body of
research, is fairly and easily dismissed. In fact, for the first time, contributors to the
literature itself are re-evaluating their positions regarding the legitimacy of using “teaching
to the test” as a bona fide reason to avoid using student scores on standardized tests in
teacher evaluation processes. Some of the recent literature seems to be adopting the stance
that teaching to the test is acceptable and even commendable if it is a test worth teaching
to (Wiggins, 1989). Apparently “teaching to the test” is no longer considered that serious
a stigma for it to be adjudicated as categorically detrimental by definition. Evidently the
literature is beginning to distinguish between “teaching to the test” and “teaching the test”

itself. Other objections, such as the lack of statistical validity and the inherent
restrictiveness of the standardized-test format have also been dealt with. Mathematically
stringent statistical validity is probably impossible in any evaluation situation and because
a test is standardized does not intrinsically imply a narrowness of either scope or format.
While there are still problems to work out in implementing such a system, here, I am
merely pointing out alternatives to the generally accepted way of thinking about these
issues.
Showing teachers their evaluation forms and apprising them of their visitation date, prior to the actual *in situ* evaluation as stated and defended in this project's data, may also detract from the expected rigor and thus the professional veracity of the evaluation. This practice almost presumes that the teachers would be unable to perform properly unless they were informed beforehand of the specific expectations of the evaluation. While a cogent argument could be made for informing teachers what output would be expected if there was no danger of an artificially manufactured *ad hoc* result specifically displayed for demonstration purposes, it is entirely another matter (1) to tell them what style of teaching to pursue and what specific practices to perform then, (2), as an administrator or as an evaluator, to observe those styles and practices on what may be, in fact, the only occasion when those repertoires of teaching skills are ever used. Even disregarding these last points, at least if the teachers were not apprised of where and when they were to be observed or if the teachers were observed more randomly and discreetly (for example, by video camera) then the evaluators would be more certain that the observations were legitimate, as opposed to being staged. As the system now operates, there is at least a reasonable chance that some of the results observed are affected. Of course, the counter claims to these points should not be summarily dismissed as totally lacking in credability. It can also be argued that this might be a completely viable way to deal with professionals. Most professions are informed of the criteria (but usually not ones that can be affected) so that they can improve professionally and be promoted.
Aggravating this problem of equivocal professional standards are the data's revelation of the number of visitations respondents say the evaluations are based on. The vast majority of principals and teachers interviewed agreed that the amount of visitations were numerically inadequate to formulate any fair or accurate data. Why then do they accept those evaluations based on an inadequate number of observations as being of any real value? In their defense, principals claim that this problem of too few visitations is compensated for by intermittent informal visits which allow them to “take-in” the appropriate information. However, as expressed in the main body of data and analyses, even if these reports are accurate, and if such “informal” visitations were intentionally and consistently carried out for observation and assessment purposes in the regular course of events, they cannot compare, in the judgement of this researcher, to the more formal official visitations. Those formal visitations, themselves flawed, are usually more substantial and require written accounts of the evaluation which at least contain formal criteria that attempt to instill a degree of discipline and accuracy in their record of detail while the informal visitations often last for less than five minutes and seem totally based on personal impressions. Combining the two types of visitations would probably act antagonistically against any scientific rigor to skew the data even further away from any type of accurate result or legitimate professional judgement and more into the realm of impressions or feelings. As an example to bolster this point, would any principal accept a teacher’s marking scheme for a student based on one test only but supplemented by a
general impression of the student from informal classroom observations? It is highly unlikely that such a scenario would be accepted. And, as mentioned previously in this research,(page 125) this is exactly what is responsible for the “halo effect” which is a well established methodological flaw that, when present, diminishes the likelihood of fair and accurate judgement.

Among the respondents presenting the data in this study, the post-evaluation interview also, generally, appears to be misused and, therefore, would seem to detract from the professional attributes of teacher evaluation. The most one could reasonably expect from that interview would probably see it being used as an opportunity to explain the reasons for any serious concerns the evaluator has expressed and to provide explanations for any questions or problems the evaluatee has raised. It may invoke setting up future meetings, scheduling re-evaluations, modifying the evaluation itself or arranging a schedule for improvement. This is the way a serious professional evaluation would be conducted. But the majority of teachers, as outlined in the data here, say that their post-evaluation interviews take five to ten minutes and are not done very seriously. Arguably, these types of deficiencies result in the loss of teacher quality, professionalism, integrity and respect for the entire evaluation process.

Besides the problems of low standards of professionalism, *per se*, already listed, the apparent lack of consequence and effective change resulting from evaluation as suggested in these data are also systemic professional weaknesses in teacher evaluation. These
conditions ultimately detract from evaluation efficiency and undermine the formative aspects while destroying the summative effects of evaluation. According to both principals and teachers, evaluations, as presently conducted, are seldom a serious consideration for promotion or even seen as very valuable for summative or formative correction. And, as the preponderance of the data suggests, most teachers claim that their evaluations would not successfully distinguish between either the competent or the incompetent teachers, and certainly not between the competent and the excellent or stellar teacher, therefore failing to meet even a basic level of consequentiality.

The information from the data also implies that teachers and principals agree that the process of teacher evaluation, already flawed in and of itself, is applied indiscriminately. This implies that all teachers, regardless of the length of their tenure in the profession or their specialized designations that places them outside the classroom, are usually subject to the same evaluation format. Yet the duties of non-classroom professional personnel are so intrinsically different from the classroom teacher that any information coming from that type of general evaluation would probably be incompatible with the actual situation, if not blatantly meaningless. Indirect evidence from other areas of the study, to be discussed later in this section, also seem to support these charges of inconsequentiality. For example, most of the teachers responding to the questions about the collective agreement's effect on evaluation said it was too limiting. The consensus from the teachers seems to indicate that either the entire collective agreement or specific clauses
within it were forcing the evaluators to restrain or completely bowdlerize their criticisms until they became so equivocal that they were almost trivial. As a result, in their cases, evaluation suffers as a proficient management tool and as an effective, consequential formative and summative instrument.

Other problems include the lack of demarcation between formative and summative processes and teacher attitudes. While most of the literature recognizes the obvious incomparability of the formative and summative processes, most of the principals and teachers in these data do not appear to be even cognizant that such a conflation of processes would be problematic. As a result, the entire evaluation procedure is flawed due to this lack of professional clarity as to exactly what the process should be accomplishing. Aggravating the entire situation are the attitudes of teachers generally and specifically with regard to derogatory evaluations. The general attitude of the teachers, although possibly merited, is still largely an unprofessional level of contempt for their evaluation system, which they see as something perfunctory and bureaucratic, if not detrimental to the actual business of teaching. This problem is further complicated by a belligerence when actually confronted by a derogatory evaluation. It would seem that these attitudes, both highly unprofessional, buttress each other and are equally detrimental to effective professional evaluation.
Communication

Communication among the various levels of the educational organizations also seems to be a major impediment to effective evaluation for the 12 respondents. As the system now exists, as described by the data in this qualitative study, lack of communication is a continuous and serious problem which pervades most of the dialogue between the study’s teachers and its principals and between the principals and higher levels of management. Some areas where this phenomenon is especially trenchant, according to the information in these data, even to the point of affecting the institutional culture, are the following: subjectivity of evaluation criteria, teacher attitudes toward evaluation, post-evaluation interviews and the criteria used for evaluation. Most of the above, in addition to the inherent weaknesses they contribute to the evaluation system, also represent serious breaches in communication between teacher and principal. These gaps reveal that the dichotomy in perception, wherein each group, principals and teachers, have diametrically opposed views, even regarding principals’ beliefs about teachers’ attitudes, is so pronounced as to make miscommunication, in this researcher’s judgement, the most reasonable explanation as an accurate description of these data. Only two categories stand out as exceptions to this communication breech: the criticism of evaluation training and the small number of evaluation visitations on which teachers and principals agree are inadequate and seem to be aware of each other’s views. Those other areas, therefore, where that agreement is lacking, and the two groups seem not even to be aware of each
other’s views, might suffer primarily because of poor communications between management levels.

While I can probably assume, then, that there is a systemic communication problem, the difficulty, as defined by my data, lies primarily in the communications between principal and teacher. In those circumstances discussed in the data, the principals often base their views on their alleged insight into the teachers’ outlook or a problem primarily affecting the teachers. An example of this type of miscommunication is the principals’ comments on teacher attitudes toward the evaluation process. There, the principals claimed the teachers had highly positive attitudes while, in contradistinction, the teachers themselves almost unanimously had profoundly negative attitudes and found the process frivolous, highly frustrating and even demeaning. Communication problems like those mentioned above may be indicative of an organization improperly managed and not well organized within the parameters outlined by modern management philosophy.

The Collective Agreement and Its Restrictive Effect on Evaluation

The near unanimity of opinion in the data, from both teachers and principals, criticizing the collective agreement’s role in the evaluation process came as a surprise to this researcher because of its strong component of teacher disapproval. Although agreeing that it protected their rights, teachers were still strongly against the collective agreement in its present form, thinking that it was too restrictive towards the evaluation process’s
summative component. Principals were even more strongly opposed to the current format of the collective agreement. This should not imply, categorically, that either principals or teachers were opposed to unions, federations or collective agreements or that they offered no benefits. What the respondents were against, however, were the types of collective agreements their boards habitually have negotiated, which are restrictive enough to render evaluation, especially summative evaluation, ineffective. They suggested that, as a result, it has become almost impossible to dismiss any teacher, even the most incompetent ones. This seems to be the general opinion, although no definitive data could be found in the literature or in accessible records delineating how many teachers have been dismissed in Ontario for poor teaching skills or technical incompetence as opposed to some form of criminal negligence. Probably with some degree of overstatement, one principal went so far as to say (based on a course she had taken) that, she had been informed, that in Ontario, there likely have not been any permanent teachers dismissed for incompetence and certainly none as a result of poor evaluations. If this is due to an overly restrictive collective agreement, it is plausibly one of the reasons why evaluation might be held in some degree of contempt by the teachers in our publicly financed schools.

Analysis of this Study and Its Discussion: Focus on the Literature Review

As already mentioned in the literature review, management philosophy in North America has gone through several philosophical evolutions which contain a range of
models. Transformations began with mechanistic models such as Taylor’s (1911) and continued into the humanistic models of Mary Parker Follett (1941) and Chester Barnard (1968). But it was the failures of these models to live up to industry’s expectations that provided the impetus for hybrid “state of the art” philosophies, influenced considerably by other cultures, for example, the Japanese culture, to become accepted as the dominant philosophies of the late ‘80s and ‘90s. These “cutting edge” philosophies, like lean production, T.Q.M. strategic planning, site-based management and management by objectives, while distinct from each other and sometimes even partially antagonistic, nevertheless advance a core of similar values. Some of these principles are manifested by new conventions like (1) external clients’ importance, (2) reciprocal dedication between the workers and the organization, (3) open communication, (4) tangible measurement of results, (5) quality circles and (6) internal clients’ consequential status have helped management focus on problem solving and on a commitment to delivering a quality product, that is, excellence. An analysis of the data of this project’s interviewees alone reveals a gap between acceptable modern management principles and those conventionally practiced by administrators at the school and board level. (For a graphic illustration comparing modern management theories and management of teacher evaluation see Figure 5., Criteria of Modern Management v. Philosophies of Modern Management and Teacher Evaluation (Data Based)). Some of these philosophies (like Deming’s classic TQM
paradigm) were opposed to individual employee evaluation. However, that rejection need not be a categorical exclusion, as Sallis has stated:

Merely because Deming is an opponent of [individual] appraisal does not mean that the two [systems] are incompatible, but it does require that special consideration is given to the way appraisal is conducted to ensure that it does not lead to the effects [infiting, loss of team spirit, loss of pride in work] abhorred by Deming. (Sallis, 1993, p.47)
Figure 5.

CRITERIA OF MODERN MANAGEMENT v. PHILOSOPHIES OF MODERN MANAGEMENT AND TEACHER EVALUATION (Data Based)

<table>
<thead>
<tr>
<th>Criteria of Modern Management</th>
<th>TQM</th>
<th>Lean Production</th>
<th>Strategic Planning</th>
<th>Benchmarking</th>
<th>Management by Objectives (MBO)</th>
<th>Management of Teacher Evaluation (Data Based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pride in the Enterprise</td>
<td>✓</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>_____</td>
</tr>
<tr>
<td>Clearly defined set of performance measures</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>☑</td>
</tr>
<tr>
<td>High risk – high reward culture</td>
<td>Only for groups</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>✓</td>
<td>_____</td>
</tr>
<tr>
<td>Employees recognized and rewarded in proportion to their personal accomplishments</td>
<td>_____</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>☑</td>
<td>_____</td>
</tr>
<tr>
<td>Constant recognition for achievement</td>
<td>Groups but not individuals</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>☑</td>
<td>_____</td>
</tr>
<tr>
<td>System-oriented evaluation &amp; planning</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>_____</td>
</tr>
<tr>
<td>Client satisfaction (internal &amp; external)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>Not focus of philosophy</td>
<td>_____</td>
</tr>
<tr>
<td>Continuous improvement</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>_____</td>
</tr>
<tr>
<td>Continuous training</td>
<td>✓</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>✓</td>
<td>Not focus of philosophy</td>
<td>_____</td>
</tr>
<tr>
<td>Consistency between policy and practice</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>_____</td>
</tr>
</tbody>
</table>

Sallis here seems to recognize an issue that this thesis itself raises. Essentially he acknowledges that even a team-based unit approach to management maintenance of quality might require individual evaluation. He recognized that in a service-oriented industry the junior employees usually render that service directly to the client, with senior staff generally having little direct contact with the customer. Thus, the quality of the "product," which in itself is often difficult to measure, becomes more dependent on that individual representative than the team and, therefore, it is the individual who needs to be evaluated even more so than in other industries.

The intangibility of the product might in itself be one of the prime reasons why most of the classic models of evaluation are not oriented directly or applied to teachers, but to programs. Nevertheless, as suggested in the Literature Review and by Sallis (1993) himself, some of the classical models could be adapted to successfully encompass teacher evaluation.

Certainly the general principle on evaluation espoused by Scriven proclaiming that "[the evaluator] has to know how to get to the bottom line," that evaluation has to be done in the particular way that it "reduces facts and values to evaluative conclusions," is applicable to almost any serious professional evaluation. While for the most part this does not appear to be fulfilled for this study's respondents, even such aforementioned evaluation models, cited previously (pages 76-85) as the Behavioural Objective Model (which compares pre-determined outcomes with actual outcomes), the Transaction Model (which
ensures the accessibility of information to all stakeholders), the Adversarial Model (where two sides might present opposing views of the teachers' skills, results, attitude and so forth) and the Art Criticism Model (where the process includes description of qualities, interpretation of employed methods and evaluation of the goals achieved by extant policies), used primarily as paradigms for program evaluation, represent potential models for teacher evaluation and for improvements over the present teacher evaluation systems in Ontario. It should also be duly noted that since these models are oriented toward some measure of result and away from process and style alone, they have the potential to offer considerable improvement on the present systems.

To further clarify how these models would work as processes of teacher evaluation, in the Adversarial Model one appointed evaluator would argue the flaws of the teacher while one would make a case for that same teacher's excellence. In the Art Criticism Model an evaluator would, using pre-determined criteria, estimate how close the teacher came to following procedures, how successful their methods were and how many of their teaching goals were achieved. In both those paradigms, there is at least some level of objectivity either by aiming towards a clearly defined goal (Art Criticism Model) or by having an opponent trying to refute a set of conclusions (Adversarial Model). Although these models do not solve all the objections raised, they at least supply a level of evidence that seems, for the most part, to be absent from the dominant model of evaluation, principal observation.
Nevertheless, if the problems discussed in the data section, which would appear to be largely due to organizational policy are not corrected, any model chosen, focused, as it is, just on evaluation alone, would provide little systemic improvement. As the modern management theories mentioned have claimed, and the analysis here of the data seem to confirm, most problems in management are systemic and therefore start at one of the management levels of the bureaucratic levels of the organization. The implication is, then, that many of the major failings within the teacher evaluation systems in Ontario originate with the management's incompetence. A review of these data reinforces that point by enumerating the following problems: inadequate visitations, lack of evaluation based promotion, lack of attention to organizational culture and lack of professionalism. These are all problems of systems management and usually beyond the control of the individual employee, teacher or principal, nor would they be solved by changing to a different model alone.

Further, management incompetence in evaluation is the marked contrast between business (which generally does not have these problems) and education. Mintzberg, as early as 1979, noted the rigorous consequential approach of business organizations. He claims that many of these aforementioned problems and attitudes are a direct result of the perspective taken on teacher evaluation, which allows it to remain process-oriented as opposed to business evaluations, which are staunchly product-oriented. This is not meant to imply that business does not or should not investigate process. However, to the extent
that it does investigate process, it does so only in the service of product. There, in business, clearly process serves product and is subservient to it. Any process that does not produce, at least, adequate product, would be discarded in any system that relied on the business model of evaluation.

While my data have not shown that this is the reason for the inconsequential attitude, they have shown a fierce resistance by both the study’s teachers and principals to the use of product-oriented data with weak reasons offered in defense of that position and an equally redoubtable systems-based inertia in favour of not firing the incompetent teacher, which would be much more difficult to maintain with a product based approach. In corroboration, Bridges (1986), commenting on the American situation, writes about the difficulties that have developed there. (So far as I know, no published data in the literature exists for the situation in Ontario.) Bridges’ main contention is that the difficulties involved with the dismissal of incompetent teachers are largely due to weak contract stipulations that define the grounds necessary for dismissal so vaguely. Specifically, Bridges maintains, as these data here also imply, that the negotiated collective agreement is a major problem. The mitigating solution he proposes is a clause that accepts clearly outlined objective outcome-oriented evaluation for summative purposes. This could result in eventually clearly outlining what conditions of academic incompetence would merit dismissal (Bridges, 1986).
In fact, collective agreements, the *Education Act* and even board policies continually dilute any rigor of evaluation processes. This process seems to have been begun by an apparently innocuous request. Initially, the unions merely insisted on inserting a "just cause" clause into the collective agreements, prohibiting unfair or capricious dismissals. As a result, since 1983, it would be almost impossible to find any collective agreement in Ontario without such a clause (Stockman, 1984). This allows any dismissal not in compliance with "just cause" to be challenged by a grievance procedure or by court action.

But this researcher's complaint does not lie with the demands of the "just cause" clause as it was originally implemented. Rather, the problem is the further limits set on evaluation restricting the employers, that is, the boards of education, from establishing a consensual "just cause." Collective agreements now restrict such basic components of evaluation as frequency, duration and criteria of evaluation. Nor is it uncommon to find collective agreements, either directly or through its acceptance of board policies, restricting what type of evaluative document can be placed in a teacher's file, the procedures for placing that material in the file and defining the manner in which a teacher can record a refutation or even dismiss an evaluative document in their file. In that same manner, some collective agreements set the number of people that must be present for a bona fide evaluation and most give the teacher an almost automatic right to grieve any negative evaluation. While many industrial unions give their members a conditional right
to grieve a negative evaluation, few give them the right to record any refutation, let alone a contemporaneous dissent on the evaluative document itself (Michaud & Richard, 1981).

Therefore, despite advances made in teacher evaluation by such systems as STAR (Hill, 1991; Million, 1987) and the Criteria of Effective Teaching (CET) (Vogt, 1984, as cited in Clark, 1993) or even the use of student grades or opinions, peer review or portfolios for evaluation, the possibility exists that the present policies, by-laws and collective agreements in Ontario tend to impede the rigorous implementation of almost any system of evaluation. As an example, instituting merit pay in Ontario, based on successful evaluations, would most likely be impossible. The unions’ strong and chronic support of a seniority-based grid system in which teachers are rewarded proportionately to the number of years of service almost certainly excludes their support for a merit-based system as a viable possibility. That in itself diminishes the potential for using almost any evaluative system in a consequential manner and also limits the possibilities of expediting any definitive action mandated as the result of a summative evaluation.

Similarly, many of the proposed important innovations in evaluation, like student evaluations of their teachers or the use of student scores on standardized tests for teacher evaluation (which admittedly should be approached with caution), so far as I know, are still not practiced even on a trial basis in Ontario. Even in the United States, where a more liberal attitude towards this type of product-focussed, or non-hierarchical, external evaluation seems to exist, educational establishments generally refuse to incorporate those
options as wholesale solutions and, therefore, even there, evaluation policies based on criteria external to the teaching process are not widespread. As I have discussed in the Literature Review, under the general guidelines for successful evaluation, the collective agreement is not the only major political or organizational impediment to effective evaluation. Conditions such as (1) the competence of evaluators, (2) cultural compatibility between the organization and the evaluation system, (3) the need for closer correspondence between evaluation criteria and teacher responsibilities and (4) consistency between practice and policy are all necessary conditions for successful evaluations but whose absence in Ontario's school management systems creates serious problems, if my data are to be believed. These criteria were derived primarily from the literature, but my own data suggest that similar problems contemporaneously exist and also accurately summarize the flaws in the evaluation systems in Ontario. If my data hold up in a larger context, it is probably safe to conclude that when something as fundamental as the training of evaluators is claimed to be inadequate by the majority of teachers and some principals discussing their own qualifications, a serious problem may exist with the technical aspect of evaluation. Essentially, our data raise questions whether confidence exists in the value or effectiveness of evaluation or in any of the organizations presently implementing them.

To be fair, however, a more balanced assessment of the collective agreements might concede at least some of their restrictive conditions as a necessary precaution against certain aspects of civil law that are weighted in favour of the supervisor, that is, the
principal. It would be derelict of this discussion not to mention those arguments as evidence in support of the protective aspects of the collective agreement that could possibly prevent potential misuse of evaluation.

The unions, at least to some extent, are reacting to legitimate fears they have about the strong authority that the law in Canada, in general, and specifically the interpretations in Ontario, have given supervisors under the aegis of “qualified privilege.” Referring back to the legal discussion in Chapter 5, slander, which when written is termed libel, to be rendered actionable only has to be defamatory (that is, it must lower the person in the estimation of others), directed at the person beyond a reasonable doubt and be false (Gatley, 1981). Nonetheless, it is still, in the case of evaluations, difficult to prove. Apparently, supervisors, because of their legal and professional duties that entitle them to qualified privilege, have much more latitude to be critical than the average citizen in a non-supervisory role (McGuigan vs. Davidson [1985]).

As a result, in order to prove defamation, a sine qua non is, in most cases, to establish malice. To do that, the plaintiff must refute the assumption of innocence the law grants evaluators because of qualified privilege (Korach vs. Moore 1991). This, apparently, can be done most successfully by showing that the supervisor lacked honest belief in the truth of the statement (Korach vs. Moore 1 OR ed Ed., [Jan.2, 1991] at 281). One does not have to prove beyond a reasonable doubt that the supervisor made statements knowingly untrue. However, the plaintiff does have to establish that the defendant was
reckless. Reckless, under Canadian law, means indifference to the truth or falsity of what was said (Horrock vs. Lowe [1975] A.C. 135, [1974] 1 All E.R. 622 [1974] 1 W.L.R. 282 H.L. at 153 A.C.). What complicates this task of proving the plaintiff was “reckless” is that reckless cannot be equated with carelessness, lack of judgement or even irrationality (in thought). Even if the written statement of evaluation was proven totally false, that still is not enough to prove defamation, unless the plaintiff can prove the defendant did not believe what was written. This being the case, the standards for proving libel and receiving damages as a result becomes almost impossibly high. Therefore, initially, it is with some justification that the unions came to insist on stringent requirements for writing negative reports and on specifying what can and cannot be said in an evaluation, having, in their opinion, almost been deprived of the legal option of a civil suit.

Nevertheless, so far as is known, it remains, that as of 2000, teachers in Ontario are almost never dismissed for technical or professional incompetence. Furthermore, the consensus among teachers and principals according to the data of this study is that the collective agreement is too restrictive to afford effective summative evaluation. Although this study’s data base is small, it would seem plausible based on the information available here and the professional literature that the balance of power may have shifted too far in favour of the teacher when it also protects the blatantly incompetent teacher. This “protection” from evaluation must then be relaxed to the point where summative evaluation can become a more effective and consequential management tool.
REFERENCES

Adam v. Ward (1917). A.C.


Johnston, R.C. (1998). In Texas, the arrival of spring means the focus is on testing. *Education Week, 29*, 20-21.


Pupil progress as a criterion: Should it be used? (1979). NJEA Review, 52, 8-10.


Revised regulations of Ontario 1990. v. 3., p. 393.


World Wide Web: http://www.oct.on.ca (June 18, 2000).


APPENDIX A

A SPECULATIVE ESSAY ON IMPLICATIONS FROM THE DATA WITH RECOMMENDATIONS FOR FURTHER STUDY

Introduction

Teacher evaluation, in Ontario, as discerned from the remarks of 12 educators, runs the risk of being largely ineffective as either a summative or formative evaluation process. It appears to be used neither summatively, formatively or for accountability purposes to any consequential degree. In support of that view, I can show from the data teachers consistently claiming that evaluation has no purpose beyond the appearances of organizational requirements and, indicative of the rectitude of their position, is what appears to be the generally accepted practice that ignores evaluation as a factor in promoting teachers into positions of responsibility such as curriculum leaders or department heads. Yet despite these possibilities of mismanagement and incompetence, a strong mythology of denial continues to persist. Among the study’s respondents, this defense mechanism is used inveterately to defend the present practice, which has been proven both from the literature and my data to be ineffective or only marginally effective.

The following questions then remain: “What should evaluation be used for?” , “How could the situation be improved?” and “How could evaluation be applied more productively?” In attempting to begin to answer these questions, I am going to use my data to elucidate the nature and scope of the problems within teacher evaluation and subsequently recommend studies or areas of inquiry, which if conducted or investigated, might help develop more effective evaluation schemes, ones capable of improving teaching performance. Still, it should be clear that the recommendations do not emerge arbitrarily or capriciously but will be based on the management problems presented from the data of this study. They are essential aspects of the demythologizing process that emerged from the data inherent to the infrastructure of this thesis. By offering alternatives, they serve to further refute any exculpating claims, not only those that assert the validity of the myths championed by the excuses used, but also those that concede the complaints against the present system and defend them by using the myths of resignation that claim a lack of any viable alternatives available to replace the extant systems despite acknowledged deficiencies. In this appendix, in the following paragraphs, I will be discussing both the problems with the management of evaluation and the recommendations for improvements. In the following essay sections of speculation, I will focus on summarizing the evidence for the implications, that is, their connection to the data and, subsequently, how the recommendations are justified from the problems the data has presented to us.
The Connection to the Data: A Summation of the Evidence

The implication that, presently, principals may lack professional training as evaluators was based on data offered by this study’s interviewees. That aspect of unprofessionalism was not seriously challenged by the majority of respondents, teachers and principals, and was allowed to stand — initially as a provocation-hypothesis and later as a statement based on the data of the legitimate research of this project. The overwhelming majority of both groups concurred that the evaluations were not administered professionally and conceded that the training for the evaluators seemed inadequate. Therefore, the derived implication that followed was almost directly based upon the data. Additional, if anecdotal, evidence to buttress this claim was also mustered to explain why the study’s principals may not be able or qualified to act as effective professional managers. The concomitant recommendation, therefore, actually did follow from the data as well, but only in the sense that it was a suggestion that might remedy the problem presented in that data.

The idea of an unprofessional culture can also be seen from this study’s data despite the fact that the interview questions do not blatantly ask about the presence of unprofessionalism. It was felt that the subjects would react too defensively if they were asked those questions directly. Arguably, however, procedures that are seen as flawed by this researcher such as previewing of evaluation criteria, unprofessional emotional responses to derogatory evaluations to the point where teachers do not consider poor evaluations as acceptable, an overly protective union that appears to exclude all but the most benign criticism, lack of meaningful professional training of the evaluator and the inability to dismiss anyone except the most blatantly incompetent could be interpreted to define an unprofessional culture because of its tolerance for apathy and mediocrity as shown by these data-based criteria.

This implication was also further supported, externally with auxiliary evidence, by the folklore of the general cultural observations that are commonly acknowledged by teachers, administrators and educational researchers in that field. As alluded to in Chapters 5 and 6 and again, but in more detail, here, allegedly accepted perquisites of professionalism as individual employee offices, on-site authority without the second-guessing challenges of superiors and access to assistants for para-professional or support work are all absent.
Regarding the use of standardized student exams for teacher evaluation, our study does provide us with data. Here, on this topic, the purpose of the interview questions was to see if the teachers and the principals were opposed to this type of evaluation, and, if so, why they rejected it and what were they advancing as an alternative? Although much of the literature supports certain of their objections, nevertheless, as I have shown in the analysis and the literature review, no reason offered by this study’s teachers or principals to justify their resistance can be said to stand up very well to the scrutiny of this project’s system of logical analysis (as defined earlier in Chapter 5). As for incentives to improve, the data inform us that, for the most part the respondents believe that any good evaluation received would not be related to better positions, promotions or monetary compensation. Furthermore, the data on the evaluators from principals and teachers overwhelmingly inform us that their training is inadequate for the task of professional teacher evaluation.

The evidence from the data for a lack of communication is the blatant dichotomies over so many crucial issues that situate the principals and teachers at antipodes (see page 302-303). Given that most modern management systems suggest the importance of attentiveness to front-line workers (see Figure 6.), there should not be so many categories crucial to evaluation that divide the workers (teachers) so categorically from the executives (principals). This discrepancy in communication, then, highlights a systemic management problem crucial to evaluation.

As discussed in the previous chapter, there is evidence from the data in this study that the respondents believe that the collective agreement in its present format, acts as a strong deterrent to efficient summative evaluation. While teachers and principals concurred that the collective agreement protected teachers’ rights, both groups strongly opposed the collective agreement as it now exists. Both groups felt that it was too restrictive and therefore was too often instrumental in protecting the incompetent. Not surprisingly, the principals were even more adamant in their opposition to the present limiting stipulations of the collective agreement. It is not, however, possible to supplement those data with strong evidence from the literature. However, mentioned previously in comparing the data with the literature review, while very few articles are written about the Canadian or Ontario situation regarding the collective agreement, what few do exist support the general views of the data here on their detrimental effects. I have, however, shown to some degree, that a certain level of restriction does exist by citing some actual collective agreements (page 215).
The recommendations that follow here are suggested to correct or ameliorate the professional culture, evaluation and effectiveness of teachers and, concomitantly, their professional status and image. Although it may be argued that the recommendations are premature, I must stress the importance of remembering that these recommendations are not utterly divorced from the data. Neither are they data-based, however, in the same sense that the outlined deficiencies in management are. The problems they attempt to address have all been outlined in the data section and those problems can be said to have emerged from those data. If there is any controversy over the authority to present recommendations, it would be directed against claiming that the recommendations themselves were data-derived. That claim, however, is not being made. The recommendations only exist as extensions of the core data — as hypotheses for future studies, and so should be taken in that light. This can be seen through a schematic representation of the entire development of the logic, methodology and findings, (see Figure 6., Flow Chart of Project Stages, page 302).
<table>
<thead>
<tr>
<th>Project Question</th>
<th>Sub Questions</th>
<th>Provocation - Hypothesis</th>
<th>Development of Questions</th>
<th>Processes of Methodology</th>
<th>Implications and Recommendations</th>
</tr>
</thead>
<tbody>
<tr>
<td>With regard to the group of principals and teachers being studied, what is the state of teacher evaluation in the Ontario public school system; what are the elements which enhance or restrict effective evaluation; and what is evaluation's compatibility with the general guidelines on education and selected modern philosophies of management?</td>
<td>What is the nature of teacher evaluation (in the schools being considered) as determined in data collected from the interviews of principals and teachers regarding evaluations' fidelity to their objectives, standards of efficiency and values?</td>
<td>Indifference of Management</td>
<td>Interview Questions (See Appendix A)</td>
<td>Lack of Professional Culture</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Do the data reveal compatibility between the administration of evaluation and the theories of modern management?</td>
<td>Absences of Incentives</td>
<td>DATA ANALYSIS DISCUSSION</td>
<td>Culture and Professionalism</td>
<td></td>
</tr>
<tr>
<td></td>
<td>What are the views of teachers and principals on the factors that restrict or enhance evaluation?</td>
<td>Dearth of Objective Data</td>
<td></td>
<td>Unprofessional Evaluators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Do principals and teachers reveal any important data about the conduct of teacher evaluation with respect to the accepted guidelines of the general principles of good evaluation?</td>
<td>Unprofessional Behaviour</td>
<td></td>
<td>Professional Evaluation And Evaluators</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Are there any important differences between these views of teachers and principals?</td>
<td>Institutional Weaknesses</td>
<td></td>
<td>Accurate and Objective Data In Evaluation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>After the analysis and synthesis of the data of this exploratory study, what suggestions can the researcher offer to help advance the present structure towards a more effective and responsive teacher evaluation system?</td>
<td>All of the Above</td>
<td></td>
<td>Poor Communication</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The Collective Agreement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>All of the Above</td>
<td></td>
</tr>
</tbody>
</table>
Development and Justification of Recommendations

As I have stated previously, no conclusions, per se, appear in this appendix. The structure and methodology of the thesis, in addition to the type of data recovered in the study, do not allow for the right to formulate general conclusions; essentially, the rigor of evidence necessary to conclude is simply not there. To compensate for that discrepancy, a collection of suggestions and ideas for future study is substituted here in Appendix C, under the title A Speculative Essay on Implications from the Data with Recommendations for Further Study.

These recommendations are based on the problems raised by the data even if they cannot be explicitly seen as emanating from the questions that elicited those data. As I have already endeavored to illustrate, what does exist more clearly is a relationship between data and recommendations, such that the data illustrate a problem or demonstrate the unacceptability of an "established" excuse based on some assumed and unchallenged reason. In essence, in almost every recommendation, a mythology of denial is challenged by the suggested solution to be studied. The recommendations, then, are being offered as possible methods or alternatives for solving the problem, but only after they have been tested through the guidelines of a more rigorous and extensively planned study. This subsection has been added to clarify the precise etiology of the source and establish the connection between the data retrieved here and the recommendations cited here.

As will be outlined in this section on implications and recommendations, essentially, based on the data, I will focus on five broad areas of systemic incompatibilities, which were subdivided from the three general categories of page 291, that seem responsible for the present weaknesses in evaluation. Specifically, they were further subdivided from the general category of professionalism because the recommendations needed to correct these systemic incompatibilities are different for each of its sub-aspects, that is, evaluation and evaluators; culture; and reasons for lack of objective data in evaluation, which cannot be included effectively under the one category of unprofessionalism Therefore, in total, these categories are delineated under the following headings: professional evaluation and evaluators; general school culture and professionalism; objective data in evaluation; communication; and the restrictive nature of the collective agreement. Their broad connections to the data will be discussed in the following paragraphs.
Overview of the Sources for the Problems with Evaluation

As I have suggested, many of the problems of evaluation appear to lie rooted in careless administration and its resulting capricious attitude toward evaluation as seen through the general categories of communication, lack of objective data in evaluation, unprofessional evaluators, unprofessional school culture and the restrictive nature of the collective agreement. That perspective might even initiate with provincial law which does not have mandatory requirements for evaluation. This probably contributes to any widespread indifference and disregard for basic principles of good evaluation — such as the competency of those doing the evaluation, the evaluation system's cultural compatibility with the organization, consistency between practice and policy or explanation of the evaluations' importance for the employees' own professional development.

As my study starts to suggest, evidence of these failings are prominent in Ontario's present evaluation systems that I have examined. Consequently, speculation, and, in some cases, suspicion and anxiety, exist over how and when the College of Teachers with their general mandate for professionalism, will try and take responsibility for implementing teacher evaluation independently from the principals. Furthermore, it seems from my data, and if that can be found to apply over a larger, more diverse population, then it could be proven that the Education Act, collective agreements and board by-laws and regulations, all of which should strengthen the evaluation systems, instead present a serious threat to professional evaluation. Not only might they be seen as anachronistic and counterproductive, but they could inherently possess a distorted view of what professional evaluation should be in light of the insights provided by modern management systems.

In order to clarify the problem of poor management techniques in the Ontario school system, I will employ one common indicative example: almost all forms of modern management view evaluation and feedback as an important component of quality control and improvement, although admittedly subsequent interpreters disagree how that evaluation should be implemented. Yet in this study on teacher evaluation, only the principals consistently claim that the post-evaluation meetings, the main vehicle for feedback, are of long duration, meaningful, rigorous and significant for professional development, while the teachers almost unanimously refute that claim and assert that these meetings are very brief and often trivial. This represents a discrepancy in both professional evaluation and management.
Regardless of which group is believed in that specific instance, overall the evaluations do not appear to be used anywhere near their potential for quality improvement. As an illustration, such management structures as quality circles, rigorous goal setting or rewards for excellence could utilize individual or group teacher evaluations to analyze and eliminate general departmental weaknesses by learning to correct them as part of a general evaluation paradigm; however, this process does not seem to be practiced stringently or consistently. Evidently, the sophistication of modern management systems and their component strategies such as modern evaluation and its corollaries exceed in rigor and complexity most management strategies that are presently exhibited in the schools and boards in Ontario, if indeed the data of this study are representative of Ontario’s entire publicly-funded school system. It is time both of these subunits, management and evaluation, were updated, integrated and channeled more effectively to improve the administration of education.

In reference to the evaluation systems’ level of sophistication, this study does not really focus on the various specific types of evaluation systems and their component criteria because of the almost infinite number of possible variations and the overwhelming dominance of some form of principal observation as the general framework for teacher evaluation in Ontario, as mentioned previously on page 13. Those factors, format and criteria, nevertheless, are probably crucial in determining the level of an evaluation’s sensitivity. To a limited degree, however, a general impression of the sensitivity of teacher evaluation can still be perceived indirectly and obliquely from other topics dealt with explicitly by this study. Based on the number of visits (too few), the lack of specialization and focus (all personnel seem to be evaluated identically regardless of duties, experience or function) and the number of people who receive a failing or even negative evaluation (almost none), it seems that the schools used as reference points by the 12 educator-subjects of my study may be employing fairly unsophisticated evaluation criteria and a format that tests only for a basic, minimal competency, which might be able to be perceived intuitively, regardlessly.

Again, given the exploratory nature of this study, it would be wrong to categorically condemn that level of evaluation which has certain beneficial properties, especially within the summative component. But it might be time to go beyond that type of minimalism and work to develop a more advanced system that holds teachers to a higher standard and rewards those teachers accordingly. Clichés like “All our teachers are excellent,” seem to flow too glibly from the mouths of the six principals, in this study and just do not carry much credibility anymore, when they are said in juxtaposition with the large numbers of
sub-standard graduates (that is, not prepared for university or community college). In the "real" world, subject to authentic competitive labour and market forces, no enterprise that kept producing that great a volume of "defective products" would remain in business long -- unless, like the schools, it remained in a quasi-monopoly situation. This problem is mentioned here just as an example of the system's inefficiency and is also subsumed in more detail by the following discussion with its recommendations to counteract these problems.

These themes just outlined here, derived from the Data and the Analysis and the Discussion sections, are the basis for the implications and recommendations in the next segment. I will, therefore, now return to focus explicitly on some of those themes in more detail in the following sub-unit. As mentioned previously, the themes to be discussed include the following: (1) Professional Evaluation and Evaluators, (2) General School Culture and Professionalism, (3) Accurate and Objective Data in Evaluation, (4) Communication, and (5) The Collective Agreement. A discussion of these themes follows immediately.

Communication

In this study, the teachers and the principals, as evaluators and evaluatees, continuously contradict each others' professional understandings. This does not bode well for any modern management system (Cummings & Schwab, 1973) but is even a more seriously pernicious flaw within the teaching profession where communication, the ability to give, accept and comprehend information and feedback is usually considered a primary skill. On examination of this dichotomy, it appears that may not just be a difference in perception but an authentic demarcation along hierarchical strata. The study's teachers usually end up as critical of the system, whereas the principals usually end up supportive or even enthusiastic about the system. Examples of this incongruity from these data here exist in the following areas crucial to evaluation: (1) Adequate Visitation, (2) Detrimental Effects of Previewing Criteria, (3) Value and Sophistication of Follow-Up Interview, (4) Discriminatory Power of Evaluation (Competent vs. Incompetent), (5) Cultural Compatibility of Evaluation, (6) General Attitude Toward Evaluation and (7) Tangible Effects of Evaluation. Perhaps even more important are the reported results of these data when teachers are in a strong position to know the actual parameters of the situation. For example, one of my interview categories, General Attitude Toward Evaluation, is actually an investigation into how teachers feel about evaluation. The principals reported that the
teachers were satisfied and even pleased with the present evaluation system. However, in contradistinction, the teachers presented themselves as frustrated, angry and generally unmotivated by a system they considered inadequate to accomplish anything tangible.

In these situations the principals seem to be a systemic victim of what Hannah Arendt designates as “totalitarian communication” where the leadership, because of the culture of fear that predominates, has its communications filtered to the point where it hears only what it wants to hear. As a result, management would not be cognizant of the profound dissatisfaction with the evaluation system (Habermas, 1994).

While even from a heuristic “common-sense” approach this communication discrepancy seems detrimental to a healthy organization, miscommunication is even more antagonistic to organizational well being in light of modern management theory. Most theories of modern management regard both employees and customers as clients and only distinguish between them by the discerning epithet of external and internal. The impetus of that philosophy is to stress the importance of feedback. That feedback, and the required attention it demands, seems to be largely, if not totally, absent in the systems of management as understood by the study’s interviewees and reported by them in the interview dialogue.

Recommendations

As I have just outlined, poor communications between different levels of an organization are a considerable problem for any modern management system. This is true for its evaluation component in particular because of the importance of the formative aspect, the necessary consensus between management and labour for the criteria chosen and the feedback required to establish congruency between practice and policy (page 123-127). Yet consistently throughout my study, I have seen educators’ perspectives that demonstrate an absence of effective communication and a resulting inability to establish the fundamental requirements essential for good evaluation.

While it is absolutely impossible within the scope of these data to say with certainty the cause of this problem, the overview of these data does suggest that a systemic orientation may exist that is incompatible with voluntary, critical and vital communication.
Such a systemic disorientation could imply a possible cultural imbalance of power in management’s favour that requires a realignment of that constituent component in order to encourage communication and eliminate the rigid, structured hierarchy preventing communication exchanges. To facilitate this development, studies should be done to determine the feasibility of lowering institutional barriers between the levels of organization. Organizational structures that welcome teachers into management councils and not just tolerate them could be created. This could generate an atmosphere genuinely accepting of their contributions, both criticisms and approbations. Under these conditions, teachers and principals might be able to state their beliefs regardless of their agreement with present management policies. If and when this is accomplished, follow-up testing could then be done to see if communications improve. Additionally, assuming modern management theory is correct, if these recommendations are successful, they should act to improve general management procedures and the cultural environment as a whole as well as evaluation.

**The Collective Agreement**

While there are aspects of the collective agreement that are necessary like the just cause clause, it is the consensus of these data here that there are many other facets of the collective agreement so restrictive that they essentially impair efficient evaluation. Even the teachers, who, for the most part, agreed that certain restrictions in the collective agreement protected their rights, were almost unanimous in their opposition to its present structure because it seemed to encourage mediocrity, incompetence and a complacent attitude toward a job that could become more like a sinecure from which you cannot get fired. The principals were even more unrelenting in their opposition, suggesting that it was nearly impossible to dismiss a teacher for technical incompetence.

**Recommendations**

The data here clearly demonstrate a dissatisfaction with the collective agreement from both the teachers’ and the principals’ perspectives. It seems probable, then, that boards of education would not have these kinds of collective agreements if the principals and teachers negotiated the contracts. It is their employers, the boards of education and the teachers’ representatives, the unions, who negotiate these contracts. If, then, both the workers’ and the management’s main front-line representative fail to see the necessity for
the strict proscriptions of the collective agreement on evaluation, it might be an opportune time to re-evaluate this aspect of the negotiation process. As mentioned earlier in support of this view, the collective agreements, either directly or through their ratification of a board’s policies, presently restrict such basic components of evaluation such as frequency, duration, criteria, documentation, language and protocol. It is therefore the recommendation of this researcher that some boards, as a test case, in order to establish statistical validity, could agree on testing a contract that has a less restrictive attitude toward evaluation while still protecting the legitimate rights of teachers. It is the judgement of this researcher that, if the above recommendations are enacted, the protection now afforded to incompetent teachers will be, at least somewhat, diminished and the summative power of evaluation could be restored.

**Accurate and Objective Data in Evaluation**

As of this writing in 2000, there has been a marked increase in standardized testing locally, by individual boards and, provincially, by the Ministry of Education in Ontario. Yet this does not necessarily present us with a reason for the lack of objective data in teacher evaluation, in fact, it may highlight the question. At the present, there has been no decision to use these data as a direct or even formal criterion for the evaluation of teachers. In this study, using data in that way was almost unanimously rejected by principals and teachers. In fact, some principals, opposing the use of this type of information, claimed that learning and teaching were two distinct, independent processes and in no way connected. But if there is no connection, the implication to this researcher suggests that these principals may be claiming a lack of correlation between learning and teaching as one principal actually did ("That’s how the student did, not the teacher"). The logical conclusion of this assumption seems to imply an indifference to bad teaching due to its irrelevance as a learning tool and even an equivalence between good and bad teaching. Yet these are the same principals who evaluate teachers.

Because of the vociferous nature of the resistance itself by teachers and principals, which manifested as arguments against using the scores of standardized exams for students as criteria for evaluating teachers, I was not able, for the most part, to present the counter arguments in the interviews proper. As such, it would not be fair to use those arguments in the data and the analysis. Therefore I have presented them here because they are essential in understanding the position I have adopted.
To this observer, teachers claiming that they would be teaching to the test, if standardized student exams are used to evaluate teachers, as several teachers in this study have, is one such weak argument that can be refuted fairly easily. It seems that since teachers, under normative testing conditions, do not know what exactly is on the test and, as this researcher has argued earlier and will continue to reiterate in this text in refutation of those ostensibly advancing that argument as an excuse, if the test is well-constructed and fairly representative of a broad spectrum of the assigned curriculum material, which would not seem to present a formidable, let alone an insurmountable problem, the teacher is supposed to teach to the test. While there are obvious problems inherent to test construction such as reliability, validity, extent of curriculum coverage, cultural bias, clarity, and so forth, the view here is claiming those flaws as exactly what the psychometrists and cognitive psychologists in consultation with departmental heads and classroom teachers are supposed to correct. If this is professionally done, they are essentially teaching the course material. This point would, then, considerably undercut Darling-Hammond’s (1984) claim that the tests are too narrow to be of real value.

The answer seems to be, then, to construct better, more comprehensive tests that cover the material in a challenging, professional manner, that don’t just assess rote skills or narrow, simple aspects of the curriculum but include the higher-order thinking skills and focus that scrutiny throughout the curriculum. Darling-Hammond’s (1984) claim could even be seen as an indictment of tests in general; the question can even be turned back on universally accepted teaching procedures themselves by asking how they can legitimately evaluate students with tests? Are tests inherently broader in scope because they are not standardized? Along these lines I can see no intrinsic connection between scope and standardization or teaching to the test and standardization or even evaluation. And, if I were to further analyze the concept, it can be seen that “teaching to the test” is common even without the standardized mode. Do not teachers who have been teaching for several years regardless of the test’s format teach to the test they are planning to give? Teachers plan the curriculum and, to the best of my understanding, can usually be expected to teach based on what they are going to test. Presumably, few teachers would test something they had not taught, nor would many teachers spend any significant amount of time teaching something they were not planning to test according to the already accepted pedagogical routine. Given those assumptions, one is left wondering what exactly is the complaint about teaching to the test?

Of course, it must be emphasized that standardization is independent of format. In the judgement of this researcher, this “law of education” -- “It’s too confining to the
teacher or too insensitive to the developing student to teach to the test" — may be a classic example of a euphonic phrase that does not stand up under any analytical scrutiny (as I have just attempted to outline through argumentation). Unfortunately, I keep finding once a mind set has hardened, it is remarkable how much energy sometimes must be expended to refute what has been accepted as “fact” even in academia.

As for Ross’s (1984) claim of different innate student abilities, he himself suggests testing the innate ability of students first and then comparing that result with their academic achievement. Although this type of comparison seems highly fraught with problems, I have already suggested ways in which that might be used to start a process that eventually could overcome many of these problems. Similarly, I have tried to show flaws in the demand for rigorous, mathematical statistical validity — an impossible standard not available in any personnel evaluation situation that this researcher is aware of. In any event, more data are needed to conclusively prove either side of that argument. In social science, evidence is not established either through scholastics or argumentation but by collecting and interpreting data.

Regarding the claim of too many external variables, teachers here generally claim that they make an important difference to the knowledge level or its rate of acquisition level by the students through teaching them, and, therefore, would probably vehemently argue against just leaving a student with a pile of textbooks. Nevertheless, these same teachers, supported by some of the literature (Berk, 1988; Millman, 1981), claim that too many variables independent of the teacher impinge on the test’s validity as a measurement of teacher effectiveness. At present, data in this study suggest that the six teachers here believe their apparent lack of responsibility for those external variables, which they assume affect learning, exonerates them from any blame over low scores on standardized student tests. If so, why does teaching, in the view of those teachers, make the aforementioned valued difference in student learning? If there are so many variables beyond teachers’ control, one could ask how can teaching make any difference at all? Ultimately, then, it could be argued that it is not the countless variables being judged; what the test does judge is how much of the subject, regardless, has been mastered. This mastery by the student is what the teacher should be evaluated on, in the final analysis, if one truly wants to gauge rigorously academic teaching in the realm of public funded primary, junior and secondary education.
Whether we consider the valid domain within or outside the parameters of this study, likely few teachers would deny that conspicuous gains in knowledge are influenced by good teaching. From my perspective, as influenced by the literature, what teachers are supposed to be doing, for the most part, is increasing the knowledge level of the students. Academically, and according to the value system this researcher has put forward, the verification of that education is how well the students know the material taught. Can the same teacher, who claims so many extraneous factors in education exist independently of the teaching capabilities of even the best teachers and beyond the control of their students, give a grade based on tests to a student and claim that those grades fairly represent student progress in learning while contemporaneously disregarding those “essential” independent factors that they claim influence grades? At the same time, can that teacher refuse to let his or her own teaching be judged by similar criteria because it would entail “disregarding” external factors and yet still claim to be innocent of the logical fallacy of “special pleading”, because the teacher is in fact, “blaming the student” by exposing those same external factors. This researcher does not see this as a reconcilable position. Statements such as these above (pages 350-352) seem based on the fallacy that most assessment tools and standardized tests are poor indicators of academic performance. However, in fact, these allegations are unsupported and contrary to the findings of Sanders’ study (Wright, Horn & Sanders, 1997).

However, even if you accept extraneous social and economic factors as beyond teacher control, students can each be given a rating based on their socio-economic status, intellectual ability, family functionality, social skills, psychological stability and so forth. Essentially, these scores would encompass any factor that could effect their scholastic performance. Teachers from classes or schools with similar socio-economic make-up and similar scores or score distributions could subsequently be compared in longitudinal studies. Then, if a statistically significant difference results in standardized-test scores between the two groups, socio-economic or other external factors can be discounted, for the most part, as extraneous because of the similarity between the two groups. This is a classic and universally accepted paradigm for controlled studies albeit the double-blind and strict random selection conditions are not being met nor are they possible to attain in any pre-assigned context like a teaching situation, where class members are designated by extraneous factors, or a business or professional situation where people are pre-assigned based on business or professional requirements. These teachers and students, by definition, are not randomly chosen. Certainly, though, this comparative model meets most of the standards for the quasi-experimental paradigm at least as stringently as most of the other contemporary systems of evaluation do.
For those who, nevertheless, persist in claiming extraneous factors beyond a teacher's control as a valid reason for rejecting the use of student scores as viable for teacher evaluations, at least one study has been published using standardized test scores as the dependent variable with socio-economic and other social-environmental factors as the independent variables. These data were then entered into a multiple regression equation. The results showed extraneous factors, no matter how abundant or severely detrimental, could only account for 80 per cent of the standardized score as a maximum, while teaching, even under the worst conditions of socio-economic distress, accounted for (a significant) 20 per cent of the mark (Bingham et al., 1991). Furthermore, "three previous studies indicate that the influence of teachers and schools on the rate of student gain are independent of the confounding of socioeconomic factors." And, in fact, the reports released in April 1993 confirm that there was no correlation between schools' cumulative gains in five subject areas and the percentage of students in that schools receiving free and reduced cost meals—which usually implies lower socio-economic status (Sanders & Horn, 1994, p. 309). However it should also be noted that these are relatively new studies that have not had time to be tested more thoroughly. Furthermore the use of student scores to judge teachers represents a minority position and are opposed by many experts in the field, because of extraneous factors beyond teacher control as well as other reasons (Berk, 1988; Darling-Hammond, 1984; Millman, 1981; Ross, 1984) (see pp. 219-224). It should also be noted that in general the majority of researchers do not accept the conceptual framework that justifies using student scores to evaluate teachers under any circumstance, thus placing that practice in a distinct minority position among scholars, researchers and educators, even though it is used in Tennessee and Texas (Dallas). This will be discussed in more detail in the following pages.

Interestingly enough, though, even without those studies, research being done in the evaluation of cardiac surgeons, as reported in the New England Journal of Medicine (Chassin, De Buono & Hannan, 1996), has shown how successful evaluation programs for surgeons were implemented. In that article, the so-called external factors usually seen as beyond the control of the surgeons were utilized as components in the calculation of the level of risk in the operation and then used to adjudicate the skill of the surgeon. Contrary to what was expected, the surgeons who operated on the higher risk patients, despite the almost infinite number of external factors, were often the most successful. Surgeons who were less successful with high risk patients were delegated to easier, less challenging surgery. In the same manner, these "external risk variables" can be factored into any formula constructed to evaluate the teacher, based on their students' results. More to the point, as true professionals, the name and rank of the surgeons were published openly because the patients had a right to know. While surgeons have more control over their
patients than teachers have over their students, it can be claimed that the very fact that the most serious conditions do not generate the most fatalities should suggest to us that teachers, to some extent, have to become more responsible for the outcomes, that is, the learning of their students. Can professional teachers claim the same level of professional veracity and accountability? Can scholars convincingly refute charges that they are using "external factors" as an excuse to avoid evaluation when another profession evaluates itself based on results despite an even greater potential influence of external factors? Regarding other claims that the entire process of test construction, annual or biennial retesting and the hiring of special staff to supervise the process would be prohibitive is valid. However, if this were to prove successful would not the improvement in teaching, literacy and numeracy more than compensate for the expense?

Furthermore, some ground-breaking work is being done in the United States, under the "value-added" assessment designation. Although many variations of this process exist, the main characteristic of value-added assessment requires a "base-line" testing of students used to predict their probable score in order to see what value the teacher has added. In Dallas, for example, the school district used hierarchical linear modeling to predict the test scores of individual students while taking into account social and economic factors, usually considered extraneous to the classroom. There, schools that surpassed expectations were eligible for financial rewards. As the superintendent for school administration commented, "It takes away the excuses" (Olson, 1998). In a corroborative study done in Boston by Bain and Co., which based teachers' ordinal rankings on students' test scores, it was discovered that the most efficient teachers were able to improve student learning by a factor of 6 when compared with the least efficient teachers, despite students' identical base-line estimates of their potential. Similar studies with comparable results were conducted by Hawley at the University of Maryland and by Sanders in Tennessee (Olson, 1998). In New York City, the Board of Education, using many of the same methodologies as those studies above, has negotiated for itself the right to dismiss up to half the teachers of failing schools, as defined largely by standardized tests (Olson, 1998).

Referring again but in more detail to the above-mentioned Sanders' study, it appears to be especially valued. It uses a statistical mixed-model methodology which allows for a multivariate, longitudinal analysis of student achievement data and was based on a massive database of longitudinally merged student, school, teacher and school system information. It has demonstrated that it is viable for producing estimates of school and teacher effects free of external (or internal) socio-economic effects. The statistical formula does not require any direct measure of these alleged external variables to arrive at its results.
The initial data for this study was based on the work of Robert McLean and William Sanders in 1983, two statistics professors at the University of Tennessee who demonstrated that the following results were among those findings:

1. Measurable differences between schools and teachers regarding their effect on student learning exist.
2. The estimates of these effects tended to be consistent from year to year.
3. Teacher effects were not site specific, that is, scores could not be predicted by location alone.

To further substantiate the credibility of these value-added studies, in 1991, when the *Education Improvement Act* was adopted in Tennessee, the TVAAS formed an integral part of that legislation (Sanders & Horn, 1994).

Furthermore, the value of the TVAAS regarding its technical capabilities has since been confirmed by the Bock Report. Current data even suggest that the system is designed for and does have the technical capacity to determine the effects of individual teachers (Stronge, 1996) Essentially, both Darrel Bock of the University of Chicago and Richard Wolfe of OISE/UT have proclaimed that using TVAAS for the evaluation of individual teachers is technically appropriate from a statistical perspective but stipulated a caveat of a three year trial period, probably in anticipation of correcting any, as yet, unforeseen problems (Mayshark, Knoxville News Sentinel, February 20,2000)

It is also true that other researchers, like Goldstein, perhaps more oriented towards establishing a rigorous statistical model, have reservations because of statistical technicalities (Goldstein, 1991; Goldstein & Thomas, 1995). Additionally, in a provocative and incisive article entitled *Fifty Reasons Why Student Achievement Does Gain Does Not Mean Teacher Effectiveness*, Berk (1988) presents a host of external factors and statistical problems that seem to preclude using student scores as a criterion for teacher evaluation. Yet, what he ignores are the following:

1. There is no system of evaluation in practice that fits all the criteria of statistical validity - nor is there very likely one soon to be developed;
(2) some of the factors that Berk lists as problematic for influencing test results such as teaching to the test or giving general instructions in test-taking methods are really valid forms of preparation which teachers should use; and

(3) most of the external factors can be neutralized by using a comparative, longitudinal approach administered between two student groups’ scores as similar along a psycho-social and socio-economic continuum in a quasi-experimental paradigm; some external factors are irrelevant and some are just theoretical possibilities that have never been shown to occur with any degree of regularity.

Nevertheless, despite those isolated improvements in the rigor of teacher evaluation just mentioned, educators still cannot claim with any consistency as high a level of standards of integrity and veracity of service to their clients as certain other professions can. This level is one that modern management actually demands for any enterprise it attempts to make truly efficient and successful. This researcher cannot find any school or board in Ontario mentioned in the literature as an example of one which can as of yet claim that type of institutional integrity, and some component of that deficiency may prove to be a result of union resistance. In fact, as a researcher, I am unaware of any union-run establishments in Ontario that allow different workers doing the same job to be paid on a differential based on quality of work rather than a grid contingency based on years of experience.

As for the allegation stated by one of the teachers that the test will not be marked fairly, there is at least equal opportunity for that type of misrepresentation and lowering of standards right now, in the present grading system. That type of grade distortion could occur if the test is prepared on site and if principals order teachers to raise the marks of their students to a pass. One interviewee, in an informal extra-mural interview, said that it was common knowledge in her staff room that a male teacher had been told by the principal about a failure rate of 50 per cent in each of two grade 11 general-level classes being “not acceptable.” Having spent days marking the exams, calculating the final marks and finishing the necessary paperwork, which included mark, comment and reason-for-failure sheets, that department head, then a grade 10 math teacher, was ordered to raise the grades for students who had not earned a passing grade. The interviewee, who had heard the story in the staff room, viewed the student evaluative process, as a result of that anecdote, as a sham and merely a top-down control to minimize any “political” repercussions at the board’s central office. Instead of using this as an opportunity to send a signal to the students that grades are only awarded for displayed knowledge and must be
earned through hard work, reportedly, the teacher was pressured to "dummy-down" standards and award grades arbitrarily.

Another, allegedly more pervasive example, also supports this view of pressure from above in the present non-standardized testing situation. Although this second example does not come from a primary source, the following memo, an internal "strictly non-circulating" directive was communicated to staff by the principal detailing a "raise the marks -- regardless" directive: "If students have marks from 46-49, please give them a clear pass of 51. If they have marks from 40-45, please consider awarding them a mark of 50, which the promotion committee would recognize as a 'gift'." This was published fairly recently in a Toronto daily newspaper (Blatchford, Toronto Sun, Jan. 19, 1995). Although this is just one memo, it had the potential to effect hundreds of exams and the concomitant standards of those exams to which students are supposed to be held. Now while such directives may or may not be commonplace, and while it may be that the principals might have had legitimate cause to believe the over-rulled teachers use inadequate instructional or assessment practices, it seems that if evaluation is to become more consequential, more professional and more objective, although I am not necessarily trying to equate these properties, then teachers must be prepared to accept certain consequences. Using excuses about unfair practices while those unfair practices already prevail only bolsters the view of this researcher that those using that claim may have ulterior, self-serving motives. Furthermore, any negative consequences that manifest as a result of this type of evaluation are apt to speak to professional responsibility even if some schools became perceived as low in achievement and even if some teachers suffer the loss of status or compensation because of their evaluations. Again, this point is in agreement with criteria I have already outlined on pages 124-128, items number two and number seven, which accept rigorous evaluation, regardless of the consequences, as a valuable part of what professional behaviour demands.

Some of the reasons mentioned by the principals in this paragraph, although expressed differently, I acknowledge as almost identical to those given by the teachers, but here they are presented from a principal's perspective, oriented towards principals' themes: "We don't want to teach to the test," "We're there for other reasons besides straight education," "Evaluating teachers on the basis of student test scores would be ludicrous" and "That's how the student did, not the teacher." As stated previously in this text, in the opinion of this researcher, the mantra of "teaching to the test" is not only banal but, when examined closely, can only be deemed as logically unsupportable. Again, if the test is well constructed and adequately represents the material designated to be covered by
the curriculum, the following question could be asked: "What exactly is wrong with
teaching to a standardized test, as long as the teacher did not have access to that test prior
to the students writing that test?"

As for other reasons given, by the principals for example, claiming that
communicating information is not the only reason the teacher is there, and therefore using
standardized tests of student scores narrows the scope of evaluation too much, certainly
teachers are there for other purposes besides straight education such as counseling, field
trips, moral instruction, advice and so forth. Just as certainly, however, a major part of
the reason teachers teach academic subjects is for students to learn them. It is not being
overly narrow-minded to expect that most of the time the teacher spends will be on
teaching the curriculum. One of the major ways to see if students have learned anything
is testing. If so, these same criteria that are applied to students should be applied to
teachers. Nevertheless, this researcher is not suggesting that this be the only measure of
teacher evaluation or even the major component, just one coefficient in a comprehensive
evaluation. Furthermore, if the principals feel so strongly that the other social criteria are
such important aspects of the students’ education, why do the students’ grades not reflect
that concern? I know of no grading system that has as part of its final mark a major
component based on any of the aforementioned non-academic criteria. Essentially, this
researcher is adopting the following position of Sanders: “Even though standardized test
data cannot provide all the information necessary to determine the effectiveness of a
teacher or a school, they can be invaluable and indispensable in linking student outcomes
to specific teachers and schools” (Sanders & Horn, 1998, p. 255).

Consequently, in the judgement of this researcher, to ignore objective- evaluation
data in the form of student scores on standardized tests as a factor of teacher evaluation (as
is now the case in Ontario’s publicly-funded education systems) is almost irresponsible and
to claim a strict dichotomy and even lack of correlation between student progress and the
skill of the teacher is counter-productive and contrary to a considerable body of research.
As one researcher stated, “Contrary to the notions commonly espoused fifteen years ago,
teachers do make a difference: Some teachers consistently elicit higher achievement from
comparable students than other teachers do and more is being discovered every day about
how they do it” (Brophy, 1986, p.1076). Confirming this, is Wright’s study that flatly
declares “the results of this study suggest that teachers do make a difference in student
achievement...Thus identifying teachers that clearly get results over time and comparing
them to teachers over time who do not, seems a logical worthwhile next step in addressing
the issues raised here...”(Wright, Horn & Sanders, 1997, p. 66). The positive aspect
inherent within the results of the principals’ views of this portion of the study is that some of the principals in this data sample were beginning, however reluctantly, to see the other side of the evaluation issue, in the form of recognition of the importance of student grades in teacher evaluation. Signs of this progress have been reported above, where, most noticeably, one principal completely reversed her position and even said that “teaching to the test” may be appropriate if the test has value. Another principal “would agree to rating teachers according to standardized tests if it was definitely a way to increase excellence.” Perhaps it should also be mentioned here that there is no definitive way of telling how reliable these studies from Tennessee are but this researcher has mention previously (pages 355-356) the rigorous statistical evidence for its validity. It should also be noted that it is already being relied upon to a considerable extent, even to the degree that Tennessee authorities are using it as support to influence changes in their own teacher evaluation systems (Sanders & Horn, 1994).

However, even excluding this more objective but controversial paradigm as the sole format for evaluation, there are other more accurate and professional forms of evaluation than those currently used. It is well documented and described in detail in the Literature Review (Chapter 3) that the correlation between teacher proficiency and principal observation is effectively zero, although this correlation can be raised somewhat with intensive pre-training (Jaeger & Busch, 1984; Latham, Wexley & Pursell, 1975; Pulakos, 1986). Yet principal observation remains the predominant method of teacher evaluation, accounting for almost 90 per cent of the evaluations in the U.S. (Lower, 1987). There is no reason to believe it is much different in Ontario, but until someone publishes that data, I can only speculate. Therefore, it is suggested that the focus of teacher evaluation could be shifted away from the observation paradigm toward either peer evaluation, student evaluation of teachers, portfolios or some combination thereof. Since each of these has its own weaknesses and strengths, they would still have to be combined synergistically with some form of product-oriented result, like student scores, to see if, ultimately, they are a success or a failure.

Recommendations

It is, therefore, the recommendation of this researcher that result-oriented evaluation in the form of student grades on standardized exams be tested in an experimental format possibly a triangulated, multi-faceted formula, using some or all of the above methods, developed to evaluate teachers. It is then that I might be able to obtain the necessary hard
data whose absence has been plaguing school systems almost since the first applications of evaluation to modern education. When this is finally done, the excuses for non-rigorous evaluation and tolerance for mediocrity and incompetence might diminish somewhat. Consequently, the mythology of evaluation, almost inherent to arbitrary formats of teacher evaluation, could evaporate, resulting in a system for the selection of the best teachers finally starting to evolve.

Professional Evaluation and Evaluators

In order to have professional evaluation, there are at least two requirements: a professional system of evaluation (process) and professional evaluators (application of process). The problems with our methods of teacher evaluation, then, exclusive of the process, seem to lie with the generally accepted choice of evaluators who are primarily principals. The principals for the most part, as shown within the limited scope of these data, are not adequately trained as evaluators or perhaps even as bona fide professional managers. It follows, therefore, that their evaluation is not likely to be professional.

Recommendations

Based on the evidence from these data here, the following seems implied: the training principals receive is deficient in evaluation training. The Principal Qualification Program, a requirement for Ministry credentials, as it now exists, is most likely not on the same level as a Master’s program, which may very well be why a separate Master’s degree is required. However, even that Master’s degree, ostensibly in educational administration, is an academic degree geared largely toward research and future academic study and not a degree that was intended to offer professional certification to a principal. Furthermore, even that mandatory Master’s degree can theoretically be in any discipline (Reg 297, R.R.O. 1990) although, it is commonly expected to be in educational administration. Regardless of the extent to which teachers are professionals, then, it appears that principals, due to their lack of management training, can be much less defined as professional managers and even less so as evaluators.
Principals today probably need a fundamental understanding of not only general management but also organizational behavior, personnel management, economics, accounting, quality control, personnel evaluation, modern management philosophy, for example, TQM, site based management or management by objectives. In addition, any program claiming to train principals should emphasize, specifically, how these subjects influence the management of a school system. Yet, time constraints alone would probably dictate that almost none of these areas are covered in the various Principal Qualification Programs with the same depth or sophistication, from a practical perspective, that a business or public service manager would receive from a Master’s of Business Administration or a Master’s of Public Administration program. Unfortunately, possibly as a result of the training system just described, there appears to be a lack of professionally-trained managers in evaluation, which leads to a less than professional atmosphere for evaluation in the schools.

It is then my recommendation that the following preliminary questions be asked so that studies can be undertaken to determine the reasons for that situation and remedial measures be taken if necessary:

(1) Why are there not more books and articles written on the Principal Qualification Programs? Why is information on that program so restricted from public and professional scrutiny, preventing it from being studied, critiqued, analyzed or improved?

(2) What percentage of the Principals’ Qualification Program is spent on evaluation and how research-based and contemporary is that section. Furthermore, as a result of their programs, degrees, in-house training and so forth, are principals adequately trained to evaluate teachers?

(3) How could a professional faculty be organized to include graduate-level professional courses to ensure professional training for principals as managers and evaluators along the knowledge guidelines previously suggested on pages 360-361?

(4) In addition to that Master’s degree, could the Principal Qualification Program be transformed into a rigorous professional-qualification exam, accompanied by a professional designation like C.A. or P. Eng. to ensure a sense of professionalism in the program?

In the absence of implementing these suggestions, perhaps evaluation should be done by an external organization whose expertise lies specifically in that area? This might
relieve the principals from performing a now essential but cumbersome service for which they do not appear presently to be qualified for and that they claim not to have the time to do properly. Furthermore, this external evaluation could add more objectivity and rigor, remove the harmful aspects of collegial indulgence and tolerance for mediocrity or incompetence and ultimately reduce the fear of upsetting the cordial aspect of professional relationships the probable rationale for such indulgence. Unfortunately, evaluation has so far often been a non-rigorous and, in many ways, self-defeating process. It should also be noted that while there are those who claim that teaching experience is the main criterion for being a good principal, there is no reason not to combine experience with a rigorous professional degree of qualification, even if that view claiming teaching experience is essential is eventually experimentally validated.

**Culture and Professionalism**

The aspects of unprofessionalism that were evident from the data-based information here were the following items: previewing of evaluation criteria, emotional reactions to negative evaluations, paternalistic unions, lack of rigorous training for evaluators, the inability to dismiss any but the most incompetent teachers and the principals’ lack of faith in implementing a purely formative evaluation system. These problems seem to reflect an attitude that may affect the entire relationship between principals and teachers and between teachers and their boards, despite this study’s principals’ claims to the contrary.

Although based on anecdotal evidence extraneous to this study, to support these data further one need only look at the teachers’ lack of authority beyond the absolute minimum and the consistency with which even that authority is questioned, to see how pervasive this cultural breach in professionalism is manifest. In the unprofessional atmosphere that may exist in Ontario’s publicly funded schools, the roles of student and teacher have become almost interchangeable, if not reversed, when considering those roles with respect to the degree of power discrepancy that now exists, especially when that is compared to what has traditionally existed between them.

These are not the only tangible indications of attacks on teacher professionalism that this account suggests. Teachers are, in the normal course of events, bombarded with duties that are not only tangential and extraneous but often so degrading that they belie their professional status. Lunchroom duties, hall duty, book stacking, library shelf dusting and
recess supervision are all aspects of this phenomenon. These “responsibilities,” which are not really the duties of teachers but are obligations foisted on them perhaps to save the board the trouble and expense of hiring support staff. These assignments, generally, are thought to be an irritation to teachers and may, as I have suggested, also be, at least somewhat, responsible for attrition of their professional status and the construction of a non-professional culture that may have become acceptable in public educational organizations. Even though, from these interviews themselves, the data leave us with hints that an unprofessional culture may have evolved within Ontario’s publicly-funded school systems, here I am attempting to support those data and also suggest that the culture discussed may have concomitant pernicious effects on other areas of school management, specifically evaluation.

How does this aforementioned portrayal influence the present situation regarding teacher evaluation in Ontario’s public and separate schools? It is the contention here that unprofessional attitudes are responsible for reciprocally low expectations of teachers in the fulfillment of their duties, possibly resulting in the laxity and inconsequential nature of evaluation. This unprofessional attitude reflected by the culture might even be based upon the anachronistic premise rooted in traditional cultural images of the teacher. With only some exceptions, those negative stereotypes view the teacher as someone whose undergraduate grades were so low that the only professional faculty that would accept them was education.

Arguably, the aforementioned negative attitude may have persisted, and, as a result, it often seems that the actual high academic requirements for entry into the profession are incongruous with the level of organizational reality, that is, the unprofessionalism of the school culture, its expectations and the assigned para-professional duties. As a result, highly qualified professionals may have been entering a system that is not yet prepared to treat them like the professionals they have been trained to be. It may continually refuse to give them the trust, authority and respect generally taken for granted by other university-trained professionals. The results of this situation can be mediocrity in teachers’ work ethics, erosion of professional expectations, apathy, lower standards for professional achievement and the inevitable lowering of educational criteria in the classroom. Even reducing this criticism to its most limited range, even omitting the supporting empirical evidence and relying totally on the interview data, it nevertheless indicates some degree of a systemic cultural problem that affects the organizational atmosphere of teacher evaluation and management of evaluation on many levels.
Recommendations

Owing to this analysis, recommendations to begin improving the situation could include the following:

(1) Investigate the possibility of treating teachers more like authentic professionals by refraining from assigning them para-professional and custodial duties. These duties demean their professional status and cause a net loss for the schools, despite any short term financial gains. Although these costs could prove prohibitive initially, it may well be the expense necessary to provide a quality, professional education system that includes professional evaluation.

(2) Experiment with a statistically significant number of schools that trust teachers to make professional decisions in their own classrooms, especially regarding student discipline. When teachers have their jurisdiction questioned and even thwarted by management, the legitimacy of their authority would seem to be undermined. These challenges by management to the teachers’ professionalism are often so blatant that the students themselves know about them, which just exacerbates any existing unprofessional atmosphere.

If that system proves valid and is implemented, instead of the present procedures, Ontario could shift the burden of proof to refute professional judgement on to management. Under these new regulations proposed to be incorporated into the Education Act, the teacher would handle all the discipline problems but could be overruled by the principal or vice-principal. Under those conditions, if the principal or vice-principal did overrule the teacher, they would be required to fill out a form defending that decision. That form could also state that the principal or vice-principal who has overruled the teacher might be asked to defend their decision in front of the board of education as to why they rejected and subsequently overruled the professional judgement of the teacher. Under those new guidelines, the principal would have to become cautious enough in their refutations of the teacher’s professional judgement to be reasonably certain that the teacher was wrong. They could not risk, under those conditions, acting either capriciously, politically, vindictively or expediently.
(3) Study the possibility of officially raising faculty of education standards to make them as rigorous as those admission requirements for graduate or other professional schools. This should enhance the professional status of the teacher both in image and substance. If this is combined with recommendation number (4) (following immediately), then the status of the profession and its economic benefits should be raised to the point of limiting, if not eliminating, the annoying and possibly pernicious teacher shortages.

(4) Determine the feasibility and benefits of raising the salaries of teachers so that they are at least congruent with the average salary scales of other university-educated professionals, that is, of doctors, lawyers and engineers. This study, however, should be done only after determining the best teachers based on rigorous, consequential evaluation that relies on student results and other more precise or product-oriented criteria than are presently employed, used as components. This study would be based on the principle that all teachers should not be paid identical amounts, just as not all hockey players are payed identical amounts. If some component of the evaluation is based on unequivocal criteria, that is product as opposed to process (for example, student scores on standardized exams) intra-faculty jealousy might be kept to a minimum. However, if the evaluation is to be authentically fair, the quality of and expectations for each teacher’s student population must be factored in when comparing the results of different teachers. This is already done in other professions, notably cardiac surgery, as mentioned earlier. In that profession, departments and sometimes individuals are evaluated on their success rates after the “risk factors” for their patients are calculated into the evaluation equation (Chassin, et al., 1996). There, innumerable “external factors,” sometimes further beyond the control of the surgeon than the external factors teachers claim to deal with, are not considered an impediment to evaluation based on production. The assumption justifying this type of evaluation is that the good or excellent teacher, like the heart surgeon, must learn to compensate for those factors with professional excellence. Those teachers who excel under this type of evaluation should be rewarded correspondingly. If shown incompetent for that situation, as in the case of heart surgeons, they should be replaced by someone with more talent for that scenario, who can produce the desired results. As in surgery, where I discovered that the surgeon with the most precarious, critical and unstable cases was often not the one with the highest failure rate, administrators may find that the teacher with the most “at risk” students is not the one whose students learn the least, have the most failures or the lowest grades. Ultimately, evaluators may discover that the excellent teacher who uses their professional expertise, when put
to the test, is capable of compensating for those external factors by means of that same professional expertise. This proposed method of evaluation could prove to be on the forefront of assessment and might place teachers in harmony with the standards of other professions in terms of evaluation, professional status and culture. Of course, none of these recommendations can be taken as proven fact within any acceptable epistemological paradigm until an actual study is done on each one of them and their effects monitored.

**Summation**

Brought to full circle now are the initial research questions with which I began this project: What is the nature of teacher evaluation with regards to its objectives, standards of efficiency and values? What are the views of teachers and principals on the factors that restrict or enhance evaluation? Do principals and teachers reveal any important data about the conduct of teacher evaluation with respect to the generally accepted principles of evaluation? Do the data reveal a compatibility with theories of modern management? Are there any important differences between these views of teachers and principals? and, What important suggestions can the researcher offer to help advance the present system towards a more effective teacher evaluation system?

The responses to these questions have become the subsequent basis for the entire thesis. The principals and teachers often reveal a wide gap on important factual issues of evaluation, particularly concerning what restricts or enhances evaluation. In addition to these data, a discrepancy seems to exist between teacher evaluation and the general principles necessary for successful evaluation. Nevertheless, aspects of the data collected from these teacher and principal interviews, revealing specific flaws in the evaluation system, were instrumental in forming a consensual view on the present state of teacher evaluation and its possible remediation. Furthermore, the nature of teacher evaluation, as practiced, according to the limited scope of the data available here, does not seem to be consistently compatible with any contemporary form of management philosophy. In effect, teacher evaluation, in the publicly-funded systems, seems to ignore basic principles of management and evaluation such as legitimate professional respect for employees, consequentiality in evaluation, effective communication, reasonable flexibility in management's role of evaluation, use of hard or product-oriented data, use of professionally-trained evaluators and a general cultural professionalism. Furthermore, as
mentioned several times, no real improvements, as can be seen by tangible results, emerge from these evaluations, according to the data in these interviews.

As provisional steps to improve the situation, several problem categories within present evaluation systems, based on those already outlined in this summation, had to be highlighted as they emerged from these interviews. These were all seen in this study’s sample population as the probable conditions that allow careless, unprofessional evaluation to exist. Furthermore, with each of these problems, after an appropriate detailed discussion, recommendations on how to correct or ameliorate the situation were given as guidelines to future study. Implied within these data was a criticism of the basic desultory management attitude with regard to evaluation and the subtle ignorance of their organizations’ cultural issues or potentials for using the culture synergistically with an efficient evaluation system for advantage in building a proficient, effective organization-friendly teaching staff. Undoubtedly, as with each individual organization, there exists profound diversities between schools, other public organizations and corporations, but the differences seem to pale with respect to the similarities. It is the implication of the information here and the recommendations thereof that these similarities be consolidated to exploit the education, business and public organization management literature to incorporate a system of evaluation compatible with the public school culture, management’s organizational goals and teachers’ professional goals. Based on the data in this project and the associated literature, such an approach may result in better, more efficiently run school systems. If the theme of this study is correct, then it is time that this knowledge was used successfully to refute the mythology employed to uphold the present system and implement an evaluation program that will improve teaching, the schools’ cultural environment and the educational institutions themselves.
APPENDIX B

INTERVIEW QUESTIONS

1) For someone like myself who has not been involved in teacher education, as you have, could you say what have been the best or most productive aspects of teacher evaluation that you’ve experienced or that you’ve heard of? (Probe)

2) By and large what have been the worst aspects of teacher evaluation that you know of? (Probe for critical incidents, memories, value judgements etc.)

3) From your knowledge of the Education Act, union by-laws, collective agreements and board policies, what influences do these policies or regulations have in enhancing or deterring effective evaluation?

4) The Ontario government seems to be going ahead with a number of changes involving teachers, such as the establishment of a College of Teachers, student representation on boards of education, school councils and province wide testing. Could you offer any insight of how each of these innovations might affect teacher evaluation? First, very briefly, the College of Teachers . . . .

5) Conceding that summative evaluation might also be necessary, nevertheless, does your evaluation system have a strong formative component? (If Yes) It’s sometimes said that teacher evaluation tries to accomplish two divergent tasks, professional growth and accountability; do you see any contradiction in this? (Probe for assumptions behind the evaluation system)

6) At your school, are the processes for performance appraisal used to evaluate ECE teachers, resource teachers, guidance counselors and classroom teachers identical? Does your school’s system of evaluation recognize the necessity to differentiate between teachers with, for example, five (5) or more years classroom teaching experience and those with less than five (5) years of classroom teaching experience; if it does recognize this difference, does your school’s system of evaluation use different processes to evaluate these different categories of teacher just described?

7) Most appraisal systems have a classroom observation component. Do you think these observed situations accurately reflect the day-to-day teaching situation? Do teachers have an opportunity to see the evaluation forms beforehand? If so, in your experience does this effect the evaluation?
8) Do you think there are an adequate number of visits to the classroom to correctly grasp what’s happening? What would be better?

9) Is there a follow up interview after evaluation? What can you say about how that follow up works?

10) Are the promotions in your school based on the evaluation system? Are these promotions uniquely administrative or do they also feed into some sort of teaching track?

11) Overall, much of the literature claims that the vast majority of teachers are acceptable or average. Does that square with your own impression of the quality of members of the teaching profession in general? (Does the evaluation system recognize competent teachers as well as identify incompetent ones?)

12) Much has been written about the importance of the evaluation being compatible with the school’s culture (clarify). What is your sense of the evaluation at your school -- is it compatible with the school’s culture?

13) Are the criteria for evaluation outlined as clearly as possible without the natural spontaneity necessary for fair, accurate and critical evaluation compromised? Are there some criteria presently used in the system that should be removed? Which are they? Are there some that should be added? Which are they?

14) How do you feel those criteria identified by the evaluators as requiring improvement or personal growth are taken by the teachers involved?

15) Have you ever felt that the evaluation system in your school has been too subjective to offer any useful information? Does your appraisal system use any form of objective criteria?

16) In your opinion, does the appraisal system provide protection for individual rights and due process? How so? Is there any opportunity for abuse of the system?

17) Should teachers be involved in setting the goals or criteria for evaluation? What role do you see teachers playing in helping to set up the scene for evaluation? Would you consider this role of primary or secondary importance?
18) How would you further characterize the roles, feelings, relationships and communications between evaluating principal and evaluated teacher in your school?

19) Can you say something about the training that principals receive to prepare them to do teacher evaluations? What happens? Is it adequate? What would be better?

20) As it is now, do you think the teachers try to avoid their evaluations or the evaluation results? Examples? Why is that?

21) This is a big question. If you had your ‘druthers’ and could redesign the whole system, what sort of evaluation for teachers would you support?

22) By way of winding down, let me get two (2) reactions. First, from your perspective, has evaluation had any significant impact on the quality of teaching in your school?

23) Second, could you comment on the following proposition. Some of the literature claims that businesses, more so than schools, are prepared to dismiss unsatisfactory employees if necessary. What’s your reaction? If you agree why is this so? Why not?

24) Is there something on the topic that you’d like to add? Something that I missed and should have asked?

25) Finally, I’d like to get some demographic data on you. How long have you been teaching/evaluating? What kind of schools have you worked in, now and in the past (rural, small town, inner city, Metro Toronto suburb, northern community)?
accountability:  
"The requirement to answer for the results of professional effort usually measured [in terms of] student achievement, outcomes and institutional effectiveness" (Herman & Herman, 1993, p. 263).

culture:  
“What is considered normal or acceptable thinking or behavior for a group of people [or an organization]” (Jacobs, 1996, p. 185). This definition coincides both with the dictionary definition which defines culture as “the ideas, customs, skills, arts etc. of a people, group [or organization] that are transferred, communicated or passed along” (Webster’s New World College Dictionary, Third Edition, 1996, p. 337). and with the more colloquial definition “The way we do things around here” (Bower, 1966 as cited in Deal & Kennedy, 1982). Focusing on schools, it usually refers to “the standard process and activities utilized by the students and employees of the school district” (Herman & Herman, 1993, p. 265).

effectiveness:  
“The achievement of some goal—even if not the original one” (Scriven, 1991), p. 129).

efficiency:  

evaluation:  
The determination of the teacher’s value worth and merit [to the students, clients or school] in as objective a manner as possible. It is done in the particular way that evaluation reduces facts and values to conclusions (Scriven, 1991). It must determine the standards or criteria, collect relevant information and apply those standards to determine [teacher] quality (Worthen & Sander, 1987).

formative evaluation:  
Typically conducted during the teacher’s development, its purpose is to improve the professional skills of the teacher. As such, if a record is made at all it usually remains on site. If the formative evaluation is successful, it all but vitiates the need for any summative evaluation. The best compliment a summative evaluation can offer in tribute to a formative evaluation is a list of the formative suggestions and a subsequent documentation of their implementation and positive results (Scriven, 1991).
institutional chauvinism:
When someone comes to believe or assert that the defects of an organization are its virtues (Whyte, 1956).

outcomes:
Defined simply, they are the effects of a program, system, algorithm or method(s) as applied to a certain situation or condition (Scriven, 1991).

professional:
[Technical] knowledge used in the service of the satisfaction of clients [overriding personal concerns for the benefit of the client and [taking] responsibility for correct standards, practices and ethics as assumed by the corresponding professional organization (Darling-Hammond, 1989).

professionalization:
“The enlargement of teacher discretion in making decisions in the immediate work place and in the larger context of school and system” (Herman & Herman, 1993, p. 264).

qualitative research:
Any method of research using a non-numerical approach, that is, one not based on statistical analysis or other statistical method but rather questions (survey, interview and so forth) direct inquiry or observations. Generally, it is any research that does not use numbers or statistical terminology and methodology to achieve or express its results and conclusions (Tesch, 1990).

quantitative study:
A numerically based study in which success or failure depends on the statistical significance of the data. That is, whether or not the relationships obtained from the data are the result of the methods applied or merely occurred by chance (Anderson, 1990).

reliability:
“The consistency of scores obtained by the same persons when re-examined with the same test on different occasions” (Anastasia, 1988, p. 109).

summative evaluation:
Evaluation done for the benefit of the organization as a whole and whose data as acted upon by some decision maker. It is the evaluation made in order to carry out a decision to punish reward or dismiss. For reasons of credibility, summative evaluation should be carried out by an external evaluator, if possible (Scriven, 1991).
system:
“A set or arrangement of things so related or connected as to form a unity or organic whole (a solar system, school system, system of highways)” (Webster’s New World College Dictionary, p. 1369).

teacher effectiveness:
“The degree to which [student] learning [of those subjects on the curriculum that are the object of teaching] is maximized [by the teacher] relative to the potentialities [and deficiencies] of each student while accounting for the normal and accepted time and resource constraints” (Scriven, 1988, p. 320) or anything done that succeeds in helping a student reach educational objectives (Medley, 1992).

validity:
“The degree to which the test actually measures what it purports to measure” (Anastasia, 1988, p. 28).