

EVALUATION AS PROTECTION:

Using curriculum evaluation to promote a just distribution of educational resources in a private post-secondary English-language liberal arts institution in Canada for Japanese students which uses a leveled, modular, skills-based mastery-learning entry programme

by

CHERYL RENE REITZ

B.A., Anthropology, The University of Washington, 1971
Professional Development Program, Simon Fraser University, 1988
Post-Baccalaureate Diploma, TESL, Simon Fraser University, 1989

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

in

THE FACULTY OF GRADUATE STUDIES
Centre for the Study of Curriculum and Instruction

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1997

© Cheryl Rene Reitz, 1997

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

Department of Centre for the Study of Curriculum and Instruction
The University of British Columbia
Vancouver, Canada

Date April 20, 1997

ABSTRACT

This thesis examines how one might evaluate the justice of educational resource distribution. It focusses on the criteria of institutional justice formulated by John Rawls: according to these criteria inequality in the distribution of resources is only allowed if it can be shown to benefit all groups, including 'the least favoured'. The thesis also demonstrates how qualitative and quantitative research methods can be combined in order to reach a more accurate and 'just' evaluation.

The research, which was conducted at a private post-secondary English - language liberal arts institution in British Columbia for Japanese students, compares annual student growth in English, both before and after the implementation of a three-to-ten-month leveled, modular, mastery-learning program for entry-level students.

The research also includes interviews to determine teacher attitudes about the previous and present programs and their effect on students. In both the qualitative and quantitative studies, program effects on high-, medium-, and low-entry ability students are looked at separately (in order to use Rawls' criteria).

The context of the research is clarified with short summaries of issues around mastery learning, leveling versus tracking, and Japanese versus western education.

The quantitative research finds that, contrary to teacher impressions, the mean improvement for students in the present program is not significantly different from that in the previous program. The qualitative research however, points out important justice implications not revealed by the other study.

The thesis concludes that (1) there are some problems with using Rawls' criteria in an educational setting; (2) looking at program effects on three separate ability groupings can reveal trends having justice implications; and (3) assessments of the justice of educational resource distribution should attempt to triangulate with both qualitative and quantitative studies which attempt to answer the same question.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgement	vii
Dedication	viii
1.0 <u>Introduction</u>	1
1.1 Definition of Terms	1
1.2 Context of the Question	7
1.3 Summary of the Thesis	16
2.0 <u>Justice as Fairness - John Rawls</u>	20
2.1 Two Visions of Democracy	20
2.2 Rawls' Principles	21
2.3 Two Questions About the Application of Rawls' Theories to Education	25
2.4 Summary of Use of the Notion of Justice as Fairness in Curriculum Evaluation	28
3.0 <u>Changing Methods of Curriculum Evaluation</u>	29
3.1 Qualitative/Quantitative Dualism	29
3.2 A Combination of Qualitative and Quantitative Methods in Curriculum Evaluation	31
3.3 Application of this Perspective to my Thesis	34
4.0 <u>Summary of Issues Related Specifically to the Research Site</u>	35

4.1	Mastery Learning	35
4.2	Use of Ability-Grouping ('Leveling') versus 'Tracking'	45
4.3	Relevant Japanese Curricular, Evaluation, and Justice Issues	49
5.0	<u>The Research Projects</u>	58
5.1	Quantitative Project - Description and Results	61
5.2	Qualitative Project - Description and Results	80
6.0	<u>Conclusions</u>	125
6.1	Quantitative Research	125
6.2	Qualitative Research	129
6.3	A Stereoscopic View	137
6.4	Conclusions about the Kind of Information Gained by Combining these Two Types of Research	141
6.5	Conclusions about the Use of John Rawls' Criteria when Addressing Issues of Educational Justice	143
6.6	Summary of Research Findings and Conclusions	144
6.7	Suggestions for Further Research	146
6.8	A Postscript	149
7.0	<u>Works Cited</u>	151
8.0	<u>Appendices</u>	
8.1	Samples of SPSS 6.1 Matched-Pairs Input and Transformed Data	157
8.2	Sample of SPSS 6.1 Output showing Means and Significance	159
8.3	Parallel 80%/50% Pass Mark Grading Schemes Used at College 'X'	160
8.4	Application: A Look at some Site-Specific Justice Issues	161
8.5	Description of SLEP Test	166

LIST OF TABLES

<u>Number</u>	<u>Title</u>	<u>Page Number</u>
Table 1	Summary of General Characteristics Defining Mastery Learning	39
Table 2	Defining Characteristics of the Three Kinds of Mastery Learning	40
Table 3	Proportion of Subjects in the Three Levels, by Matched-Pair Groupings	68
Table 4	Mean SLEP Entry Score and Gain (per Year and per Matched Pair Groupings)	71
Table 5	Student Responses to Programs by Level (as Remembered by Teachers)	130
Table 6	Summary of Quantitative and Qualitative Results	138
Table 7	Key to SPSS Input Data	157
Table 8	Sample of SPSS Input Data	157
Table 9	Key to SPSS Transformed Data	157
Table 10	Sample of SPSS Transformed Data	158

LIST OF FIGURES

<u>Number</u>	<u>Title</u>	<u>Page Number</u>
Figure 1	Mean Increase in Total SLEP Points by Matched-Pair Years	69
Figure 2.1	Mean Increase in Listening SLEP Points by Matched-Pair Years	73
Figure 2.2	Mean Increase in Reading SLEP Points by Matched-Pair Years	73
Figure 3.1	Mean Increase in Listening SLEP Points by Matched-Pair Years and Entry Level	76
Figure 3.2	Mean Increase in Reading SLEP Points by Matched-Pair Years and Entry Level	77

ACKNOWLEDGEMENT

This thesis would not have materialized had I not been the recipient of extraordinary kindnesses from many people whom it is my pleasure to thank at this time. I must thank first and foremost my family. Through some very difficult times over the past few years, they have consistently assured me it was extremely important to them that I continue working towards my educational goals.

The unconditional support of my colleagues -- both my fellow teachers at work and my fellow UBC M.A. and M. Ed. graduate students here in the Kootenays -- has been a source of great comfort and inspiration.

Dr. LeRoi Daniels of UBC first saw some value in my topic and Dr. Ian Andrews of Simon Fraser University advised me, while the project was still in an embryonic stage, that it was indeed worth pursuing.

My UBC advisor, Dr. Jerrold Coombs (Educational Studies), and committee member Dr. Murray Elliott (Educational Studies) have exercised admirable patience and restraint, frequently guiding me away from various seductive side-routes, assuring me that just because something is interesting to me does not mean it must be included in my master's thesis. Dr. Graham Mallett (Language Education) agreed to be my outside reader under extraordinary time constraints, for which I am most grateful.

Two other UBC professors outside my department deserve a special thank-you as well. Dr. Marshall Arlin (Educational Psychology) took time to make some helpful observations about my research results. Statistician Dr. Arleigh Reichl helped me import data from one program to another, and made sure I was inputting data correctly.

I must also thank several other colleagues who have assisted in the following areas: Dr. John Casey chaired the Research Committee at my institution which gave me both access to data and time to interview teachers; Dr. Carol Thew was persistent enough to locate vital data which all others thought had been lost and, along with UBC doctoral candidate Marilyn Low, provided helpful insights at crucial times. Finally I extend my deep gratitude to cultural liaison officer, Mr. Yoichi Oshima, who assured me that my description of Japanese education was fairly close to his experience as both a student and a teacher in Japan.

The thesis is still far from perfect, but it is a lot better than it might have been if not for you. I must apologize for being unable to incorporate all of your advice. (For example: Unfortunately, it is impossible to make my thesis half the size and double the size at the same time!) In those rare instances in which your advice conflicted, I hope you will agree with my choices. Taking full credit for any and all errors and omissions, I extend to all of you my heartfelt thanks and best wishes. Though I started to give up on myself many times, you never gave up; your faith and good humour always brought me back from 'the edge'. May God bless you all.

DEDICATION

To the memory of my parents

Dorothy Verne Mehl Hennen
August 16, 1921 - October 7, 1995

and

Robert Eugene Hennen
April 18, 1921 - October 19, 1996

1.0 INTRODUCTION

Because the distribution of educational resources is seen to be ultimately associated with the distribution of social, political, and economic power, the manner in which educational resources are distributed is a perennial concern in a democratic society. In this thesis, I will first present some proposals regarding ways to determine a just distribution of these resources. Next, I will suggest a way of combining qualitative and quantitative methods to aid in this determination. I will provide an example of a type of research and evaluation which uses a combination of qualitative and quantitative methods to examine how closely a curricular program conforms to a just arrangement, within the context of my own teaching environment, a private post-secondary English - Language liberal arts institution in British Columbia for recent Japanese high school graduates. The thesis will conclude with suggestions of how my research could serve as a model for further inquiries into the justice of specific curricular decisions.

1.1 Definition of Terms

The terms used in the title - *just*, *distribution*, and *educational resources* - were used intentionally in contrast to *equal*, *access*, and *education* or *knowledge*. The rationale for doing this is below and is followed by a discussion of 'input' and 'output' definitions of equality and the notion of 'outcome distribution' in education.

1.1.1 'Just'

Equal treatment, in that it can mean *identical* treatment, is not necessarily just. The word 'justice' implies a more sensitive or appropriate fit between the individual and the treatment received than does 'equality'. By 'justice' I refer to any practice in which

the *principle of equality* is followed. This principle is an assertion that "in all public matters all persons should be treated identically, except in contexts where sufficient reasons exist for treating particular individuals or groups differently" ("Justice"). In defining the antithesis of justice, John Rawls refers to 'injustice' as "inequalities that are not to the benefit of all" (62). Clearly, the important relationship between equality and justice is complex.

1.1.2 'Distribution'

'Distribution' was used instead of 'access,' because the latter denotes simply that a 'commodity' is potentially available if the 'consumer' has the desire or motivation and the personal assets to take advantage of the opportunity to acquire it. If we see the consumer and his/her assets as the sole determiners of whether the commodity is acquired, and the role of the institution as simply to provide 'access' to it, this term is sufficient. However, if we are looking at the institution's role as being more complex than this, we might consider using a different word. While the term 'distribution' can mean simply the (one-way) act of dispersing and giving out shares of a commodity, I use it here in its second meaning: "the extent to which different groups, classes, or individuals share in the total production or wealth of a community" ("Distribution"). Here, the verb 'share' implies an intentional and cooperative giving and taking. This seems to more accurately describe that active two-way relationship between educational resources (a type of wealth) and learners which results in the learners' synthesis of knowledge (a more-or-less critical determiner of future production and wealth as well as a type of wealth in itself). While the term may not be totally accurate, in that the giving and taking are at times one way, at others reciprocal, and at times unintentional, it seems to be more appropriate than 'access' in this case. It also implies a statistical construct - a 'just' distribution in contrast to a 'normal' one.

1.1.3 'Educational resources'

It was tempting to use the term 'knowledge' instead of 'educational resources', as in John Goodlad's reference to "the distribution or democratization of knowledge" (792).

However, because it fits more closely with the theories of John Rawls presented herein, I have primarily used the term 'educational resources' as interpreted by Jerrold R. Coombs ("Equal" 282) to mean "those conditions or objects which facilitate desirable educational achievements." It can mean teachers' time and attention, learning materials and methods, a comfortable classroom, and so on.

1.1.4 'Input/Output'

One way to determine whether these resources have been justly distributed (the 'input' interpretation) is through measuring how equal the programmes are which have been made available to various individuals or groups of students. *Another way* to determine whether resources have been justly distributed (the 'output' interpretation) is through measuring the degree to which students' observable educational achievements (sometimes called 'learning outcomes') are equal.

The 'output' interpretation was given support by the U.S. Supreme Court Ruling of 1954 that 'separate but equal' schools for black and white students were not just because they (by implication: the separate schools themselves rather than any genetically determined student attributes) produced *unequal effects* upon the students. As James S. Coleman noted, this landmark decision furthered the notion that "... equality of opportunity depends *in some fashion* upon effects of schooling ..." (my emphasis) ("Concept" 15). Coleman, et al. made liberal use of this output assumption in their influential *Equality of Educational Opportunity*, a report commissioned by the U.S. Civil Rights Commission in 1964.

In the first (input) case, then, it is the responsibility of institutions to give equal resources (*input*) while the learners bear responsibility for creating their own, differentiated educational achievements.

In the second (output) case, equal achievement (*output*) -- assumed to be leading to more-or-less equal power, equal satisfaction, equal income, etc.-- becomes the goal, and the responsibility for promoting these results shifts, *to some unspecified degree*, to the institution; the implication is that input may have to become unequal in order to promote a more equal output. While not discounting the factors of personal choice and responsibility on the *individual* level, this output notion correctly acknowledges that providing equal input to all *groups*, even when the principle of equality is applied, does not guarantee that all groups will equally benefit (as determined by output measures), no matter how equality of input is determined. Differential *utilization* of educational resources, in particular among racial, cultural, and gender groupings, then, is seen to be another form of inequality -- one that can serve to further additional inequality and polarization in the next generation -- which educational institutions must eventually address.

Jerrold R. Coombs ("Equal" 282) has identified fallacies integral to both input and output interpretations. The input interpretation, he points out, while correctly acknowledging the intentionality of learners, incorrectly assumes equality of learner groups -- that is, that all start out on equal footing and that therefore, given equal programmes, the groups will achieve equal outcomes. On the other hand, the output interpretation assumes that students are passive recipients of 'knowledge' -- that the institution can create achievement (and equality), neglecting the two-way intentional nature of learning and incorrectly assuming that equality of outcomes among groups is either desirable or possible.

According to the principle of equality, one could justify situations where students receive very different educational resources according to their *age* (different grade

levels), personal *interests* (elective clubs, sports, and classes), and personal *strengths* (adjustments for various handicaps). This is considered just, although most would agree that (except in cases where restitutions are being made), using *race, nationality, religion, or gender* as a determiner of the distribution of educational resources is not. However, as Coombs ("Equal") points out, minority or less-powerful groups may have inherent values and interests which prevent their members from consuming educational resources 'equally' with predominant groups. As examples, he notes groups such as Amish, various aboriginal groups, and even females (as a less-powerful group), whose 'unequal' consumption of educational resources may be their own (conscious or unconscious) choice, based on their rejection of mainstream or 'predominant' religious, cultural, or gender values. He suggests that in cases such as this, culturally-determined or gender-specific programmes could be quite different from those of the mainstream, as long as resources are distributed in such a way that "neither group has reason to envy the resources given to the other" (291). This thesis will consider the use of this principle with various *ability groups* as well.

1.1.5 'Outcome Distribution'

It is very important to distinguish between determiners of justice at the level of the individual versus that of the group. While each individual within a group is going to have very different strengths, motivations, and achievements from others in the group, making the goal of equal outcomes on the individual level ludicrous, a case can often be made for striving for equality of outcome distributions among groups. This is dependent on several factors such as whether the *learning goals are mutually and equally shared* by the groups, and whether the *instrument chosen to measure the outcomes is not biased towards any particular group(s)*.

In cases in which groups have very different attributes which would affect their learning (such as differences in prior learning or ability - criteria by which students are grouped at my institution), achievement among groups may be compared (outcome distribution determined) in terms of differences in improvement (change in achievement) instead of differences in absolute achievement -- as long as all groups can be assessed using the same scale. Otherwise, the only way the lower groups' outcomes would ever be seen as 'equalized' with those of the higher group would be by preventing the higher group from advancing, a ludicrous situation, but nevertheless, one which we will ponder in our examination of the mastery learning literature.

1.1.6 Summary - How the Terms will be Used in this Thesis

In this thesis, the entire student sample is monoracial, monocultural, and of the same general socio-economic and age group as well. The smaller groups are based simply on performance on standardized tests of English proficiency -- a function of native ability, motivation, and prior learning combined. Presumably, then, there is no significant difference among the groups in learning goals, and the test instrument is not biased towards any one group. While both input and output are considered, the emphasis of this thesis generally is on outcome distribution among the ability groups, not on specific individuals within them. However, standard deviation measures used in the quantitative study may be seen as an indication of the extent of individual variation. Note also that in the quantitative study, the distribution is that of *change* in scores among the various ability groups rather than the distribution of absolute scores.

Parts of the *qualitative* study, besides looking at the distribution of educational resources as determined by both inputs to and outputs from the various ability groups, will also look at the extent to which the curriculum is experienced as 'just' by individuals -- both teachers and students.

In short, the roles of both the institution and the learner in 'creating academic achievement' are acknowledged in this thesis. Consequently, the justice of educational resource distribution is evaluated both in terms of learners' *access to* resources (accessibility being a function of both resource and learner attributes), and learners' *achievement from* resources (again, a function of both resource and learner qualities). Both input and output will be considered, though outcome distribution will be stressed, particularly in the quantitative study. Not wishing to commodify knowledge nor trivialize the complex nature of its synthesis or expression, I hope the phrase 'just distribution of educational resources,' then, will prove useful in this case.

1.2 Context of the Question

In this section I hope, first, to show how the concerns of this thesis are not confined solely to the specific context of my institution; secondly, to establish the relevance of recurring themes of this thesis -- horizontal versus vertical power arrangements; qualitative versus quantitative research and evaluation; and standardization versus teacher professionalization -- and, finally, to demonstrate how these themes are intimately related in the present educational context and particularly in regard to determining the justice of curricular decisions.

1.2.1 Consolidation/De-centralization Trends

The political and economic reality of Canada in the late 1990s is that of balancing budgets and downsizing entitlements, programs, and bureaucracies in order that future generations not be saddled with the present generation's debts. Thoughtful people, while coming to understand this reality, are asking what the effects will be, and how the negative ones might be mitigated.

School districts in British Columbia, Alberta, and Ontario are presently being forced to consolidate (centralize) for budget reasons. The evaluation from the top down of schools, of principals, of teachers, and of the curriculum they teach will, as a result, become more centralized, but also more remote in that final decision-making will more frequently occur in places other than the local community. Private institutions such as my own, experiencing similar financial pressures, are likewise seeking to trim administrative costs; presumably this will result in less administrative time for curricular and teacher evaluation. What effect will these situations have on evaluation of curriculum?

While there will inevitably, due to the greater degree of district centralization, be some additional external constraints (such as general proclamations from unseen bureaucrats rather than more negotiable site-specific mutual agreements among personally-affected parties), these may not be enforceable at the classroom level anyway due to the increasing remoteness of administrators. Some would even question whether present dictums are truly enforced -- implementation studies have often shown that many teachers implement fully only those programs which they support, regardless of 'policy'.

In addition, constraints due to greater district centralization will probably be countered by concurrent de-centralization trends at the federal and provincial levels providing more power to -- but less 'outside' pressure on -- the district office. This situation (remoteness from consolidated district offices plus decentralization at the federal and provincial levels) will presumably provide teachers, at least on a day-to-day basis, with more latitude to conduct their classrooms as they wish.

In a recent phenomenon related to this, various provincial educational initiatives (such as British Columbia's province-wide *Year 2000* reforms) have been promoted vigorously, then suddenly dropped or changed radically; as a result, teachers abandoned on various 'bandwagons' are left with an *appetite for innovations*, but with a corresponding *cynicism towards those imposed from above*. This development, in

rejecting what William Pinar, et al. (298) describe as a "monolithic, single-voice curriculum" imposed from above, parallels the notion of 'heteroglossia' (inclusion of multiple voices, acknowledgement of *multiple* truths) first promoted by Mikel Bakhtin, popularized by James A. Whitson, and related to 'horizontal' or cooperative methods of curriculum development, research, and evaluation described below (*see also* Pinar).

Increased remoteness from administrators is potentially supportive, then, of 'horizontal' (rather than 'vertical' or 'top-down') curriculum evaluation practices such as those advocated in the 1992 *ASCD Yearbook* edited by C. Glickman: evaluation by peers, by autobiographical self-exploration, or through action research (*see also* Gitlin & Goldstein; Gitlin). With these horizontal models, qualitative evaluation methods are generally preferred to quantitative ones. My contention is that while these 'horizontal' methods have great merit, they also contain some important flaws which must be addressed.

1.2.2 Horizontal versus Vertical Methods

These educational power realignments (consolidation/decentralization) and their associated features potentially have a great impact on what educational resources will be available, how they will be valued, and, most important to this thesis, how they will be allocated and distributed among students.

Two conflicting movements characterize these realignments. *Vertical* power realignments, such as district consolidation, are hierarchical by definition; they generally are utilitarian (concerned with maximizing 'the good') in nature and lead to standardization. *Horizontal* power realignments, such as decentralization, cooperative (and individual) teacher action research, and peer evaluation, are non-hierarchical by definition; they are generally laissez-faire in nature, and lead to multiple, non-standardized results. For example, without a common provincial umbrella uniting them,

districts will tend to become more unique: less like one another -- and as Martin Carnoy (199) points out, less equal as well. Similarly, in a 'horizontal' power alignment, an individual's statement of her/his own internal experience is regularly considered valid even without the corroboration of objectively-collected 'data'. Obviously, the number of possible descriptions of one's personal internal experience (versions of reality or of 'the truth') is equal to the number of subjects queried.

There are certainly varying postures from which to regard the nature of truth and reality: Is reality/truth an illusion? If not, is there one 'unitary' reality/truth or many? Is reality/truth 'objective' or is it the synthesis of multiple 'intersubjectivities'? and so on However, the point is that significant challenges have been thrown to positivism, to the supremacy of 'objectivity', and to the *primacy* of quantification (if not to quantification itself). These challenges have had and will continue to have a major impact on curriculum research and evaluation practices. My contention is that the many educational philosophers who have put forth these challenges (*see* Pinar, et al.) are, generally speaking, supportive of the current 'horizontal' educational power realignments previously mentioned.

It must be noted that 'horizontal' does not imply 'qualitative'; neither does 'vertical' imply 'quantitative'. These are simply terms that (i.) show the general direction from which the locus of power originates and (ii.) label a collection of elements which are frequently *but not always* associated with each. While contending that 'horizontal' evaluation is primarily (though not solely) of a qualitative nature, I do not wish to imply that 'vertical' evaluation according to external standards is *necessarily* quantitative. For example, current accreditation procedures (primarily a 'top down' or 'vertical' -- and summative -- evaluation) frequently contain significant qualitative elements. My thesis is wholly supportive of this trend.

If either of these movements is allowed to progress without limits, however, unchecked by the other, injustice may result. In the schools, those on the bottom of the educational hierarchy, students, are potentially the most vulnerable to this injustice.

Students are the ones to suffer most from the excesses of vertical power, empirical evaluation, and standardization such as being inadequately taught by dispirited teachers who have no sense of ownership of the curriculum, experiencing (personally) the cumulative effects of 'failure' according to standardized testing and the normal curve, or enduring overly-standardized, boring texts.

However, students also suffer unjustly when there is insufficient vertical power and empirical evaluation. Without some means in place to monitor conditions in their schools, or compare them to other schools, students can, unbeknownst to the public, suffer from inadequate materials, lack of coordination and integration (from class to class, year to year, and school to school), disorganized or incompetent teaching, segregation and unequal treatment (according to gender, race, socio-economic class, or handicap), or unfair allocation of resources.

1.2.3 Attack on 'Vertical' Methods

Pinar, et al. note that as early as the 1960s, more than a few scholars (such as James B. Macdonald, Dwayne E. Huebner, Herbert M. Kliebard, Elliot W. Eisner, Maxine Greene, Louise M. Berman, and Paul R. Klohr) had begun to criticize what they saw as related practices such as "behaviourism, *scientism (a reduction of forms of knowing to quantifiable ones)*, dehumanizing technology, and an oppressive, alienating bureaucraticization of the schools . . . (They) . . . attacked behavioural objectives, . . . and *quantified, standardized evaluation and measurement of learning*" (my emphases) (Pinar, et al. 184; Huber). This was what Pinar, et al. refer to as 'The first stage of

Reconceptualization' of the curriculum field which was to exert ever more influence on education, continuing on into the present day.

Those who distrust vertical power, standardization, and empirical evaluation have many advocates and an attractive set of rationales. For example, a major trend in curriculum theory is to see traditional (ie. standardized, mainstream, 'top-down') education negatively as a way of assimilating and controlling lower classes while indoctrinating them to the point of view of either the most powerful or most numerous group in society (Bowles & Gintis; Apple, "Curriculum"). Schools are seen as vehicles for reproducing current socio-economic hierarchies through what Phillip Jackson calls the 'hidden curriculum.' This idea has gone through several phases: the concepts of *race and gender* were added to that of class (Apple, *Cultural*); the notion of *resistance* by the lower status or minority group members was added (Willis; Giroux) and a 'liberation pedagogy,' based on the work done earlier by Paulo Freire ("Conscientizing") was advocated. Freire's idea was that the schools could be used in a more positive manner, developing within society's victims the intellectual tools they need in order to dismantle oppressive systems.

These reformers came to advocate an evaluation of *curriculum* by such techniques as determining its value according to how well it empowers and liberates students (Freire, *Pedagogy*) or as considering the curriculum an art form which can be judged aesthetically subject to 'critical connoisseurship' (Elliot Eisner, *Enlightened*). In addition, they urge evaluation of *teachers* by self (and by peers), and of *students* by qualitative, individualized, and descriptive methods.

Note that one cannot completely separate the teacher from the curriculum any more than one can separate the medium from the message. Methods of evaluating students are also an integral part of the curriculum. Evaluation in curriculum, then, can easily come to include that of students and teachers as well.

All of these theorists were opposed to the use of top-down evaluation if it served to further the successful reproduction of an oppressive status quo. Because standardized testing has been used so extensively to sort and label students, to limit their future opportunities, to rob them of their self-esteem, or to marginalize them (House, "Justice"; Bowles & Gintis; Books), many educators have also come to oppose standardized testing of students, particularly if the tests are used to determine or limit the future academic options open to a particular individual.

However, both bureaucrats and the public continue to lend great credence to numbers, despite the protests of many evaluation critics. Districts are quite often judged by their students' performance on standardized tests. Students are admitted to university primarily on the basis of their high school grade point average. As well, as critics Blaine R. Worthen and James R. Sanders claim, "... quantitative work is still the dominant approach to educational inquiry, as even casual reading of the most influential journals in education . . . will reveal" (51).

Nevertheless, I predict that horizontal realignments, because they involve less administrator time, and are generally easy and cheap to implement (in addition to other, more idealistic rationales), will become even more popular in the coming age of 'down-sizing'. On the other hand, along with balancing budgets can come demands for greater accountability and quantitative justification of expenditures such as educational outcomes, 'payoffs' and 'dividends'. As well, economies of scale support standardization, and quantitative studies, like qualitative ones, can also concentrate on factors which are easy and cheap to measure. Though horizontal trends seem to be on the upswing, then, there is reason to predict that both trends will continue to influence evaluators of the foreseeable future.

1.2.4 Standardization - Oppression or Protection?

Note Pinar, et al.'s association (thesis, page 11) of *quantification* with *standardization*; though the two often go together, neither requires nor assumes the other. However, the two are commonly associated by many people; this common association, combined with a negative perception of the purpose of standardized testing, may well have contributed to the present popularity of qualitative (as an alternative to quantitative) testing -- and to that of curriculum evaluation and research methods, as well.

The vertical ideal of 'a standardized curriculum' has waxed and waned over the decades. For example, in the 1920s scholars such as Franklin Bobbitt, in the words of Pinar, et al. saw "curriculum standardization and centralization . . . (as) . . . goals, not oppressive realities" (33). Reformers criticized fragmented, locally-controlled school systems of disparate quality under the assumption that, as David Tyack put it, "Regulation, bureaucratization, and centralization would *equalize education by standardizing it*, delegate decision making to experts, and '*Americanize*' a diverse population" (my emphasis) (3).

In the Progressive Era, the theories of John Dewey were to conflict somewhat with Bobbitt's goals; while Bobbit was seen as an advocate for bureaucracy, Dewey was seen as one of teacher professionalism. As R. Corwin put it, "Bureaucracy, by its nature, requires a high degree of standardization, with stress on uniformity in both rules and conduct . . . Professionalization, however, is marked by a low degree of standardization" (qtd. in Glanz 162).

During the 1940s and 1950s, Bobbit's and Dewey's visions of teaching were to alternately influence and temper the proliferation of mutually supportive standardized tests and textbooks. As Kellaghan and Madaus (89) note, external exams tend to restrict what is taught -- objectives that are not to be tested, or are difficult to ~~test~~ (such as oral and manual skills) will simply not be emphasized in a class or in a text. External exams,

including university entrance exams, came to "determine much of the curriculum and circumscribe the professional role of teachers" as Pinar, et al. (1998) paraphrase Kellaghan and Madaus (1989). Here we see a relationship among bureaucracies, external standards and standardization of materials which contains elements we will see again, though in a far more extreme form, in the discussion on Japanese education (thesis, pages 50-52).

From the 1960s through the 1980s, there were various curricular reform movements, some calling for more standardization and external testing, some calling for less. In general, there was concern over both high dropout rates and poor preparation for post-secondary schooling and jobs, especially in urban public schools and among minority groups. A common response to this, noted Dennis Carlson, was the recommendation that schools return to 'basic skills' and more top-down evaluation through standardized testing, a phenomenon in evidence at my college as late as 1990, as will be seen in the qualitative section of this thesis.

1.2.5 A 'Blind Spot' of Solely-Horizontal Methods

After decades of a scientific-positivist (some would say 'utilitarian') approach to curricular decision-making, many now seek to bring a more human touch back to education. Words such as 'nurturing,' 'community,' and 'spirituality' are heard once again from academics. Many educators now recognize the mistakes of the past which resulted from the mis-use of curricular power and the unjust distribution of educational resources, often buttressed by claims of objectivity and reason, and supported by the results of quantitative research. As a consequence, these academics (and others) tend to question the exertion of top-down power, arbitrary controls, or external standards and question whether objectivity is possible or (Derrida; Deleuze & Guattari) if 'reason' actually exists.

Pinar, et al. note many other curriculum theorists today who would have educators question (or even abandon) the positivist strictures of *reason*, *logic*, and

empiricism which have directed (these theorists would probably say 'shackled') the progress of Western thought. To reject reason implies irrationality, a precarious stance at best; likewise, though we may not label ourselves 'positivists,' we devalue any of the positivist 'tools' at our own peril. This is not to say that we cannot put some new tools (e.g. qualitative, horizontal, liberating tools) into our toolbox. However, this does not imply that we need to throw out the old ones simply because people sometimes mis-use them.

While acknowledging their positive intentions, logical rationale, and well-articulated methods, I will argue that the 'blind spot' of those who see curriculum in this way seems to be their distrust of power, seen in their negative attitude towards external standards and controls and their preference for qualitative evaluation. While just allocation of educational resources, equality of opportunity, and provision of the best education possible to our young are, unquestionably, their goals, precisely these ideals -- justice, equality, and quality of education -- may be threatened by the devaluation of external standards, quantitative evaluation, and external controls. Good intentions are not enough, as school personnel themselves may not even be aware that conditions at their school are unsatisfactory or unjust, or could be better, unless they have an external standard with which to compare them.

1.3 Summary of the Thesis

How should twenty-first century educators address the problem of justly distributing educational resources? They must consider the many failures of centralization and vertical, utilitarian control; of standardization; and of using primarily quantitative evaluation methods. They can't ignore the compelling arguments for decentralization and horizontal control, for including many -- even conflicting -- points of view (heteroglossia) in evaluation, for qualitative research, and for peer evaluation.

How can they ensure *equality of educational opportunity* without creating a standardized regulatory monolith which destroys the quality of the educational experience? Is there any way to *equalize educational opportunity* for students while, at the same time, granting teachers more autonomy in the classroom?

Are vertical and horizontal arrangements mutually incompatible? Or, by incorporating the strengths of both arrangements, and ensuring that neither predominates to the point of creating injustice...can they perhaps be woven together? Is some synthesis possible? If so, what criteria shall we use to determine whether the principle of equality has been adhered to and whether a just distribution of educational resources has been achieved?

It is, perhaps, only human nature to blame the tools that people misuse rather than the motives of the people improperly wielding them. My thesis is that ironically -- despite their good intentions and the undeniable value of qualitative and horizontal methods of evaluation -- postmodern theorists' distrust of power and their devaluation (or outright rejection) of quantitative methods and external, hierarchical (top-down) evaluation could easily lead to a laissez-faire education system. Without some external standards, and without some quantitative and top-down evaluation, there is no way to ensure justice in allocation of resources, equality of opportunity, or quality of pedagogy. Ensuring justice, equality, and quality is no more likely in a laissez-faire education system than in a laissez-faire marketplace.

I am not advocating a return to utilitarian, solely-statistical, solely-top-down methods of curriculum evaluation. These methods alone are unaided by principles of justice, confined to a limited scope, and deaf to the nuances of the individual spirit; alone, neither the positivist tools of evaluation nor the new, often more qualitative tools which have been developed, are sufficient to ensure that all children truly encounter justice, opportunity, and quality education throughout their school experience.

Rather, I propose that *formative* evaluation of individuals (principals, teachers, students) should primarily be horizontal, qualitative, and mutually negotiable, as this form is potentially more meaningful and constructive on a personal level, as well as less threatening (hence, more readily acceptable) to the individual being evaluated. Furthermore, evaluation of curricula, schools, districts, 'the nation's schools,' etc. should have qualitative components as well as quantitative ones. However, a need remains for some external standards to guide all these individuals, institutions, and curricula and, it follows, for at least some quantitative assessment according to these external standards.

To determine which method to use in a given case, I propose these guidelines. A combination of vertical and horizontal, quantitative and qualitative methods *may* be used at any time to add more information to an assessment. However, a combination of vertical and horizontal, quantitative and qualitative methods *must* be used in cases in which *matters of justice, equality of opportunity, or quality of pedagogy, which impact directly upon students, are being assessed in a summative manner*. This is because when such critical factors are being assessed, as much information as possible is needed. We simply can't assume that participants either possess sufficient knowledge of the big picture or are capable of accurately situating themselves in it. We also can't assume that all evaluators have the same standards unless they are articulated.

To determine whether a just distribution of educational resources has been achieved, I originally turned to a theory of justice, sometimes referred to as 'justice-as-fairness,' propounded by Harvard University's John Rawls. Curriculum-evaluation scholar Ernest House ("Justice") had suggested that Rawls' theory might be useful in assessing the justice of curricular decisions. However, over time, I came to question (as have others such as Kenneth Strike as well as Rawls and House themselves) whether Rawls' theory was *directly* applicable to determining the just distribution of educational resources.

Consequently, then, I will urge a reconsideration of some of the reasons for external standards, quantitative evaluation, and external controls, will question the implications of their devaluation, and will suggest two ways of mitigating their negative effects while retaining their positive ones: (i.) by using some principles of justice-as-fairness, and (ii.) by using a combination of quantitative with qualitative evaluation methods.

We will now take a more scrutinizing look at previous queries into these issues, exploring in Section 2.0 the notion of justice-as-fairness as it applies to evaluation in education, particularly in the evaluation of educational resource distribution, and in Section 3.0, what research methods to use in this determination. As well, we will briefly examine literature which will help define the specific context in which the research was conducted, such as summaries in Section 4.1 of the ongoing debate over mastery learning; in Section 4.2 of the difference between 'leveling' and 'tracking' students; and in Section 4.3 of curricular, evaluation, and justice issues in the Japanese education system, with which my institutions' students are most familiar. The two research projects are presented in Section 5.0 and the conclusions of the thesis in Section 6.0.

2.0 JUSTICE AS FAIRNESS - JOHN RAWLS

2.1 Two Visions of Democracy

The principles defining democratic thought would seem to be liberty or the freedom to determine one's own life course and to maximize one's own good, tempered by rational and moral constraints; equality or "justice as regularity" (Rawls 504); and fraternity or brotherhood; a harmony of interests for mutual benefit; one-for-all and all-for one; or an expression of the feeling of " . . . not wanting to have greater advantages unless this is to the benefit of others who are less well off" (Rawls 105). John Rawls has proposed some intriguing guidelines whereby the morality of the allotment of rights and resources by democratic egalitarian institutions could be judged (to see why he recommends democratic egalitarian over natural libertarian, liberal egalitarian and natural aristocratic institutions *see* Rawls, Section 12). These guidelines or principles attempt to ensure a more genuine equality of opportunity for all through providing protection to those possessing the least resources with which to take advantage of these opportunities.

Another popular vision of democracy emphasizes liberty and equality, advocating a laissez-faire pursuit among equals for the possession and maximization of resources. The rationale for this utilitarian pursuit is that the inherent competition will result in the maximization of resources for the benefit of all (an idiosyncratic -- perhaps even paternalistic -- way of expressing 'fraternity'). This vision is, however, also related to the philosophy of social Darwinism, in which the survival of the fittest humans (and human groups) is thought to result in the eventual bettering of the human species. The fallacy inherent in both of these visions is that, once an allotment of resources has been achieved, the players are no longer equal: those who have achieved more (and their progeny as well) possess a greater advantage in further pursuits to maximize those resources. The human species is not necessarily 'bettered' by the survival of these more successful members since their 'advantage' may not be attributable to inherent (i.e.

hereditary) individual qualities, but simply, as Rawls notes, to the resources possessed or inherited:

... the institutions of society favor certain starting places over others. These are especially deep inequalities. Not only are they pervasive, but they affect men's initial chances in life; yet they cannot possibly be justified by an appeal to the notions of merit or desert. It is these inequalities, presumably inevitable in the basic structure of any society, to which the principles of social justice must in the first instance apply (Rawls 7).

2.2 Rawls' Principles

2.2.1 General Conception of Justice

Rawls, then, advocates providing a 'handicap' to at least partially redress the above inequality, thus enabling those who possess significantly fewer resources to compete on a more equal footing. These varied resources are considered by Rawls to be 'goods' (sometimes he refers to them as 'values'), as in his General Conception of Justice which states that:

All social primary goods -- liberty and opportunity, income and wealth, and the bases of self-respect -- are to be distributed equally unless an unequal distribution of any or all, of these goods is to the advantage of the least favoured (303).

Rawls particularly emphasizes self-respect not only as a good in itself but also as a prime resource upon which the ability to acquire further resources is dependent (440).

In Rawls' First Principle of Justice, he deals with *liberty* and *equality*:

Each person is to have an equal right to the most extensive total system of equal basic liberties compatible with a similar system of liberty for all (302).

2.2.2 The Second Principle of Justice (the 'Difference' Principle)

Fraternity or 'caring' has been described by Steven R. Covey (4) as a 'superordinate' or higher middle position transcending the divergent values of liberty and equality. Rawls, in a similar vein, promotes fraternity in a Second Principle of Justice whereby inequalities are only permissible when they *are to the greatest benefit of the least advantaged*:

Social and economic inequalities are to be arranged so that they are both (a) to the greatest benefit of the least advantaged . . . and (b) attached to offices and positions open to all under conditions of fair equality (302).

The rationale for social or economic inequality to the greatest benefit of the least advantaged is that it will *eventually* enrich the society as a whole (including those most advantaged).

It is easy to misconstrue Rawls' meaning here. He does not mean that inequalities must benefit 'the least advantaged' more than other groups of people. Rather, he means that any inequality must benefit those who are the least advantaged (under this unequal distribution) *more than they would be under a condition in which resources were distributed on a strictly equal basis* (Coombs, email 10 Feb. 1997). For example, Rawls would probably see justification for an unequal distribution of educational resources which equipped disadvantaged students to participate more fully as equals within societal institutions; for an unequal distribution of such resources as higher grades, advanced degrees, or income which provided an incentive to those who better society (e.g. by creating jobs or healing the sick); or for an unequal distribution of rights to those who harm society (e.g. incarceration), *so long as this inequality may be seen to benefit 'the least advantaged' groups to a greater degree than would a strictly equal distribution*.

While the Second Principle of Justice does not explicitly acknowledge the inequalities students bring into the classroom, such as varying types of home, economic,

and cultural backgrounds, intelligence, motivation, interests, and previously-attained skills, Rawls does acknowledge that in any society, some will naturally be more advantaged than others. He explains that this principle represents a social (contractual) "agreement to regard the distribution of natural talents as a common asset and to share in the benefits of this distribution . . . " (101).

Rawls compares the Principle of Redress with the Second Principle of Justice, to which it is related. The Principle of Redress:

is the principle that undeserved inequalities call for redress; and since inequalities of birth and natural endowment are undeserved, these inequalities are to be somehow compensated for . . . society must give more attention to *those with fewer native assets* and to those born into the less equal social positions. The idea is to redress the bias of contingencies in the direction of equality. In pursuit of this principle greater resources might be spent on the education of *the less rather than the more intelligent*, at least over a certain time of life, say the earlier years of school . . . " (my emphases) (100-101).

Rawls states very clearly that his Second Principle of Justice (the 'difference' principle) "is not . . . the principle of redress" (101) in that:

It does not require society to try to even out handicaps as if all were expected to compete on a fair basis in the same race. But . . . (the difference principle) . . . would allocate resources in education, say, so as to *improve the long-term expectations of the least favoured*. If this end is attained by giving more attention to the better endowed, it is permissible; otherwise not. And in making this decision, *the value of education should not be assessed solely in terms of economic efficiency and social welfare*. Equally if not more important is the role of education in enabling a person to enjoy the culture of his society and *to take part in its affairs* (my note - thus democratizing society as well), and in this way to provide for each individual a secure sense of his own worth (my emphases) (101).

Coombs warns against a possible misinterpretation of this principle which is "not about redressing disadvantages to presently disadvantaged groups, but rather about ensuring that *who ever* ends up disadvantaged by social arrangements would not be worse off than they would be if benefits were distributed on a strictly equal basis" (email, 7 Jan. 1997). Rawls adheres to the traditional convention of 'the veil of ignorance,' in

which a social contract is agreed to in the beginning by people who are 'blind' as to which socio-economic position they might personally end up filling. Therefore, they would choose a contract in which all resources were distributed equally -- unless it were clear that another arrangement maximizes "the minimum amount of goods anyone will receive, i.e. increase over what they would have under a policy that distributes goods equally" (Coombs, email 10 Feb. 1997). There is a limit, in other words, on redistribution -- it should not end up replacing one disadvantaged group with another.

Rawls' principles may be seen as a just compromise between the impulses to compete freely and to distribute fairly, tempering the excesses of each. In the material realm, of course, the extremes of 'pure' capitalism (free competition) and 'pure' socialism (fair distribution) come to mind. However, in promoting partial redress to the lesser-advantaged, Rawls is not advocating any particular economic system. He is not a Marxist; in his book (259) he considers Marxism only as an economic arrangement and as an ideal which, if carried out in its idealized form (fair distribution - a tautology for Marxism), could be 'beyond justice' (Rawls 281 refers to R.C. Tucker's *The Marxian Revolutionary Idea* chs. I and II). In addition, Rawls' principles are those of moral philosophy, not economics, and apply to the just distribution of both tangible and intangible resources. In short, to Rawls, economics should be guided by justice as well as utilitarianism, in that it should not be concerned only with the maximization of resources, but with their fair distribution as well. This is sometimes called a maximum/minimum (max-min) principle, in that "the minimum amount of primary goods any person will receive" is maximized (Coombs, email 10 Feb. 1997).

Rawls' premise is that if participants in an institution are more equally endowed with its resources, the entire institution is more *just* than an institution exhibiting extreme variations of resource endowment among its participants. It should be noted that Rawls' prime aim is to promote *justice* rather than *democracy*; in this case the two happen to intersect.

It is clear then, that Rawls, though neither a neo-Marxist nor a 'postmodernist' as such, is concerned with empowerment and the leveling of hierarchy. It is from a similar perspective that I wish to position myself and have attempted to conduct this project.

2.3 Two Questions About the Application of Rawls' Theories to Education

2.3.1 The Educationally Disadvantaged

One question came up: "Who, exactly, are 'the disadvantaged' in education?" James S. Coleman suggests that unequal *ability*, in addition to unequal (racial, cultural, familial, etc.) background, might be a natural inequality deserving of consideration when trying to equalize society (17).

In the section of his theory where he advocates a partial application of the Principle of Redress, Rawls also deals with this question, suggesting the allotment of educational resources so as to improve "the long-term expectations of the least favoured . . . (which could include) . . . those with fewer native assets . . . (such as) . . . the less rather than the more intelligent" (100-101). Clearly, Rawls is suggesting here that ability is a valid criterion, along with race, culture, socio-economic class, and gender, to consider when assessing whether educational resources are being justly distributed.

Kenneth A. Strike reaches the same point when he considers the goal of equal opportunity: "If we can assume that any social results are essentially a function of native ability plus opportunity (leaving aspirations out of consideration for the moment - see Strike's footnote #24), then when opportunity is equal the disadvantaged will be those who possess less native ability" ("Role" 7). He points out a problem with this, however: "If, then, we are to apply the difference principle to schooling, we will wish resources to be distributed and *patterns of achievement to result* (my emphasis) such that they are to the advantage of the less well endowed" (Strike "Role" 7). He proposes that only *after*

inequalities in the achievement of minority and poor children disappear should this become a goal. Interestingly, he modifies somewhat the manner in which these are prioritized in 1983 -- see thesis, page 47).

What Strike called 'patterns of achievement that are to the advantage of the less well endowed' may not mean they all have 'A's'. It is often to the benefit of the less advantaged not to limit the ways in which those naturally better endowed can advance themselves since the less advantaged, too, will share in positive ways with their achievements, regarding (as Rawls put it) "the distribution of natural *talents* as a common *asset*" (my emphases) (101). To this I would add the corollary that society, likewise, should regard the distribution of natural *disabilities* or intellectual inadequacies as a common *problem* of which all will share in the mitigation.

We will revisit the issue of the potential achievement of 'the less-endowed' versus that of 'the better-endowed' academically in the section on mastery learning.

2.3.2 Generalizability of Rawls' Principles

Another question came up: "Just how globally can Rawls' criteria be applied?" Several educational evaluators (Coleman "Equality"; House, "Justice"; Strike, "Role") have debated this and could not agree on an answer.

House was perhaps the first, in 1976 ("Justice") to use Rawls' theories in a critique of modern (both qualitative and quantitative) curriculum evaluation practice, suggesting that evaluators should be concerned with their impact on *individual* subjects' self-esteem. Also, he advocated looking not only at *mean benefits*, but at their *distribution among advantaged versus disadvantaged groups* as well. (Note that in order for evaluators to do this, their experimental design will of necessity be quite different than it would if they were looking at the group as a whole - there must be at least one more variable such as family income, socio-economic status, a standardized test

score, or some qualitative factor which could help categorize each subject according to how 'advantaged' they are.)

Strike replied to House: "One takes these [Rawls'] rules intended for such global applications and applies them to specific situations only at one's peril and against the spirit of Rawls' views." Strike (incorrectly, I contend), interpreting Rawls' 'resources' as 'wealth', claimed that "Rawls' Second Principle governs basic institutions for distributing *wealth* (my emphasis) in a society. It does not *per se* govern the distribution of test scores" (Strike "Role" 5). Strike went on to temper this with the suggestion that test scores might have some relevance to Rawls' principles if one could show a causal connection between the distribution of test scores and "the distribution of social and economic benefits" (6).

It should be noted that Rawls himself was first to point out the limitations of his theory:

There is no reason to suppose ahead of time that the principles satisfactory for the basic structure hold for all cases. These principles may not work for the rules and practices of . . . less comprehensive social groups . . . Now admittedly the concept of the basic structure is somewhat vague. It is not always clear which institutions or features thereof should be included (Rawls 8-9)

In my opinion, there are various levels on which an evaluator might use Rawls' theory. It seems logical to try to apply it in a variety of situations and see if it is of use or not. My particular research is at the level of a school, of course, and none of the students belong to a disadvantaged socio-economic group, so I am considering the 'academically' rather than the socio-economically disadvantaged. However, Rawls' theory does not particularly state the parameters of research involving his principles; it was left up to the researcher to determine whether the principles proved useful in the particular context.

2.4 Summary of Use of the Notion of Justice-as-Fairness in Curriculum Evaluation

In conclusion, Rawls' theory of justice-as-fairness has given educational evaluators who see the benefit of the new evaluation methods, but are troubled by the moral relativism that so often accompanies them, some new moral direction. Rawls' most relevant contributions to curriculum evaluation, I predict, will be this advice: (i.) to consider the effect of programmes and evaluation itself on individuals' self-esteem; (ii.) to look at the mean effects of programmes on each of the sub-groups affected, not only on the mean effect on the entire group and (iii.) to ensure that the programme is of benefit to all, including the least-advantaged sub-group. His theory, which some might perceive as an elaborate re-statement of the 'Golden Rule,' is by no means a final solution to the relativist dilemma. However, with luck, it may provide the mental scaffolding from which a future moral philosopher will be able to construct something which more closely approximates a solution.

3.0 CHANGING METHODS OF CURRICULUM EVALUATION

3.1 Qualitative/Quantitative Dualism

3.1.1 Qualitative and Quantitative Defined

During the twentieth century, logical positivism, the social efficiency movement, and behaviourist psychology were to have a great influence on curriculum research and evaluation. In each of these movements, research done by the use of quantitative methods was strongly preferred. As noted previously, while quantitative methods continue to be widely used in curriculum research and evaluation, qualitative methods (largely as a reaction against the misuse of quantitative methods -- in education as well as other disciplines) have been gaining in popularity.

How exactly are 'qualitative' and 'quantitative' differentiated in a strictly dualistic sense? Perhaps the easiest explanation is that quantitative methods of evaluation use a generally-accepted standard of measurement and measure a phenomenon according to the standard in such a way that the measurement could feasibly be replicated by others who would, under the same conditions, obtain the same results. These results could then, if desired, be analysed statistically. Qualitative methods, quite simply, are those which do not conform to this pattern. Qualitative methods in curriculum evaluation range from simple 'goal-free' and 'instrument-free' observations to interviews and autobiographical and phenomenological exploration. They tend to look in-depth at one or a small number of subjects or phenomena and seek to construct meanings which are mutually-agreed-upon by researcher and subject.

There is no sense in referring to *the* 'quantitative method' or *the* 'qualitative method'; each is a group of many different methods -- even more so, two studies which both involve 'a combination of quantitative and qualitative methods' can be extremely

different, utilizing vastly differing methods. Lynne Miller and Ann Lieberman (12), in an otherwise excellent article, ignore this distinction when they state that "both the Rand and DESSI studies depend on a combination of qualitative and quantitative measures, so we cannot attribute the difference in findings to difference in method." On the contrary, the different results may be due to very divergent methods.

3.1.2 Strengths and Weaknesses of Each Group of Methods

Each group of methods has inherent weaknesses detracting from its usefulness to curriculum evaluators. Quantitative methods focus so completely on one factor that they often distract researchers from other important, but harder (or impossible) -to-measure elements (such as internal, personal experiences). They also prevent researchers from seeing unanticipated, so unmeasured (and undetected) consequences. I have detailed many other problems with quantitative methods in the Introduction, but wish to dwell in more detail on the difficulties with qualitative methods at this point.

Qualitative methods can give such mixed results that it is difficult to know what use to make of the data. Qualitative methods also tend to lack replicability. In their focus on one or a small number of subjects in depth, they may end up with an atypical rather than a typical subject selection. With qualitative methods of curriculum (and teacher) evaluation, there are even more subtle pitfalls:

Autobiography is a form of '*self-reporting*' and '*self-assessment*' which is sometimes used as part of an accreditation or a curricular evaluation process. F. Michael Connelly and D. Jean Clandinin (141) warn against narcissistic tendencies and Hollywood-style happy endings into which autobiography -- if the work is not (intersubjectively) tempered by collaboration with others -- can fall. *Peer-assessment* also contains some potential pitfalls. William Pinar questions whether the attitude of empathy many postmodernists assume when assessing peers (and others) might not serve,

at times, to conceal or rationalize more than it reveals. Pinar, et al. note that empathy involves mentally participating in the intentions of others, "intentions which can function as self-rationalizing, self-forgiving, indeed self-deceiving ideas. Empathizing with another . . . might lead to collusion " (583). Finally, Madeline Grumet warns that 'teacher' narratives can lapse into an impotent moral relativism: a "failure to engage in some analysis . . . beyond celebration and recapitulation (which) leads to a patronizing sentimentality (consigning) the teacher's tale to myth, resonant but marginal because it is not part of the discourse that justifies real action" (324). Relativism, in its openness to many possible interpretations of 'truth' and 'the good' can open many previously-closed doors to the mind. However, it can also, in its inability to clearly show a 'correct' course, obscure routes which were previously open and led to action.

Each group of methods also has its strengths. Quantitative methods can reveal trends that casual observers would not detect, prove *causation*, and clearly *confirm or refute hypotheses*. Because they are replicable, they are not as easily subject to observers' (conscious or unconscious) personal whims, moods, and prejudices. Qualitative methods, however, can point out trends that the researcher had never known to exist nor thought to measure, or suggest questions that the researcher had never thought to ask. They help people to *form hypotheses*. They can reveal personal, internal points of view such as *motivation and reasoning* as well.

3.2 A Combination of Qualitative and Quantitative Methods in Curriculum Evaluation

3.2.1 Dualistic Nature Refuted

Until recently, these two groups of methods have been seen as antithetical dualisms, mutually unmixable, like water and oil, or (arguably) like science and religion,

predicated on two such different world views that the conclusions using one perspective could not in any way be supportive of or comparable to those using the other perspective. One advocated either one or the other, never both.

There has been a regrettable tendency, as Kenneth R. Howe points out, to retain the "rigid epistemological distinctions between quantitative and qualitative methods " (10). He claims that this dualism, which forces researchers to choose from 'value-laden' qualitative or 'value-free' (descriptive) quantitative methods is a legacy of positivist dogma which should be discarded.

Howe and others claim that objectivity is a myth -- that bias exists in quantitative methods as well as qualitative ones in that choices such as what to measure, what instruments and standards to use, and what statistical analyses to employ, etc., involve very value-laden decisions. To pretend they don't only adds to the 'hidden' bias.

Many curriculum evaluation methods contain both quantitative and qualitative elements. For example, the results of questionnaires and textual analyses both describe and tabulate (often large) numbers of 'intersubjectivities'. These tabulations, though of 'subjective' inner experiences and personal opinions, can easily be replicated *if the number of subjects is great enough*. As in the Indian folk story *The Blind Men and the Elephant* (where 'the truth' was found by synthesizing a number of quite differing intersubjectivities) the 'objective' nature of 'multiple intersubjectivities' is becoming recognized. In my mind, intersubjectivity can be seen as a synthesis of multiple personal perceptions -- perceptions moving from the subjective and personal towards the objective and impersonal -- an empirical position defying the dualism of subject/object. In other words, a description of the elephant issued by several blind men will be more accurate than that issued by one blind man, and is more accurate, in some ways, than that provided by an 'objective' (but only two-dimensional) photograph.

Elliot Eisner (*Art* 252) has suggested that standardized tests and other quantitative data can be used to supplement the qualitative methods he has developed. But how

should one go about combining the results? The notion of 'triangulation' (collecting data from more than one source about the same event or behaviour) has often been applied to combining more than one type of qualitative approach or investigating a phenomenon from the point of view of more than one stakeholder group. However, Louise H. Kidder and Michelle Fine show how triangulation can also result from combining data from qualitative and quantitative studies. This is easier, they claim, if both studies are clearly trying to investigate the same hypothesis, and if the qualitative study is not so fluid that the questions the researcher asks vary from subject to subject or over time.

For a far more detailed discussion of these topics, see Worthen and Sanders; Mark and Shotland; Kidder and Fine; Howe; Miller and Lieberman; Madaus and Kellaghan; Cook and Reichardt; Madey; and Stone, all of whom conclude that the time has come to stop thinking of qualitative and quantitative methods as mutually antagonistic and instead recognize them as complementary, mutually supportive, and best used in combination with one another.

3.2.2 Benefits of Combining Methods

The three most widely-mentioned benefits of combining qualitative and quantitative methods are (i.) that each is strong where the other is weak; thus, they fill in each others' gaps, complementing one another and strengthening the research, (ii.) that when they support one another, the results are strengthened as well, and (iii.) that when they contradict one another, both results are called into question; in this case an explanation for the contradiction (possibly requiring further research) is called for.

3.3 Application of this Perspective to my Thesis

In my thesis, I hope to follow the suggestion of Blaine R. Worthen and James R. Sanders that instead of expending energy on debating the relative merits of qualitative versus quantitative methods, scholars and practitioners' energy would "be more productively channeled into conceptualizing and testing procedures for effective integration of quantitative and qualitative methodologies, an area in which there is still very little guidance" (53). I will conduct one primarily qualitative and one primarily quantitative study, each of which tries to answer the same general question about just distribution of educational resources. I will analyze the results each study contributes -- alone, and in combination with the other -- to see how (and whether) they complement, mutually support, contradict, or inform one another. I hope to be able to verify or refute some of the claims others have made about combining qualitative and quantitative methods, and perhaps even contribute some additional observations as well.

Please note however, that this is not primarily a methodological study. While I hope to demonstrate triangulation of qualitative and quantitative data in curriculum evaluation, my primary purpose is to reach the best determination of the justice of a particular distribution of educational resources, not to prove that qualitative/quantitative triangulation 'works.' An assumption of my study is that, if conducted properly, this type of research can combine the best aspects of both quantitative and qualitative methods.

4.0 SUMMARY OF ISSUES RELATED SPECIFICALLY TO THE RESEARCH SITE

Instead of proceeding directly to the research itself, I have chosen to situate the research in a particular educational context. This is because the research site is a rather atypical Canadian liberal arts college, though those who teach adult-basic-education or English-as-a-Second Language students may find frequent parallels with their concerns. I am writing this section not for my own institutional colleagues, to whom this context is intimately familiar, nor for those who simply wish to follow the philosophical logic of my argument. This is background information for readers who want to know more clearly what sort of school, curriculum, teachers, and students we will be looking at, and what issues concern someone evaluating whether educational resources are being justly distributed through the curriculum currently in use at this particular school.

Three issues specific to my institutional setting will be summarized. First, the curriculum and evaluation methods used at my school will be situated within a mastery-learning framework, along with a short descriptive and analytical review of mastery learning. Next, the very different purposes and results of 'ability-leveling' and 'tracking' will be clarified. Finally, Japanese curricular, evaluation, and justice concerns which affect programmes and their delivery at my college will be summarized.

4.1 Mastery Learning

Mastery learning, an outcome-based curriculum model, is used at our institution for the first four ('Foundation') levels of listening, speaking, reading, grammar, and composition. The Foundation levels, while not the sole focus of my study, are the primary focus since all but the highest entry-level students spend most of their time in this programme during their first year at the college. Let's look briefly at mastery

learning's history, philosophy, and experimental base and at how it is actually practised today. The various controversies surrounding mastery learning, particularly the issues which affect our institution, will be summarized.

4.1.1 Benjamin Bloom's Suggestion

The basic premise of mastery learning is that given enough time, practically anyone can learn practically anything. The idea has been traced back as far as John Comenius' seminal *Pampaedia* in the seventeenth century and probably even further (Block, *Mastery*; Guskey), but began its most recent incarnation in 1968 when Benjamin Bloom saw the implications of a conceptual model of learning propounded by John B. Carroll.

Carroll's model posits the degree of learning to be a function of the time actually spent relative to the time needed. Time needed is seen as a function of (1) an individual's aptitude, (2) the quality of instruction (e.g. materials and methods), and (3) the individual's ability to understand the instruction (e.g. matches between student and teacher language and between learning and teaching style, plus affective factors).

Allocation of time (for learning and teaching) as well as (2) and (3) above could, reasoned Bloom, be manipulated by the school to ensure that all students succeeded in mastering the basic material to be learned. Summative tests would be given at frequent intervals. Those who didn't reach the mastery standard would re-study and then re-take an alternative form of the summative exam. They wouldn't be penalized for having taken additional time to learn, nor for mistakes made while learning.

The goal of mastery learning advocates is not simply to maximize learning through raising test scores, a purely utilitarian goal, but to better the educational achievements of those, in particular, among the least advantaged in the educational hierarchy. Echoing John Rawls, Block, et al. (220) proclaim, "... we believe in equity

in terms of student learning outcomes, not in terms of student learning opportunities (my note: inputs). Indeed, to attain outcome equity we are willing to provide unequal treatment in terms of learning opportunities and learning time for some students and especially for those who historically have been the 'have-nots' in the teaching-learning process" (220). While few would argue with the sentiment, there is great debate as to how to define 'equity in learning outcomes,' whether it is (or to what degree it is) actually possible to attain, and what the costs would be (including detrimental effects it might have on the learning of those 'more academically endowed').

4.1.2 Organization of Time versus Outcome Distribution

In mastery classrooms, time may be organized in one of three basic ways: (1) Instruction can be completely individualized as in many adult basic and computer-assisted instructional programmes. (2) Alternatively, various levels of achievement may be offered in different (homogeneous) classrooms and students may repeat, advance, or skip levels every few weeks as deemed appropriate (used in many language institutes and skill-based courses such as ballet). (3) In the typical heterogeneous classroom, however, students are not segregated according to ability or achievement as in the former two cases. Rather, the students who reach the mastery standard earlier engage in enrichment activities while their slower classmates re-study the materials and eventually re-take the exam.

While Bloom agrees that aptitude (which he would call 'learning rate') is distributed normally among students and that given uniform instruction (including equal time), their achievement (outcome) is also distributed normally, he claims that *given optimal instruction (including optimal time) based on individual needs*, most students can achieve mastery of most desired learning outcomes.

While the mastery standard at our school is only 80%, Block, et al. consider 'mastery' as a minimum achievement of 85% to 95%. This score must be attained on summative tests which are criterion-referenced to the objectives of the curriculum. Ideally, the standard is high enough to ensure that desired learning has occurred, but not so high that mastery is perceived as unattainable. In mastery learning's strictest form, students' achievement is evaluated on what, in the end, they have actually learned as shown on a summative test given about every two weeks or so, *not* by their classwork, marks on formative tests, effort, nor time required to master the material. (Note: in our school, marks on formative tests are now given some weight as well.)

For an exhaustive and research-based analysis of the relationship between time, ability, equality, achievement, and mastery learning see University of British Columbia's Dr. Marshall Arlin. Interestingly, while Arlin supports many of the claims of mastery learning advocates, he notes that their greatest weakness may be their tendency to 'hyperrationalize' (a term he credits to A.E. Wise) -- to "persist in rationalizing policy decisions that overlook means-ends relationships" (Arlin 81). He implies that the more zealous mastery learning advocates tend to minimize research which casts doubt on their claims that mastery learning does not result in slower learning for the 'faster' students in heterogeneous classrooms. Though our college uses mastery learning only in its homogeneous (Foundation level) classrooms, this issue (also called the 'Robin Hood' effect -- first, by Arlin and later, by Slavin) will surface in my research.

The following two tables summarize the characteristics of mastery learning in general (Table 1) and of the three quite different ways it can be delivered (Table 2).

Table 1

Summary of the General Characteristics Defining Mastery Learning:

- Instructional objectives are well-defined and appropriately sequenced
- Student learning is checked regularly and frequently and immediate feedback given
- Standards are criterion-referenced rather than norm-referenced
- A criterion level of performance is held to represent 'mastery' of a given skill or concept
- Corrective instruction is given to enable students who do not initially meet the mastery criteria to do so on later, parallel assessments
- Time and resources are organized to ensure most students are able to master the instructional objectives
- Students are not penalized for mistakes on 'formative' tests

Sources: Guskey; Slavin; and Block, et al.

Table 2
Characteristics Defining the Three Types of Mastery Learning:

TYPE:	Instructional Basis:	Who Paces Instruction:	Instructional Time:	Curriculum:
Heterogeneous Classroom, Year-Long Programme	Group	Teacher	Relatively fixed	Relatively fixed
Individualized Programme	Individual	Student	Variable	Variable
Homogeneous Classroom, Modular Programme (Leveled)	Group	Teacher, within module	Fixed, within module	Fixed
		Student, at module end	Variable, at module end	

(NOTE: Student can repeat, advance or skip a level at module end)

4.1.3 Grades and Mastery Learning

Bloom, while questioning the premise that grades and their sorting function are a necessary component of learning, concedes that they are socially expected and are an integral part of our culture. He advises that if grades are to be assigned in a mastery learning school, they should reflect students' actual learning rather than their standing in relationship to others and should not penalize some students for taking longer than others to learn. In mastery learning, 'failure' is not an option. Because school policies differ so

greatly, several possible ways of assigning grades in mastery-learning classes have been devised:

Many individualized and 'leveled-group' mastery-learning situations grade simply with two grades: 'A' or 'Incomplete'. The 'Incomplete' changes to an 'A' when the mastery standard is eventually met. A variation of this is to distinguish among 'passing' grades in order to give an extrinsic incentive to do more than the minimum required (e.g. 85% = C, 90% = B, 95% = A, etc.). The grading system used at our school is based on this variation, and is described in more detail in Appendix 8.3.

Another possibility, a form of 'criterion referencing', is to assign grades based on the number of course goals mastered at a particular point in time (e.g. 70% of course goals mastered = C, 80 % of course goals mastered = B, etc.).

Yet another possibility, advocated by Champlin is the 'open transcript' concept, whereby students are allowed to demonstrate and receive credit for achievement of specific units whenever they are achieved in the student's academic career.

Often a series of summative tests is given during a reporting period (all but the last of these might be thought of as 'formative', depending on how often they are given). Block, et al., advocate giving more weight to the final summative test, which is to include questions from the previous units, because it is more representative of the students' holistic learning and retention. In our school, formative ('progress') tests are given every one to two weeks, and a summative 'mastery' exam every five to six weeks at the end of a 'learning module'. This is the point at which students can change classes, depending on whether they repeat, advance, or skip a level.

The bottom line, of course, is that in mastery learning when the student leaves a particular classroom, grade, or institution, the transcript should indicate what the student has learned. With typing, it could show 'words per minute' and % accuracy; in language learning, it could show the highest 'level' achieved. (In some other subjects, writing descriptions of student learning could be far more challenging -- or anecdotal -- for

instructors). In our institution this is not a big problem because only the first four levels ('Foundation') involve the mastery grading system. Students spend from one to three-and-a-half years beyond this at our institution, being graded under a more traditional system. The registrar has devised a means of integrating grades from the two systems into one final grade-point-average (see Appendix 8.3).

4.1.4 Mastery Learning and the Slower Student

If mastery learning simply resulted in moving the normal curve of achievement in a heterogeneous class intact, but to the right on a percentage scale, the benefits to the less-advantaged students would be a sense of greater accomplishment and greater self-esteem, which could enhance self-efficacy and therefore lead to greater future achievement: 85% 'feels better' than 50%. However, Block, et al., claim that the curve does not move intact. Rather, mastery learners' rate and achievement of learning becomes less spread out and bulges far to the right rather than in the middle. Far more students become high-achievers; those few remaining on the left, while recognizing that others may be achieving more than they, still benefit through mastery learning's greater allocation of teacher time plus more time to meet learning goals, and resulting enhanced achievement and self-esteem. The cycle of failure for them becomes 'short-circuited'. While some possible questions remain (e.g. 'What about the students who simply cannot keep up with the group, even with extensive extra time?'), advocates and critics alike tend to agree mastery learning is generally beneficial to disadvantaged (low-ability) students' achievement and self-esteem.

The claim that achievement under a mastery system becomes less spread out as slower students 'catch up' to their faster peers is more likely to hold for heterogeneous classrooms. In individualized mastery instruction, as Arlin (72) points out, the spread among students tends to become greater, as 'fast' students are allowed to advance far

more quickly than they would have been able in a heterogeneous classroom and 'slower' students are allowed to take the time they need to truly master the materials. In our five-module 'levels' system, we start the year with three levels (one to three), but usually end the year with five levels (three to seven), resembling the same phenomenon found by Arlin.

4.1.5 Mastery Learning and the Faster Student

Here is where opinion diverges greatly, especially in regard to mastery learning's effect on high-ability students in a *heterogeneous* classroom. Some claim mastery learning results in teachers lowering the test ceiling (and with it, standards of success), simplifying the curriculum (neglecting critical-thinking and problem-solving, while emphasizing 'basic-skills') and spending more time with lower-ability students so that they can pass. Meanwhile, they fear, higher-ability students are being held back from progressing and deprived of challenge -- while waiting for their slower peers to 'catch up'. This is referred to by Arlin (68) as mastery learning's 'Robin Hood' effect which robs high-ability students of teacher time while giving it to the low-ability students.

Nevertheless, others such as Block, et al. cite research (Chan; Conner, et al.; Fitzpatrick and Charters) which seems to substantiate Bloom's proposal that mastery learning would lead to maximized learning for all without having a detrimental effect on faster-learning students. They noted no negative effects on critical thinking, problem-solving, or retention over time. The verdict is still out, however.

The use of mastery learning is less controversial with individualized instruction or with homogeneous, leveled classrooms such as are found at my institution (Slavin 206). In these situations, low-ability students would be expected to benefit from the high mastery-standard and the additional time provided for them to master course goals without penalty. On the other hand, the progress of higher-ability students should not,

theoretically, be affected as they are not expected to wait for their slower peers to catch up. These two assumptions will be somewhat challenged at points during my research.

4.1.6 Mastery Learning and the Language Teacher

Some mastery learning and 'outcomes-based-education' (or OBE -- which some people prefer to call the various curricular reforms that grew out of mastery learning) critics claim that its behaviourist-inspired emphasis on breaking topics down into small skills can backfire on the teacher, especially the foreign language teacher.

Gretchen Schwarz and Lee Ann Cavener (336) claim that in particular, "English . . . is a discipline that is not organized in a cumulative, sequential, linear fashion . . . The behaviourist idea of breaking down learning into bits that must be mastered before a student can go on does not work well in English. The discipline itself is more holistic, recursive, and process-oriented, to say nothing of the various ways in which students learn." The sentiment expressed here is one which is hotly debated at our college, as will be seen in the qualitative research section particularly.

They also criticise outcomes-based-education for its bureaucratic emphasis, ignoring what Dewey would call 'teacher professionalism': "Although OBE advocates claim that OBE liberates teachers, the emphasis on *standardization* and *accountability* . . . keeps teachers voiceless, yet responsible for the results . . ." (my emphasis) (Schwarz and Cavener 325). While standardization and accountability are not intrinsic to mastery learning philosophy, mastery learning does fit neatly into the 'vertical' methods described in the Introduction, and as will be shown, was in part selected for use at our school in order to ensure some standardization and accountability.

4.2 Use of Ability-Grouping ('Leveling') versus 'Tracking'

4.2.1 Rationale for 'Tracking'

This is a most important issue at many foreign language institutions, including ours, which are philosophically opposed to 'tracking' but find it pedagogically useful to place students in at least some of their classes according to ability levels. James S. Coleman and Kenneth A. Strike have both grappled with this distinction. Coleman ("Concept") showed how, over time, North American secondary schools diversified from providing only one (academic) track to offering a second, vocational track. This diversification was thought to promote a greater choice of opportunities to young people as a group. Free secondary academic or vocational education would then, it was reasoned, be available and relevant to the children of all classes.

However, as Coleman pointed out, it has come to present quite the opposite dilemma on the level of an *individual* student, who, through achievement testing or personal choice, is 'tracked' or assigned "to a curriculum . . . (which) . . . closes off for that child the opportunity to attend college" (7). "(This) . . . assignment of a child to a specific curriculum implies acceptance of the concept of equality which takes futures as given" (10). Coleman, therefore, was quite concerned with the inequality of opportunity tracking can confer to individuals. I will refer to this practice of placing an individual student onto a curriculum which forecloses significant future possibilities as 'tracking'.

4.2.2 A Sample Case: Elementary School Reading Groups -- Tracking or Ability Grouping?

Reading groups at the elementary school level have often been criticised because they are assumed to be the first stage of tracking in an educational 'meritocracy' which

continues to significantly influence one's life chances beyond graduation. Kenneth A. Strike ("Fairness") attempted to analyse this phenomenon from a justice point of view.

By 'meritocratic' Strike characterized practices such as adherence to strict medical school entrance standards which result "in the distribution of some desired but scarce benefit to those who *deserve* it. Meritocratic selection is often thought to be justified in that it results in an efficient distribution of scarce resources *to the benefit of all*" (my emphases) (127).

Critics of ability grouping have questioned "whether what is ostensibly a meritocratic decision is, in fact, based on merit or . . . whether, once students are grouped, instructional time is equally divided" (134). Strike, on the other hand, contends that these questions are irrelevant since ability-grouping, at least at the elementary level, should not be concerned with merit, but with such *non-meritocratic* criteria as the child's personal needs and ability to profit.

Strike takes issue with those who claim being placed in a particular ability group can negatively affect one's self-respect. He contends that if so, this is an unfortunate "consequence of the fact that ability grouping is so often assumed to be part of a meritocratic selection process . . . " (133) and that it can, in fact, enhance esteem, by giving students a greater opportunity to excel, since they are placed in what is truly the best situation for them personally to learn. Although Strike is specifically speaking of elementary school ability-grouping, his rationale for it is the same as that used at my school for 'leveling' students into ability-grouped classes.

4.2.3 Equality of Opportunity and Tracking

Strike implies he sees no great problem with a meritocracy or tracking in a *secondary* school when he states, "I do not believe that the fair value of liberty requires substantial equality *beyond the point of minimal competence* . . . (It) does not require that

everyone be held to a lowest common denominator of competence. It does, however, require that expertise be equitably distributed. What threatens the value of liberty is a monopoly (my note: by one societal group) on expertise" (133).

Strike supports inequality which most benefits first, those *individuals* naturally less-endowed intellectually and second, the less advantaged (cultural/racial/socio-economic/ gender/religious/etc.) *groups*. The goal in each case is different, though. The first case focuses on minimal competence on an individual level. The second focuses on equal distribution of competence (outcome distribution) among groups.

In summary, Strike supports ability grouping at the elementary level as long as the purpose is to further the two above goals, and providing that it does not serve, nor is it perceived as, the first stage of a meritocratic selection process. Adding a prerequisite of minimal competence on an individual level to the goal of equal outcome distribution among groups may be seen as a useful caveat protecting the ideal of equality against its nemesis, mediocrity.

I am reminded here of an observation by Dr. Martin Luther King, Jr. (147) who stated that compensatory programs were necessary because "It is obvious that if a man is entered at the starting line in a race three hundred years after another man, the first would have to perform some impossible feat in order to catch up with his fellow runner." King is obviously not referring to an individual African-American, because individual African-Americans have certainly 'caught up with' and surpassed most Euro-Americans in every way. He is referring here to African-Americans as a group, whose outcome distribution he wants to approximate that of other Americans. However, Strike would include a goal of minimal competence on the individual level in addition to that of equal outcome distribution among (in this case, racial) groups.

4.2.4 Leveling and Tracking

In describing the program we use at our school, a former Japanese board member also uses a racetrack analogy:

One of the characteristics is that there are two categories among the levels. Levels One to Four are called 'Foundation' and Levels Five to Seven are called 'Transition'. There are huge differences in concept between Foundation and Transition. Foundation is like, for example, going to driver's training school. If students have developed a skill, they can move on to the next level. If not, they repeat the level . . . Students only can repeat levels up to Level Four.

Because each student has a different speed of skill learning, one student will learn the skill of Level Three within two modules when the other student learns it in one module. This 'learning speed' is different for each individual, and 'faster' does not necessarily mean 'better'. This difference is like some people can naturally run faster than others; however, if the slower runners have proper training, they can run faster. This training is like Levels One to Four.

Once students are in Level Five, 'learning speed' is not an issue anymore. After they complete Level Four, all students can run faster than a certain speed (my note: Strike's 'minimal competence'). Of course there are faster runners and slower runners, but every single student can run at least 100 meters for 15 seconds or faster, and that is the minimum they will need to pass classes in Transition.

In my opinion, one of the more interesting issues at my institution has been this use of 'leveling' according to ability. I contend that tracking (which is often called 'specialization' when the various tracks are valued equally by society), even at the post-secondary level, should be avoided as long as possible. Any leveling or ability-grouping should be along the line of going through a series of 'pre-requisites' which advance one to a desired goal rather than foreclosing advancement to higher levels. Everyone is seen as climbing *the same ladder*, though people climb it at different speeds and, if they are tall enough, might even be able to start climbing it several rungs above the bottom. This is

what one sees in ballet or music lessons, or in schools such as ours. Unless it forecloses future choices, this is not tracking, but ability-grouping.

Of course, there inevitably comes a point at which a person must decide their future course -- be that in choosing a 'major', deciding to study language 'X' instead of language 'Y', or in deciding to pursue a particular trades certificate -- all of which do, in fact, take one 'out of the GENERAL race' and put one on a SPECIFIC 'track'. The point is to put this 'fork in the road' as far off as possible so as to provide the greatest selection of opportunities as possible to the greatest number of young people - to come as close as possible to the mirage we call 'equality of opportunity'.

4.3 Relevant Japanese Curricular, Evaluation, and Justice Issues

In Japan, too, there are both horizontal and vertical controls guiding people's behaviour. However, both types of control are far better established and taken far more for granted than in Canada. Japan sees itself as an egalitarian, homogeneous society in many ways and takes great pains to discourage individuality in the name of group *harmony*. Yet, its strictly-regulated meritocracy determines to a large degree what individual Japanese people do and how they relate to each other. While my western mind-set automatically sees this as a dichotomy, the Confucian mind-set sees these as complementary aspects of the *harmonious* way in which people are meant to live. Perhaps harmony can be seen as a 'superordinate' position uniting the two divergent Confucian values of merit and equality, much as fraternity (thesis, page 22) can be seen as bridging the gulf between liberty and equality in democratic thought.

Accordingly, the two first philosophies, egalitarianism and meritocracy, hold sway respectively at the elementary and secondary levels. By the time students reach the post-secondary level, they have successfully learned how to meld the two; they will have

a strong sense of loyalty to whatever group they join, while showing an equally high degree of respect for those in positions of authority.

4.3.1 Horizontal, Egalitarian Controls

'Horizontal' controls stress Japan's group mentality and egalitarianism. Japan's schools, both public and private, must teach the various curricula dictated by the federal Education Ministry, using only the textbooks they prescribe, thus contributing to the perception of Japanese schools as egalitarian agents, unifying the classes and the different parts of the country. Elementary schools do not differentiate among pupils according to ability (except for severe disabilities). Students pass on to the next grade with their cohort group, regardless of their performance or mastery of skills. To fail a student is not an option; in fact it would be an admission of poor teaching. Peer controls, occasionally including bullying, are very strong, rapidly teaching those who deviate from the norm that 'the nail that stands out will get hammered down,' as the famous Japanese proverb goes.

Individuality and creativity are not nurtured in Japan as they are in Canadian schools. Unlike their peers in Japan, students in North America are asked from an early age not only to have personal opinions, but to be able to express, defend, and justify them. The individual's opinion is sought after and his right to hold it respected, but it is open to challenge at any time. Recognition of individual performance is very important, and by secondary school, 'copying' out of a book without acknowledging it, or copying another student's work is considered 'theft of an idea'. In contrast, Japanese students may not see individual ideas as having such value, and may see copying as a sign of respect for the author, artist, etc. There are some vocal Japanese critics of this cultural value, however, as in the 1985 Provisional Council on Educational Reform's *First Report on*

Education Reform which was commissioned by the Government of Japan to deal with a perceived 'crisis in Japanese education'.

Because of their cultural ideal of homogeneity, Japanese experience discomfort in discussing individual differences, especially in ability. In elementary school, everyone of the same age is treated as being of the same ability and generally expected to be able to perform at the same level. If they do not, they are exhorted to try harder, as poor performance is interpreted as a lack of will or effort, not lack of ability or 'readiness'. Parents and schools generally frown on 'ability' testing, assuming that innate ability or lack thereof is seldom a reason for success or failure; rather, they have a deep belief that hard work will lead to success. In turn, 'lack of ability' is an unacceptable explanation for substandard achievement.

However, at puberty, once the Japanese child is thoroughly imbued with the notion of group identity, her/his group undergoes great change, and the child is placed in a new position of vertical competition with her/his peers.

4.3.2 Vertical, Meritocratic Controls

This is because after elementary (or junior high school at the latest), merit, not equality, is the driving force, as students are sifted and divided according to their ability to pass school entrance exams. Torstein Husen refers to this system as "the Great Sieve that sorts and certifies people for their slot in society" (Husen 411)." This 'Great Sieve' is also referred to as the 'Examination Hell' which determines future placement on the Japanese meritocratic ladder.

The idea of meritocratic civil service exams originated in China in the sixth century and was justified on the grounds that "this method would allow those with natural ability to enjoy equal opportunity with the aristocracy" (Pinar, et al. 797). It acknowledged that human talents of value to the state were to be found among

commoners as frequently as among aristocrats, a 'horizontal', egalitarian notion. On the other hand, the goal of this method was to place those who possessed these talents within the control of a 'vertical' hierarchy, with merit the criterion of initial placement and age the primary criterion of advancement. This idea was to take firm root in Japan, where it now affects practically every person who attends school or gets a job in the country.

Pencil-and-paper tests, which are primarily short-answer (but recently have required an occasional essay-answer), determine a student's progress throughout every stage of this system, from entry into even some kindergartens and elementary schools, to junior high, to high school, and finally, to college or university (Costniuk 147; Unks 35). From the top down, it is not the student's school record that counts; rather, performance on the entrance test determines to which school s/he can proceed, with the reputation of the preceding school determining which succeeding school (or eventually, employer) will even consider the student as a possible candidate, worthy to take the exam.

Many large businesses and government ministries hire only from among the graduates of a particular university. On-the-job education is provided by employers, who expect even university graduates to have only a good general education and very little differentiation or specific job-oriented practical knowledge (Leclercq).

Ironically, the hardest and most significant part of university is getting admitted. Once admitted, the student is virtually assured of both graduating and gaining employment based not on her/his university record, but on the prestige of the university itself. Although there are many serious university students in Japan, the university experience in Japan is sometimes referred to as the 'Four Wasted Years' or 'Leisureland' (Chapman).

Private *juku* (after-hours) schools function to assist individuals to attain a higher slot in this meritocracy. Attended by the majority of Japanese secondary pupils, they provide remediation to those with learning problems, enrichment to those who need extra challenge, and review to those students who simply want to do better on exams. Public

school teachers often lecture to large classes with little concern as to whether the majority of students understand them; this is because it is understood students will cover the subject matter again in the *juku* school.

Juku, at which many students study until midnight, are blamed for the common problem of Japanese students sleeping in their daytime classes. This is exacerbated by the fact that Japanese teachers do not regularly interact verbally with students during lectures, so students do not feel pressure to engage and can 'dose off' without penalty.

Also, 'the group,' because it doesn't want the failure of any of its members to bring it public embarrassment, often carries with it students who depend on their more diligent peers to take notes and help them with homework. Though Japanese students are highly competitive, then, they also feel significant responsibility for the success of their group.

Similarly, *juku* can be seen to assist both horizontal and vertical power structures. Thanks to *juku* schools which compensate for their shortcomings, says Kazuyuki Kitamura (161), public schools "can function according to the two principles of egalitarianism and uniformity," acting on the premise that all students have equal ability. Ironically, however, *juku* are available only to those in society whose parents can afford them, thus widening the gulf between parents with high and low incomes into even greater inequalities between their children's educational credentials -- and the respective children's future earning power.

At every level of education, then, the focus of teacher, parent, and student is on performance at the next crucial entrance exam. Junior high schools teach to the tests of the target high schools, and senior secondaries teach to the university entrance exams. At the private *jukus*, the entrance exams are the prime agenda and most learning is by the traditional rote method. Ability to demonstrate knowledge of specific facts on a specific date, then, is valued far more than ability to perform well in day-to-day classroom tasks or to organize and compose one's thoughts in either written or oral discourse.

Vertical power encompasses both the respect that one feels towards those above one in the hierarchy and the desire to rise within the hierarchy oneself. Throughout the high school years, culminating in the university entrance exam, the young Japanese person establishes her/his general position in the meritocracy, one which will probably determine much of the rest of her/his life chances. Given the importance of this position, it is no surprise that Japanese students are very anxious to excel, and fear making mistakes.

4.3.3 Amalgamation of Horizontal and Vertical Systems to Assure Harmony

Japan reflects the melding of the Confucian ideals of respect for hierarchy (including desire to rise within it) and submersion of self into the group. Strict hierarchies exist, yet equality is imperative within the group. These ideals can be seen in Japanese attitudes towards ways of showing respect to the teacher, towards ambiguity in the curriculum, towards tracking, and towards the written word. All of these attitudes are commonly exhibited in classrooms at our college.

4.3.3.1 'Respect'

The ways that Japanese students show respect to the teacher are completely the opposite of what is expected in Canada and, unfortunately, tend to stifle oral language learning. First, one normally shows respect by silence, certainly not by chattering. For another, in Japan, pupils often show respect by trying to faithfully copy their *sensei*. In language learning, the end goal is not to simply 'parrot', but to generate unique, context-appropriate utterances or writing. In a non-interactive classroom such as is the norm in

Japanese secondary schools, at least (interaction is far more common at the elementary school level), this rarely occurs.

4.3.3.2 'Ambiguity'

Japanese generally feel that respect for authority is necessary to ensure harmony, whether or not that authority is 'justified'. This abnegation of self and peers can lead to a belief that there is only one right answer, that of 'the authority'. Ironically, this way of thinking somewhat parallels that of western thinking which sees reality as unitary and more accurately (objectively) perceived from 'outside self'. The difference is that for Japanese, 'outside self' can imply another person, while in western thinking it means 'objective' science. [This 'vertical', authoritarian kind of thinking is culturally, I feel, tempered by another, more 'horizontal' Japanese thought pattern which teaches that 'the truth' instead of being 'out there' resides in multiple intersubjectivities in which reality is an "intersubjective construct to be formulated and negotiated intersubjectively," as Pinar, et al., (412) paraphrase Tetsuo Aoki.]

Nevertheless, many young Japanese students come to assume that answers are either correct or not, and that the teacher is the judge. If a student gives an incorrect answer publicly, s/he shames (embarrasses) the entire group. As a result, if a student is unsure of the answer, s/he remains silent (also a sign of respect of the teacher) or tries to consult with fellow-students rather than say the wrong thing.

Japanese students are incredulous when western teachers claim, in the Socratic tradition, not to know 'the answer,' to insist that there are many acceptable answers to a particular question, or to encourage students 'not to worry about mistakes - just talk (or write) as much as you can!' Japanese teachers, particularly in the secondary schools, are expected to follow a very explicit, prescribed curriculum which will prepare students for university exams that do not tolerate ambiguity. Students passively take in knowledge

from the teacher through listening, reading, and silent observation, do practice drills to memorize it, and reproduce it on an exam.

4.3.3.3 'Tracking'

Japanese secondary schools are mostly untracked. However, as Susan Goya (128) points out, "Japanese students *are* tracked, not into different programs within one school, but into entirely different schools. Moreover, this tracking rigidly determines a student's future career possibilities." Therefore, once the student finds her/his place within the hierarchy of secondary schools, s/he is able to find her/his place within the egalitarian, homogeneous group within that school. The placement of students into levels at our college, though certainly not intended as 'tracking' (see thesis, pages 45-49) could easily lead to student perceptions of a meritocracy. Students might consider their 'group' to be their classmates in the same level, rather than seeing the school as one egalitarian, homogeneous group.

4.3.3.4 'The Written Word'

The Chinese civil-service exams emphasized the written word. This may have been because the same written (ideographic) language was successfully used by people speaking many different spoken languages of China. Obviously, their common language was the written one, not their various oral ones. In Japan, too, written language forms are more respected than spoken language. However, tacit, mutual understanding is considered superior to either one: "Japanese tend to distrust verbal facility in communicating personal opinion as being glib and superficial . . . simplicity of expression . . . is valued more highly than elaborately reasoned explanations" (Naotsuka and Sakamoto, et al. 173-4). As S. Nakayama notes, these values clearly contrast with

the Greek and Judeo-Hebraic oral traditions of rhetoric -- dialogue, reasoned argument, and debate. Presumably, proficiency in both oral and written language is valued by modern language learners. However, Japanese cultural traditions encouraging silence, and valuing the written word over the oral tend to make this kind of learning very difficult for many Japanese students in our college.

4.3.4 Summary

The most important thing I hope to convey here is that this is the *general background* of our students. However, students who come to Canada to study English are generally not typical Japanese young people, but those who want to try a different kind of post-secondary education. They know that learning conditions will be quite different here. They expect to have their Japanese ideas about education challenged. They hope they will like and be successful with the new teaching styles. They are given extensive written and oral translations of promotional materials which explicitly describe the teaching styles of Canadian teachers, and during orientation sessions in Japan experience sample lessons taught by teachers from the college either in person or via video.

Nevertheless, for most of these students the reality of Canadian teaching assumptions and methods comes as a real shock and it is extremely difficult for many of them to overcome the classroom habits of twelve or more years. The adventure -- both for them and for us as teachers -- is the daily attempt to bridge those cross-cultural (not to mention linguistic) gaps. The fact that we are able to do so attests to the tremendous mutual effort of both students and teachers which makes our college a very exciting and satisfying place in which to grow.

5.0 THE RESEARCH PROJECTS

I have completed two research projects in my own institution, one using a primarily qualitative approach and the other using a primarily quantitative approach. The primary goal of these projects was to *determine, from two different perspectives, the degree to which the present strictly-leveled, modular, discrete-skills-based, mastery-learning programme used by the college promotes a just distribution of educational resources, according to the criteria set forth by John Rawls*. The secondary goal was to *demonstrate the type of research advocated in my thesis, which uses a combination of qualitative and quantitative methods to evaluate fairness*, in particular when determining whether a curricular programme is of benefit to all students.

As I worked into the project, I realized I had acquired three additional goals: First, as I started to wrestle with some of the implications of John Rawls' theory of 'justice as fairness,' I saw one shouldn't, as Rawls himself warned (8-9), assume that his theory would be directly applicable in any situation. Therefore, I saw that I must also *determine whether Rawls' criteria for evaluating the justice of institutions would be appropriate* for the type of research I was attempting to do.

Also, I wanted to *see what kinds of answers the two respective (qualitative and quantitative) research methods would give me*, and how the data could be synthesized. At the beginning, I did not have a clear notion of how the data would 'fit together.'

It should be noted that if the programme cannot be demonstrated to be equal or superior in effectiveness to others (e.g. the one it replaced), it cannot be considered to benefit the group as a whole, and would therefore automatically be considered an unjust innovation. Therefore, my final goal was to *compare the effectiveness of the present programme to that of the programme it replaced*. Obviously, it cannot be compared to any other possibly-superior programmes which haven't been tried at this institution using

these research methods (post-hoc test data and teacher interviews about their experience of the two programmes) which dictate that both programmes have been both tried and tested on students at the same institution.

Note that to determine the justice of the programme, one quantitative and one qualitative project were chosen from among myriad possibilities. Obviously, the greater the number of perspectives from which a programme is viewed, the more accurately it can be seen and the more fairly it can be judged. If time and funding permit two (or preferably, even more) research projects, it would seem logical to try to look at something from at least one quantitative and at least one qualitative perspective, much as a doctor not only takes patients' vital signs, but also asks them how they feel. Each perspective can be seen to inform and support the other as well as to help confirm or reinforce notions formed by using the other perspectives. Thus does the ancient wisdom of the fable, *The Blind Men and the Elephant*, continue to remind us of our personal limitations and our collective wisdom. Not only does the fable teach that a more complete notion of a whole comes from viewing it from different perspectives; it also teaches that the view from each perspective is misleading, if the viewer naively supposes that what is seen is the whole. In short, neither of the projects I have chosen gives the 'final word' on the justice of this programme change; the two projects put together are better than one, but they are still only two out of countless possible perspectives that could be taken to provide a more accurate determination.

John Rawls' principles state that an institution, policy, programme, etc. can be deemed 'just' if it benefits and distributes resources equally among all participants or, if not equally, that any unequal distribution can be shown to eventually benefit the group as a whole and benefits the least advantaged more than would a strictly equal distribution. Therefore, the distribution of resources among the lower-, middle-, and higher-entry groups will be considered separately.

It is possible that if a programme can be shown to benefit one subgroup, while not benefitting the group as a whole, its use within that subgroup only might be justified providing this limited inequality can be shown to benefit the group as a whole including the least advantaged.

As a bit of background, I noted a fair bit of questioning among my fellow first-year (of a two-to-four year programme) faculty members about the wisdom of retaining an 80% mastery pass standard, and some dissatisfaction as well with the need for summative exams after every five-to-six week module. To some teachers, the 80% standard seemed artificially high and the modules too short. I thought (and my institution agreed) it might be useful to have some formal feedback at this time, therefore, on the present programme. The 80% mastery standard and the five-to-six week modules are an integral part of the programme (but impossible to evaluate separately from one another, nor from the associated contribution of other components such as 'levelling' students).

If the research were to show the programme to be *equally or more effective* than its predecessor, its continuation would be supported. If it were to show the programme is not as effective, there would be good reason to question it and justification for experimenting with alternatives. If it were to show the programme is justly administered to all students (*provides a just distribution of learning outcomes among all ability-levels*), its continuation would also be supported. Finally, if the programme were shown not to benefit some segment of the student population, the institution might want to consider alternatives to it for that segment of the population.

I don't claim to be providing all the data relevant to making 'the right' decision. For example, the SLEP test which will be used to assess 'learning outcomes' (see Appendix 8.5 for a description of this test) does not test language 'output' (speaking and writing) skills, only the more easily measurable language input skills, listening and

reading. It is entirely possible that the output language skills are developed quite differently in the new programme, but the data to show this is not as readily available nor as clearly 'objective' as the SLEP scores.

My intent as a researcher was to create discussion, not dissension. The data provided will, hopefully, help inform and clarify this discussion in a coherent, organized fashion.

5.1 Quantitative Project -- Description and Results

This was a post-hoc study comparing matched-pairs of students before and after the institution's initiation, in the 1991-1992 school year, of a strictly-leveled, modular, primarily discrete-skills-based, mastery-learning programme. The study took place in May, 1996 through February, 1997. Listening and Reading SLEP scores of 824 students (207 in 1990, 255 in 1992, 196 in 1994, and 166 in 1996) were considered. I started out with these questions and hypotheses:

5.1.1 The Research Questions:

When Japanese students of English who are studying in a high-medium-low *track* (as distinguished from 'level' or 'ability group' in thesis, page 45), non-modular, non-mastery, and primarily content-based classroom situation (hereinafter referred to as 'the previous' programme) *are compared with* those studying under a *leveled*, modular, mastery-learning and primarily discrete-skills-based classroom situation (hereinafter referred to as 'the present' programme):

- (1) Is there a difference in the way the previous and the present programmes' mean SLEP (Secondary Level English Proficiency Test) scores increase over the year, and if so, what is it? (*Note: separate scores are given for the Reading and Listening components of this test.*)

- (2) Is there a difference in the way the two programmes' respective mean Reading SLEP scores increase versus the way their respective mean Listening SLEP scores increase, and if so, what is it?
- (3) Is there a difference in the way the two programmes' respective lower-entry SLEP students' mean scores (in both Listening and Reading) increase versus the way their higher-entry SLEP students' mean scores increase and if so, how?

5.1.2 Hypotheses:

Under a levelled, modular, skills-based, mastery-learning situation:

- (1) There will be a generally positive change in the mean increase in SLEP scores when the present programme is compared to the previous programme.

Rationale: This is because of the supposed superiority of a leveled, discrete-skills-based, modular, mastery programme in teaching basic skills and because the SLEP test in particular measures basic English listening and reading skills. As well, the teachers' growing expertise (through two to six more years' experience) in teaching the same clientele should have some positive effect on student achievement, regardless of the particular programme used.

- (2) The positive mean increase in total SLEP hypothesized in (1) is projected to have a proportionately greater impact on listening than on reading.

Rationale: Both reading and listening skills were being equally and specifically targeted in the new programme. There has traditionally been a difference between the way Reading versus Listening SLEP scores change. In general, students improve more dramatically in listening than in reading the first year. Since listening skills always

improve more dramatically, this difference will presumably continue to hold in the present programme.

(3) The lower entry SLEP students in the present programme will exhibit a greater positive change over the previous programme in both Reading and Listening mean SLEP score increase than the higher entry SLEP students.

Rationale: This is because mastery programmes are thought to benefit students with lower ability and/or achievement even more than students of higher ability and/or achievement.

5.1.3 Defining Factors:

Population: All students are recent graduates of Japanese high schools studying English in Canada at a Japanese and Canadian joint-venture institution recently granted accreditation by the Private Post-Secondary Education Commission of the Province of British Columbia. While extremely high intelligence or socio-economic status may be found occasionally among the students, the opposite extreme is never found -- the former because of entrance requirements and the latter because of the sponsor income level required to send a child to an overseas private school.

Independent Variable: This is the initiation of strictly-leveled, modular, discrete-skills-based mastery-learning instruction in Year 1991-1992 (fourth year of the school's operation). Note that a detailed description of the programme used in the first three years of the school's operation will be provided through teacher interviews in the qualitative study). There were four levels of the independent variable: previous programme (control) Reading vs. present programme Reading and previous programme (control) Listening vs. present programme Listening.

I assumed that data from the final year of the previous programme (1990-1991) would be useful for the 'baseline' since the programme was well-established by that point; as well, it was the *only* year in the previous program at which SLEP was administered at entry. In addition, I felt that data from 1991-1992, the initial year of the present programme, should not be used as teachers were just becoming familiar with it. SLEP scores from the years 1992-1993, 1994-1995, and 1996-1997, therefore, were chosen to represent student achievement within the present programme. It seemed likely that a two-year interval would give a fair indication of student progress without unduly introducing teacher and student variables (such as teacher personnel changes and changing student cohort attributes).

Dependent Variables: Change in individuals' Reading SLEP and in individuals' Listening SLEP between entry and exit.

Control Variables:

Instructor and Content Factors: Except for the first two years of its operation, the institution has had a very low faculty turnover, so it may be claimed that students in each of the four years in question (1990-1991, 1992-1993, 1994-1995, and 1996-1997) had the same school, the same general staff, and, in general, the same subjects were being taught. There were some significant departures -- more 'content' teaching characterized the previous programme, while the present programme is more 'discrete-skills'-based -- at least in the first four levels. Also, in the 1990-1991 year, many students studied pronunciation skills using computer-assisted learning ('MacEnglish'), a program which was slowly phased out over the following two-to-three years.

There is also the above-noted aspect of improved teacher expertise over the period of the study. The present programme's skills-based nature and this greater collective teaching experience would presumably favour the present programme.

Another possible teacher factor favouring the previous programme might be the greater energy and enthusiasm with which people confront new tasks and novel challenges such as the opening of a new college. On the other hand, the same factor could also have been at play during the implementation period of the present programme.

Subject Factors: Factors of gender (approximately 50% each sex) and age (primarily ages 18 and 19) were approximately the same for all four groups. Students are of the same culture and race. It is assumed that personality and motivational factors were constant for all the years (although the economic downswing in recent years in Japan may have had a motivational effect -- possibly positively on some and negatively on others). Variations in initial ability were controlled for through matching like pairs. Note: Students in the lowest entry-levels were predominantly male and those in the highest entry-levels were predominantly female, in all years in question.

Test Factors: Venue and tape quality can affect SLEP Listening scores. An attempt was made to ascertain whether this could have been a factor. The entry SLEP test is now given to students in Japan, shortly before they arrive in Canada. In 1992, 1994, and 1996 SLEP was administered in large group settings in Japan. The listening conditions and the tape quality at this venue were approximately the same in all three years, according to those who administered it.

However, the entry test in the 1990 baseline year was given at the college, in small classrooms. This (possibly) more favourable testing condition in the baseline year entry SLEP test could have made the initial scores (especially Listening) higher and thus, any gain *smaller*. Note that some of the students may have actually suffered the reverse effect since they may have been suffering from 'jet lag', having arrived just a few days earlier. This would have made the initial scores lower and thus, any gain *larger*. It is impossible to say how much these different conditions may have affected the baseline data.

I can affirm that the conditions under which the exit exam was given on campus were approximately the same in all four years.

Textbook Factors: Different texts were also introduced at the same time as the new programme. In some subjects, new texts were again introduced in the second and fifth years of the new programme. This experimentation with different textbooks and other materials characterizes the institution, in both the previous and present programmes, and is a factor to consider when interpreting the results. The texts in the present programme's second (1992-1993) year and fourth year (1994-1995) were, with minor exceptions, the same. The reading text used in 1996-1997 is different from that used in 1992-1993 or 1994-1995.

5.1.4 Kinds of Analysis Used:

Standard statistical analysis was used. 'Matched Pair t-tests' were performed to determine means of the differences between exit and entry SLEP scores (net increase) according to various conditions (and the significance of these). I used the 'SPSS 6.1 for Windows Student Version' statistics programme on a 486 Packard Bell computer to analyze the data and to make the original charts. For each of the four years, 1990, 1992,

1994, and 1996, I entered each student's student number, entry and exit Listening SLEP score, and entry and exit Reading SLEP score. From these three lists, I was able to set up three Matched-Pair groups (1990 vs. 1992, 1990 vs. 1994, and 1990 vs. 1996) in which pairs of students both had the same entry Listening and the same entry Reading SLEP score. Theoretically, the two students in each pair are considered of the same initial skill level -- they are statistically considered *one person* undergoing *two different treatments*. For a sample of the way these pairs were described statistically, see Appendix 8.1.

In order to investigate the third hypothesis, I had to create three subgroups from the data, a low-, medium- and high- entry group based on their 'total' entry SLEP score, the sum of their Reading and Listening SLEP scores. The three groups were numbered '1' to '3', from low to high. (Note that the institution does, in fact, do its initial leveling of students according to this same 'total' SLEP. While there may often be a perfect correlation between a student's actual 'assigned' entry level at the college and the level to which this statistical procedure assigns them, these must be noted as two entirely different ideas.) In order to use the same criteria for both sets of years, there was a different percentage of students composing the various levels in each set of years. For example, 20% of 1990/1992 students were in Level One, compared to 24% of 1990/1994 students and 25% of 1990/1996 students. Though the cutoff points between levels seemed arbitrary, I was attempting to create a 'low' and a 'high' range of approximately 20% each year. This was as close as it was possible to approximate that goal:

Table 3
Proportion of Subjects in the Three Levels, by Matched-Pair Groupings

<u>Entry Level</u>	<u>SLEP Score</u>	<u>No. of Matched Pairs per Group, per Level:</u>					
		1990/1992		1990/1994		1990/1996	
		%	n=*	%	n=*	%	n=*
'1'	20-28	20%	18	24%	24	25%	26
'2'	29-36	54%	72	55%	55	59%	63
'3'	37+	26%	32	21%	21	16%	17

* Note: 'n' is the number of *pairs*, so $[n = 18] = 36$ subjects, 18 from each year

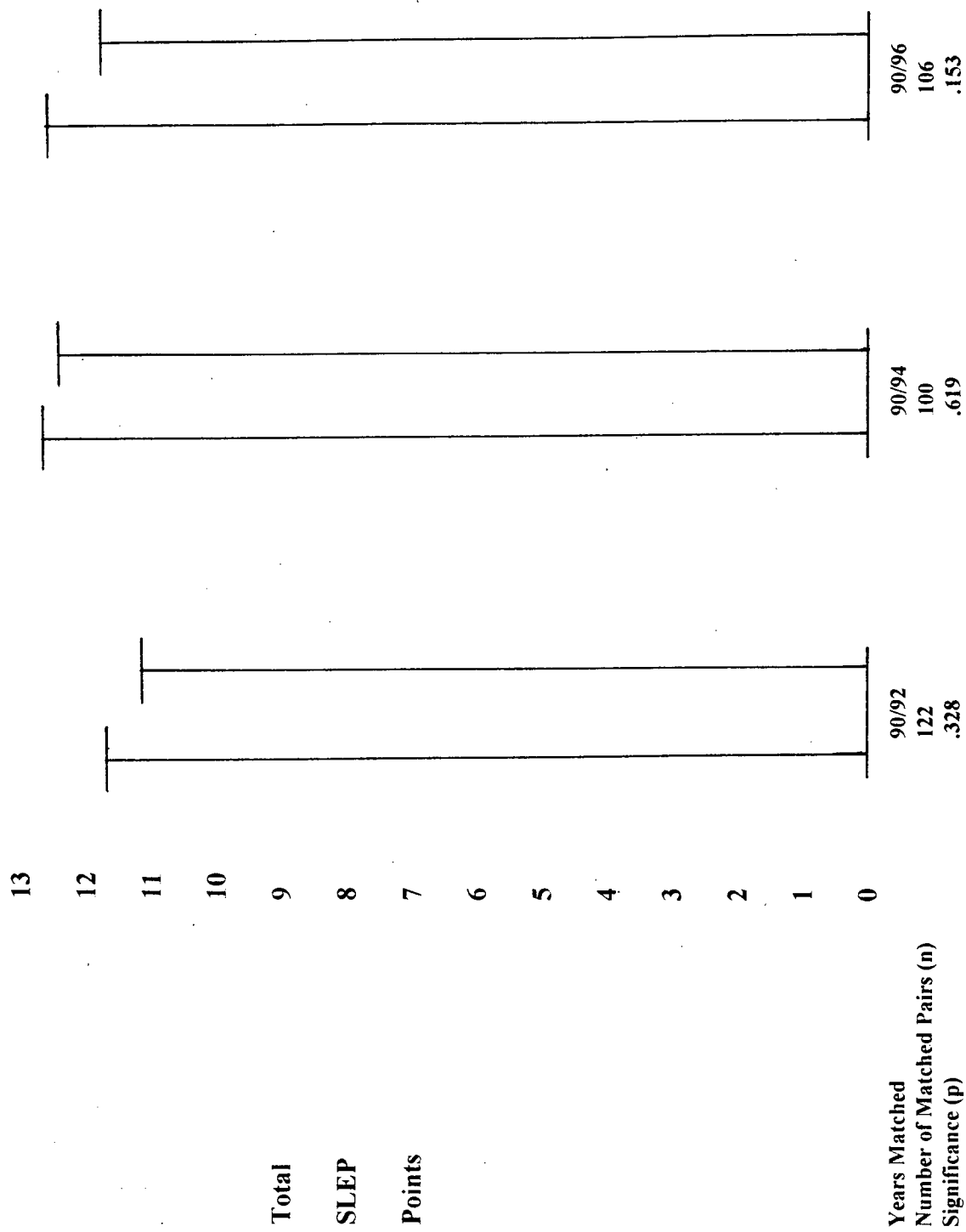
As is apparent, the numbers and percentages of pairs in each level vary greatly. However, this does not affect the means, only their variability. It is more difficult to find significance when comparing two small groups than when comparing two large groups. In general, the larger the sample, the smaller the range of possible means (i.e., the Level Two group is the largest and has the smallest range of possible means).

The statistics programme enabled me to define the three above groups. This was all the information the programme needed. With minimal direction on my part, it did the rest. The results are shown graphically in Figures 1, 2.1, 2.2, 3.1, and 3.2. The more detailed sample programme inputs and transformations plus the actual programme outputs are included in Appendices 8.1 and 8.2.

5.1.5 Quantitative Study Results according to Hypotheses

Hypothesis #1 - Figure 1:

"There will be a generally positive change in the mean increase in SLEP scores when the present programme is compared to the previous programme."



Mean Increase in Total SLEP Points by Matched-Pair Years
Fig.1

The full-group results (illustrated in Figure 1) uniformly surprised me. Comparing 1990/1992, 1990/1994, or 1990/1996, the student's mean improvement in total SLEP was greater in the *previous* programme. However, the 't'-tests showed that these differences were not significant. (1990/1992 significance was $p=.328$, 1990/1994 significance was $p=.618$, and 1990/1996 significance was $p=.153$. Generally only 'p' of equal to or less than .05 is considered significant.

The first hypothesis, then, was not supported. The lack of significance of the difference means that *neither previous nor present programme can clearly claim general superiority* over the other in terms of student improvement in basic English (reading and listening) skills over the year.

Figure 1 also illustrates an anomaly created by repeating a Matched-Pairs Design year after year using the same baseline database. Almost inevitably, the actual number of students in the baseline year is going to be greater than the number of individual cases that are found to be 'matchable' with the students in the year with which they are being 'twinning'. Therefore, three unique sets of 1990 students were created, those being matched with 1992 students ($n = 122$), those being matched with 1994 students ($n=100$), and those being matched with 1996 students ($n=106$). To see the effects of this, compare the mean increase in SLEP scores for the 1990 group matched with 1992 ($mean=11.6$) with the 1990 group matched with 1994 ($mean = 12.8$), and with that matched with 1996 ($mean = 12.7$).

Note that none of the matched-pair groups are representative of the population they are taken from. *The reason for matching the pairs is not to compare the students, or even the years, but to compare the programmes.* For example, if one wanted to compare 1992 with 1996, one would have to match 1992 with 1996 rather than

comparing 1990/1992 with 1990/1996. I have not done this as I am only comparing the two (previous and present) *programmes*.

I was interested in this phenomenon and decided to compare the mean entry SLEP and mean gain of each whole-class with the smaller matched groups to see how representative they were. Here are the results:

Table 4
Mean SLEP Entry Score and Gain (per Whole-Class and Matched-Pair Groupings)

<u>Year</u>	<u>Whole-class or Matched pair?</u>	<u>n</u>	<u>Mean Entry SLEP</u>	<u>Mean SLEP Gain</u>
1990	WHOLE-class	207	32.9	12.1
1990	Matched with 1992	122	33.7	11.6
1990	Matched with 1994	100	32.7	12.8
1990	Matched with 1996	106	32.3	12.7
1992	WHOLE-class	255	35.6	10.7
1992	Matched with 1990	122	33.7	11.1
1994	WHOLE-class	196	33.9	12.1
1994	Matched with 1990	100	32.7	12.4
1996	WHOLE-class	166	32.7	11.3
1996	Matched with 1990	106	32.3	11.8

Note that the higher mean entry SLEP scores generally show less gain. Students in 1992 entered with listening and reading skills that were on average higher than previous years, and, predictably, their average SLEP gain (10.7) was lower.

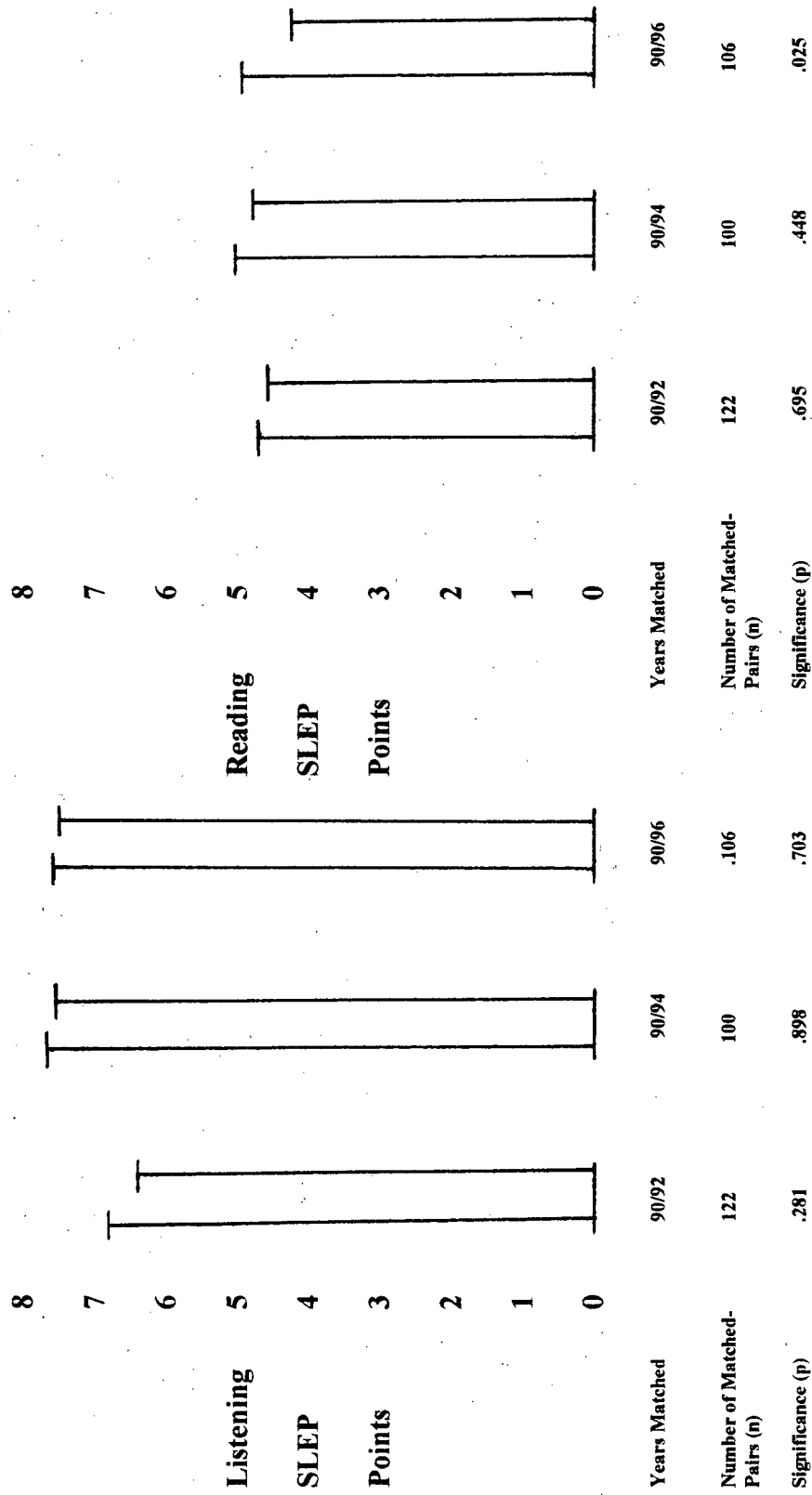
Also, in order to accommodate this difference, the 1992 Matched-Pair group had a slightly lower mean entry SLEP (and higher gain) than the whole-class, and the 1990 group it was matched with had a slightly higher mean entry SLEP (and lower gain) than the whole-class.

However, as shown in 1996, the reverse is not always true; the 1996 whole-class had the lowest mean entry SLEP, but did not show the highest mean increase.

These ideas will be further explored in the next section.

Hypothesis #2 - Figures 2.1 and 2.2:

"The positive mean increase (from Hypothesis # 1) is projected to have a proportionately greater impact on listening than on reading."



**Mean Increase in Listening SLEP Points
by Matched-Pair Years
Fig. 2.1**

**Mean Increase in Reading SLEP Points
by Matched-Pair Years
Fig. 2.2**

Again, in all six cases (1990/1992, 1990/1994, and 1990/1996 for both Reading SLEP score change and Listening SLEP score change), the students' mean improvement was greater in the previous programme. However, the 't'-tests again showed that these differences were not significant. A summary of these results is found in Figures 2.1 and 2.2.

These figures also show that *the second hypothesis appears to hold, for the group overall, at any rate*. Both Listening SLEP and Reading SLEP scores held to the same pattern through all four years in question. *In all cases* the mean Listening SLEP score improvement was two to three points greater than the mean Reading SLEP score improvement. It should be noted that the mean entry Listening SLEP is always lower than the mean entry Reading SLEP. An explanation for this is that in the Japanese secondary school system, English literacy tends to be valued, or at least emphasized, over oral English. Therefore, some of the improved Listening SLEP may be thought of as the students simply realizing what words, previously learned from books, actually sound like -- a case of listening knowledge 'catching up' to related reading knowledge.

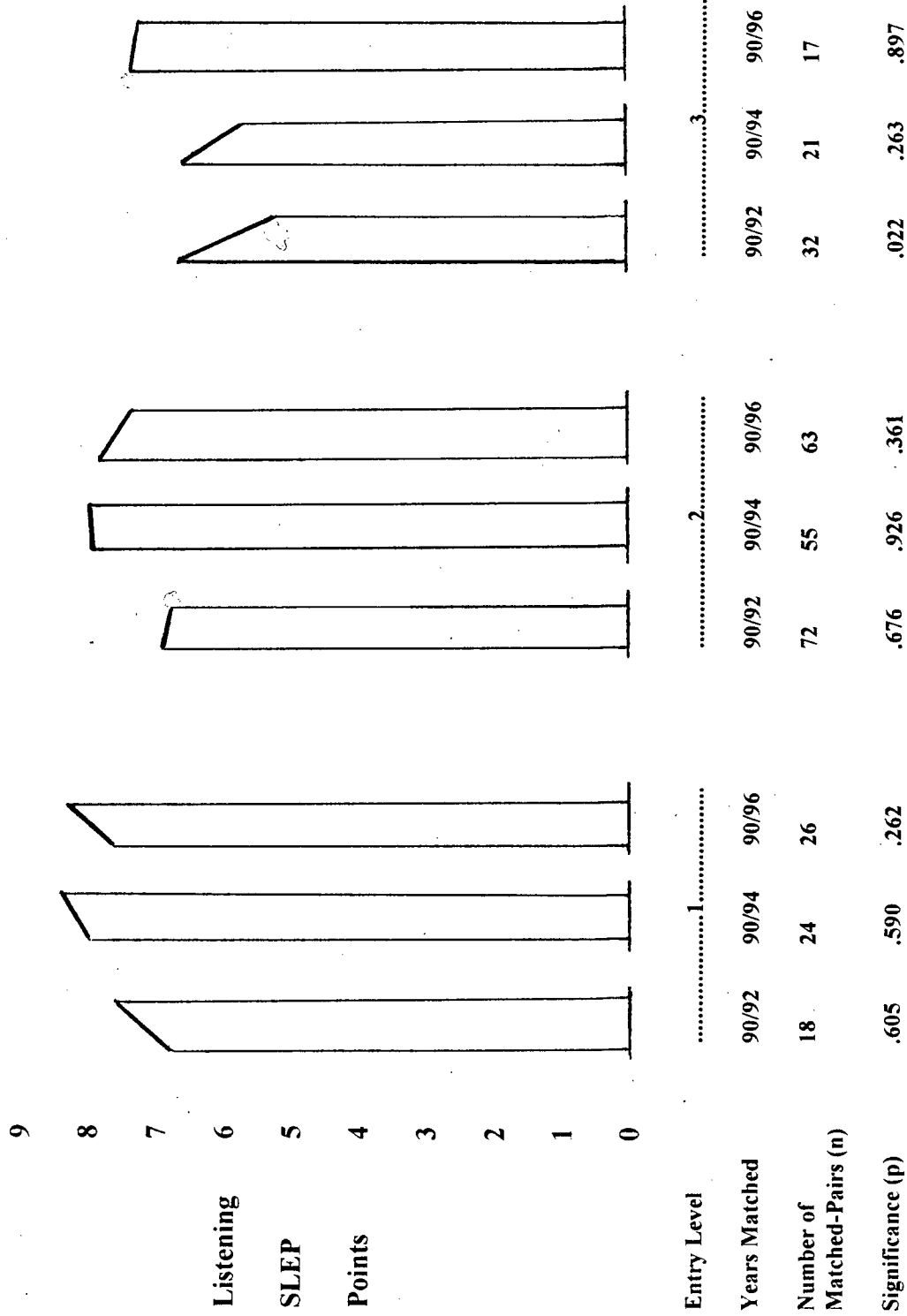
It is interesting to note that sometimes the characteristics of the three groupings of 1990 students exhibited greater differences from one another than from the other years' students. You can see the results of having three quite different groups of 1990 students if you compare the mean changes in Listening SLEP scores in the Figure 2.1 1990 group matched with 1992 (6.85) with the 1990 group matched with 1994 (7.63). In this case, the difference between the two 1990 groups was much greater than the differences between each and its matched-pair year. The important thing is to *compare the trends shown in each graph with one another, not the specific means*; as noted before, the two sets (1990/1992 and 1990/1994) are quite different in composition from one another.

The first significant result of the study ($p < \text{or} = .05$) is evident in Figure 2.2. In the 1990/1996 matched-pair group, the 1996 students' SLEP Reading improvement was significantly ($p = .025$) less than that of their 1990 'identical twins.' Since the 1996 students are using a different reading text ~~from~~ the 1992 and 1994 students, it is possibly a text effect rather than a 'present programme versus past programme' effect. This is probably something we should investigate in the next few months.

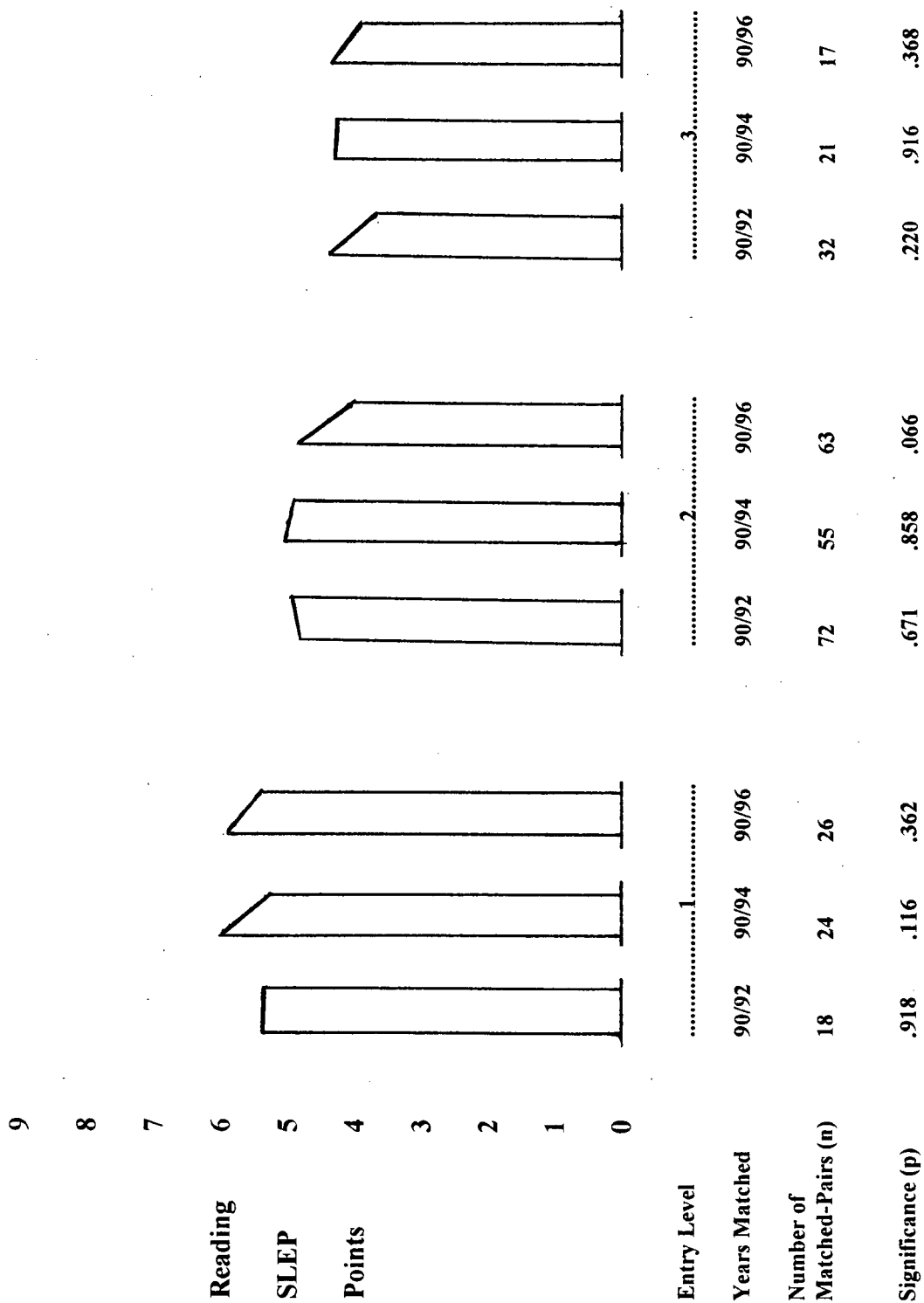
One thing this illustrated to me is the benefit of being able to chart the same standardized test over a period of years. Anomalous scores such as this can be identified as such instead of being taken to mean more than they should (i.e., 'the general trend'), incorrectly influencing important decision-making. On the other hand, tests come in and out of vogue, and there may be compelling reasons to change standardized tests over the years. Also, sometimes there is a well-meaning rush to change curriculum perceived as inadequate. It would seem illogical to wait for several years, proving through standardized tests that there was indeed something wrong with the curriculum before doing something about it. Using the same rationale, doctors often must treat extremely ill patients before all tests confirm their initial diagnoses. This is not to say only qualitative methods are to be used, only that standardized tests may not be practical or appropriate indicators in all cases; non-standardized quantitative data might be more accessible and appropriate. Each of these provides 'one more piece of the puzzle; one more blind man trying to describe the elephant'.

Hypothesis #3 - Figures 3.1 and 3.2:

"The lower entry SLEP students in the present programme will exhibit a greater positive change over the previous programme in both Reading and Listening mean SLEP score increase than the higher entry SLEP students."



Mean Increase in Listening SLEP Points by Matched-Pair Years and Entry Level
Fig. 3.1



Mean Increase in Reading SLEP Points by Matched-Pair Years and Entry Level
Fig. 3.2

The third hypothesis, in which the performance of high-entry versus low-entry students in the two programmes was compared, met with interesting and mixed results [See Figures 3.1 and 3.2]. As mentioned, matched pairs were classified as Level One, Two, or Three according to their entry total SLEP scores. To test the third hypothesis, the changes in both Listening and Reading SLEP scores were compared, giving a mean change in both Listening and Reading SLEP for 1990/1992, 1990/1994, and 1990/1996 according to each matched pairs' entry SLEP Level.

While the individual comparisons are statistically of little significance, an intriguing pattern emerges which bears considering as being more than the sum of its parts. Whether we look at Figure 3.1 from the point of view from 1992, 1994, or 1996, we see the *present programme's Level One students achieving a higher mean Listening SLEP improvement than the previous programme's Level One students.*

The present program's Level Two students, on the other hand, had almost the same (or lower) mean Listening SLEP improvement as the 1990 students in all years.

When one examines Level Three, a mirror image of the Level One performance in listening is seen; 1992's Level Three students (matched with 1990 students) *show a significant, though not great, decrease in listening improvement from 1990* ($p=.022$). Note, this is *the second of only two statistically significant differences* found in the study. Succeeding years 1994 and 1996 also showed Level Three students making less progress than they did in the previous programme, but the differences are insignificant and get successively smaller each year. (Note that the insignificance may be due partially to far lower numbers of matchable students in this level in 1994 and 1996).

Reading [Figure 3.2] follows a different pattern. Reading progress (except for an insignificant anomaly in 1992, Level Two) is consistently lower in the present

programme for all three levels. While none of these differences exhibit $p < \text{or} = .05$, one (1996, Level Two) approaches it ($p = .07$).

Note that while the only statistically significant results described here are the general drop in Reading in 1996 and the Level Three drop in Listening in 1992, other results approached significance. While the patterns are of interest and potentially significant, they should not be misconstrued or used alone to justify any decision-making.

5.1.6 Quantitative Study Results: Summary

Except for one year (1996) in Reading, there is *no statistically significant difference* in the annual increase in Listening or Reading SLEP scores between students in the previous and those in the present programme.

There appears to be *no significant difference* between programmes in the way Listening and Reading improve. Listening consistently improves more dramatically than Reading.

There is a possibility that students respond to the present programme differently according to their entry level. Of those who have entered the school with a low SLEP score (28 or less), those in the present programme tend to exhibit greater mean Listening SLEP gain than those in the previous programme. Conversely, of those who have entered the school with a high SLEP score (37 or more), those in the present programme exhibited a smaller mean Listening *and Reading* SLEP gain than those in the previous programme, though except for one year (1992) in Listening, these differences are not significant. The pattern is intriguing in its persistence, however.

I will attempt to triangulate the data (to combine my conclusions and discussion of these results with those of the qualitative project). I feel the two quite different projects, each seeking to answer the same question, shed interesting light on one another.

5.2 Qualitative Project -- Description and Results

5.2.1 Design of the Study

5.2.1.1 Introduction

This is a descriptive and comparative study using qualitative methods. The study involved semi-structured interviews of teachers. The aim was to examine, from an *instructor's* point of view, their *perceptions* of both the effectiveness (see thesis, page 58) and justice of the previous programme versus that of the present programme.

However, teachers were also asked what they perceived as *student responses* to the two programmes; results of this must be analyzed with the understanding that this data is of a hearsay nature, so quite subjective. Students were not interviewed, as they have no basis upon which to compare the two programmes, having only experienced one. Comparative student perceptions, then, were noted through instructors' imperfect interpretations and memories.

Note that for this study, unlike the quantitative one, I had no clear hypotheses. However, it should be noted that the question to be answered is the same in both studies: How justly (and secondarily, how effectively) were these two programmes distributing educational resources, according to Rawls' principles of justice?

There were several standard questions of an open-ended nature. Digressions from these were encouraged, though all questions were asked of each instructor. There were also several very specific questions formulated in order to ensure specific points were covered by all interviews.

The purpose of the college in encouraging this project was not the justice issue, nor curiosity about what kind of information one could get from qualitative versus quantitative studies; rather, it was interested in my project of seeking out teacher views in

an anonymous forum, with the purpose of (possibly) giving direction to future plans. Therefore some results I will present to the college (included in Appendix 8.4) will be more detailed and site-specific than those in the body of the thesis. The results section of the thesis, though including less detail, will include an analysis of the kind of *information about the distribution of educational resources gained* through this methodology. Similarly, the conclusions I will present at the end of the thesis will include an analysis of the kind of *conclusions about the distribution of educational resources that could be drawn* through this methodology.

The campus seems ready to re-examine itself and consider the possibility of another major curriculum shift. It is hoped that this narrative, an amalgam of about twenty hours of interviews with thirteen teachers, will enrich a thoughtful reappraisal which will contribute an even more exciting and successful episode to this unique institution's curriculum history. There were a total of thirteen subjects. Sixteen instructors taught at least two years in both the previous and the present programme (at least one full year in each). Of these, eleven (ten, excluding myself) are currently teaching at the institution. Nine of the current instructors agreed to participate. As well, four out of the five instructors meeting the criteria for inclusion, but not currently teaching at the institution agreed to participate. Of these four, two are on temporary leave; one quit voluntarily to pursue further education, and one was laid off, but hopes to return to the college.

Since I have also taught in both programmes, I had to be exceedingly careful not to interject my own biases into the interviews. Fortunately, I have not been outspoken in either attacking or defending the present programme, so my views were not generally known, nor (I think) would they have been perceived as threatening to either point of view. While a questionnaire would present less interviewer bias, the instructors are so busy that a questionnaire would probably have been answered in a cursory fashion, if at

all. However, it should be noted that one instructor, presently in Japan on leave, was willing to give a long interview (in written, questionnaire fashion) via e-mail!

5.2.1.2 The Questions

Part One: The Previous Programme

- 1) Describe (the institution's) previous entry-level or beginning ESL programme (before the Foundation programme started in Year Four: 1991-1992).
- 2) What did you perceive as the strengths of the previous programme? weaknesses?
- 3) How was the previous programme perceived by most teachers?
by most students?
by high-entry SLEP students?
by low-entry SLEP students?
- 4) Why do you think (the institution) decided to change to the present programme?

Part Two: The present 'Foundation' programme:

- 1) What do you perceive as the strengths of (the institution's) present Foundation programme?
weaknesses?
- 2) How is the present programme perceived by most teachers?
by most students?
by high-entry SLEP students?
by low-entry SLEP students?
- 3) How well, if at all, does the present programme address the weaknesses of the previous programme?
- 4) How do you personally feel about the following? (I wanted very short, specific answers here)
leveling students according to ability
five-to-six-week modular system
standardized curricula
standardized testing in general
80% pass mark
2 (formative) progress tests counting for 10-30 % of final grade
final (summative) exam counting for 50-70% of final grade
the term 'mastery'

the term 'competency-based'

Part Three: Evaluation:

- 1) What sort of evaluation is needed by
 - a) (the campus to which our students transfer in their second year) to determine student placement in Year Two programmes there?
 - b) (this campus) to determine student placement and advancement in its first year programmes?
 - c) students?
 - d) their parents?
- 2) How well were these needs met by evaluation practices used in the previous programme?
- 3) How well are these needs met by evaluation practices used in the present programme?
- 4) Were the previous program's evaluation methods fair (or perceived as fair) to the students?
- 5) Are the present program's evaluation methods fair (or perceived as fair) to the students?
- 6) Did the previous program's evaluation methods motivate or discourage students in general?
high-entry SLEP students?
low-entry SLEP students?
- 7) Do the present program's evaluation methods motivate or discourage students in general?
high-entry SLEP students?
low-entry SLEP students?

Part Four: Future:

- 1) Would you like to return to the previous programme (at this institution)? If so, in what ways? If not, why not? What would be the implications and impacts of this change?
- 2) Are there any changes you would like to make in the present programme? What are they? Why would you make these changes? What would be the implications and impacts of this change?
- 3) Is there some other kind of programme (neither the previous nor the present programme) that you would like (this institution) to use? What would be the implications and impacts of this change?

Part Five: Closing:

- 1) Do you have anything you would like to add?
- 2) Do you think the questions were fair and represent the questions that (this institution) should be asking about its programme?

5.2.1.3 Kinds of analysis used

Types of responses were loosely-tabulated and frequently-given responses were noted. Idiosyncratic responses, patterns, and relationships among responses were particularly noted. This was in many ways a 'fishing expedition' in that both the questions and the interviewers' attitude during the interviews were very 'open-ended.' I truly didn't know what would come out in the one-on-one interviews. During faculty meetings, the more vocal instructors' views were made quite clear, but the more taciturn, those inhibited by the size of the group, or those unwilling to engage in controversy hadn't made their views public.

The results of the qualitative research will be presented in the form of an historical narrative, because each interview was basically a retelling of the story of the college, and in particular of the development of its curriculum, from a different perspective. Kidder and Fine (69) refer to this practice as 'Research as Story Telling,' noting that all research, quantitative as well as qualitative, tells a story, and that in the analysis of field work (the authors are referring here especially to ethnographic methods), the researcher often is constructing "a narrative pertaining to more than one actor." As stated previously, it is imperative to realize that teachers (and only teachers from the first-year campus) were the only stakeholders consulted; consequently the story is told from their point of view only. Their perspective is only one of several possible versions of the truth and must not be misconstrued as 'the' truth. I have tried to weave all the stories into one coherent narrative, while retaining some of the contradictions and

inconsistencies, the humour and enthusiasm that made hearing it 'once again' always a new and delightful experience. Hopefully some of the flavour of these interviews will be retained . . .

My voice, however, is quite evident: in which statements I choose to include, in which I quote directly, and in how I choose to paraphrase those I do not quote. In the same vein, if given the same set of data, ten researchers would probably choose ten different ways to organize, interpret, and present it. However, I am also convinced that their ten interpretations, though 'different,' would not be contradictory; rather, they would be supportive of one another.

It should be noted that I began teaching at the institution in the beginning of Year Three, the final year of the previous programme, so was not present during the difficult startup period, but was present during the development and early years of the present programme. Because I was not able to witness the first two years personally, my interpretation of the teacher descriptions of these years is perhaps least coloured by my own personal feelings. However, I am also unable to verify any of these descriptions from my personal experience.

As well as the voices you will hear, note the missing voices that are sometimes conjectured, sometimes paraphrased, and frequently maligned - particularly those of the students, the administration, and the second-year teachers. Ideally, all of these stakeholders' points of view would be included. These are the missing perspectives which could help define far more clearly the dimensions of the elephant. Without them, we are still only thirteen blind men groping about in the dark, sharing what insights we can collectively gain.

Again, what is the nature of truth and reality? I do not present these narratives as 'reality' -- only as the intersubjective reality of thirteen teachers. Whether 'true' or not, it continues to influence -- and explain -- the way in which they have chosen (and continue to choose) the curriculum and how they teach it.

My initial feeling was that because the events took place so long ago, and occurred during a naturally experimental 'start-up' period, no one would feel hurt by the frequent 'hind sight is the best sight' criticism. Part of what I have learned from this research is that one should not assume this. For one thing, unlike teachers, of whom there are so many that no one individual need feel singled out for criticism, only a few people were responsible for management decisions; individuals could therefore be identified and unnecessarily embarrassed. For another, the reader must understand that this was a very groundbreaking cross-cultural venture. During the first two years in particular, a relationship of trust had to be established among the Japanese and Canadian board members, administrators, and staff in three cities in two different countries. This did not come about overnight. Administrators, caught in the middle, were often powerless to make changes they realized must be made until approval came down from slowly-developing trans-Pacific channels of authority.

Therefore, since my purpose was to examine a curriculum, not to present the definitive 'true' history of the institution, and certainly not to spread gossip or criticise individuals, I deleted many of these negative comments, summarizing only those teacher attitudes towards Year One and Two administration which affected curricular decisions. What follows, then, is an amalgam of the teachers' voices unless otherwise noted.

5.2.2 Qualitative Research Results - Introduction to the College's Story

Once upon a time, a group of educators and businesspeople from Japan and Canada got together to develop a private college in British Columbia for recent Japanese high-school graduates. The school had high ideals of producing graduates (after a two-year course) of 'independent spirit' who were prepared for world-citizenship in their understanding of, and in their ability to communicate (through English) with people of other cultures. As well, they would have an easy familiarity with computers and with at

least one other specialized subject area such as business, interpretation/translation, teaching Japanese as a foreign language, or environmental or multicultural studies. Amazingly enough, given these high ideals, they succeeded in their endeavour, even expanding to offer both three- and four- year programs. Over fifteen hundred graduates of this college are now working in Japan and internationally today.

However, development of the curriculum at the college, in particular that for beginning (entry) students, has had a turbulent history. In the first years of operation, the college used a 'content-based' curriculum largely based on the theories of a highly-respected educator I will refer to as Dr. V. (my note: this is not her real name. I use a pseudonym for two reasons. First, many teachers in the interviews rather vehemently malign her theories, and in reporting them, I would risk libeling her. Secondly, knowledge of the details of her theory is irrelevant to the purposes of this thesis). Heterogeneous (non-leveled) classes were to be taught on a term or year-long basis. Teachers, though bound by Dr. V.'s theory in that they had to show how every lesson met her specific criteria, were free to develop their own curricula, materials, tests, and grading schemes. Each year, in response to student and teacher demands, the curriculum changed somewhat.

By the fourth year, it had changed to providing a discrete-skills-based curriculum for its beginning (entry) students with homogeneous (leveled) classes in Reading, Writing/Grammar, and Listening/Speaking taught in five modules, each five-to-seven weeks long. For each level in each of the above three subjects, teachers developed completely standardized objectives, materials, tests, and grading schemes (which included an 80% 'mastery' pass standard in the first four levels). The strictly-leveled, skills-based component was balanced by other required but heterogeneous (non-leveled) classes delivered to students in a more content-based style (computers, presentation / study skills, experiential studies, plus a cross-cultural survey course *taught in Japanese*). As well, once students had progressed through the first four levels ('Foundation'), they

encountered 'Transition' courses: more challenging Reading (with several choices of content), Writing, and Listening/Speaking content coupled with a content 'elective' course, while being freed from the 80% mastery standard (moving to a 50% 'pass' standard). The college has continued to refine this system over the last five years.

(My note: That's the basic story, but what was really happening in those classrooms? Why was attendance such a problem? Why did 20 % of the students drop out the first year? Why did so many teachers in Years One and Two quit? Why did the teachers, in the middle of Year Three, decide to make a radical curricular change in Year Four? I continue to let them address these questions in their own voices).

5.2.3 The Previous Programme: Introduction

In the beginning . . . it clearly wasn't to be an English as a Second Language school. In the beginning of each year, went the plan, the students would be given 70% 'Bridge' classes in which English skills were taught within the context of a compelling content area such as Writing/Sociology, Reading/Newspapers, or Conversation /Communication Theory. They would also be given 30% 'elective' courses. The proportion of 'Bridge' classes was to decrease as the year progressed. Electives included a selection of business (and computer) courses; the Forest Industry in B.C., Environmental Studies, Study of Language (simple linguistics), the History of English, Canadian History, and Human Geography. In courses taught by teams, teachers agreed upon joint objectives, though each teacher was given free rein in deciding how to implement and assess them.

In the first year, students were placed at random in classes regardless of ability. Classes were either on a term (there were three terms) or year-long basis with the same teacher. By the second year, a Japanese entrance exam (not SLEP) was used to create three tracks (called 'Levels'): A, B, and C. In general, these were cohort groups which

moved through the year together in the same class. At the end of Terms I and II, teachers had a big meeting in which *a few* students were chosen to move up to the next level. The criterion was whether the teachers agreed the student could 'handle the challenge.' No one recalled any students ever having been moved downwards.

In the third year (1990-1991), though SLEP was given at entry for the first time (in the students' first week in Canada), it was not used to create the three entry tracks (now called 'Levels' One, Two, and Three). Instead, the Japanese entrance exam continued as the criterion. Towards the middle of the third year, a decision was made to start a Level Four for a group of Level Three students who needed even more challenge.

5.2.3.1 The Previous Programme: Strengths

Most teachers would agree that a lot of great things were happening.

Instructors were hired from several different countries; each had her/his 'own style', and they were 'academically exciting.' They were not hired because of their teacher-training or experience; in fact, some had neither. Instructors had been hired because of their knowledge of a content area, and they wanted to teach here primarily because it was *not* 'an ESL school'.

Free to experiment, teachers did more or less what they wanted, using their own resources, making their own tests (which were often very challenging, and custom-tailored to exactly what they had taught), innovating constantly, collaborating when they felt like it, but allowed to go their own way. One teacher successfully used elementary-school whole-language methods with the students, while another taught a 'university' course using a Canadian *sociology* textbook. The two things tying these courses together were Dr. V.'s theories and the College's Mission Statement:

... to advance students towards global citizenship as well as making them into culturally informed citizens of their home country. (College 'X') provides for the students a comprehensive learning environment designed to promote: Independence of spirit; Understanding of other peoples and cultures, and Co-existence, developing from a sense of world community.

Initially, several teachers noted, high expectations of the first-year students were generally held, "so teachers really pushed students to succeed" (which some, but not all, students did). Another positive aspect noted by more than one teacher was that over the year or term, teachers got to know students well and so got to tailor what they taught to individual student needs. There was time to really teach the material and to 'spiral' it with previous learning. As in most Japanese post-secondary institutions, student 'failure' was extremely rare, and the lack of leveling the first year gave students a sense of equality. In many ways, teachers recalled less stress (than there is now), with more continuity (fewer changes of instructor or class) and a stronger sense of teacher-student rapport.

Some electives had good, strong, challenging content. Some of the content, such as a rather sophisticated cross-cultural communication class, was "very relevant to both the school's philosophy and students' interests." Having electives start at the first of the year ensured that "all students got introduced to critical-thinking and literary-type questions right away." Several instructors noted that this content was more fun to teach and more interesting to learn than the present curriculum.

All teachers mentioned the enthusiasm of the faculty; for example: "inspired staff - always busy! ... core faculty strong, dedicated, committed, forward-looking, cooperative, had a sense of purpose and 'pulling together' ... sincerity to make this thing work."

By the third year, besides the three or four 'tracks' in use, some standardization and 'basic-skills' had been added to the curriculum, which was moving away from the initial open-ended and content-based directive. For example, in classes taught by more

than one teacher, those teaching it had to have at least 50% of their final exam questions 'in common.' As well, conversation classes had strong grammar (language structures) and pronunciation components. In recognition of differing student abilities, some curriculum materials now included suggestions as to how to adjust learning activities to a particular ability level.

Though moving in the third year towards a more standardized, skills-based curriculum, as three teachers put it (and several others echoed), "a strength of the conversation class was that it recognized the need for some basic skills; however, the movement was towards *communicative competence, not just language*," and "a strength of the school was the recognition that it *wasn't just language* in the curriculum, but a recognition of the value of the subject areas in the globalist realm," and finally, "I do not think it is a weakness that we started with *the concept of content*, even though we misapplied it." (My note: The general philosophy upon which the college was built is still supported by most of my subjects, then, though they regret the naivete with which it was initially applied.)

5.2.3.2 The Previous Programme: Weaknesses

On the other hand, teachers came up with twice as many weaknesses as strengths and expressed more emotion as they described the extremely challenging circumstances they encountered during the first three years.

It is important to note that the first year of almost any programme will have negative 'startup' effects. One could well ask if the negativity -- towards, for example, Dr. V.'s theory -- might have been misdirected; perhaps if it had been introduced to teachers after ~~they~~ had gained some experience and confidence with students and the programme, it would have been very differently received by them. However, while teachers became accustomed to, or learned to ameliorate, many negative programme

aspects by the second and third years, core weaknesses did not go away . . . Teachers continue to describe their problems:

The most glaring weakness, apparent on the first day, was a tremendous misfit between most students' ability-levels and the curriculum which the teachers had developed. In the beginning, teachers, as they recalled it, received little or no documentation about student abilities; many recalled completely rewriting their curriculum once the low English-language level of most students became apparent. As one teacher put it, "Because the curriculum was inaccessible to students (my note: due to their lack of reading skills, vocabulary, and basic idiomatic/cultural understandings), teachers often 'chucked' the official curriculum and re-wrote it on a daily basis, at least for lower-level students."

Students were equally stunned to discover how little they understood the classes, and how poorly-prepared their teachers were for them. In general, much of the materials were of far too advanced a nature; many students were barely able to learn even the 'key vocabulary,' much less the 'content.' On the other hand, when teachers tried to adapt the curriculum so that lower-ability students could understand it, they frequently felt intimidated by the school's non-ESL philosophy. Several reported using elementary-school literature in lieu of 'ESL-ed' adult materials. In this case, while the ability-level was appropriate, there was another misfit, this time between student interest and the subject matter. This situation was more intense the first year, but continued on through the third year to a lesser degree.

Several noted that in the first year, teachers often felt alienated from an administration which they perceived as overwhelmed with startup duties. Many expressed concern that administrators appeared not to have a clear concept of the students' abilities, or of the curriculum teachers were using, or of the extent to which the two 'matched' one another.

Related to this problem was the first year's total lack of student leveling, though this changed when a form of 'tracking' (which was referred to as 'leveling') was introduced in the second and third years. Sadly, the track on which one was placed often took on inordinate social importance among status-conscious students, and the previously-noted sense of equality disappeared. This was perhaps exacerbated by the fact that there was no way for most students to change tracks once they'd been placed on one. There was no way to either pass (out of), repeat, or challenge a 'level', and the fact that everyone knew the subject matter in the lower tracks would never reach the same level of sophistication as that found in the upper tracks created a self-esteem (and a possible justice) issue for lower-track students.

However, all teachers acknowledged the need for some form of tracking or leveling *at least in the beginning of the year* since many students hadn't yet acquired the language to access a 'language-based' (i.e., content) curriculum. In heterogeneous classes, teachers found upper-ability students bored or lower-ability students hopelessly confused (often simultaneously). Seldom would either end of the spectrum be satisfied. As an example of teacher-adaptation to this problem, two teachers sometimes leveled elective classes "behind administrators' backs," as one reported it, by trading students in order to form one 'high' and one 'low' ability group.

For one teacher, as well, the school, other than being "a vague, philosophical undertaking, really hadn't discovered (or developed) its true identity yet." A consensus was that there seemed to be no overall plan or coordination. Teachers noted that guidance, consistency, clarity and leadership were lacking in many areas. In the realm of curriculum, a lack of goals or a year-long scope and sequence of student learning meant that courses were planned independently of one another and "no attempt to spiral, integrate, or reinforce prior learning could be made." Skills were taught on a 'hit-and-miss' basis. "Depending on what teachers a student got, some skills could be taught several times while others were not taught at all; there was no way to ensure that all

students would be taught anything. New teachers had no idea how to proceed as nothing concrete was 'in place' to direct them."

Ironically, term- and year-end evaluation at that time centred on teachers and students, not on the courses or programme itself. A general teacher misgiving was that "students were getting insufficient training," or as others put it, "teachers felt the courses weren't helping our students" and were "random, ill thought-out". They sympathized with the many first year students who, as one teacher put it, "felt they had been lied to" in that their actual educational experience was apparently quite different from the perception they'd formed from promotional materials.

The most common complaints, however, related to *a lack of consistency and standards* in such areas as what was taught (even in the same subject in the same track), texts, tests, criteria for grades, numbers of field trips and guest speakers, rules and expectations, etc. This was confusing, demoralizing, and seemed *unjust to students* and damaged the school's credibility with them. A teacher noted that in Japanese education there is a *high degree of consistency* between classes, materials, and tests at the same grade level among all Japanese schools. Students who come to North America from Japan "want to feel they are receiving *the same education* (my note - this phrase may mean entirely different things to Japanese and Canadian educators), no matter who their teacher was, and they wanted their grades to 'mean something' -- to be tied to some meaningful 'scale.'" However, there was no way of comparing grades one got from different teachers, levels (tracks), and courses. Tests and grades were "all over the map." Teachers acknowledged this, but without clear leadership they were unable to solve this dilemma on their own. As well, standardization of objectives, materials, tests, and grading practices would mean a big trade-off with the independence enjoyed by so many of the faculty.

Probably the most demoralizing aspect for faculty members, however, was the factionalization that characterized their own ranks. Three areas of dissension arose: use

of Dr. V.'s theories, heterogeneous versus homogeneous [non-leveled (or multileveled) versus leveled] classes, and the teaching of content versus the teaching of language (skills): "The factionalization and splits among teachers was emotionally draining -- issues such as 'language versus content' and 'homogeneous versus heterogeneous group' drove people apart . . . however, mutual resentment at being forced to utilize V.'s theories in first year ESL classrooms became a source of cohesion."

To this day, a lingering bitterness is revealed in such terms used in the interviews as 'rabid V.-ism' and 'V.-ism to the extreme degree.' As one teacher noted, "some of the best 'content' teachers left the institution because of their frustration at being forced to make everything they taught meet (V.'s) criteria." Another teacher noted that, "Some very good teachers ended up quitting their jobs for some very good reasons." [My note: Ironically, while Dr. V.'s theory was meant to unite the curriculum, opposition to it seems to have ended up uniting the teachers, so inappropriate did every one of the teachers I interviewed deem its use with the first year students. Yet this also resulted in creating an uncomfortable difference between the two faculties of the first and of the second (and later) years' students as Dr. V.'s theory was -- and remains -- a useful and appropriate organizer for the second (and later) years' curriculum.]

Teachers who had a background in teaching basic language skills more readily supported the idea of homogeneous grouping, and were often upset when they had to teach a heterogeneous class. Some of them felt personally threatened by the idea of having to teach content, often unfamiliar to them, to a multileveled class (either because of feelings of inadequacy or feeling that it was inappropriate for students, or both). One of their complaints was that the content courses too often used the lecture-and-memorization style that students were familiar with from Japan. However, their basic complaint was that the curriculum didn't address students' lack of basic skills in a coherent, *systematic* manner.

At first, heterogeneous grouping was advocated by the 'content' teachers, but many of them came to the conclusion that some students just 'weren't getting it,' not because of inadequate intelligence or lack of effort, but because they lacked basic language skills. As a result, these teachers often became strong advocates of homogeneous groups, at least in the beginning of the year, and for the extreme high and low ability students in particular. Some of them, however, felt personally threatened by the idea of becoming ESL instructors (again, either because of feelings of inadequacy or feeling that it was inappropriate for students, or both). In some cases their philosophical transformation (towards favouring first year homogeneous, skills-based courses) took place over a couple of years' time. Meanwhile there was much argument and controversy.

The day-to-day reality for teachers was constant revision and creation of materials, lesson designs, and tests, 'fumbling around' to get through each day, daily (required) lunch-time meetings, 70-hour work weeks (several people noted this), struggling with constant and rapid curriculum changes, lots of developing 'by the seat of one's pants,' and "everyone 'reinventing the wheel.'" [My note: What struck me was how on one hand teachers said they were free to do as they wished, but on the other hand there were a lot of directives (i.e., to use Dr. V.'s criteria) from administration. Perhaps the directives were so frequent that, over time, overwhelmed teachers generally came to ignore them.]

For example, teachers said they "were struggling with constant, rapid curriculum changes;" "There were lots of meetings!" but "There was no coordination in the overall plan." Meanwhile, a more immediate concern was how to prevent more student dropouts, as teachers realized their jobs were dependent on retaining as many students as possible. This pressure was difficult for teachers to bear considering they were working so hard and still 'things weren't right.' Many teachers, exhausted and discouraged, simply 'dropped out' (quit) as well.

By the middle of the third year, V.'s criteria were no longer required and, in fact, rarely used at all. Those who hadn't quit came to realize that even though they felt a well-deserved sense of 'ownership' over curriculum they had developed under such trying conditions, it still needed work. The consensus was that they liked the freedom, but the curriculum was simply too difficult to teach, demanding an unrealistic amount of their time, effort, and creativity.

5.2.3.3 The Previous Programme: Teachers' Perceptions of Students' Responses to it

As the interview progressed, teachers were asked specifically to mentally reconstruct how the previous programme was, according to their memories, perceived by students -- by students in general, by the high-entry students, and by the low-entry students. In general, they said, *most students* seemed to enjoy their time in Canada about as much as they do at present; they had a good experience learning in a new way, they improved their English, and they expanded their view of the world. Each year was better than the one previous insofar as producing student satisfaction. However, recurring themes, echoed by many teachers, were that even at the end of Year Three, the programme *lacked cohesiveness, purpose, regularity (consistency), and sequence*. Depending on the course and teacher they had, they said they had 'too much homework and it was way too hard' or they had 'too little homework and it was way too easy!' Students complained of having little idea or sense of their own progress. Each year a significant group advanced to their second year with a sense of not having gotten quite what they had expected -- a vague sense of disappointment, though by Year Three, this was far less pronounced than in the first year.

The *high-entry students* found the programme either exciting and challenging or boring and too easy, depending on their teacher and whether they had been placed in the

upper track classes. In general they liked the fact that electives (unlike now) started in Term I. They sometimes felt held back by the slow pace of the non-leveled (heterogeneous) courses. Some complained that teachers 'facilitated' courses instead of 'teaching' them [Socratic-style dialogue (between teacher and students) and small group discussion - instead of lecturing]. Others expressed dislike of any skills-components in their classes (i.e., a grammar component in Conversation class), saying they had already learned it in junior high school, while others were very appreciative of specific skill instruction, particularly if they felt it was an area in which they were weak. In the first year, a large percentage of these *upper-entry* students left at the end of Term I (this was complicated by age and gender factors: they were *mostly females* who were significantly *older* than the balance of the students). In the second and third years, a better effort was made to be sensitive to these various problems -- in more explicit promotional literature, in admission practices, and in actual orientation of students.

Teachers gave very mixed answers as to how the *low-entry level students* perceived the previous programme. On the plus side, overall, most of them seemed happy with the programme. They worked hard and had no major complaints. They benefitted from upper-ability students' 'modeling' in their multileveled classes. They enjoyed the elective class *activities* and being introduced to exciting and interesting concepts, even though they realized others were understanding the subjects more thoroughly than they. Most tended to have fun with the recreation programme while basically (and uncritically) ignoring the academic programme. They knew they were all going to pass anyway. They had an experimental, playful, fun-loving attitude. Unlike most of our present students, a significant portion of our early students, particularly in Year One, had 'a distinctly separate agenda': many had a lot of spending money, were fairly 'wild,' and were often absent from classes. They, like so many of their cohorts in Japanese colleges, considered this a well-deserved 'leisureland' between 'the examination

hell' of high school and the lifetime of serious employment that would await them upon graduation.

However, on the minus side, a significant number of low-entry students were not blissfully ignoring the fact that they were struggling with the academic content. They were, as teachers described, "'lost. . . confused . . . overwhelmed . . . floundering . . . just 'here.'" Because they were unable to fail and then repeat a level, they were *consistently* dealing with new and challenging material which was 'over their head,' especially in the non-leveled electives. This experience was demoralizing for many, especially considering there was no academic support system (tutors, student-at-risk reporting and counselling, learning resource centre, etc.) like there is now. These students generally did not respond well to the lack of structure and open-endedness of the previous programme, and usually left the first-year campus dissatisfied with what they had learned. The system did not deal with the problem of how to help these students succeed; the best it could do was to put them onto a 'low' track and keep them there all year.

5.2.4 The Present Programme: Introduction

In the middle of the third year, the administration invited all the teachers to a weekend retreat at a resort to deal with all of the noted problems by developing a new programme. When asked why the institution changed to the new programme, teachers cited these as the main reasons:

- 1) Out of recognition of the *students' need for fairness* and of teachers' need for a curriculum which was *easier-to-teach*, there was a need for some *standardization* of the curriculum, of objectives, of materials, of testing, and of grading. [My note: fairness through standardization or 'justice as *regularity*' (Rawls 504)]

- 2) They were unhappy with the orientation in students' first year to content rather than *the skills they seemed to need* in order to access that content. They saw a need for curriculum which would be more *accessible* to students because it is more attuned to their ability levels. (My note: justice as *equal access to resources*)
- 3) There was a need to change how 'Levels' One to Four were being taught and delivered; a "recognition of the need to give different curriculum and content to different levels -- *not just an 'enriched', a 'regular', and a 'watered-down' (track) programme* -- using the same basic content." There was also a need for students to be able to repeat (without penalty) and skip levels. (My note: justice as *equal distribution of educational outcomes and of self-esteem*)
- 4) Out of recognition for the need (of administration and of teachers in particular) to know and *evaluate what was being taught* at the college more accurately.
- 5) *New blood*: 'Burnt-out' faculty had been replaced by a new administrator and four new teachers who were ready to experiment, and unattached to old ways of doing things. One of the teachers in particular had knowledge of a programme which sounded like an attractive alternative. (My note: This was a *modular, leveled, discrete-skills-based, mastery learning programme* successfully used to teach ESL students elsewhere).
- 6) *Exit surveys* showing *information gaps*, and *dissatisfaction* among students and teachers convinced administration a change was necessary. The school is 'market-driven'; consequently it is imperative to keep the student retention rate

high, while at the same time acquiring an ever more prestigious reputation among the highly competitive Japanese post-secondary education 'marketplace.' (My note: the rate of student retention had greatly improved by 1990, but fears of a return to the previous low retention rate were still strong)

- 7) *Evolution*: A natural desire to improve each year. The present programme was a logical outgrowth of the changes made in Year Three.
- 8) A desire to *get away from* the Japanese higher-education model of '*leisureland*' in which students cannot fail.
- 9) Administrators and instructors in the *second year* of the programme wanted students to meet *minimal entry standards* into their programme. This was impossible if the grades with which the first-year teachers provided them had no real meaning.
- 10) Finally, out of recognition that "students and teachers come to this institution because they want more than skills-based education," the two diverse 'language-skills' and 'content-teaching' camps came together to create a novel *compromise*, the *Foundation* and *Transition* programmes which effectively straddle both camps. (My note: First students are taught primarily discrete language skills in Foundation's 'mastery' programme; later they are taught primarily content in Transition's non-mastery classes. In 'Transition', Reading is no longer leveled, but Listening/Speaking and Composition remain leveled all year (though they change to a 50% pass standard after Level Four). Students can 'fail' and repeat Foundation classes without penalty, while students in Transition are penalized for

failure. As well, students take heavily content-driven, non-leveled 'elective' courses in Transition).

Much curriculum development time and interminable meetings later, the Foundation and Transition programmes were in place. Students and teachers generally seemed to be doing quite well with these programmes, though occasionally someone would comment on a glitch, ambiguity, or philosophical inconsistency - *not whether the programme itself was good, but whether its dictates were being consistently followed*. Over the years, the problems and discontent seemed to become more defined and to come ever more prominently into the foreground of teachers' attention. Therefore, it seemed to many that now might be the right time to make up *a balance sheet of the strengths and weaknesses of the present programme itself*.

Several teachers mentioned that if, indeed, the institution is to create a new programme once again, it would seem expeditious to try to retain the strengths of *both* (previous and present) programmes while addressing their various weaknesses. [My note: One important factor I think should be considered is that *in any programme, every aspect of innovation will have its inherent strengths and weaknesses (tradeoffs)*. The challenge is to build a balanced programme which acknowledges, minimizes, and mitigates the negative impacts whenever possible, while enhancing and enabling the positive ones].

5.2.4.1 The Present Programme: Strengths

This section combines the answers to several questions about the strengths of the programme, how teachers personally felt about leveling, standardized curricula, etc., and 'what they would like to add.' This is because there was so much overlap between the answers to these somewhat related questions.

The new programme definitely addressed the noted weaknesses of the previous programme. As well, teachers commented a lot on the extensive formative and summative testing used in this programme. They said it gave teachers and students lots of steady, valuable feedback, allowing teachers to locate student problems and help rectify them before they became 'fossilized errors,' and modular exams in particular "help students deal with what was previously 'down-time' in the middle of a term," or as others put it, "They're Japanese - they need and want tests!!!" . . . (though) . . . "now students can't simply (my note: in Japanese style) 'cram' at the end of the eleven-week term, then pass."

The 'Great Compromise' (between content and skills), resulting in the creation of the Transition programme -- which retained the 'content' of electives and other advanced (beyond 'Level Four') courses -- is still highly supported by most teachers.

The sequential nature of the skills-teaching, the standardization and consistency of objectives, materials, tests, and grading, and the clearly defined levels which have explicit mobility built in are generally acknowledged to be successfully countering problems of the previous programme: The specific needs of individual students are now measured and addressed. Students get a sense of their progression and can work pretty much at their own pace, repeating, progressing, or challenging (skipping) levels every five-to-six weeks as needed. Now, students don't take first year elective courses or transfer into second year courses until they have met minimal standards. Teachers noted that there is a yearlong scope and sequence chart and a common teacher understanding of what is to be taught in each course at each level, and how it is to be done and evaluated. Some lauded the institution of 'Curriculum Heads,' teachers given extra time to help oversee that the curriculum in a given subject area is kept up-to-date and standards agreed-upon (and followed) by the entire team of teachers. New teachers can step into the programme with minimal preparation. As well, teachers have developed and now share a fairly large body of quality supplemental materials. Courses taught many times by

many people can be compared and slowly improved over time. As one teacher noted, the curriculum is seen as "a work in progress - not written in stone."

5.2.4.2 The Present Programme: Weaknesses

Here again I combined answers to the specific query about present-programme weaknesses with answers to other questions which addressed areas teachers perceived as weak, plus the additional questions about what other programmes they would like to try or changes they would like to make. Here are the major complaints and concerns:

Too Much Standardization: While there was 100% support for the standardization of course objectives, several teachers felt standardization of materials and tests might have gotten carried too far; they feared that there is too much 'teaching to the test,' and decried the 'loss of creative juices' amongst faculty who had become lost in the 'safe mediocrity' of the modular system. Students, one claimed, were being led on an educational 'forced march.' The 'lockstep' system is perceived to be so inflexible teachers can neither take advantage of 'teachable moments' nor address individual students' needs. A good question was asked: "Does 'standardized' always mean *higher standards*?"

At times teachers confused 'standardization' with 'mastery' (i.e., "Mastery tests must be standardized") since standardized objectives, materials, tests, and grading standards were introduced at the same time as mastery learning. (My note: However, the two are unrelated issues as mastery tests in other institutions are not necessarily standardized among teachers).

There is a problem with 'fossilization' of tests and materials. Teachers noted that it is very difficult to make needed changes to courses since all the standardized materials and tests have to be changed (this is particularly difficult with listening exams, for which scripts must be written and cassettes made); if there were not so much standardization, it

would be much easier. Or, as one teacher stated, "the system tries to maintain itself rather than addressing students' needs."

80% Mastery Pass Mark: The 80% mastery standard and the term 'mastery' itself (with the false expectations of 'perfection' it connotes) came in for the teachers' toughest criticism. They seemed to object more to the term than to the actual philosophy and practice of 'mastery' learning such as giving people extra time to learn the material without penalizing them; in this institution, this means allowing them to repeat a level without penalty. In our system, only the grade they receive when they pass -- greater than or equal to 80% -- goes on their final transcript. Teachers rarely criticised other 'mastery' learning practices such as using frequent formative testing; using closely parallel course objectives, materials, and test items; or grading according to how well criteria are met instead of according to a bell curve. Giving a summative grade based primarily on a final exam came in for some criticism; but requiring a fairly high (80+ %) pass standard was definitely questioned by many. A large number of teachers made statements like "an 80% standard implies they've learned 80% of what they're supposed to know (but in reality haven't)," or the "concept of 'mastery' of something within six weeks is not practical nor is it educationally sound . . . often things they're taught in Module One don't really get learned until Module Three."

Problems noted were that with such a high pass standard and with lots of outside pressure to pass the majority of students, teachers were inclined to 'teach to the test,' to scale marks (or "tailor marking so that there are not too many failing or getting A's"), and, over time, to remove difficult items from the tests "so that most of the students who complete the level in Module Five (note, these are the 'lowest entry' students to take any given level) can pass." As one teacher muses, "Unless we change the administrative posture of the college, an 80% will never be a real 80%. We forgot who we and our

students were. We signed up for the Guided Tour to El Dorado . . . but does it really matter?"

Students, on the other hand, are, as one teacher put it, "forced by the mastery concept to memorize rather than to internalize." Others noted that with an 80% pass standard, there aren't "many numbers to play with" - it seems strange to call 79% 'unsatisfactory.' Others noted that the changeover from an 80% pass in Foundation to 50% pass in Transition is awkward for teachers and confusing for students. In one week, a good essay is given 85%; in the next it is given 70%, a 'Fail' the week before (See Appendix 8.3).

Not enough Levels, but Modules are Too Short: A few teachers suggested there should be one or two more entry levels -- four to five levels minimum -- to accomodate the extremely low- and (possibly) the extremely high-entry students, and one more added at the end of Term One for ambitious upper level students to challenge into. (My note: This would require the development of curricula for two or more additional levels for the first and last module of the year. Also, any of these changes would be rather dependent on the number of students. There would have to be a minimum of one class of students at each level in order for this to be a viable option).

While most teachers felt that the 'challenge' option was positive, one warned that in some cases it can be very damaging; students who skip a level can miss out on important information, developing incomplete schemata. [My note: In order to 'challenge' (or skip) a level, students must receive a 95% final mark in their present level, score 80% on the final exam of the level they wish to skip, and get their teacher's recommendation.]

One of the biggest complaints was that the modules are too short. For one, teachers complained that it is very difficult to complete a full evaluation procedure with summative exams and reports in such a short time (some modules are only separated by

three days). A teacher allowed that the modular system ensures that teachers keep 'a tight ship', but fears that "sometimes the ship is 'too tight!'" and went on to complain that "continual evaluation takes time from teaching." [My note - by 'teaching,' I surmise the interviewee meant instructing the class, as evaluation (particularly the 'formative' testing used in mastery learning) is certainly a function of teaching and is generally assumed to have pedagogical value].

Another complaint many noted was that the modules are not long enough for teachers to adequately cover (or for students to synthesize) all the objectives in each level, especially considering how many progress (formative) tests must be given in the class time allotted. There isn't enough time for "experimentation, individualization, enrichment, and creative activities!" . . . and another: "No room to manoeuvre, no room for creativity or teacher strengths . . . No time!" Various teachers recommended deleting objectives, particularly in grammar and composition such as recommending less sophisticated rhetorical foci in lower levels and "less grammar - period!" -- this was followed by the comment that "All students, even especially-low-level students, should have at least one (some?) content courses . . . [and in an only half-joking vein:] If grammar has failed them for so long, why not try something else . . . electricity? carpentry?" This leads appropriately into the next topic.

Content Issues: Content issues were raised as well. Many teachers felt that listening and speaking classes should be allotted more time per week and that pronunciation instruction was being neglected. Others felt that vocabulary wasn't being specifically targeted; for example, some students leave the institution "without knowing the numbers, months, or days of the week." Various teachers wanted to add more 'content' courses (electives) to the curriculum (my note: presumably this would entail changing the current elective requirements and/or minimal criteria for taking electives). One found the grammar programme "boring -- students have already had six years of

grammar . . . they should be ready to have it applied in another way. There are good programmes out . . . we haven't looked far enough." One claimed that upper level students should be given more challenge at the beginning of the year, not the 'review' they are given now (a review, nevertheless, as several teachers noted, of what they've often been inadequately taught and which they have never been asked to apply in contextual, genuine, extemporaneous, oral/aural or interactive situations).

Various individuals noted the need for a more interesting reading text, a better language lab, more discussion groups with Canadians, more field trips, more 'interactive activities' in general, longer classes in computer studies, and 'values clarification' and 'intercultural competency' taught across the curriculum. Finally, one teacher criticized the lack of a Curriculum Head for the electives offered in Transition, a further tribute to the effectiveness of Curriculum Heads.

Miscellaneous Doubts: Some doubt and ambiguity came to light, such as "our purported coherence and sequentiality are in many ways only apparent; they are insubstantial, focussed only on how the institution *appears to students*." Placement and testing are areas of great concern: "We claim to and appear to be competently placing and advancing students into appropriate levels, but are we?" Several teachers would like to have more input into initial placement of students (such as administering a speaking test and seeing a short writing sample before students are initially placed), and for students to be able to begin in different levels in different subjects according to their specific abilities in each area (a 'finer-tuning' of our present practice in which each student starts all his/her Foundation classes at the same level).

One teacher criticized some tests as having poor questions: "Tests need to be analyzed item by item." Another subject questioned how objective teacher-produced tests really are. S/he worried that scaling and adjusting marks was a sign of 'fudging' and

'dishonesty.' Derogatory words like 'bogus,' 'arbitrary,' 'not legitimate,' 'incompetent,' and 'inflated' came up many times, especially when teachers were discussing teacher-developed tests and standards. (My note - it struck me that perhaps these teachers had too much faith in 'professionals' and too little in themselves; they didn't consider that commercial standardization also takes time and that even commercially-made standardized tests are regularly scaled). However, one teacher, while not denying that scaling goes on, suggested that "no matter what the standard was, we would still have 'borderline' cases and some scaling." One teacher suggested that final exams should be the only ones standardized; progress (formative) tests should be made by individual teachers.

Examination practices advocated by individual teachers included more one-on-one interview-testing of students' actual 'communicative competence' (especially in listening/speaking and grammar) and a "comprehensive exam at the end of Foundation covering Levels One through Four -- a more holistic measure, not simply looking at discrete skills -- *to keep students out of Transition that don't really belong there.*" (My note - The implication here is that the Level Four exit standards are too low, allowing students into Transition who are unable to succeed at that level).

Another wasn't happy with the way the 'core body' of (Levels One to Four) skills was defined, and suggested that this area be re-examined.

Another, concerned that upper-level students aren't getting enough intellectual stimulation, suggested that electives in the final two modules be leveled, to enable teachers to present more highly-challenging content to these students. This teacher recommended that some Reading electives such as *Anne of Green Gables* might be best 'reserved' for these higher level students as well.

Another wasn't sure if the present programme is any more successful for teaching English than the previous one, claiming, "Nobody knows if it is more successful . . . " In the same vein, a teacher wondered, "How much of our current success is due to 'student-at-risk' protocol (my note: tracking student progress, regular meetings among teachers and interviews with students regarding 'at risk' status) and the Learning Resource Centre (free tutoring service), and how much is due to the modular, leveled programme itself? It's hard to determine which factor is helping students more."

5.2.4.3 The Present Programme: Teachers' Perceptions of Students' Responses to it

In general, teachers said that students seem pretty happy with the curriculum. Some aren't content with the speed with which they are learning English, nor with what they consider to be unnecessary review of high school grammar in the beginning of the year. Dropout rates have plummeted since the first year (though Year Three was very low as well), and exit surveys show students have a high degree of respect for the programme.

The *high-entry students* generally like the programme because they perceive Foundation as a challenging but short route to the more interesting Transition courses. They seem happy in their own 'prestigious' group. As one teacher noted, "They need to work with successful peers. They don't want to be 'teachers' (peer tutors, i.e., in a multileveled class); they want to be learners." However, some see the Foundation courses as too easy -- the grammar is perceived as 'the same' as what they learned in junior and senior high, though, noted several teachers, most of them can't see their own weakness: that they have only learned how to *pass grammar exams*, not how to *use good grammar* in their writing or speech. Several teachers said students would like another

level into which the most ambitious and hard-working students in the highest level could challenge into at about mid-year.

The *lowest of the low-entry students*, much as in the previous programme, seem to have to work pretty hard to succeed. However, they appreciate being able to work at their own speed, repeating a level if necessary, but also having the opportunity to 'challenge.' Two teachers noted some concern over their lower social status, though others also noted many of these students used other opportunities to gain status through sports, music, etc. In fact, some of these students may be in this category *because of* their 'other agendas.' A few teachers have heard more than one low-level student grumbling that some courses are too difficult, especially Foundation Listening/Speaking and Reading.

One teacher felt that these students are "plugged through' a lockstep system which doesn't adequately address their learning problems;" as a result, they are "frustrated with their learning experience, which is actually very Japanese in its way of testing and sorting students." However, another teacher, while acknowledging that these students "find it very difficult to move at the pace we've established" claims that, "Most are happy to use all the extra help and personal attention we've provided (learning resource centre, tutorials, etc.). They perceive a lot of extra effort is being put forth on their behalf." The consensus among teachers would seem to be that most low-entry students perceive the present programme to be challenging but satisfying.

5.2.5 Qualitative Research Results: Teachers Compare Present Versus Previous Programme

5.2.5.1 Evaluation

First, I asked teachers what they perceived as the evaluation needs of various entities (information about students from the first-year teachers needed by the second-year campus, by the first-year administrators, by students themselves, and by their parents) and how well each programme met their needs. Here are their answers.

Second-Year Campus Needs: Teachers noted that the second year teachers expect us not only to prepare students for the next year, but to 'sort' students for them. They want a general sense of students' oral and written language competencies, of their ability to research and to meet Canadian classroom expectations (e.g., are they 'active' learners?), general social skills, and of any notable character traits. Teachers noted that the second-year teachers expect consistency and standardization from us in our evaluation practices, but some noted that they were unsure if the second year teachers agree on minimum criteria for their programme; another said it would be "very helpful if lots more faculty [from the second year programme] could visit here."

The consensus was that we are doing a very good job now in providing them with information they requested, but that the inconsistency of evaluation practices in the previous programme had made transcripts they received from us worthless. However, one teacher proposed that the second year programme should develop and have us administer a standardized 'Exit Year One/Entrance Year Two' exam which would more accurately reflect what they were to be doing in the second year. (My note - All students 'advance' to the second year campus; however, some of them go into an alternative programme if they are judged ill-equipped language-wise to handle the regular

programme; generally, students who do not complete Level Four by the end of the first year go into this programme.)

First-Year Faculty Needs: Teachers felt the evaluation information needed at this institution for placement and advancement within the first year programme is, as well, pretty much what we are giving ourselves now. Since 1991 we have used SLEP for placing students in levels initially. We give marks five times a year and supplementary anecdotal comments at least twice a year (end of modules two and four) and more if students are having difficulties. With this information, students advance within the levels in what teachers judge is a fairly satisfactory manner.

One teacher, however, would like to see initial placement improved with a formal 'test of motivation' administered in Japanese. (My note: Motivation is informally tested in the entrance evaluation procedure in Japan. It is unclear if a valid test of motivation exists in any language).

There is also a concern that too much of our testing is written; several advocated more oral testing based on 'genuine communicative competence.' Some teachers mentioned the need for greater standardization among writing and speaking tests, and in electives. Again, compared to the previous programme, teachers felt we have significantly improved our placement and advancement evaluation practices.

Students' Needs: What do students need to know about their academic achievement and progress? Teachers thought that their particular needs were for frequent (weekly was often recommended), clear feedback on how they are meeting specific course objectives -- formative information, in other words. They need to know if they are in danger of failing, and how they can improve. They need to know how the evaluation system 'works' and that it is *not biased*.

One teacher felt we gave too much evaluation to students, others recommended more self-evaluation and 'communicative' testing. One spoke of the motivating effect of evaluation: "There is a fine line between criticism and encouragement . . . we should set a standard high enough that they have to reach." Feeling was mixed somewhat here, as some teachers thought that students really don't care all that much about evaluation: all they want to know is if they are going to pass and graduate.

Parents' Needs: To teachers, parents were assumed to want to know basic information such as if their child is having severe academic, life-style, social, or health problems; they need warning that their child may not pass a level, whenever possible. They definitely need warning if their child may be placed in the alternative programme for the second year. They need to have a clear understanding of what criteria are being used to make decisions. Other than that, they want enough information to feel secure that they've "turned their child over to an institution that will take good care of him/her -- because they're so far away." Culturally, said one, they "can accept poor behaviour as a reason for failure, but not 'inability to learn.'" (My note: effort, consequently, is emphasized far more in our anecdotal comments than ability). They need to have anecdotal comments translated by the Japanese staff. The consensus was that over time, we have improved a lot in reporting to parents, but there are still some problems. To say your son/daughter has 'successfully completed' or 'successfully mastered' something is a little inaccurate (and redundant) in English -- how does it translate into Japanese? "This term -- successfully completed," said one teacher, "doesn't address underlying issues of communicative competence and personal growth."

Fairness: Next, I asked some questions about whether evaluation methods were *fair or perceived as fair*, and then how the evaluation methods used affected students in each of the two programmes -- specifically, whether the methods *motivated or*

discouraged students -- students in general, the high-level students, and the low-level students. The answers, and the reasons teachers gave for them, were rather interesting.

Predictably, nine out of thirteen teachers clearly rated the previous programme as unfair in that lots of students complained about *inconsistencies* from teacher to teacher, and class to class. Students, remembered several teachers, felt they were graded quite subjectively, and often didn't understand teachers' evaluation methods. Of two 'maybe/not sure' answers, one remembered students as realizing and accepting that some teachers graded more strictly than others, that the system was imperfect but 'not bad all in all.' Another pointed out the internal consistency of each teacher -- that each teacher evaluated in a fair manner according to her/his own individual criteria, objectives, and tests. One also noted that "all students were part of 'the same system' and equally subject to its whims," (in my opinion, a convoluted form of fairness!). Of the two who felt the previous programme evaluated fairly, two didn't remember any complaints. One said, "Students only cared if they passed and got a diploma. If so, they felt it was fair." In another part of the interview, one teacher also pointed out that the evaluation methods used in the initial year, in particular, were very close to those found on most North American university campuses, where the professor is given a large measure of autonomy. However, the general verdict was 'Unfair, and perceived as such.'

Also, quite predictably, ten out of thirteen teachers felt the present programme is both fair and *is perceived as fair*. They cite well-stated goals and objectives, and standardization as the basis, though note that in addition "team meetings help build consensus [and hence, greater consistency] about marking." One teacher observed: "Errors are usually that students who shouldn't, do move up (pass), not the other way around." Critics, however (those who saw bad points as well as good), saw some inconsistencies, such as in how items on the same test are marked, or the amount of time one class versus another might be allotted to spend on a progress test. They did note, though, that these were pretty minor compared to student complaints about the previous

programme. Several said that complaints about inconsistent evaluation were far more common, understandably, in the least standardized classes, experiential studies and the electives. One teacher said that the present evaluation methods are unfair because *too much of the final grade* is based on 'testing.' Others, however, thought present methods are unfair because *not enough of the final grade* is based on the final exam: too much of the grade is based on doing homework, going to conversation groups, and taking progress tests. (This is clearly not an area of teacher agreement!)

Motivation vs. Discouragement: When asked whether *evaluation practices motivated or discouraged students*, teachers said that in general the present programme's practices are far more motivating to students, though slightly less motivating for the lower-entry than for the higher-entry students. The teachers gave very mixed reviews to the previous programme; they tended to say it discouraged more than it motivated, but many were unsure or said that it had done both or neither, or that they didn't remember. They also, however, gave the previous programme *a slightly higher rating for motivating higher-entry than lower-entry students*.

A reason one teacher gave for students in general being unmotivated in the previous programme was that there were "tremendous 'lulls' in the middle of the first and second terms which allowed students to become lazy for longer -- it didn't 'keep them on their toes' . . . feedback, even when it is negative, can be encouraging." The present modular system, on the other hand, provides a module-end exam during what used to be mid-term One and mid-term Two.

Some previous-programme students, a teacher said, also suspected that "grades were probably inflated and didn't reflect achievement." High-entry students, some claimed, were motivated because they were quite happy and challenged, and not as apathetic as lower levels. However, others said they were discouraged because of *inconsistencies* in grading and because of a sense that "they had 'arrived' and had no

motivation to knock themselves out." Many dropped out, complaining, noted their teachers, that standards weren't very high and that "they wanted their excellence recognized."

Previous-programme low-entry students "had more problems with the curriculum; whether this was motivating or discouraging depended on the student." One said there was lots of apathy in lower-levels; another referred to a sense that they (the low-entry students) would always be 'at the bottom.' In multileveled courses they knew that "they'd get low marks anyway, no matter what they did . . . (and) that if you failed, it didn't matter." (My note: Students apparently perceived few, if any, consequences should they not meet course expectations.)

General reasons given for the present program's motivating students included desire to move up to the next level (there are now more levels than previously, and students have four opportunities to move up a level, plus chances to 'challenge'). Students who want to take more 'interesting' courses (e.g. electives) realize they must go through Level Four first. 'Status' was cited as another motive for advancement. Some teachers said the present programme is more motivating because it is easier; some said it is more motivating because it is more difficult, some said it is less motivating because it is easier. (Obviously this is quite a subjective area as well!) One said in the first part of the year the high-entry students are motivated, but s/he is not sure if this continues into the latter part because there are no levels for them to challenge into.

Lower-entry students in the present programme were characterized as "motivated -- they work their tails off!" by one teacher, but another said, "while those moderately low are encouraged because they know they can 'do it' with hard work, those very low are discouraged." Another teacher claimed the low-entry students are "terrorized, not motivated, and fear does not promote language learning." An interesting point was a conjecture that "middle-entry students are less motivated, perhaps, than the higher or lower students because (unlike the lowest levels) they can 'fall behind' and . . . (unlike

the highest levels) it's not such a 'fall from grace.'" Another interesting question was posed: "Does the present programme motivate the lower-level students negatively (through avoidance of failure) or positively (through attraction to success)?"

5.2.5.2 A Balance Sheet

When asked how well the present programme addresses the weaknesses of the previous programme, or whether they would like to return to the previous programme, most teachers reacted strongly in favour of the present programme. They said the present programme *addresses lack of standardization*: "There is a comfortable balance between standardization and creativity . . . scope and sequence enables us to know what students have been taught so we don't have to start at 'square one' all the time." It also *addresses students' lack of basic English skills*: "We're acknowledging that ESL is an important component of the first year. More students are participating actively; there's more discussion and less lecture-style." Finally, it *addresses the need for content-learning*: "Content is being taught, but in a much more logical way . . . the programme clearly defines the parameters of language and content so that teachers and students know better what to expect from course to course." One teacher noted that students seem to take on more responsibility for their own learning progress when they see it laid out so clearly. Another claimed the materials are more respecting of students' maturity (no more infantile reading materials) and diverse ability levels. Also, students are given more chances (to succeed, or to repeat levels without penalty) than before.

However, the present programme is seen as flawed as well. As one teacher noted, "in the process of addressing the previous program's weaknesses, we also created some new problems and needs." (Examples of these follow.) Another noted that there is a difference between 'addressing skills' and demonstrably improving them. This subject isn't really sure if or how students' skills, especially grammar, have improved: "Some

Level Five students still cannot write a paragraph." Another questioned if the very high-entry-level students are being truly challenged, noting that "we have a greater spread of ability levels and rate of learning than we account for or admit." This is related to another comment that, "Our clientele may have changed somewhat" (since initiation of the present programme).

"Would I want to return to the previous programme? . . . not a chance . . . not even in my dreams . . . definitely not," Ten out of thirteen teachers were adamant on this, but one wanted to return:

I would like to get back to three terms rather than five modules, I'd like to explore different ways of evaluating students that aren't so fear-producing, and I'd like to see all students getting into content courses earlier than they are now;

one wanted to combine the two:

Ditch the module system, put less weight on the final exam, give teachers a little more flexibility, and include more 'content' learning (but not 'V.-ism!'); however, retain some of the present skill expectations;

and one was nostalgic for certain aspects of the past:

I don't miss the lack of direction and goals. (V.) structures are great for organizing, but you need goals. However, we had some wonderful themes that we've lost -- deemed 'unreachable' for our students, and perhaps we went overboard, simplifying too much. Perhaps the 80% pass standard caused us to be a little too 'bare-bones', boring, and simplistic.

5.2.6 Qualitative Research Results: Proposed Changes

This leads us to the question of what changes teachers would like to make in the present programme. When trying to visualize an improved programme, the areas they most frequently cited are listed below in the same order in which teachers prioritized them. For possible justice implications, see Appendix 8.4.

1) *Avoiding using the term 'mastery' incorrectly:* Most teachers object to the term 'mastery' which creates unrealistic expectations and claims. 'Mastery' implies *control* of a skill or *comprehensive knowledge* of a subject, neither of which, they claim, our students can realistically be said to have attained by achieving 80% on a fairly simple exam at the end of a five-to-six week module.

A proposal was made by one teacher which would enable students to truly 'master' the subjects, enabling students to progress 'in their own time' as the proponents of mastery learning advocate. This was the idea of 'continuous intake' of students and allowing students to take more than one year to complete the 'entry' programme. (My note - Without this freedom to take as long as necessary to master the objectives, we are following a mastery system 'imperfectly' -- even, perhaps, as many teachers noted, 'dishonestly')

One alternative to this rather drastic step is to use another term for what we do. 'Competency-based' learning is sometimes used to denote evaluation according to how well a student has achieved course objectives (instead of according to how they stand in relation to other class members on a 'normal' curve) However, most teachers were unfamiliar with the term. One said it "more accurately describes what we're doing" while another said that "though we attempt to do it to some degree, it is not an accurate description." Another said it's good because it implies an 'application' of skill: "I can do it!"

2) *Lowering the 80% pass standard for Foundation:* Interestingly, several teachers' rationales for lowering the pass standard were similar to that used for creating it - *consistency* and *raising standards*. They would like to see a *consistent pass standard* (60% perhaps) used in both Foundation and Transition, and they feel the 80% standard, instead of motivating students to achieve a high standard, has resulted in teachers actually *lowering their standards* by over-simplifying exam questions and 'teaching to the

test' to enable 'an acceptable' (e.g., acceptable to the administration) number of students to pass instead of addressing actual student needs. They also feel the 'double standard' (separate grading scales for Foundation and Transition - see the GPA table in the Appendix 8.3) is confusing to students and parents as well.

3) *More opportunities for listening, speaking, instruction in pronunciation, and (structured) interaction with Canadians.* Particularly advocated were longer Listening/Speaking classes, possibly with specific pronunciation and grammar skills built into the course objectives; this is related to the next suggestion . . .

4) *Combining listening/speaking class with grammar class and combining experiential studies class with the study/presentation skills class.* This is partially in answer to the need many teachers have expressed to decrease the number of classes and increase the number of hours for oral and interactive activities. The major reasoning here, however, is that these are pairs of related subjects which should be integrated for their mutual enhancement, to reinforce skills taught in one that are directly applicable to the other. Advocates especially wanted verb tenses and articles to be in listening / speaking (instead of writing) class in order to make grammar more contextual and less like it was taught (often unsuccessfully) in Japan.

5) *Decreasing the number of objectives for each level (particularly grammar), having fewer and longer modules, and increasing the number of levels:* These are all related to the same problem of having too many objectives to teach properly in one five-to-six week module.

6) *More challenge for upper level students:* Teachers advocated *more challenge for upper level students* such as adding an additional level for them to challenge into at mid-

year, the opportunity to audit university-level courses, a year-long voluntary Honours Seminar (noted on the transcript in some way), the option to take more than one elective course in a term, and, possibly, leveled electives.

5.2.7 Summary: Back to the Future?

Teachers often waxed philosophical towards the end of the interview. The theme of '*getting back to what we have lost*' seemed to surface for a lot of teachers as we neared the end. When asked if s/he had anything to add, one teacher said, "It was interesting - it made me think, especially about the past," and another said, "I realized doing this interview how much I favour, support, and enjoy content learning and how much students benefit from it." One teacher claims, "I don't believe Japanese students are nearly as committed to sameness for everyone as we have presumed that they are. I believe we can be more creative and do more with students than modularize them 100%." Another questions whether teachers who are 'jacks of all trades' (rather than content specialists) are really what students want, noting that a truly fine school "should aim for a team of specialists." This same teacher wants "more freedom to teach outside a team at the upper levels." Another says:

I would like to get into more depth, content, and academic material. Students want new information; we should bring more research and issues to them rather than just slipping along with a few almost 'stereotypical' assumptions . . . people aren't interested in reading about what they already know. They lose the spark of motivation. *In our 'compromise,' perhaps we swung a little away from the junior college and a little too much towards the junior high school in terms of content!*

Thinking about the past and the future opens up *the theme of change* - of our students, of ourselves, of our world. Several teachers noted that, "We should never rest and be totally complacent. As our clientele and (their) employment requirements change, we must keep our eyes open and change to meet their needs," and "It's not just a matter

of the curriculum 'then and now,' we -- teachers and the college -- have changed and matured, too." Another noted that the present programme "isn't written in stone; since we've instituted it, we've actually made substantial changes such as initiating and standardizing 'progress' tests (my note - as compared to more summative final exams) : . . . and generally (and incrementally) improving most of our individual courses."

Others talked about how working at the college has deeply affected them personally and emotionally, in both positive and negative ways, such as one teacher: "We should focus on actual student learning rather than how we appear to others. [On the other hand], it is important for students and parents, especially of four-year students, to keep 'a core' of very good, contented, committed, trusting and trusted faculty who feel they have a personal stake in the institution's success -- this is Japanese style." Another teacher's bottom line was that compared to the past, "life is a lot easier as a teacher now!" This contrasted with the thoughts of a teacher whom contentment has clearly eluded:

I think the teamwork we have been put through has been necessary as *we have broken new ground in a teaching area for which we had no role model*, and that it has trained and helped both us as teachers and the programme in general. However, it has also been difficult. It is difficult to create curriculum for teachers with different teaching styles and requirements; it is equally difficult (no, more difficult) to follow another's half-completed or experimental curriculum, especially when that teacher has a different style. Having been through that treadmill, I believe it would be good now to try something free-er. Also, I do believe that our system of teaching -- so many hours, so many preps, so many meetings, so many tests, so much curriculum planning, teaching new areas, teaching in areas that are not one's strengths -- does lead to burnout and maybe not doing one's best. I would like to see *more time*, and encouragement to present, publish, and generally take part in the wider world of teaching -- all requiring time (which is nibbled away by the factors mentioned above) -- but *this is professional development*.

Another echoes these concerns with, "This is not a system for encouraging people to grow. There are structural weaknesses."

As if conversing with an unseen comrade, another teacher contributes an additional perspective to the same phenomena:

Are these good questions? Yes, but perhaps we should be asking how the resources have changed and been developed. I was hired to develop a computer course ten days before the first 280 students arrived. The first keyboarding programme was Shareware, the computers hadn't arrived, and the computer courses were taught from 6:00 - 10:00 pm. In addition, there was only a half-hour lunch, plus a half-hour faculty meeting every day! I would like to express a sense of gratitude to the dedication of my fellow faculty members, for their high expectations and high achievement. From Day One and continuing to now, never has there been ample time to get what needs to be done, accomplished. Faculty have developed curriculum with minimum hours, taken on leadership with no recognition, jumped in wherever a need was perceived, and evidenced tremendous teamwork consistently.

With all of its weaknesses, when asked how the present programme is perceived by most teachers, the answer was usually a mildly qualified 'good;' for example: "A big improvement - but not perfect!" or "O.K., but could stand improvement," or "Generally -- a fair degree of confidence that 'it works' for most students at a pace that helps them develop," but some said feedback from students is "Mixed: some seem happy; others are wondering why students aren't happy and want to re-assess it." Another ventured that "There is general dissatisfaction that it might not be doing what we thought it would . . . administrators like it, but teachers are getting disillusioned."

One teacher summarized it thus:

Teachers can't do many creative things (in 'Foundation'), but realize when they put the *whole* programme into perspective (including classes in computers, presentation/study skills, experiential studies, Transition and elective courses plus a cross-cultural 'survey' course *taught in Japanese*) the students' needs are being quite well-met. All in all, it's meeting students' needs and preparing them to get successfully into content issues. One strength of our faculty is that we are flexible and always looking for ways to improve.

I think in the final sentence above this teacher has defined our two common denominators, flexibility and the desire to improve. This is probably one of the few possible statements with which I am convinced every teacher at the college could agree.

6.0 CONCLUSIONS

6.1 Quantitative Research

6.1.1 Site-Specific Conclusions of the Quantitative Research

Within the parameters of the quantitative study, no consistently significant differences were demonstrated between the two programmes. In other words, no significant differences showed up *as significant* for more than one matched-pair year, contributing to the impression that these may be anomalies, due possibly to factors other than the difference between the two programmes.

Arguably, the most important weakness of this research is that it did not include other important skills that are taught at the college such as speaking or composition. Without this information, the results are somewhat narrow in implication.

6.1.2 Site-Specific Justice Implications of the Quantitative Research

As noted previously, we will consider *learning outcomes as a special form of potentially distributable 'resources'* like self-esteem, because they are indicative of what knowledge the individual has incorporated through an educational experience, and because they (both the knowledge itself and its demonstration through testing) help to determine an individual's future access to other resources.

The first question is, *according to John Rawls' criteria, and the quantitative data alone*, whether learning outcomes (mean increase in SLEP Listening and Reading scores) in the previous programme were justly distributed among students. It was just in that it was (1) of benefit to the group as a whole (at least as much as the present programme), and (2) it was of benefit to the least advantaged subgroup -- in that those with lower entry

SLEP scores made the greatest gains (not *significantly* different, however, from the present programme) of the cohort group.

According to the same criteria and data, the present programme would also be considered just. That is, it was (1) of benefit (or not *significantly* worse, at any rate) to the group as a whole, and (2) it was of benefit to the least advantaged subgroup -- again those with lower entry SLEP scores made the greatest gains of their cohort group.

One question from a justice point of view is: Is there any way of reversing the negative trend in mean listening SLEP gain for the upper-entry students without affecting the gains being made by the lower-entry students? SLEP gain is not a scarce resource in that one group's gain is not necessarily another's loss. Therefore, this should be possible. Note that this trend may be reversing anyway, as the difference between the two programmes has been decreasing each year.

It is possible that if the programme for upper-entry students were changed to give them significantly greater enrichment and challenge, this might have detrimental effects on the lower-entry students such as fewer educational resources (i.e., teacher time and materials), including decreased self-esteem. These justice implications will be explored more fully when qualitative data are added to the discussion. (See also Appendix 8.4).

6.1.3 Conclusions about the Kind of Information Gained from this Kind of Quantitative Research

This kind of research can give very specific answers to very specific questions. It is not concerned with questions of equality of access to education by different ability groups (or even to 'knowledge'), but of how learning outcomes are actually distributed among different ability groups by different programmes. It assumes that the difference between achievement on a pretest and a post-test is a function of the distribution of learning outcomes (which may, arguably, be indicative of access to knowledge itself or to

education). It also assumes that in a matched-pair group of students from two different programmes (starting with the same pretest scores), any difference in their mean post-test scores is a function of the difference between the two programmes. Given those assumptions, the research can give:

- 1) general information about differences in student achievement in the two programmes.
- 2) specific information about differences in the achievement of students in different ability groups in the two programmes.
- 3) accurate estimates of the significance of differences found in achievement in the two programmes.
- 4) (if comparing more than two years) evidence of patterns which can further confirm the significance of the above differences.

While some very specific useful information can be gained, there are some important limitations that must be understood by researchers using post-hoc testing and a matched-pair design to compare students' mean improvement in two different programmes. For example:

- 1) You need pre- and post-test scores for the same test, given during the same time of year (if comparing full-year programmes), under the same conditions. It needn't be a commercial standardized test. It could be made by a single teacher (or group of teachers) for her/his/their own action research, comparing different teaching methods or materials with different classes.
- 2) You cannot extrapolate that because a particular programme demonstrably improves Skill 'A' (e.g., Listening) that it also improves Skill 'B' (e.g., Speaking). You would need separate tests to demonstrate this. Therefore . . .

- 3) The greater the variety of skills being tested, the more generalizable are your results to the entire programme; the fewer the skills, the less generalizable are your results.
- 4) The more years you can compare with the baseline year, the better.
- 5) The baseline year is very important: The fewer anomalous conditions that year which can get confused with the programme effects, the better.
- 6) The fewer the changes in factors *other than programme changes* (e.g., student or teacher characteristics) among the baseline and other comparison years, the better.
- 7) The larger the groups you have from which to draw the matched pairs, the more likely you are to have large enough matched-groups to give you significant results.
- 8) You won't be able to tell which aspects of a programme are responsible for any observed differences, especially if the skills being measured are those general skills which (like 'listening' and 'reading') are developed and reinforced by many different aspects of the programme.

In summary, the strength and the weakness of quantitative data is its specificity. It cannot answer a lot of questions, but it can answer a specific question or set of questions quite well. In fact, it can even tell what the chances are that the results it gives are true.

I have one more piece of advice, particularly for teachers who might want to do this as action research and are uncomfortable with statistics. They can test for significance and make graphs with one of the new computer statistics programmes (for

Windows or Mac) made especially for the social sciences or education. They are relatively inexpensive and easy to learn and use, without an extensive knowledge of the mathematics. Be sure to input your data into these from the very beginning; importing data from a spreadsheet into one of these programmes can be very challenging, as I will attest.

6.2 Qualitative Research

6.2.1 Site-Specific Conclusions of the Qualitative Research

Here are the most important findings of the qualitative research:

While the final judgment was 'mixed', teachers generally characterized the previous programme as interesting, exciting, and ambitious in intent, but lacking consistency and standards, boring to some high-entry students, marginally accessible to many lower-entry students, and damaging to some of the latter students' self-esteem.

Teachers (and, by conjecture, students) find the *present* programme generally (though not 100%) effective in its distribution of educational resources to students. Most teachers felt it teaches language skills very effectively, but perhaps it doesn't teach enough content. Most teachers indicated they were ready to consider making some major changes in the present programme such as changing the 80% mastery standard or giving more hours to the teaching and practice of oral/aural skills.

Table 5**Student Responses to Programmes, by Entry Level (Remembered by Teachers)**

Summary of Student Responses to Programmes (remembered by teachers), by Entry Level		
<u>PREVIOUS PROGRAMME</u>		<u>PRESENT PROGRAMME</u>
<u>In general:</u>	feeling upset by lack of standards and consistency	trying to 'level up' or challenge -motivated!
<u>Level 1</u>	either 'having fun' or demoralized by 'track' system and 'lost' in difficult content	most working very hard, but a few demoralized by how difficult it is to achieve 80% pass
<u>Level 2</u>	either lost or challenged; depending on teacher, individual	system is basically good, but sometimes boring
<u>Level 3</u>	either challenged or 'coasting' some bored, many upset by lack of direction	many challenged, a few 'coasting', a few bored because there is nowhere to challenge to

6.2.2 Site-Specific Justice Implications of the Qualitative Research

According to the qualitative data alone, the previous programme was unjust because it lacked standards, was delivered inconsistently, and was somewhat inaccessible to the lower-entry students. Inequalities that occurred were generally seen to benefit no one, and lower-entry students were especially noted as suffering from curricular practices; the programme did not provide them, in other words, with the educational resources they needed to take advantage of the knowledge that was being presented or offered to them.

Tracking practices, inaccessibility of the curriculum, and the perception of inflated grades for the lower tracks all could have resulted in decreased motivation and *decreased self-esteem* for the lower-entry students.

Also, lack of consistency in what was taught from class to class and from teacher to teacher meant that, *in general*, resources were not being distributed equitably. Teachers seemed to accept that standardization and consistency can be carried to an extreme, bypassing teacher professionalism and stifling creativity, resulting in widespread boredom and the elevation of mediocrity to the norm. However, they generally felt that in the first programme, there was so little standardization and consistency that many students were actually being treated inequitably. This kind of unfairness is described by John Rawls as failure to meet the criteria of the concept of equality. The first of three levels of the concept of equality applies

to the administration of institutions as public systems of rules. In this case equality is essentially *justice as regularity*. It implies the impartial application and consistent interpretation of rules according to such precepts as to treat similar cases similarly (my emphasis) (504).

While the inconsistency in evaluation practices and in what was taught from teacher to teacher and class to class could have adversely affected students from all entry-levels, all these practices considered together probably resulted in an unjust distribution of resources in the previous programme favouring the upper-entry students.

According to the qualitative data alone, the present programme justly distributes educational resources to students. This is because teachers perceive that (1) it is of benefit to the group as a whole, (2) it is of benefit to the least advantaged subgroup (the lower-entry students), and (3) the unequal distribution of resources (in this case extra tutoring, opportunity to repeat levels without penalty, and the balance between skills-based versus content learning) is to the benefit of the least advantaged subgroup.

In other words, while those teachers interviewed tended to question whether upper-entry students in the present programme are getting enough challenge (particularly towards the latter part of the year), they are convinced that the lower-entry students need the skills-based approach. The programme is considered generally more just for the entire group, as well, because of its consistent standards (from class to class, teacher to teacher, and level to level) and in its *leveling* practices which, because of the ease and regularity with which students move on to upper levels, were felt to be less destructive of self-esteem than previous *tracking* practices. Therefore, the programme was thought to be generally just, but perhaps slightly in favour of the lower-entry students because of a possible decrease in challenge and interesting content offered to upper-entry students.

Again, we ask ourselves if it might not be possible to retain the positive effects on the lower-entry students while increasing the challenge and giving more interesting content to the upper-entry students. However, let us postpone this discussion until we amalgamate this data with the quantitative data.

6.2.3 Conclusions about the Kind of Information Gained from this Kind of Qualitative Research

The aim of these interviews was to examine from an instructor's perspective: (1) how *effectively* the previous and present programmes have distributed learning resources to students, (2) how *justly* each has distributed learning resources to students, and (3) what teachers perceived as student impressions of (1) and (2). The results to all three were to be in terms of 'students in general,' 'upper-entry students,' and 'lower-entry students.' While the interviews were 'open-ended,' in many ways, I was also looking for (and asked for) some very specific information from all teachers, knowing that a lot of the answers would overlap with others. Yet, I found the variety and scope of the answers was much greater than I had anticipated.

In addition to the specific information I wanted, I got these additional kinds of information:

- a) historical context - of events, of thoughts, and of feelings
- b) causality and order of events
- c) reasons for curricular decisions
- d) 'behind the scenes' actions, including evasive and compensatory ones
- e) opinions that are/were controversial, *especially* from people who generally avoid public controversy - the normally-outspoken people tended to be rather low-key in these interviews
- f) 'the flavour' of the past -- and of 'behind the classroom/office door' in the present as well
- g) unresolved contradictions and unanswered questions -- a very large portion of the information was in this category

- h) guesses and conjecture, though interviewees generally labeled these as such, I surmise

Were the questions answered clearly; are they neatly packaged and graphed? No, this was qualitative research: "You can't always get what you want, but you'll surely get more than you need . . ." I had intended for this to be a 'fishing expedition' with open-ended questions the bait, intending to 'catch' the unexpected but valuable new idea or insight, allowing it to 'surface' so that others might encounter it as well -- Just because an idea is good doesn't mean everyone has encountered it; just because everyone has an idea doesn't mean it's the best.

However, teachers (especially *before* the interview) tended to look upon what I was doing as a poll -- that the view that was stated the most often, 'won.' Several teachers wanted to participate, but hadn't taught in the previous programme. When told they weren't eligible to participate, two of the teachers expressed a feeling of disenfranchisement.

In reality, I ended up putting almost every idea, even if it was stated only once, into the 'results' section. This was because I didn't think my role was to be the judge of the worth of particular ideas at this point in the project, and because part of the value of this kind of research is in seeing the range of opinion within the subject group. However, if I saw an inconsistency, a negative consequence or implication, I may have pointed it out in a note or in the 'results' section -- not to the interviewee -- but tried to leave it to the reader to make the final decision.

In this vein of wanting to include all ideas, whenever I put prevalent ideas into the 'results,' I noted them as such. However, if a single person's idea was unique or represented an extreme point of view, shed interesting light on the topic, was very well-stated, or was what I judged to be a positive contribution to 'the dialogue' -- one which

others might 'pick up on' and incorporate into their own points of view -- I added these ideas with the caveat that 'one teacher said'

People sometimes gave several different results of one phenomenon, or several different causes of another. One person said 'A' was *bad* because of 'B,' while another person said 'A' was *good* because of 'B.' It would be interesting to see how they would react to inconsistencies such as this which cannot be pointed out during an interview.

Because of thoughts like this, and acknowledging that the interviewer's role is very subjective, I thought it would be a good idea to take the results I had written up back to the teachers to ask for their written comments as to where the results 'rang true' and where they didn't. When I was told this is a technique often used in interviewing, I decided I would definitely try it. However, there was no time to re-interview teachers; instead, I made several copies of the 'results' available in the teacher's lounge and asked teachers who had participated in the interviews to read them, then write notes and comments onto these copies over a two-week period. After that time, I incorporated their 'second opinions' into the results as well. Unfortunately, only two teachers responded to this format.

In contrast, teachers from the second-year (and beyond) campus, when shown the qualitative data, felt that their voices and that of other stakeholders should have been included because the limited point of view presented herein is not 'correct'. While I sympathize with this point of view, acknowledging that it would have been very interesting and that the reality this enhanced perspective would have provided would be closer to 'objective truth' (if I may use this term), I feel that the (albeit limited) view of the stakeholders I have interviewed is every bit as valid as that of any of the other stakeholders. Their insider point of view certainly helps explain why various decisions were made. Often, as has been shown, justice issues such as concerns about equality, consistency, standards, accessibility, and self-esteem were at the root of their decision-making.

A similar response to interview results has been described by Kidder and Fine. Trying to answer the question, 'Whose story shall prevail?' they note that quite often, those who observe others and the actors themselves have very different notions of the causes of the actors' behaviour: "Observers are likely to locate causes within the actor . . . and actors are likely to locate causes in their surroundings (71)." Note that teachers and administrators can each, at various times, be either actor or observer. From my teacher interviews, Kidder and Fine's claim is often (but not always) borne out -- teachers frequently justify their own actions in terms of the situation in which they found themselves, but explain administrators' actions in terms of their respective personal strengths or weaknesses. These personal attacks, while interesting, can be very hurtful to administrators who, unlike teachers, can be easily identified since there were so few of them. I have attempted to delete most of these statements -- they do not help evaluate the curriculum and they cause unnecessary pain to people who were working under great duress.

Teachers *generally* described the previous programme with consistency, logic, and clarity, in sharp contrast to the way in which they describe the present one. Teachers generally spoke 'with one voice' describing the previous programme; only a couple of teachers were enthusiastic about it very often. As a consequence, it was easy to say, "The consensus was" (and very tempting to say, "The programme was . . . !")

However, when teachers started describing the present programme, their answers started to seem more and more subjective, expressing many different points of view and shades of opinion. Teachers had various complaints about the present programme, but there seemed to be no clear or coherent statement, neither of the problems nor of possible solutions to them. Here I found myself quoting more, in order to let the variety of voices be heard -- I couldn't label something as a 'consensus' which clearly was not.

Now, there is no clear battlecry such as was heard last time the programme changed: "Consistency! Standards! Levels! Modules!" However, I hope that the process

of participating in my research may enable teachers to start formulating a more coherent statement of current problems, possible solutions, and a clear, collective vision of the future course of curriculum at the institution.

In summary, anonymous interviews of people who taught in both a previous and a present programme can be very helpful in programme evaluation through comparison. They can give collective (not necessarily 'correct') answers to specific questions. In open-ended questions, many new and interesting points of view emerge, as in a 'group brainstorm.' It gives participants a chance to examine (and explain) the present in terms of the past; and to re-examine the past in terms of the future; indeed, the process itself may lead to problem-definition and problem-solving. The narrative form that people so often select to give their answers in naturally includes *causal information*. However, as noted in the introduction, these narratives each told a somewhat different story. This leads us to the notion that . . .

Inherent weaknesses of this research technique include the subjective nature of both (1) the data and (2) its interpretation by the interviewer. Returning a transcript (of the notes taken) to the interviewee for corroboration could help correct the latter, but not the former, subjectivity. *On the other hand*, the multiple perspectives it provides (the multiple *subjectivities*, if you will) are also its greatest inherent strength. The more blind men describing that elephant, the better.

6.3 A Stereoscopic View (Getting Three Dimensions from Two Perspectives)

I have contended in this thesis that in order to make a more accurate determination of whether knowledge is being justly distributed in an educational programme, both quantitative and qualitative research should be pursued. Therefore, let us briefly examine what happens when we put the results of these two projects together.

Table 6
Summary of Quantitative and Qualitative Results

[Quantitative Results in Bold, Qualitative Results in Italics]

Almost no significant differences were noted between the two programmes, though assessment was confined to listening and reading proficiency.

The previous programme had positive features -- it was probably of greater interest to upper-entry students, but all students suffered from the lack of standards and consistency. Lower-entry students in the previous programme in particular experienced low self-esteem and low motivation due to the inaccessibility of the content and/or to the limited opportunities for advancement available within the three 'track' system

There is a possibility that the present programme may be slightly more effective in Listening for the lower entry students than the previous programme, and slightly less so for the higher-entry students.

The present program is more effective in general (for both Listening and Reading), particularly for lower-entry students, who are given extra time and resources, if necessary, to complete it; perhaps it isn't challenging enough for higher-entry students.

Both programmes seem to be 'just' according to Rawls' criteria. Both programmes seemed to be equally effective in their distribution of educational resources.

The present programme seems to be more 'just' according to Rawls' criteria. It is generally more effective as well.

How does the amalgamation of these two perspectives alter the view of the programmes we would get by using only one?

6.3.1 Stereoscopic Conclusion #1

Let us begin by looking at how the combined results affect *how just and effective the two programmes are in general*

IF WE ONLY USED THE QUALITATIVE RESEARCH, we would conclude that the present programme is far more just and effective than the previous one, though the previous programme contained elements some teachers would like to re-incorporate.

IF WE ONLY USED THE QUANTITATIVE RESEARCH, we would conclude that there was little difference in either the justice or effectiveness of the two programmes.

STEREOSCOPIC VIEW: The quantitative research informs the qualitative research that the previous programme was far more effective (or the present one far less effective) than widely supposed by teachers. On the other hand, the qualitative research informs the quantitative research about the high dropout rate of both students and teachers in the previous programme (a factor which could be looked at quantitatively, but which was brought to light by the qualitative interviews), the general unhappiness, lack of motivation, low self-esteem, and particularly the *lack of consistency and standards*, all of which resulted in the creation of the present programme.

STEREOSCOPIC CONCLUSION #1: The two programmes are both very similar in the effectiveness with which they distribute Listening and Reading knowledge to students, but the present programme does it more justly.

To arrive at this conclusion, we had to reject a finding (actually, an extrapolation of a finding) of the quantitative research -- that there was little difference in justice -- as based on incomplete information, since the quantitative research did not look at factors (such as 'consistency' and 'self-esteem') which also determine justice, but may have no discernible effect on test outcomes.

6.3.2 Stereoscopic Conclusion #2

Now let us look at how the combined results affect *how justly and effectively the two programmes distribute educational resources to low-entry and high-entry students*.

IF WE ONLY USED THE QUALITATIVE RESEARCH: we would conclude that the present programme is probably slightly more effective and just for the lower-entry than for the higher-entry students (for whom it provides less challenging content, at least in the first part of the year, than did the previous programme). However, given Rawls' criteria, we could conclude that this inequality was just (though there are some problems using Rawls' criteria here).

IF WE ONLY USED THE QUANTITATIVE RESEARCH: we would conclude that the present programme is probably slightly more effective for lower entry, and slightly less effective for higher entry, students in its distribution of Listening knowledge. However, given Rawls' criteria, we could conclude that this inequality was a just one -- though again, there are problems with this conclusion. We would also conclude that there

was no apparent difference in the two programs' distribution of Reading knowledge to the three entry levels.

STEREOSCOPIC VIEW: The tentative conclusions of the quantitative research support those of the qualitative research as far as Listening knowledge, but do not support them for Reading knowledge. In both cases, Rawls' criteria for justice are probably met.

STEREOSCOPIC CONCLUSION #2: Exactly the same as the quantitative conclusion above. In this case, we must stick with the more specific results of the quantitative research which are supported in part by the qualitative research. We must reject the qualitative results as applied to Reading because the quantitative research clearly and convincingly refutes this.

To arrive at this conclusion, we had to reject a finding of the qualitative research as based on teachers' *overgeneralization to Reading* of what was basically a valid observation about lower-entry versus upper-entry students' *Listening* progress.

6.4 Conclusions about the Kind of Information Gained by Combining these Two Types of Research

Hopefully this protracted exercise has served to demonstrate the usefulness of combining the two approaches to research, emphasizing their complementarity rather than their mutual exclusiveness. In Section Three (thesis, p33) I listed the three most widely-mentioned benefits of combining qualitative and quantitative methods:

(i.) that each is strong where the other is weak; thus they fill in each others' gaps, complementing one another and strengthening the research, (ii.) that when they support one another, the results are strengthened as well, and (iii.) that when they contradict one

another, both results are called into question; in this case an explanation for the contradiction (possibly requiring further research) is called for.

My research supports (i.) above. The qualitative research definitely complemented the quantitative in that it brought issues to light and allowed programme justice issues to surface which were not detected through quantitative methods. However, the quantitative research was particularly good at verifying programme effectiveness and demonstrating how educational resources (as shown via learning outcomes) are actually distributed among the various ability levels. That these two perspectives can act as checks on one another should now be apparent.

In addition, my research supports (ii.) in that when results overlap, conclusions are clearly corroborated. A good example of this is when both the qualitative and quantitative conclusions supported the idea that lower-entry students' mean Listening improvement *may be* superior in the present programme (though the data was insignificant).

However, my research gives particular support to (iii.), giving three examples of it: a) by pointing out when logical but incorrect conclusions have been made by using incomplete information -- e.g. that there was no difference between the two programmes in how justly resources were distributed; (b) by pointing out when overgeneralizations have been made from essentially correct observations -- e.g. that lower-entry students' mean Listening and Reading improvement are both superior in the present programme; and (c) by clarifying whether differences in achievement are real or imagined -- e.g. that the present programme is superior in its distribution of Listening and Reading knowledge.

6.5 Conclusions about Use of John Rawls' Criteria when Addressing Issues of Educational Justice

I found the notion of educational 'values' or 'resources' which included access to teacher, social rewards (such as grades, diplomas, etc.) and self-esteem to be very useful in determining the various dimensions of justice in education. It isn't enough to look at 'how much money is spent', certainly.

I also found very interesting and helpful the idea of comparing the distribution of resources to 'the least advantaged' (person, group) versus more advantaged (people, groups). The idea of justifying a certain amount of inequality if it can be shown to eventually better the whole of society seems to make a lot of sense.

If not for John Rawls' theory, I would not have deemed it so important (in either the quantitative or qualitative study) to see if the two programmes affected students in the various levels differently, and so been unaware of one of the two significant findings of the research.

However, dealing with students who are Japanese nationals studying in Canada begs the notion of 'the whole of society'. Should we consider it to be 'the college as a whole?' Canadian society? Japanese society? Or 'global society?' The answer here is unclear, though my natural tendency would be to try to determine the eventual effect on global society (of having more Japanese young people who, being able to speak English, can communicate with 'foreigners').

My most serious misgiving when trying to apply John Rawls' theories to evaluating the justice of educational resource distribution pertains to his notion that no one should be worse off in an unequal distribution than they would have been had the resources been distributed equally. How to determine (a) what an 'equal' distribution of educational resources would actually look like and (b) what the results of this equal distribution on the various groups of students would be is a big question for me still; I

have no clear solution for this and hope that those who argue for and against, for example, affirmative action programmes will eventually help clarify these issues.

6.6 Summary of Research Findings and Conclusions

6.6.1 Summary of Research Findings and Conclusions - For the Site (in-house)

Getting back to the theoretical base from which we began this journey, the research taught me several interesting new things about my institution.

First, it showed me that what teachers perceived as a big general increase in student learning due to initiation of a modular, skills-based mastery learning programme was only imagined. The previous programme had glaring faults from fairness and 'equality of access' points of view, but it certainly distributed knowledge (of listening and reading at any rate) no worse than the present programme.

Secondly, the phenomenon of the greater mean increase in Level One Listening SLEP, along with that of the smaller mean increase in Level Three Listening SLEP was reminiscent of the 'Robin Hood' effect, sometimes hypothesized for Mastery Learning, in which low-ability students make proportionally more progress, but high-ability students make proportionally less progress, than they did previously. Of course, the effect was small, it only held for Listening, and it was only significant for Level Three -- and that only for one year. Nevertheless, having it *corroborated independently by teacher interviews as well* was quite exciting.

Next, there was the issue of V.'s theories. They were almost totally out of use by the time I started teaching at the college, so I was unprepared for the vitriolic memories of them which my interviews brought up. My conclusion is that because of the timing and the way her theories were presented, no sense of teacher-ownership of the theories had a

chance to develop. In the end, V's theory may be seen to have served three unintended functions: first, as a symbol and convenient scapegoat for teacher frustration with the entire programme; next, as a unifying factor, in that opposition to it served to enable the new and disparate teaching staff form common cause; and finally, as a catalyst for change.

Finally, and perhaps most interesting: in the beginning, I defined two sets of attributes of power as 'horizontal' and 'vertical'. Horizontal power I saw as diffused, egalitarian, unpredictable, individualistic, tolerating of many 'correct' answers and ambiguity, soliciting active cooperation from those within its influence, encouraging creativity, and evaluating primarily by qualitative means. Vertical power, on the other hand, I saw as centralized, authoritarian, predictable, standardized, assuming that there is only one true 'reality', soliciting passive receptivity from those within its influence, encouraging uniformity, and evaluating primarily by quantitative means. In apparently typical Cartesian fashion, I had set up a series of dualisms. Yet, I hope it is clear to the reader that I see each as representing an extreme on a continuum rather than a binary choice.

Using these sets of attributes, it struck me that teachers in the first three years of this institution's history, though they were being asked to conform to some extent by use of V.'s theories, were basically in an extremely 'horizontal' situation. They were quite free to develop their own curricula including their own materials, methods, tests, and evaluation systems. Perceiving this extreme situation's effects to be a lack of standards, of coordination, of logical sequence, and of monitoring; confused and demoralized students; and inadequate materials, teachers voluntarily resituated themselves in a strongly 'vertical' configuration: the leveled, modular, skills-based, mastery learning programme we now have.

Ironically the power did not come from above; rather, the teachers imposed a high degree of standardization upon themselves. When asked why they were doing this, they

often cited a justice rationale: fairness (consistency and equity) in the teaching and evaluation of students. A few years later, when the negative effects of extreme vertical power (teaching to the test, overstandardization of materials and tests, individuality and creativity stifled) started becoming apparent, teachers began questioning if they had perhaps gone too far in their pursuit of fairness through standardization.

They are at this crossroads now, seeking to strike an acceptable balance between these two extremes.

6.6.2 Summary of Research Findings and Conclusions - Of The Thesis

While my primary intent at the beginning was to use the criteria for justice formulated by John Rawls to determine the justice of the distribution of educational resources at my institution, I think it is clear that I ran into some important limitations which made it impossible to apply these criteria fully in this context. While justice was my initial interest, it soon became obvious to me that effectiveness was requisite to a just programme, so effectiveness became a secondary focus. The idea of looking at how a programme affects both those students with low- and high-entry skills was inspired by Rawls, however, it was useful in helping determine both effectiveness and justice, and was equally appropriate for the quantitative and the qualitative studies. I hope this thesis has also succeeded in its goal of demonstrating use of a combination of quantitative and qualitative research methods in the determination of the justice of educational programmes.

6.7 Suggestions for Further Research

6.7.1 Suggestions for Further Research - In-House

I will only give three of numerous suggestions that come to mind. First, as stated, the implications of the quantitative research are really limited to reading and listening. I would suggest collecting random samples of student writing and student speech (through audio or video cassette recording) at the beginning and end of each year. This would be more difficult, and the samples by practical necessity, smaller. Each student's identification number would have to be attached to the sample for correlation with SLEP (so that a 'level' could be assigned to the student). Through these samples, a data bank would be created, allowing researchers to match pairs and compare achievement in these 'output' forms of English over the years and over different programmes at the institution.

Secondly, the institution should continue with beginning- and end- of year SLEP testing, if at all possible, in order to compare the listening and reading achievement of students among the various years and programmes.

Lastly, while the institution does administer anonymous year-end programme evaluation questionnaires to first-year students (to which they must reply in English in writing) and does have informal, one-on-one interviews between Japanese staff and students towards the end of their first year, it might consider making the latter into more standardized exit interviews to give students a voice in an ongoing qualitative analysis of faculty and student perceptions of programme effectiveness and justice. Care should be taken to note responses according to entry-level. These could also be correlated with various qualitative and quantitative data the college already collects, but has not, to my knowledge, tried to combine, such as achievement in Year Two, graduation exit interviews, and eventual employment data.

6.7.2 Suggestions for Further Research - Theoretical

There are three big questions remaining in my mind after this thesis. The first is the aforementioned exploration of John Rawls' idea: how to define '*the whole of society*'

in the late twentieth century. While Rawls himself recommended applying it on the level of country (or below), in this rapidly forming 'global society' I wonder if at some point Rawls' theory could be seen as applicable at this level.

The second is also related to Rawls' theories: how to define what an equal distribution of educational resources would be, and how to conjecture, *without actually putting it into effect*, what the effects of this would be on each group -- in order to determine whether any proposed unequal distribution would distribute to any group fewer resources than they would have had under this hypothetical situation of strict equality.

The third is an extrapolation of a criterion of equality proposed by Jerrold R. Coombs (thesis, page 5) that any two groups could have two very different programmes, both of which could be considered 'just' if "neither group has reason to envy the resources given to the other (Coombs 291)." In the case of extra resources being given to low-entry level students (for example, more time to complete their study of Levels One to Four), one could look at a possible tradeoff available for upper-entry level students who get to study in more depth and pursue a greater variety of topics. However, how are we to know if one group envies the resources given to the other?

One possibility would be to specifically address fairness issues in the term-end 'student satisfaction' surveys. A researcher would have to decide what percentage of students would have to be unhappy, and how unhappy they would have to be, before programme changes would be made. S/he would also have to decide what to do if lower-level students were happy, but upper-level students weren't. This is an interesting proposal which does not throw extra weight towards the well-being of the least-advantaged groups as does Rawls. Thus it is more egalitarian than Rawls' proposal, though arguably not as 'just'. It poses many unanswered questions, but is intriguing and, like Rawls' proposals, should be explored.

6.8 A Postscript

Three major developments of interest have recently occurred. In December of 1996, the teachers decided to remove grammar from the writing course and put it in the listening/speaking course, adjusting hours accordingly.

On January 2, 1997, the teachers of the college, in a typically 'horizontal' act, took a vote on whether to continue the 80% mastery standard for Foundation courses. They voted to make the new standard 50% starting in April, 1997, the beginning of the new year. The two most prevalent rationales for this were (1) unhappiness with two parallel grading schemes -- a desire for consistency -- and (2) to raise standards, in that out of a desire to enable most students to gain 80% on final exams, teachers had created exams that were too easy. Levels, modules, and the use of (presumably more difficult) standardized exams by a team of teachers will continue.

Interestingly, a desire for consistency (among teachers, not among grading schemes) and a desire to raise standards had been two of the major rationales for the change from the previous to the present programme. Consistency and high standards, then, seem to be very important values to our teaching staff. Hopefully the consistency and higher standards that were gained by changing to the present programme will not be lost by the change to a 50% pass standard.

Finally, though the school has made no plans to discontinue SLEP testing, it will be administering as well an instrument more popular in the Japanese market -- the TOEIC exam. If the SLEP testing were to be discontinued as 'redundant,' a valuable means of comparing our programmes over time ~~would be lost~~.

My concern here is mainly with justice, particularly, as John Rawls has taught me, with that of the least-advantaged students. Hopefully, teachers and others who evaluate these new programmes will look specifically at how each programme affects these students. Hopefully, as well, they will not depend *solely* on either teacher intuition

nor on test scores alone. Rather, they will thoughtfully combine these two kinds of results. If they do this, not only the least-advantaged, but the entire college, will benefit.

WORKS CITED

- Aoki, Tetsuo. "Toward a dialectic between the conceptual world and the lived world." *Contemporary curricular discourses*. Ed. William Pinar. Scottsdale, AZ: Gorsuch Scarisbrick, 1988. 402-16.
- Apple, Michael, ed. *Cultural and economic reproduction in education: Essays on class, ideology, and the state*. NY: Routledge, 1982.
- "Curriculum and reproduction." *Curriculum Inquiry* 9 (1979): 231-52.
- Arlin, Marshall. "Time, Equality, and Mastery Learning." *Review of Educational Research* 54 (1984): 65-86.
- Bakhtin, Mikel. *Speech genres and other late essays*. Trans. and Ed. V. McGee, et al. Austin, TX: U Texas Press, 1986.
- Block, J.H., H.E. Eftthium and R.B. Burns. *Building Effective Mastery-Learning Schools*. NY: Longman, 1989.
- Block, J.H. ed. *Mastery Learning: Theory and Practice*. NY: Holt, 1971.
- Bloom, Benjamin S. "Learning for mastery." *Evaluation Comment* 1 (1989): 1-12.
- Books, Sue. *Critical authority as a terrain of struggle: What is gained and what lost in the struggle on this terrain?* Proc. of Bergamo Conference on Curriculum Theory and Practice. Dayton, OH: October, 1992.
- Bowles, Samuel and Herbert Gintis. *Schooling in Capitalist America: Educational Reform and the Contradictions of Economic Life*. 1976. NY: Basic Books Paperback, 1977.
- Brecher, Janice I. "Secondary Level English Proficiency Test." *Reviews of English Language Proficiency Tests*. Ed. J. Charles Alderson, Karl J. Krahne, and Charles W. Stansfield. Wash., D.C.: TESOL (1987): 68-70.
- Carlson, Dennis. *Teacher and crisis: Urban school reform and teachers' work culture*. NY: Routledge, 1992.
- Carnoy, Martin. "School improvement: Is privatization the answer?" *Decentralization and school improvement: Can we fulfill the promise?* Eds. J. Hannaway and M. Carnoy. San Francisco, CA: Jossey-Bass, 1993. 163-201.
- Carroll, J.B. "A model of school learning." *Teachers College Record* 64 (1963): 723-33.

- Champlin, J.R. "Is creating an outcome-based program worth the extra effort? A superintendent's perspective." *Outcomes* 1.2 (1981): 4-8.
- Chan, K.S. "The interaction of aptitude with mastery versus non-mastery instruction: Effects on reading comprehension of grade three students." Diss. U. Western Australia, 1981.
- Chapman, William. *Inventing Japan: The Making of a Postwar Civilization*. NY: Prentice Hall, 1991.
- Coleman, James S. "The Concept of Equality of Educational Opportunity." *Harvard Educational Review* 38 (1968): 7-22.
- "Equality of Opportunity and Equality of Results." *Harvard Educational Review* 3 (1973): 129-37.
- Coleman, James S.; E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfield and R. York. *Equality of educational opportunity*. Washington, D.C.: Government Printing Office, 1966.
- Comenius, John, a.k.a. Johann Komensky. "Pampaedia." c. 1630. *Comenius' Pampaedia or Universal Education*. Trans. (f. Latin) A.M.O. Dobbie. Dover: Buckland, 1986.
- Connelly, F. Michael and D. Jean Clandinin. "Narrative inquiry: Storied experience." *Forms of curriculum enquiry*. Ed. E. Short. Albany, NY: SUNY Press, 1991. 121-154.
- Conner, K, I. Hill, H. Kopple, J. Marshall, K. Scholnick and M. Shulman. "Using formative testing at the classroom, school, and district levels." *Educational Leadership* 43 (1985): 63-67.
- Cook, T. and C. Reichardt, eds. *Qualitative and quantitative methods in evaluation research*. Beverly Hills, CA: Sage, 1979.
- Coombs, Jerrold R. E-mail to the author. 7 Jan. 1997.
- E-mail to the author. 10 Feb. 1997.
- "Equal access to education: the ideal and the issues." *Journal of Curriculum Studies* 26 (1994): 281-295.
- Corwin, R. *Militant professionalism: A study of organizational craft in high schools*. NY: Appleton-Centry-Crofts, 1970.

- Costniuk, Bill. "Education in Japan." *History and Social Science Teacher* 23 (1988): 147-50.
- Covey, Steven R. "Whole New Ball Game." *Executive Excellence* August 1996: 3-4.
- Deleuze, G. and F. Guattari. *Anti-oedipus: Capitalism and schizophrenia*. NY: Viking, 1977.
- Derrida, Jacques. *Of grammatology*. 1967. Trans. (f. Fr.) G. Spivak. Baltimore, MD: Johns Hopkins UP, 1976.
- "Distribution." *Concise Oxford Dictionary*. NY: Oxford UP, 1990.
- Educational Testing Service. *SLEP Test Manual*. Princeton, NJ: ETS, 1988.
- Eisner, Elliot. *The art of educational evaluation: A personal view*. London, UK: Falmer, 1985.
- *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. NY: Macmillan, 1991.
- Fitzpatrick, K.A. and W.W. Charters. *A study of staff development practices and organizational conditions related to instructional improvement in secondary schools*. Eugene, OR: U of Oregon - College of Education's Center for Educational Policy and Mgmt.: 1986.
- Freire, Paulo. "Conscientizing as a way of liberating." *Contacto*. March, 1971. Ed. A. Hennelly. *Liberation Theology*. Maryknoll, NY: Orbit Books, 1990.
- *Pedagogy of the oppressed*. 1968. NY: Seabury, 1970.
- Giroux, Henry A. *Theory and resistance in education: A pedagogy for the opposition*. South Hadley, MA: Bergin and Garvey, 1983.
- Gitlin, A. "Educative school change: Lived experiences in horizontal evaluation." *Journal of Curriculum and Supervision* 4 (1989): 322-39.
- Gitlin, A. and S. Goldstein. "A dialogical approach to understanding: Horizontal evaluation." *Educational Theory* 37 (1987) 17-27.
- Glanz, J. "Beyond bureaucracy: Notes on the professionalization of public school supervision in the early twentieth century." *Journal of Curriculum and Supervision* 5 (1990) 150-70.
- Glickman, C., ed. *Supervision in transition: 1992 Yearbook of the Association for Supervision and Curriculum Development*. Alexandria, VA: ASCD. 1992.

- Goodlad, John. "Access to knowledge." *Teachers College Record* 84 (1983): 787-800.
- Goya, Susan. "The Secret of Japanese Education." *Phi Delta Kappan*, October 1993: 126-9.
- Grumet, Madeline. "Retrospective: Autobiography and the analysis of educational experience." *Cambridge Journal of Education* 20 (1990): 321-6.
- Guskey, T.R. "Defining the essential elements of mastery learning." *Outcomes* 7 (1987): 30-34.
- House, Ernest. "Justice in Evaluation." *Evaluation Studies Review Annual, Volume One*. Ed. G.V. Glass. Beverly Hills: Sage, 1976. 75-100.
- Howe, Kenneth R. "Two Dogmas of Educational Research." *Educational Researcher*. Oct. 1985: 10-18.
- Huber, M. "The renewal of curriculum theory in the 1970s: An historical study." *JCT* 3 (1992): 14-84.
- Husen, Torstein. "Problems of Securing Equal Access to Higher Education: The Dilemma between Equality and Excellence." *Higher Education* 5 (1976): 402-22.
- Jackson, Phillip. *Life in Classrooms*. NY: Holt, 1968.
- Japan. Provisional Council on Educational Reform. *First Report on Education Reform*. June 26, 1985.
- "Justice." *Fontana Dictionary of Modern Thought*. London, UK: Fontana-Harper Collins, 1988.
- Kellaghan, T. and G. Madaus. "National testing: Lessons for Americans from Europe." *Educational Leadership* (1991): 87-90.
- Kidder, Louise H. and Michelle Fine. "Qualitative and Quantitative Methods: When Stories Converge." *Multiple Methods in Program Evaluation*. New Directions for Program Evaluation 35 (Fall, 1987). Ed. Mark W. Lipsey. San Francisco, CA: American Evaluation Assn., Jossey-Bass, 1987. 57-75.
- King, Martin Luther (Jr.). *Why We Can't Wait*. NY: Harper, 1963.
- Kitamura, Kazayuki. "The Decline and Reform of Education in Japan: A Comparative Perspective." *Educational Policies in Crisis*. Eds. William K. Cummings, et al. NY: Praeger, 1986. 153-170.

- Leclerq, Jean-Michel. "The Japanese Model: School-Based Education and Firm-Based Vocational Education." *European Journal of Education* 24 (1989): 133-49.
- Madaus, G. and T. Kellaghan. "Curriculum evaluation and assessment." *Handbook of research on curriculum*. Ed. Phillip Jackson. NY: MacMillan, 1992. 119-54.
- Madey, D.L. "Some benefits of integrating qualitative and quantitative methods in program evaluation, with illustrations." *Educational Evaluation and Policy Analysis* 4 (1982): 223-36.
- Mark, Melvin M. and R. Lance Shotland. "Alternative Models for the Use of Multiple Methods." *Multiple Methods in Program Evaluation. New Directions for Program Evaluation* 35 (Fall, 1987). Ed. Mark W. Lipsey. San Francisco, CA: American Evaluation Assn., Jossey-Bass, 1987. 95-100.
- Miller, Lynne and Ann Lieberman. "School improvement in the United States: nuance and numbers." *Qualitative Studies in Education* 1 (1988): 3-19.
- Nakayama, S. *Science in Japan, China, and the west*. Trans. J. Dusenbury. New Haven, CT: Yale UP, 1984.
- Naotsuka, Reiko and Nancy Sakamoto, et al. *Mutual Understanding of Different Cultures*. Tokyo: Science Education Institute of Osaka Prefecture, 1981.
- Pinar, William F. "Whole, bright, deep with understanding: Issues in qualitative research and autobiographical method." *Contemporary Curriculum Discourses*. Ed. William Pinar. Scottsdale, AZ: Gorsuch Scarisbrick, 1988. 134-53.
- Pinar, William F., William M. Reynolds, Patrick Slattery, and Peter M. Taubman. *Understanding Curriculum: An Introduction to the Study of Historical and Contemporary Curriculum Discourses*. NY: Counterpoints-Peter Lang, 1995.
- Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap-Harvard UP, 1971.
- Slavin, Robert E. "Mastery Learning Reconsidered." *Review of Educational Research* 57 (1987): 175-213.
- Stansfield, Charles. "Reliability and Validity of the Secondary Level English Proficiency Test." *System* 12 (1984): 1-12.
- Stone, L. "Results from a global curriculum project evaluation: Practical problems - theoretical questions." Paper presented at the annual meeting of the American Educational Research Assn., 1984, New Orleans, LA.

- Strike, Kenneth A. "Education, Justice and Self-Respect: A School for Rodney Dangerfield." *Philosophy of Education, 1979: Proceedings of the 35th Annual Meeting of the Philosophy of Education Society*. Ed. Jerrold R. Coombs. Normal, IL: PES, 1980. 41-49.
- "Fairness and Ability Grouping." *Educational Theory* 33 (1983): 125-34.
- "The Role of Theories of Justice in Evaluation: Why a House is not a Home." *Educational Theory* 29 (1980): 1-9.
- Tucker, R.C. *The Marxian Revolutionary Idea*. NY: Norton, 1969.
- Tyack, David. "School governance in the United States: Historical puzzles and anomalies." *Decentralization and school improvement: Can we fulfill the promise?*: Eds. J. Hannaway and M. Carnoy. San Francisco, CA: Jossey-Bass, 1993. 1-32.
- Unks, Gerald. "Three Nations' Curricula: What Can We Learn from Them?" *NASSP Bulletin* 76 (1992): 30-46.
- Whitson, James A. "The politics of 'non-political' curriculum: Heteroglossia and the discourse of 'choice' and 'effectiveness'." *Contemporary Curriculum Discourses*. Ed. William Pinar. Scottsdale, AZ: Gorsuch Scarisbrick, 1988. 279-331.
- Willis, Paul. *Learning to Labour*. Farnborough, UK: Saxon House, 1977. Hampshire, UK: Gower, 1981.
- Wise, A.E. "Minimum competency testing: Another Case of Hyper-Rationalization." *Phi Delta Kappan* 59 (1978): 596-8.
- Worthen, Blaine R. and James R. Sanders. *Educational evaluation, alternative approaches and practical guidelines*. White Plains, NY: Longman, 1987.

APPENDIX 8.1 Samples of SPSS 6.1 Matched-Pairs Input and Transformed Data

Table 7

Key to SPSS Input Data (1990 matched with 1996)

id96	=	Student ID# (1996)
list196	=	Listening SLEP score, beginning of year (both members)
read196	=	Reading SLEP score, beginning of year (both members)
list296	=	Listening SLEP score, end of year (1996 student)
read296	=	Reading SLEP score, end of year (1996 student)
id90	=	Student ID# (1990)
list290	=	Listening SLEP score, end of year (1990 student)
read290	=	Reading SLEP score, end of year (1990 student)

Table 8

Sample of SPSS Input Data (1990 matched with 1996)

id96	list196	read196	list296	read296	id90	list290	read290
962097	17	21	26	23	902163	20	24
962136	17	22	26	25	902034	22	23

Table 9

Key to SPSS Transformed Data (1990 matched with 1996)

chslep90	=	[list290 + read290] - [list196 + read196]
chslep96	=	[list296 + read296] - [list196 + read196]
chlist90	=	list290 - list196
hread90	=	read290 - read196
chlist96	=	list296 - list196
hread96	=	read296 - read196
slep1	=	list196 + read196
slep1lev	=	1 IF slep1<29; 2 IF 28<slep1<37; 3 IF slep1>36

Table 10
Sample of SPSS Transformed Data (1990 matched with 1996)

id96	list196	read196	list296	read296	id90	list290	read290	chslep90	chslep96	chlist90	chread90	chlist96	chread96	step1	slep1lev
962097	17	21	26	23	902163	20	24	6	11	3	3	9	2	38	3
962136	17	22	26	25	902034	22	23	6	12	5	1	9	3	39	3

APPENDIX 8.2Sample of SPSS 6.1 Ouput Showing Means and Significance -Partial Data Used for Figure1

t-tests for Paired Samples

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
chslep90	122	.235	.009	11.6393	4.835	.438
chslep92				11.1148	4.694	.425

Paired Differences

Mean	SD	SE of Mean	t-value	df	2-tail Sig
.5246	5.896	.534	.98	121	.328

95% CI (-.532, 1.581)

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
chslep90	100	.071	.480	12.7600	5.131	.513
chslep94				12.4400	4.205	.421

Paired Differences

Mean	SD	SE of Mean	t-value	df	2-tail Sig
.3200	6.397	.640	.50	99	.618

95% CI (-.949, 1.589)

APPENDIX 8.3

Parallel 80% / 50% Pass Mark Grading Schemes

Used in Students' First Year at College 'X'

(Initiated April, 1991 and to be Discontinued March, 1997)

Letter Grade: Hour	80% Pass 'Foundation' % Equivalent	50% Pass 'Transition' % Equivalent	Grade Points per Credit
A+	98-100	95-100	4.33
A	95-97	90-94	4.0
A-	92-94	85-89	3.67
B+	89-91	80-84	3.33
B	87-88	75-79	3.0
B-	85-86	70-74	2.67
C+	83-84	65-69	2.33
C	80-82	60-64	2.0
NC	0-79	-----	Incomplete
C-	-----	55-59	1.67
D	-----	50-54	1.0
F	-----	0-49	0

APPENDIX 8.4

Application: A Look at some Site-Specific Justice Issues:

Given the 'Stereoscopic Conclusion' that the present listening program may well be better for low-entry (and worse for high-entry) students than the previous program, the next logical step would be to continue the program 'as is' for the low-entry students and formulate one with more challenge for the high-entry students. A number of intriguing suggestions were made along these lines by teachers in the qualitative interviews.

However, there are justice issues to consider when changing a program: in particular, possible unintended negative effects on lower-entry students. The following principle, based on John Rawls' theory, should be kept in mind when considering changing a program: *An institution would not normally want to reverse an increased mean gain for its lower-entry students as the cost of increasing a mean gain for its higher-entry students.*

Consider the following possible unintended effects *adding enrichment and challenge for higher-entry students* might have on lower-entry students:

- (a) Funds that might have been spent on remedial tutoring and materials might be shifted to an enrichment program.
- (b) If Levels Three through Five (or elective) requirements or standards were made more challenging, this could cause hardship to students who began in Level One when they got to those levels (in the latter part of the year). [On the other hand, changing Levels Six or Seven (Listening/Speaking and/or Writing) would probably have no effect

on lower-entry students; neither would adding an additional level for upper-entry students to challenge into at mid-year.]

(c) If electives and Transition Reading became leveled (homogeneous), there would be further separation of students by ability-levels even at year end, in contrast to the present, where most 'Transition' classes (except Listening/Speaking and Writing) are multileveled (heterogeneous). Lower-entry students would not experience the positive effects (such as positive peer modeling and feelings of equality) they now receive towards the end of the year from being in the same reading and elective courses as higher-entry students. Note that in students' second year, there is only rarely leveling of students.

(d) In addition to the obvious detrimental effects of the above, all three could also feasibly result in *lower self-esteem* for entry-Level One students.

As for *the creation of an additional 'pre-Level One' entry level*, there seem to be three problems to consider:

(a) there would be no chance for a student to fail and repeat a level. If they failed, they would be unable to complete Level Four, and automatically end up in the Second Year's alternative program. This might result in a greatly increased number of students in the alternative program, an impact possibly unacceptable to parents, students, and the second year faculty.

(b) besides the lowered self-esteem they might experience being put in the alternative program, the students (the 'lowest' of the low) placed in this level at entry would experience, from the first day of their first year, even lower self-esteem than do

entry-Level One students now, . (What happens now is that students who fail Level One in Module One in effect create this class, though it doesn't start until Module Two -- the 'Level One Repeat class'). It is perhaps debatable whether there is more loss of self-esteem for a student who is initially placed in a 'Pre-Level One' class versus that of a student who is initially placed in a Level One class and proceeds to fail it.

(c) another associated problem is that it is still quite difficult to predict from student Listening (or total) SLEP scores which students will fail Level One. Therefore, to place a student in a pre-Level One class at entry might be to prematurely judge that student as less capable. Of course, this is exactly what one does by leveling students. However, this addition of a further ability-demarcation at entry might be excessive.

Solutions to the problem of 'not enough time in a five-to-six week module to teach all the objectives' included (1) *decreasing the number of objectives taught* at each level or (2) *increasing the number of weeks in each module* [thus decreasing the number of modules -- presumably from five to four]. What would be the implications of these to lower-entry students?

(a) To allot fewer objectives to each level would limit how many total objectives would be presented to each student by years' end, thus altering the present distribution of knowledge in all levels and lowering the entrance criteria for the second year program. It would also require wholesale reorganization of the first year curriculum. Note: lowering the minimum Exit Year One/Entry Year Two requirements would particularly impact upon the lower-entry students; it wouldn't have much effect on upper-entry students who only spend two modules in Foundation courses.

(b) Lengthening the modules creates fewer of them in a year, limiting how many levels could be taught to any one student. At present, if a student enters at Level One and fails one module, s/he is only able to complete Level Four, the minimum for entry into second year programs, by the end of the year. If there were fewer modules, a low-entry student who failed once would be unable to enter the second year program. A possible solution to this would be to decrease the required number of levels to be passed in order to enter second year programs. It is important to note that this option, which limits the number of levels any student can take, would have a more significant impact on high-entry students than would option (a).

(c) The idea of a continuous intake and exit of students would enable increasing the number of times students could repeat levels without penalty. However, as the teacher who suggested this also noted, there are social and logistical advantages to having cohort groups. The lack of a cohort 'support' group might have more negative impact upon lower-entry students.

Any of these options would conflict with the present requirements of the second year administration and/or faculty as well. There are, then, no obvious satisfactory solutions to this problem. Clearly this is a complex issue that would have to be very carefully implemented, considering all potential implications, particularly for the low-entry level students.

Finally, what would be the effect on lower-entry students of lowering the 80% pass standard? The effects of this are not so obvious. As noted in the qualitative interviews, some of the reasons for doing this are compelling. The only caution I would have is to think about the possibility of a program 'being more than the sum of its parts'. Proponents of various programs often say that in order to get the optimal effect of their

program, it must be taken as a whole. Mastery learning proponents tend to say this, certainly, though some argue that the major cause of the success of mastery learning is its use of frequent formative testing.

As noted earlier, the quantitative research done here does not explain what parts of a program are responsible for what effects; similarly it cannot predict what effect piecemeal changes (such as lowering the pass standard) would have. One would have to seriously consider all of the ramifications of this sort of change and arrange to monitor its effects, particularly the impact it had upon the distribution of educational resources to lower-entry students.

APPENDIX 8.5Description of SLEP Test

The Secondary Level English Proficiency (SLEP) Test was developed for use with high-school or junior college students by Educational Testing Service, the makers of the more famous TOEFL (Test of English as a Foreign Language). It is standardized and norm-referenced. It is a multiple-choice test with four options for each of its 150 questions -- 75 each in listening and reading comprehension.

Brecher (69) states that "validity studies indicate that SLEP is a valid test of English language proficiency . . . (and) . . . largely due to its multiple choice format, the reliability of the test is quite high (see also Stansfield)." It comes in three forms, which we administer in the beginning of the year in April, in mid-year in October, and at year-end in February.

We give it for the purposes of initial placement, to provide feedback on student progress to supplement that given by teachers, and to help evaluate our programmes over the various years.

However, the SLEP test does not measure proficiency in output skills -- speaking and writing. As well, notes the Educational Testing Service (41), "The test is not designed to provide information about scholastic aptitude, motivation, language-learning aptitude, and cultural adaptability."

SLEP can be a predictor of TOEFL achievement. The Educational Testing Service (Table 4, p39) provides this table of score equivalency:

<u>SLEP Total Scaled Score</u>	<u>Expected TOEFL Total Scaled Score</u>
64	600
58	550
53	500
47	450
42	400
37	350
31	300