Comparing ESL Electronic and Face-to-Face Pre-Writing Conferences and First Drafts: Discourse, Participation, and Idea Transfer

Michael James Fitze, B.Sc., B.Ed.

Department of Graduate and Undergraduate Studies in Education

Submitted in partial fulfillment of the requirements for the degree of Master of Education

Faculty of Education, Brock University
St. Catharines, Ontario

© January, 2004
Abstract

This study was a comparative investigation of face-to-face (i.e., proximate) and computer-mediated written (i.e., graphic) pre-writing conferences. The participants in this study were advanced English as a second language students. The 2 types of conferences were compared in terms of textual features, participation, and the degree to which they were on topic. Moreover, drafts written after the 2 types of conferences were compared in terms of textual features, and the degree to which they were related to the conferences. Students produced an equivalent amount of discourse in an equivalent amount of time in the 2 types of conferences. The discourse in graphic conferences displayed greater lexical range, and some evidence suggests that it was less on-topic. Both these results likely occurred because the graphic conferences contained more discourse demonstrating interactive competence.

Participation in graphic conferences was found to be as balanced or more balanced among students, and among students and the group leader combined. Overall, the drafts produced after the 2 types of conferences were of equivalent length and topical range, but some evidence suggests that drafts written after proximate conferences were more related to the conferences.
Acknowledgements

The Four Noble Truths* of Writing a Master's Thesis:

1. Writing a Master's thesis has many occasions of suffering. In fact, writing a Master's thesis has so many occasions of suffering that one might be tempted to say that writing a Master's thesis is suffering.

2. The cause of suffering while writing a Master's thesis is craving: craving for clarity, craving for accuracy, craving for brevity, craving for completion, craving for publication.

3. The release, forsaking, and abandonment of Master's thesis-related craving can overcome Master's thesis-related suffering.

4. One can overcome Master's thesis-related suffering by following a noble eight-fold path.

The Noble Eight-Fold Path of Writing a Master's Thesis:

1. Right advisor: First and foremost, I wish to thank my advisor, Dr. Hedy McGarrell, for her patience, understanding, encouragement, and timely assistance throughout this entire project. If it were not for these wonderful qualities in Dr. McGarrell, I doubt this thesis would ever have come to completion.

2. Right committee: Dr. Coral Mitchell and Dr. Vera Woloshyn were invaluable as committee members. They provided feedback that was extremely helpful towards producing the final draft and preparing for the defence.

3. Right colleagues: I wish to thank the administration of the IELP program, particularly Dr. Glenwood Irons, Jackie Cleland and Patricia Brillinger for allowing me to work with IELP students. Moreover, I wish to thank Dana and Jean, the teachers who believed in me enough to avail valuable class time for this study.

4. Right students: My thanks go to the students, now scattered around the world, who participated in this

* My method of organizing this section is intended to be unique, and perhaps humorous. I have nothing but the utmost respect and admiration for Buddhism and Buddhists.
study. Their hard work and willingness to participate made this project possible.

5. Right friends: Two of my friends deserve special mention, although for very different reasons. First, thanks go to Andrew, both a friend and a colleague, for helping me to focus, and for taking some of the fear out of inferential statistics. Second, thanks go to Chris for always being fun, and there, and for very rarely asking me about my thesis.

6. Right family: Between the outset and completion of this project, I became an uncle, a husband, and a father. Welcome to the world, Katherine and James! To my wife, Ayano, I cannot express enough thanks. My accomplishment here can hardly match your patience, understanding, and sacrifices. I also wish to thank my mother for her boundless compassion, sense of curiosity and wonder, and penchant for waxing philosophical. Our many discussions were an inspiration. Finally, for my father, even though he could not see the completion of this work, his patient and good-natured help in the proposal stage of this project, as in my life in general, put me on the right track.

7. Right concentration: In addition to meeting my wife, getting married and having my first child, during the course of this research, I moved from my home and native land, Canada, to Japan, where I am presently teaching. While many things in my life have changed all around me, the one thing that has remained constant is this thesis. In a deep sense, concentrating on this research has provided me with stability during an otherwise unsettled time. As a result, I move into the future with some hesitation, certain only that there will be other studies, other challenges.

8. Right attitude: Considering all the help and support that I have received, and the stability I have gained by focusing on this work, it is perhaps a small wonder that I have, finally and at long last, completed this thesis. However, even with all this help and support, I doubt I would have done so were it not for an attitude which in a word is preserved as the motto for the school that granted me this Master's degree: Surgite!

^ push on; never surrender
# Table of Contents

## Abstract

Page ii

## Acknowledgements

Page iii

## List of Tables

Page x

## List of Figures

Page xii

## CHAPTER ONE: THE PROBLEM

Page 1

- Background of the Problem
  - Page 1
- Statement of the Problem Situation
  - Page 2
- Objectives
  - Page 3
- Rationale
  - Page 4
- Importance of the Study
  - Page 6
- Scope and Limitations
  - Page 6
- Outline of the Remainder of the Document
  - Page 8

## CHAPTER TWO: REVIEW OF THE LITERATURE

Page 11

- The Social Context of Computer-Mediated Communications
  - Page 12
- General Approach to Studying Communication
  - Page 13
- Investigating Electronic Conferences
  - Page 20
- Face-to-Face and Electronic Written Conferences
  - Page 25
- The Process of Writing
  - Page 35
- The Effect of Electronic Conferences on Writing
  - Page 40
Measuring the Relation of Conferences and Drafts......49
Latent Semantic Analysis..........................................................52
Creating A Semantic Space.........................................................54
  Selecting a corpus of texts.....................................................55
  The term-document matrix.....................................................57
  Information theoretic weighting............................................63
  Singular value decomposition..............................................64
Comparing Documents in Semantic Space..............................72
Validity and Reliability of LSA...............................................76
Analytical Measures Derived From LSA..................................78
Summary of Research Direction.............................................80

CHAPTER THREE: METHODOLOGY AND PROCEDURES.......................83
Research Methodology............................................................83
Participants.............................................................................84
Ethics......................................................................................87
Procedures..............................................................................88
Research Design and Instrumentation....................................97
  Investigating Conference Textual Features.........................98
  Investigating Conference Participation..............................100
  Investigating Draft Textual Features.................................101
  Investigating the Relation Between
  Conferences and Drafts......................................................103
  Conference average maximum cosine.................................103

vi
| Conference mean average cosine | 105 |
| Comparing conference average maximum cosine and conference mean average cosine | 107 |
| Draft average maximum cosine | 110 |
| Draft ratio total words to new words | 112 |
| Data Collection and Recording | 113 |
| Data Processing and Analysis | 114 |
| Methodological Assumptions and Limitations | 121 |
| An Assumption About the LSA Measures | 121 |
| External and Internal Validity | 128 |
| Restatement of the Problem | 139 |

**CHAPTER FOUR: DATA ANALYSIS**

| Conference Measures Correlations | 141 |
| Comparisons of Conference Textual Features | 144 |
| Amount of Discourse and Length of T-units | 144 |
| Lexical Range | 152 |
| Comparisons of Conference Participation | 154 |
| Researcher Participation | 154 |
| Equality of Participation Among Students and Among Students and the Researcher | 158 |
| The Effect of the Researcher on the Equality of Participation | 165 |
| Draft Measures Correlations | 166 |
Comparisons of Draft Textual Features .....................................168
The Semantic Relation of Conferences and Drafts.............169
The Relation of Conferences to Drafts.................................171
The Relation of Drafts to Conferences.................................174

CHAPTER FIVE: SUMMARY, CONCLUSIONS, AND IMPLICATIONS.........181
Summary of Rationale and Research Objectives.................182
Participants........................................................................185
Conclusions.........................................................................186
Research Question 1 - Conference Textual Features..............186
Amount of discourse..........................................................187
Length of t-units................................................................188
Lexical range......................................................................188
Research Question 2 - Conference Participation......................189
Researcher participation......................................................190
Balanced participation in conferences..................................191
The effect of the researcher on balanced participation..............193
Research Question 3 - Draft Textual Features.........................194
Research Question 4 - The Relation of
Conferences and Drafts.......................................................195
On-topic discourse in conferences.......................................196
The relation of drafts to conferences............199
Implications for Practice...........................................201
Implications for Theory............................................204
Implications for Further Research...............................206
Conclusion....................................................................210

References...................................................................212
Appendix A: Glossary of Terms.......................................227
Appendix B: Semantic Space Information.........................247
Appendix C: Excerpt of Graphic Conference Sorted by Average Maximum Cosine.................................251
Appendix D: Excerpt of Proximate Conference Sorted by Average Maximum Cosine.................................253
Appendix E: Excerpt of Graphic Conference Sorted by Mean Average Cosine............................................255
Appendix F: Excerpt of Proximate Conference Sorted by Mean Average Cosine............................................258
Appendix G: Ethics Approval...........................................261
Appendix H: Consent Form............................................262
Appendix I: Topics for Discussion..................................265
Appendix J: Summary of Procedures in Temporal Order.....268
Appendix K: Meaning and Calculation of the Gini Coefficient.................................................................273
Appendix L: Conference Transcription Procedures............276
Appendix M: Example of Conference Text Transcription...........282
Appendix N: Excerpt of Graphic Conference Sorted by
Average Maximum Cosine.............................................287
Appendix O: Excerpt of Proximate Conference Sorted by
Average Maximum Cosine.............................................291
Appendix P: Example of Researcher Participation During
a Proximate Conference.............................................295
List of Tables

Table | Page
1. Example Term-Document Matrix 1 | 58
2. Example Term-Document Matrix 2 | 62
3. Composition of Class Groups | 86
4. Scheduled Participation | 90
5. Extract of Conference Data | 115
6. Intercorrelations of Conference Measures | 142
7. Comparison of Conference Textual Features Across Conference Condition | 146
8. Comparison of Standardized Type-Token Ratios Across Conference Conditions | 153
9. Comparison of Researcher Participation Across Conference Conditions | 156
10. Comparison of Equality in Participation Across Conference Conditions | 161
11. Intercorrelations of Draft Measures | 167
12. Comparison of Draft Textual Features Across Conference Conditions | 170
13. Comparison of On-Topic Discourse Across Conference Conditions | 172
14. Comparison of the Similarity of Drafts to Conferences Across Conference Conditions | 177
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Breakfast in three-dimensional space</td>
<td>66</td>
</tr>
<tr>
<td>2.</td>
<td>Interaction of week and condition for conference words per t-unit</td>
<td>150</td>
</tr>
<tr>
<td>3.</td>
<td>Gini by class - students only</td>
<td>163</td>
</tr>
<tr>
<td>4.</td>
<td>Gini by class - students and teacher</td>
<td>164</td>
</tr>
<tr>
<td>5.</td>
<td>Interaction of week and condition for conference mean average cosine</td>
<td>173</td>
</tr>
<tr>
<td>6.</td>
<td>Interaction of week and condition for conference average maximum cosine</td>
<td>175</td>
</tr>
<tr>
<td>7.</td>
<td>Interaction of week and condition for draft average maximum cosine</td>
<td>178</td>
</tr>
<tr>
<td>8.</td>
<td>Interaction of week and condition for draft ratio total words to new words</td>
<td>180</td>
</tr>
</tbody>
</table>
CHAPTER ONE: THE PROBLEM

Whole-class discussions, or in other words, class conferences, are often used in second language writing classrooms to generate language and ideas about a topic. This pre-writing stage is usually viewed as particularly important for second language writing students who typically have a relatively restricted vocabulary. The pedagogical rationale behind this activity is that second language students can use the language and ideas generated by group members during conferences afterwards in their writing. This study compared the amount and complexity of language, as well as the patterns of participation observed during face-to-face and computer-mediated pre-writing conferences. Furthermore, this study compared the length and topical breadth of drafts written after the two types of conferences. Moreover, this study compared the degree to which discourse in the two types of conferences was on topic, and the degree to which second language students made use of language and ideas generated during face-to-face and computer-mediated pre-writing conferences when writing first drafts.

Background of the Problem

Because pre-writing conferences precede first drafts, researchers investigating the process of writing have
typically read a cause-and-effect relationship into the temporal gap between these activities and have assumed that the pre-writing activity must have some effect on the writing. First language (e.g., Bartholomae, 1993) and second language (e.g., Kern, 1995) writing teachers continue to use whole-class pre-writing conferences and claim that students do seem to use ideas and language generated in pre-writing conferences in first and subsequent drafts, even though no systematic investigations of this process exist to suggest that this is indeed the case.

Statement of the Problem Situation

In part, the assumption by researchers that writing students do use language and ideas generated in conferences in subsequent writing has undermined closer investigation of the process by which this is presumed to occur. In addition, perhaps equally, investigation of this process has not been undertaken because the tools necessary for a topical analysis of ideas generated in large numbers of texts have not existed until recently. A quantitative investigation of the transfer of words and ideas from conferences to drafts at an acceptable level of statistical power would require that a large number of texts be analyzed.
In other words, part of the difficulty in investigating the transfer of language, and particularly ideas, from pre-writing conferences to first drafts has had to do with the seemingly Herculean task of transcribing conferences, identifying and categorizing all the ideas in pre-writing conferences and first drafts, and then matching those ideas across stages of the writing process. Consequently, evidence to support the assumption that language and ideas generated in conferences are indeed transferred to writing has not been forthcoming.

Objectives

This study was a comparative investigation of idea generation activities in face-to-face and computer-mediated conferences. There were four objectives motivating this study. The first objective was to compare the textual features of these two types of conferences in terms of lexical range, or in other words, lexical repetition, and the total numbers of words and the total number and length of statements generated by students in equivalent time frames. Is there a difference in the amount of discourse students generate, and in students' tendency to repeat the same words in the two types of conferences?

The second objective was to provide evidence to support, or refute, the claims previously made by other
researchers about the differences between face-to-face and computer-mediated conferences in terms of teacher and student participation. For instance, is participation more equally balanced among the students, and among the students and the teacher combined, in one type of conference than in the other?

The third objective was to compare the textual features of first drafts written after the two types of conferences in terms of length and topical breadth. Do students tend to write drafts that are longer or that cover more topics after one of the two types of conference?

The fourth objective was to compare the degree to which the two types of conferences give rise to on-topic discourse, and to compare the degree to which students incorporate what is said in the conferences when writing their drafts. To put it another way, the questions are: Is there a difference in students' ability to generate relevant, on-topic discourse in the two types of conferences; and Is there a difference in students' ability to use the language and ideas generated in the two types of conferences in their drafts?

Rationale

Research into face-to-face and written electronic conferences in second language education has begun to
reveal how these two types of conferences impact participation patterns and discourse. However, such studies have been very limited in number, have used small sample sizes, and have remained largely unreplicated. As a result, part of my rationale for this study was to partially replicate other studies because research in this field stands in dire need of additional evidence about discourse and communication in computer-based written conferences.

In addition, powerful new computer-based tools for text analysis make it possible to carry out relatively large-scale investigations of the relationship between the pre-writing and drafting stages of the writing process. As mentioned, previous research has not provided evidence to support the assumption that ideas generated in conferences are indeed transferred to writing. Part of my rationale for this study was to investigate the transfer of language and ideas from pre-writing conferences to drafts. By tracing language and ideas generated by group members during face-to-face and computer-mediated pre-writing conferences to drafts written by students afterwards, data can now be gathered to investigate and compare how communication in these two types of conferences influences students' writing.
Importance of the Study

This study will add to existing knowledge of the impact of computer-mediated written conferences on group communications. Moreover, this study will provide direct evidence to support or refute the notion that students make use of pre-writing conferences when writing first drafts. In sum, this study will enhance theoretical understanding of how the conditions for communication in pre-writing conferences impact conference communications and the transfer of ideas from conferences to drafts.

Scope and Limitations

As mentioned, in part, this study compared the textual features and participation patterns between computer-mediated and face-to-face pre-writing conferences. Moreover, this study compared the textual features of drafts written after the two types of conferences. Finally, this study compared the degree to which these two types of conferences prepared students for the drafting task via idea and language generation and transfer.

Although it would have been interesting and valuable to explore the influence of factors such as gender, learning style, and cognitive style on conference communication and idea transfer, it was decided that the
number of variables already under consideration were quite enough for one study.

Moreover, since the focus of this study was, in part, on the transfer of language and ideas from conferences to drafts, no measures of draft quality were included. Given my current focus, I was not particularly interested in the quality of the students' writing, but merely with their ability to incorporate the language and ideas generated in the two types of conferences into their writing. Although writing quality is of crucial importance to studies of writing pedagogy, I leave comparisons of draft quality for a future study.

Twenty-seven intermediate to advanced English as a second language students were involved in the study, and observations were made for a total of eight conferences (i.e., four face-to-face, and four computer-mediated) over 4 weeks. Consequently, rather than being generally applicable to the population of intermediate to advanced English as a second language students at large, the results of this exploratory study were expected to provide a picture of the differentiating effect of conference conditions for the two groups of students involved. The results of this study were expected to be applicable to
similar groups of students studying under similar conditions.

This research was designed to be particularly valuable to teachers who are contemplating introducing computer-mediated pre-writing conferences into their English as a second language writing class. This study was expected to provide such teachers with an idea of what they can expect when introducing this pre-writing activity into existing class groups.

Outline of the Remainder of the Document

The next chapter begins with a general discussion of computer-mediated communication within a wider social context. Following this, the formal features of traditional communications are contrasted with the formal features of electronic communications in terms of two key concepts: presence and bandwidth. This discussion is used to develop terms to allow clarity in thinking and speaking about new communication environments. For definitions of technical terms used in this study, please see the Glossary of Terms presented in Appendix A of this study.

After providing the terms necessary to speak clearly about the underlying formal features of electronic communications, a review of empirical studies of how the formal features of communication in electronic conferences
have been observed to affect the communication process is offered. Following this, a review of process writing theory is used to prepare the way for an assessment of comparative studies of the effect of face-to-face and electronic pre-writing conferences on subsequent writing. Following this methodological critique of previous studies, the approach to be used in this study is outlined. The chapter concludes with a description of the analytical instrument used in this study to gauge the semantic relation between conferences and subsequent drafts, and a restatement of the research directions for this study.

The third chapter begins with an explanation of the research methodology used in this study. Following this comes a description of the participants in the study, which includes an explanation of how participants were assigned to groups. Next, ethical considerations in the design of this study are addressed. After that, the procedures followed with participants are described. Next, a detailed explanation of the research design and the measures used in the study is followed by a description of the data collection procedures. Finally, methodological assumptions and limitations are discussed, and the research questions under investigation are reiterated.
The fourth chapter contains the data analysis section of this study. Chapter 4 is organized around the four research objectives listed in the above section on objectives: Each research objective is addressed in turn. Moreover, sections on correlations of conference measures and correlations of draft measures are also included to examine the interrelationships between, respectively, conference measures and draft measures for each of the two types of conferences.

In the final chapter, the results of this study are summarized and the implications for practice, for theory, and for additional research are discussed.
CHAPTER TWO: REVIEW OF THE LITERATURE

This chapter begins with a general discussion of computer-mediated communication within a wider social context. Following this, the formal features of traditional communications are contrasted with the formal features of electronic communications in terms of two key concepts: presence and bandwidth. This discussion is used to develop terms to allow clarity in thinking and speaking about new communication environments. After this, a review of empirical studies of how the formal features of communication in electronic conferences have been observed to affect the communication process is offered. Next, a review of process writing theory is used to prepare the way for an assessment of comparative studies of the effect of face-to-face and electronic written pre-writing conferences on subsequent writing. Following this methodological critique of previous studies, the approach used in this study is outlined. The chapter concludes with a description of the analytical instrument used in this study to gauge the semantic relation between conferences and subsequent drafts, and a restatement of the research questions to be addressed in this study.
The Social Context of Computer-Mediated Communications

Given the widespread use of electronic communications in the technologically developed world, it is perhaps unsurprising that researchers drawing theoretical roots from a multitude of traditions and interests have investigated various manifestations of computer-mediated communication. Information technologies, including computer-mediated communications technology, impact our notions of literacy, indeed our daily lives, in ways that are of historical, social, economic, and political interest.

For instance, some suggest that electronic modes of communication provoke anxiety that young people are at risk, endangered by a rising tide of information over which the traditional controls of print media and the guardians of knowledge hold no sway (Birkerts, 1998; Merchant, 2001). Accordingly, Merchant argues that when the popularity and power of new media become more widespread and start to impact on the wider culture, dominant groups often meet the new media with concern and contempt.

For example, placing the label "chat" on a communication tends to frame the interaction as trivial through association with popular and frivolous communication (Bordieu, 1992). As a result, in this study,
I primarily refer to what has been called chat as graphic conference. As shall become clearer, I use this term in part to draw attention to this activity's formal features. Moreover, as just suggested, since the term chat is typically associated with frivolous communication, I use the term graphic conference to emphasize the legitimacy of this activity as part of second language writing curricula.

Although any study of electronic communication must ultimately be concerned with wider social attitudes, the traditions of most direct relevance to this study are first and second language pedagogy, psycholinguistics, and cognitive and social psychology. Research in these fields has begun to illuminate how social, psychological, and linguistic processes in written electronic conferences tend to change patterns of teacher and student interactions in ways that might be beneficial to second language students generating and sharing ideas for writing (e.g., Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Slatin, 1998; Sullivan & Pratt, 1996; Warschauer, 1996; Warschauer, Turbee, & Roberts, 1996).

**General Approach to Studying Communication**

Historically, beginning with ancient Greek philosophers such as Plato, much of Western thinking about communication has focused on differences in form and
function between speaking and writing (Derrida, 1972). For instance, speaking is aural, often informal and social, and typically associated with face-to-face interaction. In contrast, writing is graphic and often associated with more official communications that take place over a distance in space and time. An analysis of the term "written conference" will help to reveal how pervasive this tendency still is in contemporary Western thinking.

At first the words might seem at odds, even contradictory. Before being acquainted with electronic media, one reason along lines such as these: For the communication to be a conference, participants must be able to engage each other in synchronous conversation. So why would participants choose to write? Under usual circumstances, assuming that the participants are all capable of encoding and decoding aural information, speaking seems to be a more efficient and more naturally expressive form of communication.

The term written conference provides a good example of how electronic media blur traditional thinking about the formal features and language functions with which speaking and writing are traditionally associated (Merchant, 2001). For instance, in electronic media, voices are not tied to a particular time and place. Speech, in other words, need
not be bound to face-to-face interaction in the here and now. While in a corresponding fashion, as rapid real-time conferencing allows users to interact with whoever happens to coexist in "cyberspace," writing has taken on some of the communicative functions and formal features traditionally associated with speech (Chun, 1994; Kiesler, Siegel & McGuire, 1984; Merchant, 2001).

When researching how the formal features of some new electronic media might influence the communicative process (e.g., the group dynamic, process, and content of discourse), the terms used in analysis are inevitably derived from talk about analogous situations drawn from experience with the most familiar modes of communication. Yet terms derived from traditional analysis of communication are laden with traditional associations that may impede clarity in thinking about the purposes for which language tends to be used, or in other words, the functions of language, and the unique formal features of a given electronic environment.

For instance, it can be quite misleading to refer to electronic conferences as written speech. First, one is not informed if the communication is to be "synchronous" or "asynchronous." Second, one is left wondering what "paralinguistic cues," or in other words, nonverbal
communications, may be in play. Third, one may question if the discussion is to be thought of as a formal or an informal communication. Finally, the polysemy of the word *speech* may make it difficult to tell if the communication is to be a conversational dialogue, or a monologue that was prepared beforehand in writing. As a result, clarity in the terms used to describe electronic communications media is necessary to avoid ambiguity.

In fact, in contrast to traditional associations, participating in a written electronic conference is an experience unlike either writing or speaking as conventionally understood (Kiesler et al., 1984; Merchant, 2001; Perkins & Newman, 1999). For instance, although such computer-mediated communication is synchronous, it is graphic rather than aural. In other words, although participants can communicate with each other at the same time by exchanging text and pictures, they cannot see or hear each other. As a result, it is more accurate to describe what has been called "written speech," "electronic conferences," or "written electronic conferences," as "graphic synchronous conferences," or more simply, "graphic conferences," since the word *graphic* implies the exchange of text and other graphical forms of information, and the
word conference implies synchronous, dialogical communication.

To ensure that my theoretical talk about the formal features and functions of language in graphic conferences not be ambiguous, I will begin by briefly reviewing traditional thinking about the formal features and functions of language in spoken and written communications. I do this to show where theoretical thinking inherited from traditional analysis of speaking and writing might warrant revision, to avoid confusion, and to more conveniently hold forth clear discourse about some distinctive aspects of graphic conferences.

In terms of function, historically, speaking has been clearly distinguished as the principle and superior form of communication. For instance, In Phadreus, Socrates argued that proper instruction is only possible in speaking because speaking is the only medium through which one might impart not only the truth, but also the words to defend the truth, the specific content of which is idiosyncratic to the mind of the listener (Plato, c. 427-347 BC). This sort of complete interaction is not deemed possible in writing because of the formal features of writing. Socrates says that writings, like paintings,
...look like living beings, but if you ask them a question they maintain a solemn silence... you might suppose that they understand what they are saying, but if you ask them what they mean by anything, they simply return the same answer over and over again. Besides, once something is committed to writing it circulates equally among those who understand the subject and those who have no business with it; writing cannot distinguish between suitable and unsuitable readers. And if it is treated unfairly or abused it always needs its parent to come to its rescue; it is quite incapable of defending or helping itself. (p. 97)

In the quote above, writing is depicted as a radically problematic form of communication: In the hands of an unsuitable audience, written words may be misinterpreted; they may seem authoritative without necessarily being so, or trivial when they are profound. Writing is "silent," in that the author of written works is not ordinarily co-present with the readers. Yet, as writing exists in a relatively permanent corporeal form, it can always be misinterpreted or reinterpreted -- as a result, the author
needs to come to the writing's defence (e.g., see Barthes, 1968).

On the other hand, speech is superior in form and function because with the art of dialectic, a speaker,

... fastening upon a suitable soul, plants and sows in it truths accompanied by knowledge. Such truths can defend themselves as well as the man who planted them; they are not sterile, but contain a seed from which fresh truths spring up in other minds; in this way they secure immortality for it... (Plato, p. 99)

The preceding quote suggests that in contrast to written words, misunderstood spoken words have a "voice" with which to defend themselves. This voice is the speaker's full presence to the communication, made possible by the temporal and spatial co-presence of the speaker and addressee. In contrast, in traditional written communications, the author is not present to the reader, and as a result, is not equally thought to interact with the reader.

Thus, traditionally, speaking is impermanent, present: synchronous and proximate (Derrida, 1972; Heidegger, 1927), and appropriate for education. On the other hand, writing is relatively permanent, and absent: asynchronous and
distant (Derrida, 1972; Heidegger, 1927), and problematic as a means of transmitting instruction. However, electronic communications media cast the orderly binary distinctions found in traditional thinking about the form and functions of speaking and writing into disarray. For example, graphic conferences can give written words an interactive voice, which can be as fleeting as speech if no one chooses to electronically save the conversation. As another example, electronic voice mail can allow for asynchronous verbal communications, that do not have an interactive voice, and that are permanently recorded. As a result, traditional conceptions about the formal features and functions of communication in speaking and writing must change to accommodate the new reality of electronic, computer-mediated communications.

Investigating Electronic Conferences

In this section, I will discuss in more detail how traditional distinctions between the formal features of speech and writing are restructured by electronic communications. In the process of more clearly articulating this restructuring, I will introduce some terms that are important to understanding the discussion in remainder of this thesis. The formal characteristics of speech and writing derived from analysis of traditional
communications were relative permanence or impermanence of communication, and relative presence or absence of communication.

As mentioned, permanence is traditionally a linear variable that distinguishes speaking from writing (Derrida, 1972). Writing, being permanent, allows information to pass through time, whereas speech, being ephemeral, is bound up with the speaker -- it comes and goes with the interlocutor. Of course, certain forms of writing are more timeless than others. For example, a classic work of literature is more enduring than a daily news article, weather report, or shopping list. However, in contrast to the traditional association of permanence only with written interactions, in electronic environments permanence is applicable to verbal communications as well as written ones.

As I mentioned in my example in the previous section about electronic voice mail, since the computer mediates communication, the potential for recording and archiving communications always exists. As a result, the underlying characteristic permanence in electronic media is not so much tied to a particular type of communication, as permanence is to most writings, but is rather an underlying
potential characteristic of all communications in electronic media.

Proximity of communication refers to spatial distance between interlocutors and is traditionally viewed as a binary variable that distinguishes speaking from writing: Speakers and listeners are proximate while writers and readers are not (Derrida, 1972). In most instances of spoken interaction, assuming interlocutors are neither deaf nor blind, proximity is associated with paralinguistic elements of communication.

One category of paralinguistic cues, one might call it the "aural" one, is conveyed through the pitch, rhythm, and intonation of the voice. Another category, which can be labelled the "visual" category of paralinguistic cues, is conveyed through visual channels, such as through gesture or facial expression. Since electronic communication can be written and synchronous, yet nonaural and nonvisual, in the sense that interlocutors cannot hear or see each other, a third category of paralinguistic cues that have emerged in graphic conferences are graphical emotion icons, or emoticons (Kiesler et al., 1984). Emoticons can be considered a "graphical" paralinguistic cue used in graphic conferences to indicate boredom, interest, or irritation, for instance, in the way that body gestures or tone of
voice might in face-to-face communication (Kiesler et al., 1984).

The linguistic and paralinguistic features that make up an electronic communication are dependent upon the formal features or bandwidth of the communication rather than the proximity of the interlocutors (Kiesler et al., 1984). In this context, bandwidth can be defined as the degree to which electronic communication allows participants to approximate the paralinguistic aspects of face-to-face interaction conventionally made possible by spatial proximity -- being able to see the other person and hear the pitch, rhythm, and intonation of his or her voice (Kiesler et al., 1984). Although interlocutors in a high bandwidth dialectical electronic communication, such as a videoconference, may be able to see and hear each other, the interlocutors need never be spatially proximate to do so. As a result, referring to what has been called a face-to-face conference as a "proximate conference" also clearly distinguishes the traditional spoken form of communication from electronic forms of communication.

Temporality of communication refers to the amount of time that separates the sending and reception of messages. Like proximity and permanence, it has historically been thought of as a binary variable that distinguishes speaking
from writing (Derrida, 1972). Speaking is usually considered to be synchronous because interlocutors interact together with each other. On the other hand, writing is considered to be asynchronous because first someone writes, and later someone reads. Thus, synchronous communication can be defined as communication that occurs above a minimum temporal threshold beyond which interlocutors can be said to be communicating with each other at the same time. In light of this definition, and in contrast to traditional thinking about communication, writing in electronic media can be as synchronous as speech.

In sum, electronic media blur traditional distinctions about the formal features of language with which speaking and writing are traditionally associated. Traditionally, speaking is impermanent, proximate, and synchronous. On the other hand, traditionally, writing is permanent, distant, and asynchronous. However, as illustrated in my example in the previous section about graphic conferences, written communications in electronic media can be synchronous, distant, and impermanent. On the other hand, as illustrated in my example about voice mail, spoken communications in electronic media can be asynchronous, distant, and permanent. In terms of function, with reference to the quotes from Plato cited in the previous
section, in electronic media, both spoken and written communications can have a voice with which to defend themselves, and can thus qualify, in Socrates' terms, as suitable for proper instruction.

**Face-to-Face and Electronic Written Conferences**

Having clarified how electronic communications media force one to rethink traditional formal and functional distinctions between writing and speaking, I now turn to empirical studies of conference communication in written electronic, or graphic, and spoken face-to-face, or proximate, conferences. Through this discussion I will continue to use the terms developed in the previous sections to represent the way that underlying formal characteristics in graphic conferences have been observed to influence communication.

Researchers have begun to address how underlying communicative features in graphic conferences seem to affect the communication process (e.g., Beauvois, 1992, 1998; Chun, 1994; Coleman, Peternite, & Sherman, 1999; Cooper & Selfe, 1990; Hayne, Rice, & Licker, 1994; Kelm, 1992; Kern, 1995; Kiesler et al., 1984; Perkins & Newman, 1999; Pinsonneault & Kraemer, 1990; Siegal, Dubrovsky, Kiesler, & McGuire, 1986). For instance, because graphic conference communication occurs above the temporal
threshold necessary for synchronous communication yet takes written form; researchers have noted that there is no need for participants to take turns because interruption is impossible (Beauvois, 1992, 1998; Kelm, 1992).

Moreover, since communication in graphic conferences does not take place in the spoken mode, aural linguistic information must be signified some other way (Kelm, 1992, Kiesler et al., 1984). As a result, in graphic conferences, reading and writing (i.e., typing) skills, as well as skills that allow participants to accurately express and interpret attitudes expressed in graphic form, through emoticons or explicit language, are central to the interaction (Beauvois, 1992, 1998; Kelm, 1992; Kiesler et al., 1984).

In addition, since the communication is graphic-only, participants lack the aural and visual paralinguistic cues that in the proximate environment nonlinguistically signify the listener's orientation to the discourse: for example, annoyance, agreement, or indifference (Kiesler et al., 1984). Moreover, conventional aural and visual paralinguistic methods of dramaturgically conveying meanings through such acts as taking the head seat, speaking loudly, staring, or gesturing, are impossible (Kiesler et al., 1984). Finally, since there is no aural
or visual paralinguistic information, it may be impossible to distinguish social status and position cues such as gender, race, ethnic background, or socioeconomic status, especially if participants use pseudonyms (Hayne et al., 1994; Marjanovic, 1999; Selfe & Meyer, 1991; Sullivan, 1998).

As a result of the different conditions underlying the communication, the skills involved in expressing one's self and interpreting others in the graphic environment are different from the skills required for communicative competence in the proximate milieu (Chun, 1994; Perkins & Newman, 1999). Communicative competence refers, in a general way, to a person's ability to communicate appropriately within the context of the culture within which the language functions, and to make himself or herself understood through verbal and nonverbal language (Celce-Murcia & Olshtain, 2000). Communicative competence in the graphic environment requires proficiency in strategies that enable participants to successfully communicate with others using a relatively limited or restricted conduit for communication and interaction -- a graphic-only one (Chun, 1994; Perkins & Newman, 1999).

The words limited and restricted are not meant to imply that communication in the graphic environment is
inherently inferior, in the sense that some ideas cannot be communicated at all. However, these words are used to suggest that information in proximate communication that is ordinarily communicated through aural or visual paralinguistic cues; attitudes like boredom, curiosity, or agreement for instance, must be expressed in explicit language or with graphic paralinguistic cues such as emoticons (Kiesler et al., 1984). In other words, the content of the communication is not necessarily restricted in graphic conferences, but the bandwidth for the communication is. The terms limited or restricted are also misleading because, as shall become evident in a moment, especially for second language students, communication in the graphic medium seems to be opened up in comparison with communication in the proximate milieu (Beauvois, 1992, 1998; Chun, 1994; Cooper & Selfe 1990; Neuwirth et al., 1993; Sullivan, 1998; Warschauer, 1996; Warschauer et al., 1996).

The literature has begun to paint a picture of how the formal characteristics underlying communication in graphic conferences seem to shape the patterns of teacher and student participation in ways that may be beneficial to second language students engaged in an idea-generation conference (Beauvois, 1992, 1998; Braine, 1997, 2001; Chun,

For example, Chun (1994) found that second language students interacted directly with each other rather than mainly with the teacher and that students took the initiative, constructed and expanded on topics, and took a more active role in discourse management in comparison with students in proximate classroom discussions.

Taking an active role in controlling the discourse through speech acts such as expanding on topics, taking turns, giving feedback, capturing attention, steering or avoiding topics, starting and ending conversations, requesting confirmation or clarification, and apologizing are components of communicative competence known as interactive competence (Chun, 1994). Chun's conclusion that second language students display more interactive competence in the graphic medium is consistent with the findings of other researchers (e.g., Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996). For example, Slatin (1998) noted that even at the initial stages of foreign language learning, students who participate in graphic discussions make more of these types of higher order communications, or in other words, communications that demonstrate interactive competence.
Researchers' observations of student participation in graphic conferences have led many to suggest that the environment facilitates communication for second language students. As an example, Marjanovic (1999) and Sullivan (1998) have observed that the anonymity of the network can empower second language students to participate more fully in discussions with native English speakers. They report that second language students in their studies communicated more because they found themselves defined by their communicative skills in a particular type of discussion, rather than by their race or native language background. By obscuring names, as well as aural and visual paralinguistic cues, the graphic conference environment can facilitate anonymous interaction, thus removing a disadvantage that minority students often face when attempting to get included into a discussion with native speaker participants (Marjanovic, 1999; Sullivan, 1998).

In conference groups that did not contain native speakers, and that did not use pseudonyms, others (e.g., Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995) have also reported more balanced participation in graphic as opposed to proximate conferences. However, the evidence for more balanced participation in these studies, although
quantitative, was not subjected to analysis that could demonstrate statistical significance.

Rather than attributing the more balanced participation to anonymity associated with obscured social status and position cues, Kelm (1992) and Beauvois (1992, 1998) proposed that more balanced participation may have been due to the lack of need for consecutive turn-taking patterns. Students in graphic conferences need not wait for others to finish writing before beginning to express language, but can begin to do so even while other participants are writing. Thus, assuming the pace of the graphic conference is not too overwhelming, students in graphic conferences may have more time to compose and evaluate their contributions in the midst of interaction, thus tending to reduce the fear of interruption or failure often experienced by less orally fluent students during an oral exercise (Beauvois, 1992, 1998; Kelm, 1992).

Like the studies just mentioned, Warschauer (1996) also noted more balanced participation in graphic conferences compared to proximate conferences. In fact, by using the gini coefficient (i.e., a measure of inequality; see Appendix K), he was able to demonstrate that, on average, participation was more balanced (i.e., less inequality) for his small groups of English as a second
language students in graphic compared to proximate conferences. Although Warschauer claimed that participation was twice as balanced in the graphic conferences because the numerical value of the average gini coefficients for the graphic conferences was double the numerical value of the average coefficient for the proximate conferences, one might question the meaning of this claim. In particular, he failed to provide a mathematical proof to show that gini coefficients represent a ratio scale. Moreover, as Warschauer admits, the number of students in his exploratory study was quite small, so the results can hardly been seen as conclusive.

In addition to participation being more balanced, Warschauer (1996) also noted that the discourse in graphic conferences was more lexically complex. Warschauer based this claim on a comparison of the type-token ratios of conferences. Type-token ratios are defined as the total number of unique words, or types, divided by the total number of words, or tokens (Scott, 1998; Warschauer, 1996). Thus, a higher type-token ratio indicates that a text has greater lexical complexity, or in other words, has a greater lexical range (Ortega, 1997). As Warschauer and others (e.g., Ortega, 1997) have noted, the need for more studies with groups of various sizes that include measures
such as the gini coefficient and type-token ratios would be very helpful towards constructing a theory of communication in graphic conferences. Accordingly, I include both of these measures in my study.

Since graphic conferences are generally seen as a useful way to stimulate second language students' active participation, teachers and educational researchers motivated by a social-constructivist pedagogical model (e.g., Beauvois, 1998; Peterson, 1997; Warschauer, 1996; Warschauer et al., 1996) have generally been quite enthusiastic about using graphic conferences in their second language writing classrooms. The term social-constructivist stands for a shift in educational theory from an interest in cognitive and developmental models of learning to a social and collaborative view of learning. These researchers have noted more balanced participation between second language students, albeit in most cases through anecdotal reports rather than quantitative data, and have maintained that communication in the graphic medium seems to be opened up to second language students.

Yet the communication conditions in graphic conferences that tend to encourage groups of students to participate are the same conditions that can also result in an inability of groups to share information efficiently and
reach consensus (Coleman et al., 1999; Kiesler et al., 1984; Marjanovic, 1999; Selfe & Meyer, 1991). In other words, while the lack of need to take turns, the lack of visual cues that usually signal hegemony within a group, and the increase in direct student-student interaction can allow students to express themselves more freely, these same conditions might negatively impact a group's ability to maintain an orderly and structured approach to the conversation and come to a group consensus (Hightower & Sayeed, 1995; Kiesler et al., 1984).

Although there is no real need for pre-writing discussion groups to reach consensus for ideas generated during conferences to be useful to students when writing their first drafts, collaboration is desired because it has generally been found to positively influence the number and quality of ideas generated by groups (Neuwirth et al., 1993; Pendergast & Hayne, 1999). As a result, for the activity to be considered beneficial, although participants need not be in agreement in terms of, for example, their viewpoints regarding the topic, what to include or not include in the essay, or how to most logically organize content, pre-writing activities should permit communication and collaboration about ideas among participants. More specifically, pre-writing conferences should facilitate
groups to generate language and ideas that can be transferred to subsequent individual writing (Hayes & Flower, 1986; Zamel, 1983).

**The Process of Writing**

Before beginning to discuss research into the effect of electronic conferences on student writing, a review of contemporary notions in writing research and pedagogy will help to clarify the terms of discussion and to place the forthcoming assessment of research within the broader scope of process writing theory.

Process writing theory can be seen as a critique of the earlier product approach to writing pedagogy. Teachers using the product approach generally assigned students a topic and asked them to create a composition. The teacher then corrected the writing and assigned a grade (Reid, 1993). This pedagogical approach focused on critiquing the final written product, hence: the product approach. The research focus of this approach was on effective teaching and learning environments, and writing was generally conceptualized as a bipartite task -- *think first, then write* (e.g., Braddock, Lloyd-Jones, & Schoer, 1963; Hillocks, 1986).

In contrast, the process approach places the research focus on the cognitive processes involved in writing and
the stages in which they occur. The process approach began in the expressive and cognitive schools of first language composition in the 1970s and was largely adopted by second language writing instructors in the 1980s (Reid, 1993). In the process approach, writing is conceptualized as a problem-solving activity and a process of discovery (Elbow, 1973; Hairston, 1982; Sommers, 1980; Taylor, 1981; Zamel, 1980, 1982, 1983) that consists of several cognitive processes -- planning, sentence generation, and revision -- that are used to address several goals for composition (Flower & Hayes, 1980; Hairston, 1982; Zamel, 1983).

The first stage, planning, or pre-writing, is typically associated with cognitive processes directed at gathering ideas about the topic (Reid, 1993). This can be accomplished through, for a few examples, reading about the topic, discussing or brainstorming the topic with others, outlining, cubing, clustering, or free-writing (Reid, 1993). For example, in free-writing, the writer writes continuously about whatever comes into his or her mind regarding the topic for up to 20 minutes, without being concerned about grammar, and trying not to lift pen from paper (Elbow, 1973).

Pre-writing activities are believed to be beneficial to writers because an overall lack of pre-writing planning
has been linked to the strategies of unskilled writers in studies of both first (Perl, 1979; Pianko, 1979) and second language (Raimes, 1985, 1987; Zamel, 1983) writing development. On the other hand, increased planning has been linked to more proficient first and second language writers' cognitive processes (Hayes & Flower, 1986; Zamel, 1983). As a result, pre-writing activities such as pre-writing conferences are assumed to be beneficial to novice writers in general.

Particularly for second language writing students, collaboration during the pre-writing stage is usually viewed as important. These writers, who typically have a relatively restricted vocabulary, are assumed to benefit from being given the opportunity to use the language and ideas generated by group members in their writing (Reid, 1993). As a result, collaborative pre-writing conferences are generally accepted as a particularly appropriate pre-writing activity for second language writing students.

The next stage in the writing process, associated with sentence generation or drafting processes, consists of writers making further decisions about which ideas to include and which to leave out, and deciding how ideas should be organized (Reid, 1993). At the conclusion of this stage, the writing is often fairly complete with
respect to content, although the writing may be full of grammatical and surface feature errors, such as errors in punctuation, capitalization, indenting, subtitling, or even be partially in point form (Reid, 1993; Zamel, 1983).

At the third stage, cognitive processes related to revision come to the forefront. Writers return to their drafts and rethink the inclusion of content and the pattern of organization (Faigley & Witte, 1981; Reid, 1993). Having someone else, such as a teacher or a peer writer, read and respond to the writing is often seen as an effective way to facilitate writers at the revision stage (Reid, 1993). The final part of the revision stage is sometimes referred to as editing, especially in the context of second language writing. In the editing stage, the focus shifts to correcting errors in grammar and surface features (Reid, 1993).

The final stage of the writing process is publishing. In the publishing stage, the focus is on meeting requirements of form and style such as those set out in the Publication Manual of the American Psychological Association (American Psychological Association, 2001). In the publication stage, the writing is polished and shared with its intended audience.
Although I have presented these cognitive processes as occurring in a linear series of stages, in fact, writing process theorists have generally conceived of these activities as recursive (e.g., Carey & Flower, 1989): Any process can turn up at any time in the writing process and any activity can precede or follow any other. However, as some (Schilperoord, 1996; Van Den Bergh & Rijlaarsdam, 2001) have more recently pointed out, these processes are perhaps not equally likely to occur at any time during the writing process. As the writing progresses, the task situation, or in other words, the cognitive representation of the text, changes. Accordingly, certain activities seem to be more likely to occur at some stages of the writing process than at others (Schilperoord, 1996; Van Den Bergh & Rijlaarsdam, 2001).

The cognitive processes under investigation in this study are the ones most likely to occur early in the writing process: planning and drafting. However, as it turns out, the majority of process writing research has focused on cognitive processes involved in revision, perhaps at least partially because early attempts to understand the writing process by comparing novice versus expert writers revealed the most prominent differences at the revision stage (Reid, 1993). For instance, several
studies have demonstrated that professional writers revise more frequently on a global level than novices (Faigley & Witte, 1981).

As a result, relatively few studies within the realm of process writing theory are of direct relevance to my study. The studies that are of direct relevance have generally come from researchers interested in using computer-mediated communication as a pre-writing tool in the process of writing. As has been discussed, motivating this interest is the belief that the graphic, computer-based mode of communication serves a democratizing function: tipping the balance of participation patterns towards the students and away from the teacher; facilitating balanced participation among students; and provoking active, spirited, and critical discourse.

**The Effect of Electronic Conferences on Writing**

Researchers have begun to explore how participation in graphic pre-writing conferences might benefit second language students' writing. For instance, in an exploratory study of second language students' participation in graphic pre-writing conferences, Kern (1995) claimed that increased second language student participation and direct student-to-student interaction stimulated peer learning and led to a "...significant
improvement in the depth and strength of students' arguments in assigned compositions'' (Kern, 1995, p. 470). Although this is a compelling conclusion, only anecdotal evidence was offered to support this claim.

Others have used holistic scores of drafts produced after participating in pre-writing discussions in graphic and proximate environments to quantitatively compare the value of these two types of conferences. For example, Sullivan and Pratt (1996) compared students' drafts using holistic measures. In addition, both Braine (1997, 2001) and Ghaleb (1993) used similar holistic instruments taken from the Test of Written English Scoring Guide (Educational Testing Service, 2003). Unfortunately, taken together, these studies failed to demonstrate a clear trend of effect of pre-writing environment on writing quality of either first or final drafts.

Brain (1997, 2001) and Ghaleb (1993) both reported higher scores for first drafts written after graphic conferences. Sullivan and Pratt (1996), however, reported higher scores for first drafts that followed proximate conferences. Moreover, although student writing in Brain's and Ghaleb's classes improved in quality over the course of several months, the degree of improvement for students who
used computers was less than the improvement for students who studied writing in lecture-style classes.

Ghaleb (1993) attributed the greater improvement of writing quality in lecture-style classes to the considerable amount of time spent on grammar instruction and correction of errors. This suggests that students who use electronic media in the way described in these studies may actually be at a disadvantage, compared to their counterparts in the traditional classroom, at least when it comes to producing grammatically accurate work. Nevertheless, in Sullivan and Pratt’s (1996), Ghaleb’s (1993), and Braine’s (1997, 2001) studies, papers written in networked classes, or in other words, classes in which students used the network for pre-writing discussions, as well as revision activities, improved significantly from first to final draft.

However, in these studies, several methodological issues make it difficult to determine the impact of the two types of conferences on first and final drafts. The first methodological issue involves the procedures. The second methodological issue involves the analytical instruments used. The third, and most significant methodological issue involves the lack of evidence offered to clearly demonstrate the influence of conferences on first drafts.
With regard to procedures, as previously mentioned, pre-writing activities in Sullivan and Pratt’s (1996), Ghaleb’s (1993), and Braine’s (1997, 2001) studies were followed by revision activities in the same environment. In other words, in these studies, students who participated in pre-writing activities in the graphic condition also participated in peer-review activities in the graphic condition, and vice versa. As a result, both the conference condition and the peer-review condition were assumed to impact the quality of final drafts. Thus, it is difficult to tease out, specifically, the impact of pre-writing conference condition on final drafts.

Second, as mentioned, Sullivan and Pratt (1996), Ghaleb (1993), and Braine (1997, 2001) evaluated first drafts with holistic scales. However, holistic scales are ordinarily reserved for the evaluation of final drafts because holistic scales evaluate writing in terms of features not normally incorporated into first drafts. For instance, holistic scales award higher scores for features such as high level discourse organization, and grammatical accuracy and complexity (Educational Testing Service, 2003). However, contemporary writing teachers ordinarily instruct students writing first drafts merely to get ideas on paper and to begin to narrow and organize them (Reid,
1993). In other words, according to process writing theory, when writing first drafts, students should primarily be concerned with the inclusion of appropriate content, and should not be overly concerned with grammar or organization (Elbow, 1973; Reid, 1993; Zamel, 1982, 1983). As a result, one must question if these holistic instruments were appropriate for evaluating first drafts; and indeed, one must question the finding that first drafts written following one type of conference were superior to first drafts written following the other.

Moreover, since holistic measures award higher scores for features that are essential in final drafts, but not particularly important in first drafts, Sullivan and Pratt's (1996), Ghaleb's (1993), and Braine's (1997, 2001) findings that writing improved from first to final draft following either type of conference was most likely an artefact of the way they assessed writing quality. In other words, given the analytical instruments used, the finding that writing improved from first to final drafts is unsurprising enough to be considered trivial.

In light of process writing theory, it seems reasonable to argue that an analytical instrument suitable for assessing the quality of first drafts should focus on the inclusion of content. Moreover, since teachers do not
expect content in first drafts to be particularly well organized (Elbow, 1973; Reid, 1993; Zamel, 1982, 1983), it also seems reasonable to argue that the measurement of content should be derived from analysis of small bits of discourse, such as individual sentences, or even from lexis, rather than from analysis of the extended discourse of texts taken as cohesive wholes.

Accordingly, I chose to assess the content of first drafts by analyzing t-units, rather than by analyzing texts holistically. Hunt (1965) originally proposed a concept of the t-unit (i.e., minimally terminable unit), and defined it structurally, as a main clause including all subordinate clauses and other constructions that go with it. Hunt developed the t-unit as a way of measuring the development of syntactic complexity in children's writing. In general, the greater the number of words in a t-unit, the greater the syntactic complexity. Hunt could not use sentences as the boundary node and simply count the number of words in sentences because, like second language writers, children do not always accurately segment their texts into sentences.

Since Hunt's (1965) purpose was to gauge the syntactic complexity of well-formed t-units, a structural definition was sufficient for him. However, in my study, I used t-
units as a way of dividing writing into individual ideas. Consequently, I required a definition of the t-unit that conceptualizes t-units in psycholinguistic terms, rather than in purely structural terms. Accordingly, I used the Keep It Simple, Stupid (hereafter: KISS) definition of the t-unit (Vavra, 2000), which is based on Hunt's original conception of the t-unit, with some modifications. In practice, the main difference between Hunt's definition and Vavra's definition is that while Hunt discarded sentence fragments (i.e., incomplete clauses) and garbles (i.e., incomprehensible statements) from analysis, Vavra retained them. Vavra writes,

As with garbles, fragments represent concepts in short-term memory. An orthographic fragment counts as a t-unit. The assumption here is that an orthographic fragment is the result of an overloading of the writer's short-term memory. The overloading causes the writer to put down an ending punctuation mark or end their contribution and begin a new "sentence" or contribution [italics added]... A student's average words per main clause (t-unit) is thus a reflection of the average number of words that the writer can
process in short-term memory before dumping to long-term memory. (Vavra, 2000)

In other words, fragments and garbles can and should be counted as individual t-units if t-units are conceptualized in psycholinguistic terms as complete thoughts. For example, since a fragment such as agree represents a student’s complete thought, it should be counted as a t-unit. In sum, since my focus is not on t-unit complexity, emphasizing linguistic skill, but rather on generating semantic propositions, emphasizing idea formation, Vavra’s (2000) KISS psycholinguistic definition of the t-unit is a better fit.

One of the advantages of dividing texts into t-units is the reliability of the analysis. Hunt (1965, p. 22) writes, “Any two grammarians should be able to agree on an analysis, whether their denominations be transformational or structural or traditional.” Indeed, given the simplicity and assumed reliability of identifying t-units, studies that employ this text analysis measure often do not use a second rater, nor report estimates of inter-rater reliability (e.g., Schneider & Connor, 1990). Moreover, finding volunteers willing to perform t-unit analysis can be exceedingly difficult, particularly when volunteers are
required to analyze a large amount of text without remuneration. When estimates of inter-rater reliability are reported, they tend to be very high (e.g., 94%; Reynolds, 2001). For these reasons, I did not use a second rater for the t-unit analysis in my study.

To return to my critique of Sullivan and Pratt's (1996), Ghaleb's (1993), and Braine's (1997, 2001) studies, finally, and perhaps most dubiously, none of these studies offered any direct evidence that the language and ideas generated in either type of conference were indeed transferred to drafts. In these studies, higher holistic scores on first and final drafts were merely assumed to reflect the relative superiority of conducting conferences in one environment compared to the other. As a result, the claim that conferences had an influence on drafts was not evaluated with a test that might tend to falsify it. In other words, since none of these studies offered any evidence to show that the same ideas generated in conferences were also used in drafts, the methodology of these studies begs the question: Did either type of conference have any influence on drafts at all, much less a relatively superior or inferior influence?

Although I was also concerned with comparing the usefulness of conferences to students' subsequent writing,
because of the methodological issue just mentioned, I did not attempt to gauge the usefulness of conferences to drafts based on some measure of structural or conceptual quality in drafts. Instead, I focused on directly comparing the usefulness of conferences to drafts by comparing how much of the language and ideas generated in the two types of conferences were transferred to the drafts.

**Measuring the Relation of Conferences and Drafts**

Typically, to measure how many of the ideas generated in a conference were transferred to a draft, an experimenter would need to examine each sentence or t-unit in the conference and the draft and then attempt to match the information contained in the draft to the information contained in the conference. More generally, such an analytical model assumes that if the terms and ideas in the draft are similar to the terms and ideas in the conference, then the conference influenced the draft.

Discovering how many words two or more documents have in common is a rather effortless and exceedingly accurate and reliable undertaking, thanks to recent developments in the application of computer technology to applied linguistics. The application of computer technology to linguistic analysis, and the collection of computer
programs that make such application possible are generally grouped under the appellation Corpus Linguistics (Nelson, 2000; Scott, 1997; 1998). I used the Wordlist component of the computer application Wordsmith Tools (Scott, 1998) to determine how many words drafts had in common with the conference that preceded them.

The matching of ideas between two or more documents, however, is a considerably more problematic matter that has been approached in several ways. First, sentences have been linked semantically to each other based on sharing identical or similar lexical items. For instance, Hoey (1991) has argued that when sentences contain a sufficient amount of lexical overlap with each other, they also tend to be semantically related. Lexical overlap for Hoey does not mean only exact word overlap. Hoey divides lexical overlap into two categories: simple lexical overlap and complex lexical overlap. Simple lexical overlap includes exact word repetition as well as variants of the same root word (e.g., the same word in noun, adjective, or adverb form: happiness, happy, and happily). Complex lexical overlap includes synonyms (e.g., happy and glad), hyponyms (e.g., school and building), and phrasal definitions (e.g., "classroom" and "the place where we have our classes").
Hoey's (1991) analytical method is not without technical hitches. First of all, nonlexical cohesive devices, such as referential pronouns, must first be replaced by the content words to which they refer. This, in itself, can become problematic when words such as this can be used to replace a large part of the preceding sentence, as this at the beginning of this sentence was used to do. Moreover, particularly in the context of second language writing, such a method requires a good deal of subjective analysis when deciding which lexical items should be linked. For instance, should the term cloning be considered a hyponym of genetic manipulation? Furthermore, should the researcher's opinion prevail over the student writer's?

Aside from this, in a study like mine in which a rather large number of texts are subjected to analysis, matching information in this way would be immensely time consuming. Indeed, research using this sort of lexical cohesion analysis has typically used a rather limited number of very short texts (e.g., Reynolds, 2001).

Alternatively, propositional analysis (e.g., Bovair & Kieras, 1984; Turner & Green, 1978) has been used to provide a set of semantic primitives that describe the information contained in a text. Nevertheless, again, the
amount of time and effort involved in categorizing the propositions contained in texts according to their semantic content has typically limited research using this method of analysis to a few texts of less than 1,000 words in length (Foltz, 1996). Unfortunately, computerized methods of identifying and categorizing semantic propositions are not currently feasible since they would require the computer to accurately parse the text and interpret the correct meanings of all the words (Foltz, 1996).

The primary advantage of the two methods of comparing texts just discussed is that textual comparisons are based on semantic similarity rather than on surface features such as choice of words alone. However, in both cases the primary disadvantages are that the techniques require subjective judgement and are not feasible for the analysis of large amounts of text. To overcome these obstacles, in this study I used an automated semantic matching technique called Latent Semantic Analysis (LSA; Laundauer & Dumais, 1997).

**Latent Semantic Analysis**

Latent Semantic Analysis is a corpus-based statistical model of word usage that allows researchers to make comparisons of semantic similarity between pieces of textual information. LSA is an outgrowth of Latent
Semantic Indexing (LSI), which was first developed at Bellcore in the late 1980s (Yu, Cuadrado, Ceglowski, & Payne, 2002). The original purpose of LSI was to improve search word-based information retrieval systems. In recent years, LSA has demonstrated remarkably human-like abilities in a variety of language tasks. For example, LSA has taken the Test of English as a Foreign Language and performed as well as second language students who were successful college applicants (Laundauer, Laham, Rehder, & Schreiner, 1997). It has shown an ability to learn words at a rate similar to humans (Laundauer & Dumais, 1997). It has even graded papers as reliably as human graders (Foltz, Britt, & Perfetti, 1994). As mentioned, I used LSA to match the semantic content of conferences and subsequent first drafts. Although an in-depth understanding of the mathematics behind LSA is not necessary for the reader to understand and interpret the LSA measures used in this study, a qualitative overview of the technique follows.

LSA can be divided into two main stages. The first main stage consists of four substeps which all involve the creation of a semantic space. In the second main stage, the semantic space is used to compare the semantic similarity of texts. In general, a semantic space can either be thought of as a matrix on paper with many rows...
and columns, or as a geometric space with many dimensions. These descriptions of "semantic space" will become clearer as this discussion progresses, particularly in the section on singular value decomposition. For now, the reader need only recognize that a semantic space is eventually used, in the second main stage, as a basis for comparing the semantic similarity of texts.

The four substeps of the first main stage are: 1) selection of a corpus of texts, 2) creation of a term-document matrix from the selected corpus, 3) information theoretic weighting of the term-document matrix, and 4) singular value decomposition (hereafter: SVD) of the term-document matrix. As mentioned, the second main stage in LSA is the actual comparison of texts using the semantic space created in the first main stage.

Creating a Semantic Space

The first main stage in LSA involves the creation of a matrix that will be used for the comparison of texts in the second main stage. To put it another way, the overall purpose of the first main stage is to create a matrix that represents a semantic space that contains information about words that relate to a given topic, and about how words and documents related to the topic relate to each other. To
clarify, each of the four substeps of the first stage of LSA will be discussed in turn.

Selecting a corpus of texts. The choice of the collection of texts, or corpus, used in LSA depends on the content of the writing that will be compared in the second main stage. The corpus of texts should relate semantically to the texts that will eventually be compared. An example of the use of LSA will help to clarify why.

Foltz, Britt, and Perfetti (1994) used LSA to determine where student writers had found the information they wrote in summaries about the events leading up to the construction of the Panama Canal. In preparation for writing summaries, students read 21 texts that included excerpts from textbooks, historians' and participants' accounts, and primary documents such as treaties and telegrams. To determine where the information that students wrote had come from, the sentences that students wrote in their summaries were compared to the sentences that students had initially read, to see which were most similar.

The semantic space used in LSA to compare texts is dependent on having many examples of the co-occurrences of words related to the topic. As a result, for their corpus, Foltz et al. (1994) used the 21 texts the subjects read,
along with paragraphs from eight encyclopaedia articles on the Panama Canal and excerpts from two books. The addition of these other textual materials helped to provide LSA with additional examples of words related to the Panama Canal to help define the semantics of the domain (Foltz, 1996).

As discussed, Foltz et al. (1994) made comparisons between students' texts, and articles from the corpus that were used in the construction of the semantic space. However, this does not mean that the semantic space used in LSA must always contain some of the texts that will eventually be compared. For instance, this same semantic space could have later been used to compare the semantic similarity of two essays written by two different students about the creation of the Panama Canal. To emphasize, the articles selected in the first main stage for the corpus used to create the semantic space are used to define the semantics of the topic under investigation — in this case, the creation of the Panama Canal. These articles do not necessarily have to be among the articles that will be compared in the second main stage.

In my study, I required a semantic space that would allow me to analyze several different topics of general interest to the students in my study. As a result, the corpus for the semantic space that I used consisted of a
wide variety of articles on various topics. The semantic space that I used, entitled "general reading up to first year college" is freely available on the University of Colorado at Boulder Web site. A description of the corpus used to create this semantic space along with the URL and details regarding how the Web site was used can be found in Appendix B. This semantic space was chosen because it contains general reading materials related to the topics discussed by students in this study and because it was believed to reflect the level of vocabulary acquired by the students in my study.

The term-document matrix. After an appropriate corpus of texts has been selected, a single matrix of occurrences of each word by each document of all the texts in the corpus is created. A "document" can be, for instance, an entire paper, a paragraph, a sentence, or a t-unit. This matrix is called a term-document matrix because each column in this matrix represents each document in the corpus and each row represents each word in the corpus (Yu et al., 2002). Each cell in the matrix contains the number of times that each word occurred in each document (Laundauer & Dumais, 1997). For an example of a very small term-document matrix, see Table 1.
Table 1

Example Term-Document Matrix 1

<table>
<thead>
<tr>
<th>Term</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>D8</th>
<th>D9</th>
<th>D10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oedipus</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Antigone</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note: in this table, D1 stands for Document 1; D2 stands for Document 2, etc.*
Creating this matrix yields a grid containing mostly zeroes with a sparse scattering of positive non-zero integer entries. This grid represents everything that LSA "knows" about the corpus. One could list all the words in any given document by looking for non-zero entries in the appropriate column, or one could find all the documents containing a certain word by looking across the appropriate row. In its current manifestation, LSA does not make use of word-order information (Laundauer & Dumais, 1997; Laundauer, Foltz, & Laham, 1998; Laundauer et al., 1997; Wiemer-Hastings, 1999).

In general, LSA relates words by exploiting a property of natural language, namely that words with similar meaning tend to occur either together, or in similar contexts. As a result, LSA relates words in two ways. First, if two words appear together in many documents, the words are semantically related (Laundauer & Dumais, 1997; Laundauer et al., 1998; Wiemer-Hastings, 1999; Yu et al., 2002). Second, if two words both appear in the context of other words, the words are semantically related (Laundauer & Dumais, 1997; Laundauer et al., 1998; Wiemer-Hastings, 1999; Yu et al., 2002). Before moving into my discussion of the next substep of LSA, it is worth pausing to explain
in more detail how and why semantic relation is established by LSA as described in this paragraph.

First, I reported that if two words often co-occur together in many documents, the words are semantically related. With respect to the term-document matrix, this means that if two rows mostly have non-zero entries in the same columns, and zero entries in the same columns, the words represented by those rows are semantically related.

An example of a corpus compiled to include many texts about the writings of Sophocles will help to explain why LSA relates such words. In such a corpus, one might expect that the words Antigone and Oedipus would co-occur together in several of the same documents, since in one of Sophocles' plays, Oedipus Rex, Oedipus had a faithful daughter named Antigone. One might also expect that neither Antigone nor Oedipus would occur in other documents that did not deal with this play, and that either Antigone or Oedipus would less frequently occur alone in some documents (e.g., see Table 1).

In Table 1, when either Oedipus or Antigone occurred, they occurred in the same document 5 out of 8 times. When either Oedipus or Antigone did not occur, they did not occur in the same document 2 out of 5 times. There were only 3 cases out of 10 where one of these words occurred
without the other. Using all this information about the mutual occurrence and non-occurrence of *Oedipus* and *Antigone*, LSA would relate these terms to each other (Foltz, 1996).

Second, I mentioned that if two words often appear together in the context of other words, the two words are semantically related. To put it another way, LSA associates words that occur in similar contexts in different documents (Skoyles, 2002). For clarity of the following explanation, I shall momentarily narrowly define documents as sentences.

In terms of the matrix, this means that LSA relates two words even though they do not have non-zero entries in any of the same columns representing the sentences. In such a case, LSA would relate the two words to each other because they often appear in columns that have matching non-zero entries for many of the other rows. This case is illustrated in Table 2 for the two sentences: (1) "Oedipus solved the riddle of the Sphinx" and (2) "Oedipus solved the puzzle put to him by the Sphinx." In Table 2, for reasons that will become clearer in my discussion of information theoretic weighting, I have only included content words, or, in the case of these sentences, words that are not prepositions, articles, or pronouns. For the
Table 2

Example Term-Document Matrix 2

<table>
<thead>
<tr>
<th>Term</th>
<th>Sentence 1</th>
<th>Sentence 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oedipus</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Solved</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Riddle</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Sphinx</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Puzzle</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
above sentences, LSA would associate riddle with puzzle because both words occur in sentences that contain other words such as Oedipus, solved, and Sphinx (e.g., see Table 2). In sum, since LSA can identify the semantic similarity between different words, matching can be done between two texts even if they have no words in common (Foltz, 1996).

**Information theoretic weighting.** I mentioned in my discussion about Table 2 that I had eliminated non-content words for reasons that would become clearer in my discussion of information theoretic weighting of the term-document matrix. In fact, the information theoretic weighting step serves a similar function to eliminating non-content words from the term-document matrix, although in the case of LSA, the process is completely automated (Wiemer-Hastings, 1999).

Non-content words such as *the* are quite common in English, and thus occur in many sentences. In terms of the matrix, this means that the rows for non-content words like *the* would have non-zero entries in many of the columns representing the sentences. However, since non-content words are not very important in determining the semantics of a sentence, this weighting step mathematically reduces the numerical value of each cell entry for rows with very frequent non-zero cell entries to a very small number
approaching 0. In this way, the information theoretic weighting step reduces the importance of words like the in defining the semantics of the sentences in which they occur. Correspondingly, this step increases the influence of words that occur in a small number of sentences in defining the semantics of the sentences in which they occur (Laundauer et al., 1998; Wiemer-Hastings, 1999). An explanation of how reducing the numerical cell entries associated with non-content words reduces the significance of these words in defining the semantics of a sentence is reserved for my discussion of SVD.

The information theoretic weighting step is important because of the way LSA matches words. In light of the first way LSA matches words discussed above, without information theoretic weighting, LSA would tend to speciously associate many words with the because the consistently co-occurs with many other words. With respect to the second way LSA matches words, because many words occur in the context of sentences containing the word the, without this weighting step, many words would be erroneously matched to each other based on this similar context.

*Singular value decomposition.* In the previous discussion, I described how LSA relates words to each
other. This relating is further enhanced by a matrix transformation called SVD. Although the mathematics of the SVD transformation are beyond the scope of this discussion (for a rigorous treatment, see Huang, 1998), one can get an intuitive grasp of what SVD does by thinking of the term-document matrix spatially, or geometrically. An analogy taken from Yu et al. (2002) will help.

One might be curious about what people typically order for breakfast at a local diner, and one might want to display this information in visual form. One might decide to examine all the breakfast orders on a Saturday morning, and record how many times the words bacon, eggs, and coffee occurred in each breakfast order.

One could graph the results of the survey by setting up a chart with three orthogonal axes -- one for each keyword. The choice of direction is arbitrary -- perhaps a bacon axis in the x direction, an eggs axis in the y direction, and a coffee axis in the z direction. To plot a particular breakfast order, one would count the occurrences of each keyword, and then take the appropriate number of steps along the axis for that word. When finished, one would have a cloud of points in three-dimensional space, representing all of that day's breakfast orders (e.g., see Figure 1).
Figure 1. Breakfast in three-dimensional space.

Figure adapted from Yu et al., 2002
If one drew a line from the origin of the graph to each of these points, one would obtain a set of document vectors in \textit{bacon-eggs-and-coffee} space. The length and direction of each document vector describes how many of the three key items were in any particular breakfast order, and the place where the points clump together in this space would tell one something about the kind of breakfast people favour at this restaurant on a Saturday morning.

What the graph in Figure 1 shows is called a \textit{term} space. Each breakfast order forms a document vector in this space, with its direction and magnitude determined by how many times the three keywords appeared in it. Each keyword corresponds to a separate spatial direction, perpendicular to all the others. Because this example uses three keywords, the resulting term space has three dimensions, making it possible for human beings to visualize it. It is easy to imagine that this space could have any number of dimensions; depending on how many keywords one chose to use. If one were to go back through the orders and also record occurrences of sausage, muffin, and bagel, one would end up with a six-dimensional term space, and six-dimensional document vectors.

Applying this procedure to a real document collection, where one notes each use of every word, results in a term
space with many thousands of dimensions. Again, mutually orthogonal axes represent the terms, and a single vector that equals the sum of the vectors for the words it contains represents each document. Although one cannot possibly visualize such a hyperspace, it is built in the exact same way as the whimsical breakfast space just described. Documents in such a space that have many words in common will have vectors that are close to parallel, while documents with few or no shared words will have vectors that are close to orthogonal (Yu et al., 2002).

I mentioned in my discussion of information theoretic weighting that in this section I would explain how reducing the numerical values associated with non-content words in the term-document matrix reduces the significance of such words in defining the semantics of a document. At this point, the explanation should be clear. Although non-content words still have rows in the term-document matrix, or dimensions in the term space, since the numerical values associated with non-content words are reduced by the information theoretic weighting, the vector for a document containing a non-content word will not move far along the axis that represents the non-content word. As a result, the non-content word will not have much effect on the overall direction or length of the vector representing the
document. When the information-theoretic weighting step is added to the creation of a term-document matrix, rather than being called a term space, the spatial representation is called a document space (Foltz, personal communication, May 1, 2002).

So far I have described, through analogy, a spatial representation of the term-document matrix. I shall now move on to an explanation of SVD. LSA matches terms semantically by projecting the large, multidimensional term-document matrix down into a smaller number of dimensions. In reducing the matrix, document vectors that are similar get squeezed together, and are no longer distinct. Another analogy taken from Yu et al. (2002) will help to explain how this takes place.

One might want to submit a picture of a prize aquarium to Modern Aquaria magazine. To get the best possible picture, one would want to choose a good angle from which to take the photo. One might want to make sure that as many of the fish as possible are visible in the picture, without being hidden by other fish in the foreground. One also would not want the fish all bunched together in a clump, but rather shot from an angle that shows them nicely distributed in the water. Since the tank is transparent on all sides, one could take a variety of pictures from above,
below, and from all around the aquarium, and select the best one.

In mathematical terms, one is looking for an optimal mapping of points in three-dimensional space onto a two-dimensional plane (Yu et al., 2002). Optimal can mean many things -- in this case it means aesthetically pleasing. But one's goal might also be to preserve the relative distance between the fish as much as possible, so that fish on opposite sides of the tank do not get superimposed in the photograph to look like they are right next to each other. Here one would be doing exactly what the SVD algorithm does with a much higher dimensional space (Yu et al., 2002).

Instead of mapping a three-dimensional space onto a two-dimensional space, however, the SVD algorithm goes to much greater extremes. A typical term-document matrix might have tens of thousands of dimensions, and be projected down onto 300 dimensions. Nevertheless, the principle is exactly the same. The SVD algorithm preserves as much information as possible about the relative distances between the document vectors, while collapsing them down into a much smaller set of dimensions (Yu et al., 2002). In this collapse, information is lost, and words are superimposed on one another; in geometric terms, words
that once had separate orthogonal axes now have very similar or identical axes. In sum, if two terms often co-occur or are often used in similar contexts, they will have similar axes, and thus similar vector directions in the reduced dimensional representation (Foltz, 1996).

Information loss may seem like a bad thing, but here it is a benefit. What one loses is noise from the original term-document matrix, revealing similarities that were latent in the corpus -- hence the term Latent Semantic Analysis. Similar things become more similar, while dissimilar things remain distinct. This reductive mapping is what gives LSA its seemingly intelligent behaviour of being able to correlate semantically related terms (Laundauer & Dumais, 1997; Laundauer et al., 1997; Laundauer et al., 1998; Skoyles, 2002; Wiemer-Hastings, 1999; Yu et al., 2002). Again, all LSA really does is exploit a property of natural language, namely that words with similar meaning tend to occur either together, or in similar contexts.

For instance, to return to the examples I presented above, one might suppose that in one of the articles in the corpus used to create the semantic space about Sophocles, Oedipus and Freud co-occurred in a document that discussed Freud's use of the story of Oedipus as an eponym for one of
his stages of psychosexual development. Since Oedipus and Freud only co-occurred in one document, their axes would still have distant or orthogonal directions after the dimensional reduction. To put it another way, in performing the dimensional reduction, the SVD step would map Oedipus and Freud onto distant or orthogonal axes, keeping the association between the terms Oedipus and Freud in the semantic space rather distant.

On the other hand, since Oedipus and Antigone co-occurred in many documents, their dimensionally reduced vector representations would have similar or identical directions. In other words, in performing the dimensional reduction, the SVD step would map Oedipus and Antigone onto the same or similar axis, thus geometrically associating these terms in the semantic space. Likewise, since the terms puzzle and riddle often occurred in similar contexts, SVD would collapse their vector representations onto similar or identical axes. In this way, the SVD step serves to capture regularities in the patterns of co-occurrence data while filtering out less common co-occurrences (Foltz, 1996).

Comparing Documents in Semantic Space

As I mentioned in the first part of my discussion of LSA, after the semantic space has been created, it can be
used to compare the semantic similarity of other documents related to the corpus used to create the semantic space. I shall now discuss in more detail how this is accomplished.

In document space, the direction of the vector for a given document such as a t-unit, a sentence, a paragraph, or an entire text, is equal to the sum of the vectors of the terms it contains, adjusted by the information theoretic weighting (Skoyles, 2002, Wiemer-Hastings, 1999). Semantically similar documents, or documents that use the same or semantically related words, will have close to parallel vectors; while dissimilar documents will have close to orthogonal vectors. As a result, to determine if two documents are semantically similar, all one need do is determine the direction of each of the document vectors in document space, and compute the cosine of the angle between them (Laundauer & Dumais, 1997; Laundauer et al., 1998).

As the angle between two vectors decreases to 0, the cosine of the angle approaches 1. Similarly, as the angle between two vectors approaches 180 degrees, the cosine approaches -1. Conversely, as the angle between two vectors approaches 90 degrees or 270 degrees, the cosine approaches 0. Thus, mathematically, the value of a cosine can vary between -1 and 1. However, in practice, angles between vectors in document space are almost never obtuse; and
therefore, cosines computed by LSA generally vary between 0 and 1 (Laundauer & Dumais, 1997; Laundauer et al., 1998).

Cosines close to 0 signify vectors that are close to orthogonal, with little or no semantic similarity. On the other hand, cosines close to 1 signify vectors that are close to parallel, with a high degree of semantic similarity. A cosine of exactly 1 would only occur if the two documents were identical; while a cosine of exactly 0, or slightly less, would only occur if one or both of the documents mostly contained terms that did not exist in the corpus used to create the document space (Laundauer & Dumais, 1997; Laundauer et al., 1998).

One more useful piece of information about documents can be gleaned from LSA. As I discussed, the length of the vector for a given document in document space is equal to the sum of the vectors of the terms it contains, adjusted for length by the information theoretic weighting (Skoyles, 2002, Wiemer-Hastings, 1999). Thus, the length of a document's vector is a function of the number of content terms in the document, and the distance in document space between the content terms contained in the document (Laundauer et al., 1997; Laundauer et al., 1998).

After the dimensional reduction mapping performed by the SVD, terms that are similar, as was defined above, will
lie along similar axes, while terms that are different will lie along close to orthogonal axes. As a result, if, for example, a sentence contains terms that were often used together or in similar contexts in the documents used to create the original term-document matrix, the vector for that sentence will travel along an almost straight path through the dimensionally reduced document space. On the other hand, if a sentence contains terms that were rarely or never used together or in similar contexts in the documents used to create the original term-document matrix, the vector for that sentence will travel between distant locations in the document space, resulting in a longer document vector. Thus, the length of a document’s vector is a measure of the number of terms, and the semantic diversity, or breadth, of terms contained in the document (Laundauer et al., 1997; Laundauer et al., 1998).

As with the semantic matching techniques discussed in the section on measuring the relation of conferences to drafts, one of the principle advantages of LSA is that comparisons of textual material are not dependent on surface features such as the choice of words alone. LSA’s comparisons are based on a derived semantic relatedness measure that reflects semantic similarity among synonyms, antonyms, hyponyms, and other words that tend to be used in
similar contexts (Foltz, Kintsch, & Laundauer, 1998).
Moreover, unlike the previously discussed semantic matching techniques, LSA requires no subjective judgement; and since it is completely automated, it is suitable for analyzing large amounts of text (Skoyles, 2002).

**Validity and Reliability of LSA**

Technically, the validity of LSA is dependent on two main factors. First, as discussed, validity depends on choosing a suitable corpus of texts to create the semantic space for the textual material to be analyzed. Second, validity depends on choosing a suitable number of dimensions for the SVD (Foltz, 1996; Laundauer & Dumais, 1997, Wiemer-Hastings, 1999). Both of these factors influence the ability of LSA to match semantic information.

Pertaining to the second point, if too few dimensions are used in the SVD dimensional reduction mapping, invalid associations between terms are created; if too many dimensions are used, not enough associations are created to capture the latent semantic relationships between words. There is a general consensus in the literature that SVD to 300 dimensions is a suitable reduction for representing human linguistic knowledge (Foltz, Gilliam, & Kendall, 2000; Laundauer & Dumais, 1997; Skoyles, 2002).
Reports of validity for LSA offered in the Humanities and Social Sciences literature have principally been empirical. In general, validity is established by comparing the results of LSA semantic matching to similar judgements by human raters (e.g., Butcher & Kintsch, 2001; Foltz, et al., 2000; Wiemer-Hastings, 1999). Such reports are similar to my use of Appendices C, D, E, and F to show that the LSA measures seem to be doing what they are supposed to be doing in terms of sorting textual material given my choices of dimensionality and semantic space. For a more detailed explanation of Appendices C, D, E, and F, see the appendices themselves, and the research design and instrumentation section in chapter 3 of this study.

Because the validity of LSA has typically been established empirically by comparison with human graders, the validity of LSA measures is tied to inter-rater reliability. An example will help to clarify. LSA has been used to grade student papers (Foltz et al., 2000). To do so, Foltz et al. awarded each sentence in each student's paper with a cosine equal to the maximum cosine it reached with sentences in an expert's paper. Then, all these maximum cosines for each of the student's sentences were averaged to yield a single score: an average maximum cosine. Student papers with higher average maximum cosines
were deemed to be more semantically related to the expert’s paper, and thus were awarded a higher mark (Foltz et al., 2000).

Judgements of writing quality by LSA have been compared to judgements by expert human raters, and LSA has achieved scores of inter-rater reliability to human raters as high (Foltz et al., 2000; Wiemer-Hastings, 1999) or higher (Butcher & Kintsch, 2001) than the human raters achieved with each other. Of course, a human rater might not give the same paper the same mark, if, for instance, a human rater were asked to rate a paper, and then to rate the identical paper 3 years later. However, such test-retest reliability would always be perfect for LSA, assuming that the same semantic space and expert paper were used. Because LSA is as reliable as human graders, it is inferred to be measuring the same thing as human graders, and therefore it is assumed to be valid for the intended purpose.

Analytical Measures Derived from LSA

Two analytical measures derived from LSA have been used in previous research. I introduce these measures as part of my review of the literature; however, a detailed description of the LSA measures I used in my study can be
found in the research design and instrumentation section of the next chapter.

As mentioned in my description in the previous section of Foltz et al.'s (2000) use of LSA to grade student papers, one analytical measure derived from LSA involves the calculation of average maximum cosines. A draft with a high average maximum cosine was taken to have a high degree of semantic similarity with the expert's paper.

Another measure that has been used in LSA analysis is the vector length of a text (Foltz et al., 2000). Rather than using the cosine between two vectors, the length of a single vector that represents the entire text is computed. In other words, unlike average maximum cosine, vector length does not compare, for example, essays against other texts. In contrast, the vector length of a document indicates the amount and breadth of information conveyed in the document. The vector of a document takes into account the number of words in the document, the semantic relation between the words, and whether the words were low frequency in the original corpus used to create the semantic space (Foltz et al., 2000). As was discussed in the section on comparing documents in semantic space, if a document uses semantically dissimilar content words, it will traverse
much of the semantic space created by the SVD, and therefore have a long vector.

**Summary of Research Direction**

To conclude, I will summarize how my review of the literature has led me to several research directions that provide the rationale for my current study.

First, as Ortega (1997) and Warschauer (1996) have suggested as a direction for future research, I wish to compare the textual features of the conferences. In particular, I wish to determine if students will tend to generate more discourse in one of the two conference conditions given equivalent amounts of time. I also want to know if conference discourse will differ in terms of the length of statements, or more precisely, the number of words per t-unit. Moreover, I wish to investigate if discourse will be more lexically complex in one of the two types of conferences.

Second, I wish to investigate participation in conferences. Research suggests that student participation will increase, teacher participation will decrease, and participation will be more balanced among students in the graphic conferences compared to the proximate conferences (Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Slatin, 1998; Sullivan & Pratt, 1996; Warschauer, 1996). I
wish to attempt to verify these findings. Moreover, in line with Warschauer’s (1996) suggestion of a direction for additional research, I wish to use the gini statistic to determine if more balanced participation in graphic conferences will be evident with my larger groups of students. To avoid misunderstanding, I should clarify that although the total number of students in my study (i.e., 27 students) is not decisively larger than the total number of students in Warschauer’s study (i.e., 16 students), my conferences take place with more students in each group. As a result, I wish to test if more balanced participation in graphic conferences will also occur with larger conference groups.

Third, I want to know if there is a difference in the textual features of drafts written after the two types of conferences. In other words, I wish to examine the textual features of the drafts in terms of total length, t-unit length, and topical range. No studies to date have compared these aspects of drafts written after graphic and proximate conferences.

Finally, as I mentioned in my critique of Sullivan and Pratt’s (1996), Ghaleb’s (1993), and Braine’s (1997; 2001) studies of the influence of graphic and proximate conferences on first and final drafts, none of these
studies offered any direct evidence that the language and ideas generated in either type of conference were indeed transferred to drafts. I concluded that the methodology of these studies begs the question: Did either type of conference have any influence on drafts at all, much less a relatively superior or inferior influence? As a result, I wish to use LSA and Corpus Linguistics measures to compare the two types of conferences in terms of the linguistic and semantic relationship between the conferences and drafts. More precisely, I wish to use these analytical techniques to compare the degree to which conferences were on topic, and to compare the degree to which drafts were lexically and semantically similar to the conferences.
CHAPTER THREE: METHODOLOGY AND PROCEDURES

This chapter begins with an explanation of the research methodology used in this study. Following this comes a description of the participants in the study, which includes an explanation of how participants were assigned to groups. Next, ethical considerations in the design of this study are addressed. After that, the procedures followed with participants are described. Subsequently, a detailed explanation of the research design and the measures used in the study is followed by a description of the data collection procedures. Finally, methodological assumptions and limitations are discussed, and the research questions under investigation are reiterated.

Research Methodology

This study used a quasi-experimental, counterbalanced, repeated-measures design. As shall be discussed more fully in the forthcoming section on participants, the participants in this study were not randomly assigned to groups. As a result, the counterbalanced design was chosen to ensure that group would not be a confounding variable in comparisons across conditions. In other words, each group was assigned to participate in each condition because, if each group had simply been assigned to a single condition, it would have been difficult to say with certainty that an
observed difference across conditions was due to condition rather than group. The measures were repeated in order to gain a sufficient number of trial comparisons upon which to base meaningful conclusions.

Participants

The participants in this study were high intermediate to advanced English as a second language students who were placed in level 4 of the Intensive English Language Program (hereafter: IELP) at Brock University based on their Michigan Test scores at the beginning of the term. The range of Michigan scores for these students was from 56 to 77. These reported scores are the average of the listening and grammar/reading components of the Michigan Test.

As mentioned, subjects were not randomly assigned to groups for this study; the groups of students were intact class groups. The assignment of individual students to class groups was a decision made by the administration of the IELP, and as such, was beyond the control of the researcher. These particular class groups were chosen to participate in this study because they were at the desired proficiency level. Since these classes (i.e., class 4 and class 4a) were at the same proficiency level in the IELP program based on their Michigan scores (i.e., level 4 in
the IELP), all students in the study were deemed to be of comparable overall language proficiency.

As can be seen in Table 3, the two class groups, classes 4 and 4a, were approximately balanced in terms of language proficiency, number of students, gender, and language and cultural background. At the beginning of the term in which this study was conducted, the students in classes 4 and 4a had average Michigan Test scores of 68.72 and 68.50 respectively. Class 4 had 13 students of which 8 were males and 5 were females. Class 4a had 14 students of which 7 were males and 7 were females. Regarding language and cultural background, Class 4 had 5 Koreans (i.e., South Korea), 2 Japanese, 2 Russians, 2 Spanish (i.e., Mexican), 1 Taiwanese, and 1 French (i.e., Quebecois). Class 4a had 4 Koreans (i.e., South Korea), 4 Spanish (i.e., Mexican), 3 Japanese, 1 French (i.e., Quebecois), 1 Russian, and 1 Cantonese.

Although the usual classroom teacher of class 4 and the usual classroom teacher of class 4a attended all the conferences, they did not take part in the conference discussions. The two teachers were not asked to participate because research (Yagelski & Grabill, 1998) has shown that a teacher’s style of leading face-to-face and computer-mediated class discussions, as well as his or her
Table 3

Composition of Class Groups

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Class 4</th>
<th>Class 4a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Michigan Test Score</td>
<td>68.72</td>
<td>68.50</td>
</tr>
<tr>
<td>Total Students</td>
<td>13.00</td>
<td>14.00</td>
</tr>
<tr>
<td>Males</td>
<td>8.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Females</td>
<td>5.00</td>
<td>7.00</td>
</tr>
<tr>
<td>Korean (South Korea)</td>
<td>5.00</td>
<td>4.00</td>
</tr>
<tr>
<td>Japanese</td>
<td>2.00</td>
<td>3.00</td>
</tr>
<tr>
<td>Russian</td>
<td>2.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Spanish (Mexican)</td>
<td>2.00</td>
<td>4.00</td>
</tr>
<tr>
<td>French (Quebecois)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Taiwanese</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cantonese</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
level of computer proficiency, and his or her attitude towards the use of computers in writing, may influence the quality and quantity of students' participation. As a result, in order to keep experimental conditions consistent between classes, it was decided that the researcher, rather than the regular classroom teachers, would lead all the conference discussions.

**Ethics**

Since this study used human participants, before any aspect of this study was conducted, ethics approval was sought and obtained from the Senate Research Ethics Board at Brock University. A copy of the ethics approval form can be found in Appendix G of this study. Although I have modified the title of this thesis from the title reported on the ethics approval form in order to better reflect the results of my research, the data collected and procedures followed with students in this study were identical to the information provided in my application for ethics approval.

A copy of the consent form participants signed before participating in this study can be found in Appendix H. As can be seen from the consent form, students' data were used in this study with students' informed consent. Students were free to withdraw their consent to allow their data to be used in the study at any time, and without any adverse
consequences. All the students in classes 4 and 4a signed the consent form before participating, and no students withdrew their consent to participate.

**Procedures**

The students participated in the pre-writing and drafting sessions as part of their regular writing class requirements. Face-to-face conferences for each class were held in that class's usual classroom at the university. All electronic pre-writing sessions were held in each class's usual computer lab. The computer software used for the graphic conferences was Web-CT. The topics used each week were selected from a number of possible topics by the class teachers based upon the teachers' estimate of overall interest to the students. Copies of the topics selected by the teachers and used in this study are included in Appendix I.

The topics chosen by the teachers and used in this study were new to the students, within the context of this course, at the beginning of each conference. In other words, the topics were not based on past work or assigned readings because I wanted the conferences to be the first time students had the chance to discuss these topics within the context of the course. To put it another way, I wanted the topics to be "fresh" to the students at the beginning
of the conferences. If students had been exposed to these particular topics beforehand, it would have been more difficult to argue that ideas presented in the students' drafts had come from the conferences in this study, rather than from some other source. Moreover, giving students a fresh topic and requiring them to write an essay is a common practice in English as a second language writing classes in the IELP, particularly for writing examinations. As a result, I believed that basing the conferences on fresh topics and asking students to generate ideas and begin writing in a limited amount of time would be good preparatory practice for the students.

A graphic representation outlining the two groups' scheduled participation can be found in Table 4. Table 4 shows that there was a total of 4 weeks of observations in the study, and groups alternated from one conference condition to the next, week-by-week. For instance, during weeks 1 and 3, class 4 was in the graphic condition, while class 4a was in the proximate condition. On the other hand, during weeks 2 and 4, class 4a was in the graphic condition, while class 4 was in the proximate condition.

As previously mentioned, in the proximate condition, the students and the researcher discussed the writing topic while seated around a table, face-to-face, as a whole-class
Table 4

Scheduled Participation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Week 1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphic</td>
<td>4</td>
<td>4a</td>
<td>4</td>
<td>4a</td>
</tr>
<tr>
<td>Proximate</td>
<td>4a</td>
<td>4</td>
<td>4a</td>
<td>4</td>
</tr>
</tbody>
</table>
discussion. On the other hand, in the graphic condition, although the students and the researcher also discussed the writing topic as a whole-class discussion, they did so by typing messages that could be viewed by all other participants while each participant was seated at his or her own computer in a computer laboratory.

All of the observations took place between January and April 2000. The 4 weeks of observations (i.e., weeks 1 to 4 in Table 2) were held during the 9th, 10th, 12th, and 13th weeks of the term, respectively. These particular weeks were chosen in consultation with the instructors of the two classes.

At this point, I will describe the specific procedures used in this study, and give my rationale for why the procedures were designed the way they were. Since the remainder of this section is organized to highlight the rationale for my procedures, for ease of reference, the procedures are also summarized in temporal order in Appendix J.

One of my key concerns was to ensure that procedures remained consistent between class groups, conference conditions, and weeks of the study. Since conference condition was the only independent variable of interest in this study, I wanted to make sure that all other possible
variables, such as differences between class groups or weeks of the study, were controlled or eliminated. One important threat to my ability to control possibly confounding variables was brought to my attention through the work of Yagelski and Grabill (1998). These researchers found that the amount of student participation and the nature of that participation are related to the way the instructor frames and manages the use of computer-mediated communications technologies. As a result, I took several steps to minimize the possibility of that particular threat. In total, I took six steps to help ensure consistency in procedures throughout the study.

First, in order to ensure that both classes would have a similar understanding of the importance of their participation in this study, I took the following preliminary step. Before the data collection portion of the study, during the 6th and 7th weeks of the term, I visited with the students in both of the classes to describe the nature of the study and what was required of them. With the sanction of both teachers, I informed students that their participation in both types of conferences would constitute a portion of their participation mark for the class. In consultation with the teachers, students were also informed that submitting
drafts for this study would count towards the homework portion of their mark. Moreover, students were told that the first drafts written for this study would need to be developed into final drafts, and that the final drafts would count towards their final mark in the course. Students were also informed that their regular classroom teacher would attend but not participate in the conferences, and that the regular classroom teacher, rather than the researcher, would retain responsibility for assigning all marks. During this meeting, to further ensure that students understood what was required of them, I distributed the consent forms (see Appendix H), and students read and signed them.

Second, I needed to make certain that students in both classes would be prepared to use the graphic conferencing software. I did not want students to perform differently in the two types of conferences, particularly at the beginning of the study, due to unfamiliarity with the computer software. Consequently, during the 8th week of the term, before data collection began, I took each class group into the computer lab for an introductory session on the computer. During this training session, students learned how to log in and out of the system, and a short
rehearsal conference was held so that students could practice sending and receiving messages.

Third, to make sure that both class groups would have similar experiences during all of the conferences, I, rather than the regular class teacher, led all the conferences. Although this decision introduced a threat to internal validity in the form of a researcher effect into my methodology, this was deemed preferable to the alternative. See the methodological assumptions and limitations section of this chapter for a more detailed discussion of the expected researcher effect. In other words, given Yagelski and Grabill's (1998) findings, a researcher effect was considered preferable to the amount of variability that would have been introduced into my study if I had asked the two teachers to introduce the activities and lead the conferences. For instance, the two teachers may have had two different styles of leading face-to-face and computer-mediated class discussions, two different levels of computer proficiency, and two different attitudes towards the use of computers in writing, all of which may have influenced the quality and quantity of students' participation (Yagelski & Grabill, 1998).

Fourth, to ensure that all conference groups would be equally prepared to discuss each of the assigned topics, on
the day before each of the pre-writing conferences, a written copy of the writing topic for the week was given to the students in each class by the classroom teacher. At that time, the teacher reviewed the topic with the class to ensure that students understood the vocabulary and general meaning. At this time, the teacher also reminded students that on the following day, they would be required, as a whole-class activity, to discuss the issue for approximately 20 minutes, after which they would have about 25 minutes to individually begin writing a first draft that would be due for homework on the following day.

Fifth, to ensure that all conference groups knew what was expected of them, at the beginning of each conference I reminded students that they had 20 minutes to discuss the topic as a class, after which they were to individually write first drafts. I also reminded students that they would be required to complete the drafts for homework, and that the drafts would be collected in class on the following day.

Finally, in order to maintain consistency in my role in the conference discussions, I took the following approach. I decided that I would facilitate the discussions by restating the assigned topic and questions, as well as by summarizing students' contributions and
questioning for clarification when necessary, but as far as was possible, I would attempt to restrain myself from offering new ideas or giving examples about the topic.

Although I tried as much as possible to minimize my participation and let the students lead the discussion, as shall be revealed more fully in the data analysis and results sections of this study, despite my best efforts, there were differences in my participation across conditions. This is not particularly surprising since, as I discussed in my review of the literature, substantial evidence suggests that teacher participation in graphic conferences tends to be less than teacher participation in proximate conferences (Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996). I shall return to discuss this point as a possible methodological limitation in the final section of this chapter.

Before moving into the data collection and recording section, it is worth mentioning that throughout the portion of the term preceding this study, students had already become accustomed to using small-group and pair-work face-to-face pre-writing discussions to generate ideas for writing. These sorts of discussions are widely used in IELP writing classes because pre-writing activities such as pre-writing discussions are an integral part of the process.
approach to writing pedagogy (Elbow, 1973; Hayes & Flower, 1986; Reid, 1993) employed in the IELP. As a result of students' familiarity with using discussions to generate ideas for writing, in addition to their familiarity with the use of computer conferencing software gained as a result of the training session, it was determined that students were adequately prepared to engage in, and make constructive use of the graphic and proximate pre-writing conferences used in this study.

**Research Design and Instrumentation**

This study aimed at detecting differences between face-to-face and computer-mediated conferences as pre-writing activities. As such, the key independent variable was conference condition. In general, then, the null form of all research hypotheses could be expressed as no difference across conference conditions for any given dependent measure. However, specific statements of null hypotheses will be reserved for the data analysis chapter of this study and, where appropriate, stated with specific reference to the statistical tests used.

Several dependent variables were considered in comparisons between conference conditions. The way in which these variables were operationalized as measures is described in this section. Details about the statistical
tests used are reserved for the next chapter of this study. However, in this section, where necessary, I will go into some detail to explain the measures themselves. The dependent variables used in this study are perhaps best organized with reference to the research directions I identified at the end of the last chapter.

Before moving into this discussion of comparisons between conferences, I should point out that proficiency measures were also used to compare groups. Measures of general language proficiency were obtained from students' Michigan Test scores taken at the beginning of the term in which the data collection portion of this study was carried out. These measures were used to verify that the groups were balanced in terms of language proficiency.

**Investigating Conference Textual Features**

The first research direction I identified at the end of the last chapter dealt with comparisons of conference textual features. In particular, I wanted to know if both types of conferences would allow students to produce an equivalent amount of discourse in an equivalent amount of time, if there would be a difference in the length of t-units contributed by participants in the two types of conferences, and if the two types of conferences would be of equivalent lexical complexity.
To address the first part of this research direction, both the total number of words and the total number of t-units contributed by all the students collectively in each conference were calculated. This information was then used to compare the total amount of discourse produced by students in each of the two types of conferences.

To address the second part of this research objective, the mean length of t-units contributed by each student in each conference was calculated. These data were subsequently used to compare the lengths of t-units produced by students in each of the two types of conferences.

To address the third part of this research objective about conference textual features -- to see if students in one type of conference tended to repeat the same words more than they did in the other type of conference -- the standardized type-token ratios (Scott, 1998) of students' discourse in each conference was computed. Standardized type-token ratios were calculated by using Wordsmith Tools (Scott, 1998) to determine the ratio of unique words to total words for the first x words, and then again for the next x words, until the end of the transcript. The mean of the obtained ratios was then recorded as the standardized ratio (Scott, 1998). For this study, since the conference
transcripts ranged in length between 1,041 and 2,508 words, the standardized ratios of conferences were based on 500 word intervals (i.e., x = 500).

**Investigating Conference Participation**

The second research direction was to determine if the amount of researcher participation was equivalent in both types of conferences; if participation was equally balanced among students, and among students and the researcher combined, in both types of conferences; and if the addition of the researchers' discourse to the students' discourse tended to create greater inequality of participation in either type of conference.

To address the first part of this research direction, for each conference, the total number of words contributed by the researcher as a percentage of the total number of words contributed by all participants, and the total number of t-units contributed by the researcher as a percentage of the total number of t-units contributed by all participants was calculated. These percentages were then used to determine if the amount of the researcher's participation was balanced between the two types of conferences. Moreover, the average length of t-units contributed by the researcher in each conference was calculated. This measure
was used to find out if the researcher's contributions were of equivalent length in the two types of conferences.

To address the second part of this research objective, the number of words contributed by each participant was first calculated. Then, the number of words contributed by each participant during each conference, both including and excluding the researcher's participation, was used to calculate gini coefficients of inequality for each conference. For an explanation of the meaning and calculation of gini coefficients, see Appendix K. These gini coefficients were subsequently used to determine if participation was equally balanced among students only, and among students and the researcher combined, in both types of conferences.

With respect to the third part of this objective, these gini coefficients were also used to determine if the addition of the researcher's discourse tended to create greater inequality of participation in each of the two types of conferences.

Investigating Draft Textual Features

The third research direction was to compare the textual features of drafts written after the two types of conferences. In particular, I wanted to know if drafts
written after the two types of conferences would be of equivalent length, t-unit length, and semantic breadth.

To compare the length of drafts, both the total number of words and the total number of t-units in each student's draft were calculated. This information was then used to compare the lengths of drafts following the two types of conferences.

With respect to the second part of this research direction, to compare the length of draft t-units, the mean length of t-units in each student's draft was first calculated. These data were subsequently used to compare the mean lengths of t-units in drafts produced by students following the two types of conferences.

As for the third part of this research objective, the measure draft vector length was used to compare the semantic breadth of drafts written after the two types of conferences. As was discussed in the section on analytical measures derived from LSA in the previous chapter, the vector length of a document indicates the amount and breadth of information conveyed in the document.

As was discussed, longer documents have longer vectors. However, this need not be the case if, for example, a document uses a lot of non-content words. Such a draft could be relatively long in number of words, yet
have a relatively short vector. Such a draft might be characterized as "wordy" in style. Conversely, a shorter draft could have a relatively long vector if it covered many different topics without going into much depth. Such a draft might be characterized as incoherent or lacking in cohesion (Foltz et al., 2000). Since vector length can detect when a long draft is wordy and when a short draft covers a lot of semantic ground, it was thought to be a better measure of semantic breadth than draft length.

**Investigating the Relation Between Conferences and Drafts**

The final research direction I identified concerned comparing the degree to which conferences and drafts were lexically and semantically related. A Corpus Linguistics measure and several LSA measures were used to address this research objective. The LSA measures were: the average maximum cosine and the mean average cosine of conference t-units to draft t-units; and the average maximum cosine of draft t-units to conference t-units. The Corpus Linguistics measure was the ratio of total words to new words in the drafts. Each of these measures will be discussed in turn.

**Conference average maximum cosine.** The measure average maximum cosine, which was described in the section on analytical measures derived from LSA in the previous
chapter, was used in this study to determine how similar conference t-units were to the draft t-units following that conference. In other words, this measure was used to determine how on topic conference t-units were. This measure was ultimately used to determine if there was a difference between the two types of conferences in terms of how on topic the conference discussions were.

In the preceding paragraph, I claimed, somewhat implicitly, that a conference t-unit that is semantically related to one or more draft t-units can be characterized as on topic. This claim can only be valid if it is based on the assumption that the draft t-units are all on topic. The logic here is syllogistic: Assuming that the draft t-units are all on topic, if a conference t-unit is semantically related to a draft t-unit, the conference t-unit must also be on topic. I will provide evidence to support the validity of the assumption upon which this syllogism is based in the methodological assumptions and limitations section of this chapter.

To obtain the maximum cosine of a conference t-unit, the conference t-unit was first assigned a value equal to the largest cosine value that it reached with any draft t-unit (Foltz et al., 2000). For instance, if a conference t-unit yielded cosines of 0.23, 0.11, 0.34, and 0.22 with
the draft t-units following that conference, the maximum cosine of that conference t-unit would be the highest value -- in this example, 0.34.

To obtain a student's average maximum cosine of conference contributions, or for short, a student's conference average maximum cosine, the maximum cosine values for all the t-units contributed by that student in the conference were first summed, and then divided by the total number of conference t-units contributed by that student (Foltz et al., 2000).

*Conference mean average cosine.* Like average maximum cosine, the measure mean average cosine was also used to determine how similar each student's conference contributions were to the draft t-units written by all the students following that conference, or in other words, to determine how on topic each student's conference t-units were. Moreover, this measure was also ultimately used to determine if there was a difference between the two types of conferences in terms of how on topic the conference discussions were.

This characterization of a conference t-unit that is semantically related to the drafts as on topic rests on the same assumption as was discussed in the previous section with reference to conference average maximum cosine. I
will provide evidence to support this assumption in the methodological assumptions and limitations section of this chapter.

Although mean average cosine has not been used in previous research, it was included in this study to determine when terms and ideas introduced in the conferences were on topic in that they were widely used in students' drafts in general. In contrast, as shall be discussed further in a moment, average maximum cosine is more a measure of how on topic conference t-units were based on close or exact repetition of a conference statement in one of the drafts written afterwards.

To obtain the average cosine of a given conference t-unit, cosines were first computed for that conference t-unit in relation to every draft t-unit written afterwards by all the students. Next, summing all these draft cosine values and dividing by the number of draft t-units gave the average cosine of the conference t-unit. For example, if a given conference t-unit yielded cosines of 0.23, 0.11, 0.34, and 0.22 with the draft t-units following that conference, then the average cosine of that conference t-unit would be the mean -- in this case, \([0.23 + 0.11 + 0.34 + 0.22] / 4\) = 0.225.
To obtain the student’s mean average cosine of conference contributions, or for short, the student’s conference mean average cosine, the average cosines for all the t-units contributed by that student were first summed and then divided by the total number of conference t-units contributed by that student. In essence, the measure is an average of averages, hence the term mean average cosine.

Comparing conference average maximum cosine and conference mean average cosine. An example will help to clarify the difference in meaning between conference average maximum cosine and conference mean average cosine, and to clarify why both of these measures were important to my study. What would happen if a t-unit contributed by a student during the conference was repeated verbatim or closely paraphrased in one student’s draft but ignored by all the other students? Since the conference t-unit in this example was very similar to one of the draft t-units, it would obtain a high cosine with that draft t-unit, and consequently, it would have a high maximum cosine. On the other hand, since the conference t-unit in this example was not similar to the vast majority of draft t-units, it would have relatively low cosine values with most draft t-units. As a result, even though the conference t-unit had a high cosine value with one or a few of the draft t-units, when
the average cosine was calculated, it would end up having a relatively low value.

Conversely, for similar reasons to those just discussed, if a conference t-unit contained terms and ideas that were relevant to many students drafts, yet the conference t-unit was never closely paraphrased or reproduced verbatim in any of the drafts, it would have a low maximum cosine and a relatively high average cosine. As a result, these two measures can be conceptualized as two different ways of measuring semantic relation, and accordingly, two different ways of measuring how on topic conference t-units were. On the one hand, maximum cosine is particularly well-suited for matching specifically related t-units to each other (Foltz et al., 1996; Laundauer et al., 1998). For instance, to see how this measure was able to match discourse between conferences and drafts, see Appendices C and D. On the other hand, average cosine is uniquely appropriate for identifying more general relationships between t-units. To see examples of how this measure was able to distinguish between on-topic discourse that was generally related to the drafts and off-topic discourse that was not, see Appendices E and F.

In sum, although using both of these measures may seem excessive, each of these measures assesses the degree to
which conference t-units were on topic in a different way. On the one hand, while conference average maximum cosine is sensitive to when a conference t-unit is on topic in that the language and ideas in the conference t-unit are repeated very closely in at least one of the draft t-units, it is relatively insensitive to when a conference t-unit is on topic based on more widespread semantic relation of the terms it contains to the draft t-units. On the other hand, while conference mean average cosine is sensitive to a conference t-unit that is on topic in that it introduces terms and ideas that are relevant to many students' drafts, it is relatively insensitive to when a conference t-unit has a strong and specific relationship to a draft t-unit. As a result, using both, rather than one or the other of these measures in my study better enabled me to accurately detect and compare how on topic conferences were.

Although I have accentuated the differences between these measures, I want to be careful not to overstate the point. Overall, one would expect these measures to be positively correlated since both assess the relation of conference t-units to draft t-units. For example, one would expect that a student with a low conference mean average cosine would also have a low conference average maximum cosine since both would result from a student
making mostly off-topic comments. Conversely, one would expect that a student with a high conference mean average cosine would also have a high conference average maximum cosine since both would be the likely result of a student making mostly relevant, topic-related comments.

*Draft average maximum cosine.* In addition to measuring, as discussed, the maximum semantic relation of each conference t-unit to all of the draft t-units, I also approached the comparison from the other direction. I measured the semantic relation of each draft t-unit to all of the conference t-units. Draft average maximum cosine was one of the measures used to gauge the semantic relationship of students' drafts to the conference, and ultimately to compare the two types of conferences in terms of the semantic relation of drafts to conferences.

The calculation made to obtain draft average maximum cosine and the interpretation of this measure are analogous to those described above for conference mean average cosine. A student's draft average maximum cosine is equal to the sum of the maximum cosines of the student's draft t-units to all of the conference t-units, divided by the total number of draft t-units the student wrote. A draft with a high average maximum cosine would be one in which
each draft t-unit repeated or closely paraphrased one of the conference t-units.

Although conference average maximum cosine and draft average maximum cosine are similar, these measures are not the same. For instance, a student might have had a relatively high conference average maximum cosine if that student's conference comments were closely paraphrased or repeated in his or her or another student's draft; yet this same student might have had a low draft average maximum cosine if his or her draft did not closely paraphrase or repeat t-units made during the conference. Conversely, a student might have had a relatively low conference average maximum cosine if that student's conference t-units were not closely repeated or paraphrased in his or her or any other student's draft; yet this same student might have had a high draft average maximum cosine if his or her draft closely paraphrased or repeated t-units made during the conference.

In sum, while conference average maximum cosine is a measure of how semantically related conference contributions were to all of the drafts, draft average maximum cosine is a measure of how semantically related drafts were to all of the conference contributions. However, interestingly, if a student's draft were to
closely paraphrase or repeat verbatim only what he or she contributed during the conference, that student's conference t-units would likely have maximum cosines with his or her own draft t-units, and conversely, that student's draft t-units would likely have maximum cosines with his or her own conference t-units. In such a situation, conference average maximum cosine and draft average maximum cosine would be measuring the same thing, and would therefore yield identical values.

A draft with a low average maximum cosine is one that did not tend to rely much on what was said in the conference. However, such a draft need not necessarily be of low quality, or off topic. In fact, it may be a superior treatment of the topic at hand. Low scores for draft average maximum cosine merely indicate that the writer did not tend to repeat or paraphrase the ideas presented in conference t-units in his or her draft.

**Draft ratio total words to new words.** Finally, keyword analysis using Wordsmith Tools (Scott, 1997; 1998) was used to produce a measure of word-overlap of drafts with the conference. The main difference between this measure and the LSA draft measure is that while the LSA measure can detect semantic overlap when words are not identical, this
measure is only sensitive to similarity based on exact vocabulary overlap.

To obtain this measure, first, wordlists of all the words used in the conference and all the words used in each draft were created. Next, the wordlists were compared to see which words used in each draft were not used in the conference. The ratio of total words in the draft to new words used in the draft but not used in the conference was then calculated. A higher ratio indicates more exact lexical overlap between a draft and the conference.

**Data Collection and Recording**

The proximate conferences were videotaped and transcribed and the graphic conferences were saved as a file and printed out so that verbal communications could be subjected to analysis. Detailed information about the transcription of graphic and proximate conferences can be found in Appendix L, and an example of transcribed conference text can be found in Appendix M.

After the conclusion of each week, videotape recordings of the proximate pre-writing sessions, and copies of the graphic transcripts were made available in the IELP Self-Access Centre. The Self-Access Centre is a facility where students can go after classes to engage in self-study activities. For instance, the Self-Access
Centre contains pedagogical books, audiocassettes, videotapes, and computer software, as well as the desks and equipment necessary for students to use these materials for self-study. The written and videotaped transcripts were made available in the Self-Access Centre so that students who participated in the study could review the ideas presented during the conferences, after submitting their first draft, to support further sustained writing about the topic, or for more general review of language skills.

Data Processing and Analysis

In this section, two explanations regarding how the data analysis was approached are warranted. First, a brief description of how conference scores were tabulated will help to clarify how missing data were handled and why they were handled the way they were. Second, a general discussion of statistical issues related to the design of this study will serve to clarify the choice of statistics.

An extract of conference data is presented in Table 5. The tabulation of scores for student 12 was quite straightforward since the student was present and participated in all conferences. On the other hand, student 28 was physically present in all the conferences, but in week 1, this student made no contributions. As a
# Table 5

**Extract of Conference Data**

<table>
<thead>
<tr>
<th>ID</th>
<th>Week</th>
<th>Conference Condition</th>
<th>Number of T-Units</th>
<th>Number of Words</th>
<th>Average T-Unit Length</th>
<th>Mean Average Cosine</th>
<th>Average Maximum Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>1</td>
<td>Graphic</td>
<td>1</td>
<td>14</td>
<td>14.0</td>
<td>0.1016</td>
<td>0.4800</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Proximate</td>
<td>5</td>
<td>78</td>
<td>15.6</td>
<td>0.2062</td>
<td>0.5760</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Graphic</td>
<td>1</td>
<td>13</td>
<td>13.0</td>
<td>0.2130</td>
<td>0.0750</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Proximate</td>
<td>6</td>
<td>57</td>
<td>9.5</td>
<td>0.1423</td>
<td>0.4850</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>Proximate</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Graphic</td>
<td>11</td>
<td>119</td>
<td>10.8</td>
<td>0.1571</td>
<td>0.5909</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Proximate</td>
<td>1</td>
<td>7</td>
<td>7.0</td>
<td>0.0487</td>
<td>0.3300</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Graphic</td>
<td>12</td>
<td>94</td>
<td>7.8</td>
<td>0.1140</td>
<td>0.5075</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>Proximate</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Graphic</td>
<td>43</td>
<td>312</td>
<td>7.3</td>
<td>0.0942</td>
<td>0.5367</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Proximate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Graphic</td>
<td>35</td>
<td>221</td>
<td>6.3</td>
<td>0.0853</td>
<td>0.4549</td>
</tr>
</tbody>
</table>
result, for student 28, week 1, no data were entered for average t-unit length, mean average cosine, or average maximum cosine. No scores were entered for these measures because entering zero for average t-unit length in cases of no contributions would seriously bias a comparison of mean t-unit length across conditions. Since the only cases in which students attended a conference without making any contributions occurred in the proximate condition, entering zero would make it appear that, on average, t-units in the proximate conference were shorter than they really were.

Similarly, although it may be tempting to enter a value of zero for mean average cosine and average maximum cosine for students who said nothing in the conference, after all, students who contributed nothing could not be said to be on topic, this would make it impossible to distinguish between students who made off-topic comments and students who made no comments. In other words, since off-topic comments received very small positive scores on the LSA measures (Foltz et al., 1994; Laundauer & Dumais, 1997; Laundauer et al., 1998), entering a value of zero for the LSA measures when a student said nothing would be to infer that a student who made only highly off-topic comments during a given conference was more on topic than a student who made no comments at all. Moreover, for the
same reason as discussed with reference to average t-unit length, entering zero would seriously bias a comparison of means for the LSA measures across conditions, making it appear that, on average, t-units in the proximate conference were much more off topic than they really were.

Finally, the data for student 21 in Table 5 illustrate the difference between making no contribution while being physically present at a conference (i.e., week 1), and not being there at all (i.e., week 3). When a student did not attend a conference, his or her scores were simply left out of the analysis for that week.

With regard to the second preliminary section outlined above, given the design of this study, it was not statistically sound to simply run a one-way ANOVA on the scores for any given measure across conditions (Howell, 1982). For example, to determine in which conference condition students tended to produce higher conference mean average cosines, I could not run an ANOVA on all the conference mean average cosine scores for the electronic condition versus all the conference mean average cosine scores for the graphic condition because such an analysis would only be appropriate for a design in which eight groups participated in each condition once. In other words, such an analysis would only be appropriate for a
design in which all the groups in the ANOVA were independent (Howell, 1982).

To look at it another way, if such an ANOVA were run, assuming no absences from any conferences, the N for the ANOVA would have been the total number of students in this study multiplied by 4. Such a tactic would have greatly increased the risk of a type I error, or the risk of finding a statistically significant difference where such a difference is not, in fact, justified (Howell, 1982).

Since two groups of students participated in each condition twice, the most suitable way to make comparisons across conditions was to run four independent-sample t tests across conditions week-by-week (Howell, 1982). It was decided that 2 out of 4 statistically significant differences in the same direction at an alpha level of 0.05 would be taken to suggest moderate evidence of a main effect. Three out of 4 statistically significant differences would be considered persuasive evidence of a main effect; and 4 out of 4 statistically significant differences, very strong evidence of a main effect.

With regard to the dependent measures analyzed with independent sample t tests, preliminary analysis using Levene's test yielded statistically significant values for all the measures (i.e., p < 0.05) across conditions and
groups week-by-week. This indicates that the variance in these measures was not homogeneous across conditions and groups. Since the variances were heterogeneous, variances were not pooled for the independent sample t tests used in this study (Howell, 1982). As a result, to calculate degrees of freedom for these tests, the Welch-Satterthaite solution was used (Howell, 1982):

\[
\text{df} = \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} - \frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}
\]

In this formula, \(s\) refers to the variance in the measure for each group and \(N\) refers to the number of students in each group. Degrees of freedom were obtained by rounding the result of this formula to the nearest integer (Howell, 1982).

Another way of testing differences across conditions and groups was used when comparisons were made of single scores rather than means of scores. In such cases, a matched-sample, or one-sample t test was performed (Howell, 1982). For example, when comparisons of the standardized type-token ratios of the conferences were made, in the next chapter, the matched-sample t test was used because each conference was assigned a single score rather than the test
being based on a comparison of means and standard deviations calculated from several scores. It was permissible to use a matched sample t test in these cases because an equal number of observations were being compared across conditions for groups matched on discussion topic, as well as on the language proficiency, gender, native language, and cultural background of the group members (Howell, 1982).

The matched sample t test was also used when comparisons were made of scores for the same individual (Howell, 1982). For example, when the difference in the researcher's participation between graphic and proximate conferences was compared, the matched sample t test was used because scores for the same person were being compared across conditions. An explanation of how the test was run will help to clarify.

To make comparisons of the percentage of words contributed by the researcher, a two-columns matrix of scores was formed so that scores for the proximate condition were in the first column and the scores for the graphic condition were in the second column, while the scores for the same week were in the same row. The scores for each week in the graphic condition were then subtracted from the corresponding scores for the same week in the
proximate condition resulting in a single column of difference scores. A one-sample \( t \) test was then run on the difference scores with the null hypothesis being that the mean of the difference scores was equal to zero (Howell, 1982).

**Methodological Assumptions and Limitations**

In this section, the methodological assumptions and limitations of this study are addressed. A discussion about an important assumption made regarding the interpretation of the LSA measures will be followed by a few words about the external validity of this study, and then a discussion of several possible threats to internal validity.

**An Assumption About the LSA Measures**

In the research design and instrumentation section of this chapter, in my discussion of conference average maximum cosine and conference mean average cosine, I reported that since these measures evaluate the semantic similarity of conference t-units to draft t-units, these measures evaluate the degree to which conference t-units were on topic. I mentioned that this characterization of the meaning of these measures rests on the assumption that the draft t-units were all on topic, and that I would return in this section to provide evidence to support the
validity of this assumption. However, before doing so, I will explain in more detail why this assumption is necessary.

A conference maximum cosine is the largest cosine value that a conference t-unit obtained with a draft t-unit. As a result, if some of the draft t-units were off topic, it would be possible for an off-topic conference t-unit to obtain a high maximum cosine with one of the off-topic draft t-units. In such a case, it would make no sense to claim that conference t-units with high maximum cosines are on topic. Nor would it be logical to assume that students with high conference average maximum cosines were more on topic than students with low conference average maximum cosines. As a result, the assumption that all draft t-units were on topic is necessary to characterize high conference average maximum cosines as indicating on-topic discourse.

A conference average cosine is the average cosine value that a conference t-unit obtained with all the draft t-units. As with the last measure, if some of the draft t-units were off topic, it would be possible for an off-topic conference t-unit to obtain a high average cosine due to similarity with the off-topic draft t-units. In such a case, it would make no sense to claim that conference t-
units with high average cosines are on topic. Nor would it be logical to assume that students with high conference mean average cosines were more on topic than students with low conference mean average cosines. As a result, the assumption that all draft t-units were on topic is necessary to characterize high conference mean average cosines as indicating on-topic discourse.

As has been shown, both of the LSA conference measures can be used to characterize conferences as on topic or off topic, if the assumption that all draft t-units are on topic is met. One way to provide evidence to support the assumption that all the draft t-units were on topic would be to allow the reader to examine all the draft transcripts and determine for him or herself whether or not this was indeed the case. However, since there was a total of 74 drafts, this would be quite a labourious exercise. Fortunately, as shall be discussed in a moment, there is an easier way for the reader to make this determination.

I have closely examined the draft transcripts, and although I do not expect the reader to simply take my word for it, my reading of the drafts has led me to believe that the draft t-units were all on topic. When I write that the draft t-units were all on topic, I do not mean to imply that the drafts were all of high quality. In fact, there
are a few drafts that I feel are of somewhat low quality because I disagree with the student's reasoning and claims, or because I think the student has provided a shallow analysis of the topic. However, what all the drafts have in common is that each and every t-unit reflects the student's attempts to address the topic.

I mentioned that there is a less strenuous way for the reader to be convinced that draft t-units can be used as a baseline to judge whether or not conference t-units are on topic. For the reader to decide if conferences t-units that were more semantically related to the drafts were also more on topic, the reader need only compare conference t-units with high cosines with the drafts, to conference t-units with low cosines with the drafts. If the conference t-units with high cosines are consistently more on topic than the conference t-units with low cosines, then my claim about draft t-units all being on topic is empirically rendered valid.

Appendices E and F show conference data sorted by mean average cosine. In these appendices, conference t-units that have high mean average cosines with the draft t-units are more coherently on topic than conference t-units that have low mean average cosines with the draft t-units. As a result, the assumption that conference t-units with higher
mean average cosines are more coherently on topic than conference t-units with lower mean average cosines appears to be valid, *quad erat demonstrandum*.

Appendices N and O show conference data sorted by average maximum cosine. In these appendices, conference t-units that have high average maximum cosines with the draft t-units are more coherently on topic than conference t-units that have low average maximum cosines with the draft t-units. As a result, the assumption that conference t-units with higher average maximum cosines are more coherently on topic than conference t-units with lower average maximum cosines appears to be valid, *quad erat demonstrandum*.

Before moving into the discussion of external and internal validity, it is worth explaining why I could use the measure draft average maximum cosine, but not the measure draft mean average cosine to gauge how closely the drafts were related to the conferences.

Draft mean average cosine would not have been a good measure to compare the degree to which drafts were related to the conferences. First, as defined above, all draft t-units were on topic. Since all the draft t-units were on topic, the draft t-units would all be more related to the on-topic conference t-units and less related to the off-
topic conference t-units. As a result, if conference a contained a high proportion of on-topic t-units, while conference b contained a lower proportion of on-topic t-units, and if draft t-unit x, written following conference a was as related to the on-topic t-units in conference a as draft t-unit y, written following conference b was related to the on-topic t-units in conference b, when the average cosine for each draft t-unit was calculated, the draft t-unit following conference b, draft t-unit y, would always have a lower average cosine than draft t-unit x because of the higher proportion of off-topic t-units in conference b. However, this does not mean that draft t-unit y is less related to conference b than draft t-unit x is related to conference a. In fact, they are both equally related to the on-topic t-units in their respective conferences. However, the higher proportion of off-topic t-units in conference b would force draft t-unit y’s average cosine down relative to t-unit x’s average cosine. By extension, the draft mean average cosine for an on-topic draft following a more on-topic conference would always be higher than the draft mean average cosine for an equally on-topic draft following a more off-topic conference.

As a result, to use draft mean average cosine as a measure to compare the degree to which drafts were related
to the conferences, one must assume that both conferences have equal proportions of on-topic and off-topic t-units. However, I had no reason to make such an assumption a priori. In fact, one of the research questions in this study involved comparing the degree to which the two types of conferences were on topic.

In sum, draft mean average cosine would not have been a good measure to compare the degree to which drafts were related to the conferences because this measure would have involved circularity: Because the conferences were more on topic, drafts would be found to be more related to the conferences. Thus, it was necessary to use measures of the similarity of drafts to conferences that were not influenced by the degree to which conferences were on topic, or in other words, that were not influenced by the proportion of on-topic t-units to off-topic t-units in conferences.

Draft average maximum cosine was an appropriate measure for comparing the similarity of drafts to conferences because it avoided the circularity just discussed. Unlike draft mean average cosine, draft average maximum cosine is not based on the average similarity of a draft t-unit to many conference t-units. As a result, the proportion of on-topic to off-topic t-units in a conference
has no effect on this measure. Draft average maximum cosine is based on the maximum cosine a draft t-unit reaches with any single conference t-unit. If a draft t-unit is similar to a conference t-unit, it will achieve a relatively high maximum cosine, regardless of the number of off-topic statements in the conference. If a draft t-unit is not similar to a conference t-unit, it will achieve a relatively low maximum cosine, again, regardless of the number of off-topic statements in the conference. By extension, if draft x is more similar to conference a than draft y is similar to conference b, draft x will receive a higher average maximum cosine than draft y, as desired.

External and Internal Validity

With regard to external validity, as was mentioned in the introduction, since this study made use of a relatively small sample size, the results of this exploratory study were not expected to carry a great deal of external validity, although the results were expected to be applicable to similar groups of students studying under similar conditions.

Several possible threats to internal validity were predicted in the methodology of this study. A practice effect was predicted because as participants gained experience in participating in conferences and writing
subsequent drafts, their skill in these procedures was predicted to improve. Moreover, since students were aware that they were being observed, this awareness was expected to result in an observation effect (e.g., students might feel more shy or inhibited). Furthermore, because the researcher rather than the classroom teacher led the pre-writing discussions, a researcher effect was predicted (e.g., students might either feel more shy or inhibited, or work harder to please the researcher). In addition, since students discussed various topics that may have been more or less motivating for them, a topic effect was predicted. Also, since the final week (i.e., week 4) of this study coincided with the examination preparation week of the program, a history effect was also predicted. Finally, as mentioned in the section on procedures, because the researcher's participation was different in the two conditions, this difference may have introduced a threat to the internal validity of conference comparisons. All of these possible threats to internal validity will be discussed in the remainder of this section.

While the practice effect was expected to increase over the 4 weeks of the study, the researcher and observation effects were expected to decrease over the course of the study as a result of students' increased
familiarity with the experimental conditions and the new personality facilitating class discussions. It is easy to imagine that these effects could be rather detrimental to the validity of comparisons of students' performance between weeks, for instance between weeks 1 and 4. However, one might also imagine that these effects would not be so harmful to the validity of comparisons of students' performance in the same week, for instance between group 4 and group 4a during week 1. In this latter case, these effects would likely have an equivalent impact on both groups during each week of the study, and should therefore have no adverse impact on the validity of the comparisons. Since comparisons in this study were only made across classes and conditions for the same weeks and topics, as in the latter example just mentioned, these effects should have had an equivalent impact on both groups and conditions, and no adverse impact on the validity of comparisons.

Since different topics were used each week, a topic effect was also predicted. In other words, participation in one week might be stronger than in another week because, for example, students were more interested in one topic than another. However, like the possible threats just discussed, since comparisons were made between classes
discussing the same topic, each topic was likely equally interesting or less interesting, on average, to both groups of students. As a result, any topic effect should not have had an adverse effect on the validity of the comparisons.

Another threat to the internal validity of this study that resulted from the repeated-measures design was a history effect. Since the final week (i.e., week 4) of this study coincided with the examination preparation week of the program, although the majority of students still attended the conferences, very few students chose to submit first drafts. As a result of the rather small number of drafts received, and correspondingly small n for draft comparisons in t tests of dependent measures across conditions for this week, statistically nonsignificant results for week 4 were not considered conclusive given the low power of the statistical analysis (Howell, 1982).

I mentioned in the section on procedures that I would return, in this section, to discuss the possibility of differences in the amount of my participation in the two conference conditions as a possible methodological concern. I believe that this difference is not a threat to the internal validity of the comparisons made in this study. To understand why, it will be helpful to begin with an
example of when differences in the amount of teacher participation can be a methodological concern.

In some studies, differences in the role and level of involvement of the group leader in the two types of conferences have been portrayed as a threat to the internal validity of comparisons between conference conditions. For instance, as I discussed in chapter 2, Ghaleb (1993) argued this when she attributed the greater improvement of writing quality in lecture-style classes in her study to the considerable amount of time spent by the teacher on grammar instruction and correction of errors. In sum, in a study like Ghaleb's, where the quality of drafts is compared, and the extra input from the teacher involves grammar or writing instruction, the extra input from the teacher can be seen to be unfairly beneficial to students' writing in the condition in which students received more instruction.

However, in my study, I did not make comparisons of writing quality. Moreover, the content of my increased participation in the proximate condition did not consist of discussing grammar or any other aspect of writing with the students. My increased participation in the proximate condition primarily consisted of guiding the discussion by summarizing students' contributions, and to a lesser
extent, contributing my ideas about the topic (e.g., see Appendix P).

To explain why I believe differences in my participation should not be treated as a threat to internal validity, I will discuss exactly what aspects of the conferences and drafts were compared. First, I made comparisons of conference textual features: total amount of words and t-units contributed by students during the conferences, length of students' conference t-units, and lexical complexity of conferences. Second, I compared balanced participation in the conferences: amount of researcher participation, balanced participation among students, and among the students and the researcher combined, and increased inequality of participation due to the researcher. Third, I compared the textual features of drafts: number of words and t-units, t-unit length, and semantic breadth. Finally, I compared the lexical and semantic similarity of conferences to drafts, or the degree to which the conferences were on topic; and the lexical and semantic similarity of drafts to conferences, or the extent to which the drafts incorporated what was discussed in the conferences. I will discuss each of these comparisons in the remainder of this section.
First, with reference to comparisons of conference textual features, since I am a fairly well educated native speaker of English, one might argue that the t-units I contributed to the conferences were more complex, and thus longer than the students' t-units, and that the discourse I contributed was more lexically complex. As a result, one might argue that the addition of my discourse to the comparison would bias the comparison to favour longer t-units and greater lexical complexity in the proximate condition. However, where I made a comparison of the t-unit length and lexical complexity of the two types of conferences, I did not include the discourse that I contributed in the comparison.

Also with reference to comparisons of conference textual features, since I participated more in the proximate conferences, it is easy to see why someone might point out that the addition of my conference discourse to the students' conference discourse would bias comparisons of the total number of words and t-units contributed by students to favour more discourse in the proximate conferences. However, adding my conference discourse to the students' conference discourse to make comparisons of the total number of words and t-units contributed by students would be boldly illogical. For that reason, where
I made a comparison of total numbers of words and t-units contributed by students in the two types of conferences, I did not include the discourse I contributed in the comparison; but rather, I only compared the students' discourse.

However, since I contributed more in the proximate conferences, one might argue that I took time away from students to contribute. Consequently, even though I only compared students' contributions, it must be kept in mind that if I had contributed less in the proximate conferences, students might have contributed more -- although this is far from certain. Nevertheless, one might argue that by taking time away from the students to participate in the proximate condition, my comparison of the total number of words and t-units in the two types of conferences is biased to favour more student discourse in graphic conferences.

However, as was discussed in the review of the literature, several studies have established that teacher participation in proximate conferences tends to be greater than in graphic conferences (Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996). Therefore, it seems reasonable to argue that increased teacher participation in the proximate condition and
decreased teacher participation in the graphic condition is part of the "furniture" of such comparisons. Seen this way, rather than being a prejudicial threat to the validity of comparisons between types of conferences, increased teacher participation in the proximate conference condition is a characteristic feature of the comparison.

In other words, if I had reduced my participation in proximate conferences more than I already did, comparisons of the total amount of students' discourse would have been made under atypical or even unnatural classroom conditions. Moreover, I feel that further reducing my participation in proximate conferences would have created a rather unusual classroom situation that might have been unduly stressful for the students.

In sum, although I will keep this issue in mind when I discuss the results of my study, I do not see my increased participation in proximate conferences as a threat to the validity of comparisons of the total amount of student discourse in graphic and proximate conferences. On the contrary, I see my increased participation in the proximate condition as a natural, even unavoidable aspect of the comparison.

Second, with regard to comparisons of balanced participation in the conferences, different amounts of
researcher participation in the two types of conferences could obviously not bias these comparisons because determining if the amount of teacher participation was different in the two types of conference was part of the objective of these comparisons.

With regard to comparisons of draft textual features and comparisons of the extent to which students incorporated what was discussed in the conferences in their drafts, I do not see my increased participation in the proximate condition as a methodological threat to these comparisons. I understand that one might argue that my increased participation in the proximate conferences unfairly biased comparisons of draft textual features to favour longer drafts with more semantic breadth in the proximate condition. Similarly, one might argue that my increased participation in the proximate conferences unfairly biased comparisons of the relation of drafts to conferences to favour more relation in the proximate condition. However, to make either one of these arguments, one must assume that my ideas on every topic were consistently superior and singularly influential to students' drafts. I have no convincing grounds to make such an assumption.
To support this claim, I have included a copy of all my comments during one of the proximate conferences in Appendix P. Examination of Appendix P suggests that my increased participation in the proximate condition primarily consisted of guiding the discussion by summarizing students' contributions. I contributed my ideas about the topic to these conferences to a much lesser extent. As a result, it seems unlikely that my increased participation in the proximate conferences gave students more ideas to write about in their drafts, or such noteworthy ideas as to bias comparisons of the semantic similarity of drafts to conferences.

Finally, I will consider the notion that my increased participation in the proximate conferences was a threat to the validity of comparisons of how on topic conference discussions were. As was described in detail in the research design and instrumentation section of this chapter, to compare the two types of conferences in terms of how on topic they were, students' conference mean average cosines and conference average maximum cosines were first calculated and then compared across conference conditions. Importantly, no measures of conference mean average cosine or conference average maximum cosine were calculated for the researcher. As a result, the degree to
which the researcher's comments were on topic did not enter into the comparison between conference conditions. Consequently, the additional participation of the researcher could not have biased comparisons of the degree to which students' conference discourse was on topic.

**Restatement of the Problem**

Having described the methodology of this study, a more concise statement of the research questions can now be made. In total, four research questions will be addressed.

First, is there a difference in textual features between the two types of conferences? In particular, do both types of conferences allow students to produce an equivalent amount of discourse in an equivalent amount of time and is there a difference in the length of t-units contributed by participants in the two types of conferences? Moreover, is the discourse produced in both types of conferences equivalent in terms of lexical range, or do students in one type of conference tend to repeat the same words more than they do in the other type of conference?

Second, with regard to conference participation, is the amount of researcher participation equivalent in both types of conferences; is participation equally balanced among students, and among the students and the researcher
combined in both types of conferences; and does the addition of the researcher's discourse to the students' discourse tend to create greater inequality of participation in one or both of the two types of conferences?

Third, are the textual features of drafts produced after the two types of conferences equivalent in terms of total length, t-unit length, and topical range?

Fourth, are both types of conferences equivalent in terms of students' ability to generate on-topic discourse in conferences, and to use the words and ideas that were generated in the conferences in their drafts?
CHAPTER FOUR: DATA ANALYSIS

For the most part, this chapter is organized around the research questions that were presented at the end of the previous chapter: Each research question is addressed in turn. However, sections on correlations of conference measures and correlations of draft measures will also be included to examine, respectively, the interrelationships between conference measures and draft measures for each of the two types of conferences.

Conference Measures Correlations

Pearson product-moment correlation matrices of conference measures were created in order to examine the relationships among the conference measures when considering graphic conferences only versus proximate conferences only. For these matrices, only the contributions of students were considered. The matrices were constructed using all available conference scores. As previously mentioned, each class participated in each conference condition twice. As a result, \( n \) for each correlation was the total number of students in the study (i.e., \( N = 27 \)) multiplied by 2, minus the total number of absences for that condition in all 4 weeks. The matrices are presented in Table 6.
Table 6

**Intercorrelations of Conference Measures**

<table>
<thead>
<tr>
<th></th>
<th>Number T-Units</th>
<th>Number Words</th>
<th>Words / T-Unit</th>
<th>Mean Average Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proximate Conference (n = 35)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Words</td>
<td>0.964**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words / T-Unit</td>
<td>-0.121</td>
<td>0.110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Average Cosine</td>
<td>-0.010</td>
<td>0.088</td>
<td>0.308</td>
<td></td>
</tr>
<tr>
<td>Average Max. Cosine</td>
<td>-0.003</td>
<td>0.045</td>
<td>0.158</td>
<td>0.420*</td>
</tr>
<tr>
<td><strong>Graphic Conferences (n = 49)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Words</td>
<td>0.948**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words / T-Unit</td>
<td>-0.304</td>
<td>-0.113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Average Cosine</td>
<td>-0.243</td>
<td>-0.096</td>
<td>0.662**</td>
<td></td>
</tr>
<tr>
<td>Average Max. Cosine</td>
<td>-0.012</td>
<td>0.100</td>
<td>0.464**</td>
<td>0.871**</td>
</tr>
</tbody>
</table>

* *p < 0.05  ** p < 0.01 (two-tailed)
In graphic conferences, both mean average cosine, and average maximum cosine were correlated to words per t-unit. This means that in graphic conferences, longer t-units were associated with higher cosine measures, or in other words, more relation to drafts. In contrast, no such correlation was found for proximate conferences. This was likely the case because many of the short t-units in the graphic conferences were either of a social nature (e.g., greetings), were used to express attitudes towards the communication (e.g., it’s interesting, I’m tired, etc.), or were used to concisely express agreement or disagreement (e.g., yes, I agree, no way, etc.; see Appendices C and E). On the other hand, these sorts of communication functions were not as frequently verbalized in proximate discussions, and if they occurred, they were usually conveyed paralinguistically. Even short t-units in proximate conferences tended to contain ideas related to the topic of discussion (e.g., see Appendices D and F).

In both types of conferences, conference mean average cosine and conference average maximum cosine were positively correlated. In other words, in both types of conferences, when a student’s conference t-units were related to at least one draft in a specific way, they were often related to other drafts in a general way. This
finding also suggests that both these LSA measures were capable of detecting the semantic similarity of students' contributions in both types of conferences with the drafts that followed them.

Comparisons of Conference Textual Features

In this section, the research question about comparisons of conference textual features was addressed. The question was, first, did both types of conferences allow students to produce an equivalent amount of discourse in an equivalent amount of time and was there a difference in the length of t-units contributed by participants in the two types of conferences? Second, was the discourse produced in both types of conferences equivalent in terms of lexical range, or did students in one type of conference tend to repeat the same words more than they did in the other type of conference? Each part of this question is addressed in turn.

Amount of Discourse and Length of T-units

The first part of my research question about conference textual features was to determine if there was a difference in the total number of words, the total number of t-units, or the average length of t-units generated by students in the two different types of conferences.
The total number of students in the proximate condition was different from the total number of students in the electronic condition each week. As a result, running t tests to compare the mean number of t-units or the mean number of words per student across conditions would have been inappropriate since the intention was to compare the total number of t-units and the total number of words produced by students in each condition.

To clarify, I will describe a case in which only two conferences were being compared. If there were 5 participants in one condition and 10 participants in the other, and if the total number of t-units were the same in both conferences, the mean number of t-units in the t test (i.e., the total number of t-units divided by the number of scores) for the conference with 5 students would be double the mean for the conference with 10 participants. Such a test would be a comparison of the average number of t-units per student rather than a comparison of the total number of t-units in the conferences. As a result, it was necessary to obtain totals of t-units and words contributed by students for each conference and run matched-sample t tests on the totals by condition. The totals are presented in Table 7; class is indicated in parenthesis.
Table 7

Comparison of Conference Textual Features Across Conference Condition

<table>
<thead>
<tr>
<th>Week</th>
<th>Total T-Units</th>
<th>Total Words</th>
<th>Words per T-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graphic</td>
<td>Proximate</td>
<td>Graphic</td>
</tr>
<tr>
<td>1</td>
<td>170 (4)</td>
<td>119 (4a)</td>
<td>1183 (4)</td>
</tr>
<tr>
<td>2</td>
<td>277 (4a)</td>
<td>111 (4)</td>
<td>2012 (4a)</td>
</tr>
<tr>
<td>3</td>
<td>180 (4)</td>
<td>251 (4a)</td>
<td>1343 (4)</td>
</tr>
<tr>
<td>4</td>
<td>245 (4a)</td>
<td>160 (4)</td>
<td>1575 (4a)</td>
</tr>
<tr>
<td>Mean</td>
<td>218.00</td>
<td>160.25</td>
<td>1528.00</td>
</tr>
<tr>
<td>S.D.</td>
<td>51.50</td>
<td>64.19</td>
<td>360.42</td>
</tr>
</tbody>
</table>
The test on total number of words in the conferences revealed no statistically significant main effect across conference conditions, $t(3) = -1.13, p > 0.05$ (two-tailed). In other words, the total number of words generated by students was not found to differ between the graphic ($M = 1528.00, SD = 360.42$) and proximate ($M = 1615.00, SD = 627.88$) conference conditions. However, because of the low power of this statistical test due to the small number of comparisons and the rather large amount of variability in number of words within each of the two types of conferences (Howell, 1982), this nonsignificant finding must be interpreted cautiously.

First, the entries in the middle of Table 7 for total words reveal a noticeable interaction between class and condition in that the students in class 4a consistently produced more words than the students in class 4, regardless of the conference condition. This illustrates the strength of the counterbalanced design used in this study: If the two classes had simply been assigned to different conditions, this difference, apparently due to class, may have been mistaken for a difference due to condition. In other words, if each group had participated in the same condition all 4 weeks, this design would have tended to systematically confound differences due to
condition, the variable under consideration, with pre-existing differences due to class, a potentially confounding variable.

Second, the difference in the means for total words between the two types of conferences over all 4 weeks is only 87 words, a mere 5.7% and 5.4% of the total words in the electronic and proximate conferences respectively. This strongly suggests that the conference condition did not have a significant effect on the total number of words produced by students.

The matched-sample t-test on total number of t-units in the conferences revealed no statistically significant main effect across conference conditions: \( t(3) = 0.24, p > 0.05 \) (two-tailed). In other words, no statistically significant difference was found in the total number of t-units produced by students in the graphic \((M = 218.00, SD = 51.50)\) and proximate \((M = 160.25, SD = 64.19)\) conferences. However, as shall become clear in a moment, interpretation of this nonsignificant finding will be more meaningful after considering differences in number of words per t-unit across conditions.

Unlike the comparisons of total t-units and words, unequal numbers of scores across conditions had no undue adverse effect on the comparison of average t-unit length
because a comparison of the mean of average t-unit length for students in one condition versus the mean of average t-unit length for students in the other condition was desired.

As a result, it was possible to use four independent-sample *t* tests, one for each week of the study, to compare the average number of words per t-unit contributed by students across conditions. Accordingly, the means for average t-unit length reported in Table 7 and displayed in Figure 2 cannot be obtained by dividing the total number of words contributed by all students in a given conference by the total number of t-units contributed by all students. In contrast, totaling the average t-unit length scores of the students in a given conference and dividing by the total number of students in that conference derived these means.

Although preliminary stem and leaf plots of words per t-unit scores for both conditions week-by-week revealed that all sampling distributions were mildly positively skewed, the parametric test was used because the unpooled independent-sample *t* test is robust to minor deviations in the assumption of normality, especially when the distributions being compared are skewed in the same direction (Howell, 1982).
Figure 2. Interaction of week and condition for conference words per t-unit.
In the analysis of words per t-unit, or t-unit length, across conditions, three out of four tests reached statistical significance. Since I did not predict which condition would result in longer t-units, two-tailed tests were used. The results of the tests from weeks 1 to 4 were, respectively: $t(18) = 2.09, p = 0.05$; $t(11) = 1.85, p > 0.05$; $t(14) = 3.48, p < 0.05$; and $t(8) = 3.48, p < 0.05$ (all two-tailed). A plot of the means is presented in Figure 2.

The difference in means during the week that did not reach statistical significance (i.e., week 2) was still in the direction of the main effect. These results suggest that, on average, students' spoken t-units (i.e., in proximate conferences; $M = 10.33, SD = 1.06$) were longer than their written ones (i.e., in graphic conferences; $M = 7.58, SD = 0.71$).

Given this result, the means for total number of t-units presented in Table 7 can now be interpreted in a clearer perspective. Since statistical analysis suggested that there was no appreciable difference in the number of words in the conferences across condition, and that there was an appreciable difference in t-unit length across conditions, one would expect that the condition that produced longer t-units (i.e., the proximate condition)
must have had fewer t-units. Still, as the t test on number of t-units across conditions revealed, this difference was not statistically significant. Moreover, it should be noted that since the t-units across conditions were not of comparable length, in this study, number of words should be considered a better comparative measure of conference participation than number of t-units.

**Lexical Range**

The other part of my research question about comparisons of conference textual features was to determine if one type of conference tended to be more lexically complex, or in other words, to exhibit greater lexical range, than the other. Wordsmith Tools (Scott, 1998) was used to calculate the standardized type-token ratios, or the proportion of new words to total words for all conferences. The standardized type-token ratios of the conferences are presented in Table 8. The class that took part in the conference is indicated in parenthesis.

Since previous research had revealed higher type-token ratios for the electronic condition (Warschauer, 1996), a one-tailed test was planned. As such, it was decided a priori that the null hypothesis of no difference between means would only be rejected if the mean for the electronic discussion was higher than the mean for the proximate
Table 8

*Comparison of Standardized Type-Token Ratios Across Conference Conditions*

<table>
<thead>
<tr>
<th>Week</th>
<th>Type-token Ratio</th>
<th>Type-token Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graphic</td>
<td>Proximate</td>
</tr>
<tr>
<td>1</td>
<td>43.50 (4)</td>
<td>34.73 (4a)</td>
</tr>
<tr>
<td>2</td>
<td>39.35 (4a)</td>
<td>34.90 (4)</td>
</tr>
<tr>
<td>3</td>
<td>42.20 (4)</td>
<td>36.69 (4a)</td>
</tr>
<tr>
<td>4</td>
<td>40.40 (4a)</td>
<td>36.85 (4)</td>
</tr>
<tr>
<td>Mean</td>
<td>41.36</td>
<td>35.79</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>1.85</td>
<td>1.13</td>
</tr>
</tbody>
</table>
discussion. The matched-sample t test, described in the previous chapter, revealed a statistically significant difference across conditions, \( t(3) = 4.89, p < 0.01 \) (one-tailed), suggesting that the discourse produced in graphic conferences \( (M = 41.36, SD = 1.85) \) was more lexically complex, or in other words, exhibited a greater lexical range, than the discourse produced in proximate discussions \( (M = 35.79, SD = 1.13) \).

**Comparisons of Conference Participation**

The second research question in this study dealt with conference participation. I wanted to know if the amount of researcher participation was equivalent in both types of conferences, if participation was equally balanced among students and between students and the researcher in both types of conferences, and if the addition of the researcher's discourse to the students' discourse tended to create greater inequality of participation in one or both of the two types of conferences. Each part of this question is addressed in turn.

**Researcher Participation**

The first question of interest regarding conference participation was to determine if there was a difference in the amount of discourse generated by the researcher in the two types of conferences. To examine this question,
comparisons across conference conditions of the researcher's participation as a percentage of total participation were made. As mentioned in the procedures section of this study, for all conferences, I led the discussion groups, and in each discussion, I attempted to contribute as little as possible. However, analysis of the conference transcripts clearly shows inequality in my participation. A summary of the data for my participation is presented in Table 9.

Since I attempted to contribute as little as possible in both conditions, and thus did not predict higher participation in one type of conference compared to the other, two-tailed probability levels were used to test the differences across conditions. Moreover, matched-sample t tests were used to examine the contributions of the researcher across conference conditions because in each week, it was the same person (i.e., the researcher) participating in both conditions. As a result, one would expect that if the researcher contributed a lot about a given topic (i.e., in a given week) in one condition, the researcher might also tend to contribute a lot about the same topic (i.e. in the same week) in the other condition.
Table 9

Comparison of Researcher Participation Across Conference Conditions

<table>
<thead>
<tr>
<th>Week</th>
<th>% T-Units</th>
<th>% Words</th>
<th>Words per T-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graphic</td>
<td>Proximate</td>
<td>Graphic</td>
</tr>
<tr>
<td>1</td>
<td>1.7</td>
<td>24.2</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>5.8</td>
<td>27.5</td>
<td>3.4</td>
</tr>
<tr>
<td>3</td>
<td>12.6</td>
<td>28.9</td>
<td>13.6</td>
</tr>
<tr>
<td>4</td>
<td>7.5</td>
<td>26.6</td>
<td>7.9</td>
</tr>
<tr>
<td>Mean</td>
<td>6.92</td>
<td>26.79</td>
<td>6.50</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>4.51</td>
<td>1.97</td>
<td>5.56</td>
</tr>
</tbody>
</table>
To perform the tests for each measure, my scores for each week in the graphic condition were subtracted from my scores for each week in the proximate condition. This resulted in a single column of difference scores for each measure. Two-tailed one-sample t tests were then run on the difference scores with the null hypothesis being that the means of the difference scores were equal to zero.

The percentage of t-units, percentage of words, and average length of t-units contributed by the researcher were all statistically significantly higher in the proximate condition ($M = 26.79$, $SD = 1.97$; $M = 28.55$, $SD = 3.47$; and $M = 11.06$, $SD = 1.26$ respectively) than the graphic condition ($M = 6.92$, $SD = 4.51$; $M = 6.50$, $SD = 5.56$; and $M = 5.78$, $SD = 2.03$ respectively). The t tests revealed the following results for percentage of t-units, percentage of words, and average length of t-units contributed by the researcher: $t(3) = 14.17$, $p < 0.001$; $t(3) = 19.60$, $p < 0.001$; and $t(3) = 5.54$, $p < 0.05$, respectively. In summary, despite my best efforts to minimize my participation in both conditions, I ended up using more words, more t-units, and longer t-units in the proximate conferences.
Equality of Participation Among Students and Among Students and the Researcher

The second part of the research question about conference participation was to see if participation was equally balanced among students and among students and the researcher in both types of conferences.

*Prima facie* evidence for greater equality of participation among students in graphic conferences could be found from two sources. First, the raw participation data revealed that there were 9 cases where students attending proximate conferences did not contribute a single t-unit: a total of 7 different students, 3 from one class, 4 from the other. Two students, one from each class, exhibited this behaviour both times they were in the proximate condition. In contrast, in each graphic conference, every student contributed at least one t-unit, including the students who said nothing at all in the proximate conferences.

Second, more evidence for equality of participation among students in the graphic condition was obtained by examining the standard deviations of number of words per student across the two conditions for all 4 weeks combined. The standard deviation for number of words contributed by students in the proximate condition was 161.3, compared to
84.7 in the graphic condition. Much more variation in participation was apparent in the proximate conference condition, which suggests that participation in the graphic condition was more balanced.

To obtain a more definitive measure of the equality of participation in conferences, the gini coefficient was calculated for each of the conferences. The gini coefficient is a measure of inequality that yields a value between 0 (i.e., no inequality) and 1 (i.e., maximum inequality). Since this coefficient is not in common use in Educational literature, a general discussion of the calculation and rationale underlying this coefficient is offered in Appendix K.

For the purposes of this study, the number of words per student was used to calculate the participation percentage per speaker, which was then used to determine the gini coefficient for each group. Number of words, rather than number of t-units, was used to calculate the participation percentages because, as mentioned in the section on amount of discourse and length of t-units in this chapter, number of words was determined to be a better comparative measure of conference participation. Coefficients obtained for participation in the two types of conferences when considering only the students'
participation, and when considering the students’ and the researchers’ participation combined are presented in Table 10. Classes are indicated in parentheses.

Since previous research (Warschauer, 1996) had revealed lower gini coefficients (i.e., more equality) for the graphic condition, and since single scores were being compared, for the analyses of balanced participation among the students only and among the students and the researcher combined, one-tailed matched-sample t tests were planned. Since the test was one-tailed, it was decided a priori that the null hypothesis of no difference between means would be rejected only if the mean coefficient for the graphic discussions was lower (i.e., less inequality) than the mean coefficient for the proximate discussions.

The matched-sample t tests revealed a statistically significant difference in gini scores across conditions considering the participation of students only, \( t(3) = -6.02, p < 0.01 \) (one-tailed), and considering the participation of the students and the researcher combined, \( t(3) = -4.37, p < 0.05 \) (one-tailed). This suggests that participation in the graphic conferences for students only \((M = 0.371; SD = 0.081)\) and for the students and the researcher combined \((M = 0.377; SD = 0.079)\) was more
Table 10

Comparison of Equality of Participation Across Conference Conditions

<table>
<thead>
<tr>
<th>Week</th>
<th>Gini Coefficients for Students Only</th>
<th>Gini Coefficients for Students and Researcher Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Graphic</td>
<td>Proximate</td>
</tr>
<tr>
<td>1</td>
<td>0.438 (4)</td>
<td>0.498 (4a)</td>
</tr>
<tr>
<td>2</td>
<td>0.301 (4a)</td>
<td>0.446 (4)</td>
</tr>
<tr>
<td>3</td>
<td>0.445 (4)</td>
<td>0.578 (4a)</td>
</tr>
<tr>
<td>4</td>
<td>0.300 (4a)</td>
<td>0.419 (4)</td>
</tr>
<tr>
<td>Mean</td>
<td>0.371</td>
<td>0.485</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.081</td>
<td>0.070</td>
</tr>
</tbody>
</table>
balanced than in the proximate conferences for students only ($M = 0.485; SD = 0.070$) and for the students and the researcher combined ($M = 0.517; SD = 0.066$).

However, Table 10 reveals that the variance in gini coefficients for class 4a was much greater than the variance for class 4. In fact, Levene's test (i.e., a test of homogeneity of variance that is suitable for any distribution) of the student-only and student-plus-teacher coefficients across classes yielded $p$ values of 0.002 and 0.007 respectively, indicating a statistically significant difference in variances across classes. To understand and visualize this effect more easily, these data have been plotted in Figures 3 and 4. In Figures 3 and 4, the data labels G and P stand for graphic and proximate conferences, respectively.

As these figures depict, the effect of conference condition on equality of participation interacted with class. Specifically, these figures suggest that the effect of conference condition on equality of participation was greater for class 4a than for class 4. In other words, for class 4, conference participation was almost equally balanced, or equally unbalanced, regardless of the conference condition, while for class 4a, participation in
Figure 3. Gini by class – students only.
Figure 4. Gini by class - students and teacher.
conferences were considerably more balanced when they were held in the graphic condition.

The Effect of the Researcher on the Equality of Participation

The third question of interest in this section on conference participation was to see what influence the researcher had on balanced participation in the two types of conferences. In order to test the influence of the researcher on balanced participation, two matched-sample t tests were performed. To run these tests, the gini scores for the teacher and students combined were subtracted from the gini scores for the students only for all 4 graphic conferences and for all 4 proximate conferences. This resulted in a single column of difference scores for the 4 graphic conferences and a single column of difference scores for the 4 proximate conferences. Two-tailed one-sample t tests were then run on the two sets of difference scores with the null hypotheses being that the mean of each set of difference scores was equal to zero.

For the graphic conferences, the two-tailed matched-sample t tests revealed no statistically significant difference \( t(3) = -0.35, p > 0.05 \) between gini coefficients for the students only \( (M = 0.371, SD = 0.081) \), compared with gini coefficients for the students and
researcher combined ($M = 0.377$, $SD = 0.079$). In other words, the addition of the researcher’s contributions to the graphic conferences did not tend to create greater inequality in participation.

On the other hand, for the proximate conferences, the two-tailed matched-sample $t$ test revealed a statistically significant difference ($t(3) = -4.90$, $p < 0.05$) between the gini coefficients for the students only ($M = 0.485$, $SD = 0.070$), compared with the gini coefficients for the students and researcher combined ($M = 0.517$, $SD = 0.066$). In other words, the addition of the researcher’s contributions to the proximate conferences tended to create greater inequality in participation.

**Draft Measures Correlations**

Before moving into the next research question regarding the textual features of drafts, it will be advantageous to examine the intercorrelations of draft measures. Pearson product-moment correlation matrices of draft measures were created in order to examine the relationships between the draft measures when considering drafts written after graphic conferences (hereafter: graphic drafts) and drafts written after proximate conferences (hereafter: proximate drafts). The correlation matrices are presented in Table 11. No differences in the
Table 11

Intercorrelations of Draft Measures

<table>
<thead>
<tr>
<th></th>
<th>Number Words</th>
<th>Number Words</th>
<th>Words / T-Unit</th>
<th>Vector Length</th>
<th>Average Max. Cosine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T-Units</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Words</td>
<td>0.846***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Words / T-Unit</td>
<td>-0.323</td>
<td>0.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vector Length</td>
<td>0.814***</td>
<td>0.929***</td>
<td>0.114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Max. Cosine</td>
<td>0.158</td>
<td>0.047</td>
<td>-0.167</td>
<td>0.035</td>
<td></td>
</tr>
<tr>
<td>Total Wds. / New Wds.</td>
<td>0.004</td>
<td>-0.092</td>
<td>-0.229</td>
<td>-0.069</td>
<td>0.596***</td>
</tr>
</tbody>
</table>

Proximate Drafts (n = 37)

|                                |              |              |                |               |                     |
|                                | Number Words |              |                |               |                     |
| Number Words                   | 0.945***     |              |                |               |                     |
| Words / T-Unit                 | -0.048       | 0.225        |                |               |                     |
| Vector Length                  | 0.937***     | 0.948***     | 0.147          |               |                     |
| Average Max. Cosine            | 0.270        | 0.252        | -0.001         | 0.348         |                     |
| Total Wds. / New Wds.          | 0.183        | 0.125        | -0.170         | 0.167         | 0.685***            |

Graphic Drafts (n = 31)

***p < 0.001 (two-tailed)
patterns of correlations were observed between these matrices.

For drafts following both proximate and graphic conferences, the ratio of total words to new words was strongly correlated with average maximum cosine. This suggests that the exact linguistic overlap ratio measure obtained by Corpus Linguistics was consistent with average maximum cosine. In other words, if when writing their drafts, students tended to repeat verbatim or closely paraphrase what was contributed during the conference, using some of the exact same words, students received high scores on both of these measures.

For both correlation matrices in Table 11, draft vector length correlated significantly with number of words and number of t-units. This implies that when students wrote longer drafts, they generally tended to include more topics.

Comparisons of Draft Textual Features

My third research question dealt with the textual features of drafts produced after the two types of conferences. I wanted to know if students produced drafts that were longer, had longer t-units, or that covered a wider range of topics after one of the two types of conferences.
Two-sample t tests on each of these measures week-by-week revealed no significant differences in drafts across conditions in terms of number of t-units or words, t-unit length, or breadth (i.e., average vector length). The results for these tests are presented in Table 12.

The statistically nonsignificant findings for all weeks for the measures in Table 12 suggests that there is no reason to believe that conference condition had any systematic influence on number of t-units or words, t-unit length, or vector length of drafts. In other words, there is no reason to believe that students tended to produce drafts that had more t-units, more words, longer t-units, or that covered a wider range of topics after one type of conference compared to the other.

The Semantic Relation of Conferences and Drafts

The final research question in this study asked if both types of conferences were equivalent in terms of students' ability to generate on-topic discourse in conferences, and in terms of students' ability to use the words and ideas that were generated in the conferences in their drafts. Each part of this question is discussed in turn.
<table>
<thead>
<tr>
<th>Vector Length</th>
<th>Words / T-Unit</th>
<th>Words</th>
<th>T-Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.160</td>
<td>13.830</td>
<td>13.250</td>
<td>18.066</td>
</tr>
<tr>
<td>1.160</td>
<td>12.650</td>
<td>10.600</td>
<td>15.130</td>
</tr>
<tr>
<td>1.160</td>
<td>3.820</td>
<td>24.800</td>
<td>2.000</td>
</tr>
<tr>
<td>1.160</td>
<td>2.820</td>
<td>2.800</td>
<td>1.984</td>
</tr>
<tr>
<td>1.160</td>
<td>2.880</td>
<td>2.800</td>
<td>1.840</td>
</tr>
<tr>
<td>1.160</td>
<td>2.800</td>
<td>1.800</td>
<td>0.800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Graphic Proximate</th>
<th>Mean</th>
<th>dp</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Week 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of Draft Textual Features Across Conference Conditions

Table 12
The Relation of Conferences to Drafts

The question of interest in this section was to determine if one of the two types of conferences tended to be more on topic than the other. To investigate this research question, two-tailed independent-sample t tests were run on the data collected in each week of the study across conditions for mean average cosine and average maximum cosine. The results of these tests are presented in Table 13.

Table 13 reveals no significant differences for any of the weeks for mean average cosine. This finding suggests that students were able to focus equally well on the topic in both conference conditions. That is, in both conference conditions, students were equally able to introduce t-units that were on topic in that the terms and ideas contained in conference t-units were equally likely to be used in many drafts. A plot of the means for both classes and conditions for this measure are reported in Figure 5, data labels indicate the class. Figure 5 shows that for 3 out of 4 weeks, the means for this measure were higher in the proximate conference condition.

As Table 13 indicates, comparisons of conference average maximum cosine across each week reached statistical significance for weeks 1 and 2, but not for weeks 3 and 4.
<table>
<thead>
<tr>
<th>Week</th>
<th>Condition</th>
<th>Graphic Proximity</th>
<th>Proximity</th>
<th>Mean</th>
<th>df</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Week 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Week 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Week 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Week 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparison of on-topic discourse across conference conditions

Table 13
Figure 5. Interaction of week and condition for conference mean average cosine.
These results suggest that there is moderate evidence to propose that students in the proximate conferences were better able to introduce t-units that were on topic, in the sense that proximate conference t-units were more likely to be closely paraphrased or repeated verbatim in at least one of the drafts. A plot of the means for average maximum cosine is presented in Figure 6. Figure 6 reveals that for 3 out of 4 weeks, the means for this measure were higher in the proximate conference condition.

In sum, for both LSA conference measures, the means were higher in the proximate condition 3 out of 4 weeks. In addition, for conference average maximum cosine, the higher means in the proximate condition were statistically significantly higher 2 out of 3 times. Thus, considering both these measures, there is modest evidence to suggest that, compared with t-units made during the graphic conferences, t-units made during the proximate conferences were, on average, more on topic.

**The Relation of Drafts to Conferences**

The question of interest in this section was to determine if the drafts written after one conference condition were more closely related to the conference than the drafts written after the other conference condition.
Figure 6. Interaction of week and condition for conference average maximum cosine.
To investigate this research question, two-tailed independent-sample t tests were run on the data collected in each week of the study across conditions for draft average maximum cosine and ratio total words to new words.

The results of these tests are presented in Table 14. Table 14 shows that the means for average maximum cosine across conference conditions revealed a statistically significant difference only for week 4. These results provide little statistical evidence to suggest that drafts written after one type of conference were more related to the conference than drafts written after the other type of conference. A plot of the means for both classes and conditions by week is presented in Figure 7. Figure 7 shows that for 3 out of 4 weeks, the means for this measure were higher in the proximate conference condition.

As Table 14 depicts, the t tests for ratio total words to new words reached statistical significance across conference condition for weeks 1 and 3, but not weeks 2 and 4. These results suggest that there is moderate evidence to believe that the drafts written after the proximate conferences were more related to the conferences. More specifically, these results suggest that proximate drafts tended to rely on words that had been used in the proximate
<table>
<thead>
<tr>
<th>Class</th>
<th>Mean</th>
<th>Mean</th>
<th>Mean</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>0.498</td>
<td>0.510</td>
<td>0.520</td>
<td>0.531</td>
</tr>
<tr>
<td>Week 2</td>
<td>0.471</td>
<td>0.483</td>
<td>0.490</td>
<td>0.497</td>
</tr>
<tr>
<td>Week 3</td>
<td>0.468</td>
<td>0.482</td>
<td>0.490</td>
<td>0.497</td>
</tr>
<tr>
<td>Week 4</td>
<td>0.467</td>
<td>0.480</td>
<td>0.490</td>
<td>0.497</td>
</tr>
</tbody>
</table>

Comparison of the Similarity of Digits to Conference Across Conference Conditions

Table 14
Figure 7. Interaction of week and condition for draft average maximum cosine.
conference more than graphic drafts tended to rely on words that had been used in the graphic conferences. A plot of the means for ratio total words to new words for both classes and conditions is presented in Figure 8. Figure 8 reveals that for 3 out of 4 weeks, the means for this measure were higher in the proximate conference condition.

In sum, for both draft measures discussed in this section, the means were higher in the proximate condition 3 out of 4 weeks. Furthermore, for draft average maximum cosine, the higher means in the proximate condition were statistically significantly higher 1 out of 3 times. Moreover, for draft ratio total words to new words, the higher means in the proximate condition were statistically significantly higher 2 out of 3 times. In sum, considering both these measures, there is modest evidence to suggest that the drafts written after proximate conferences were more related to the conferences than drafts written after the graphic conferences.
Figure 8. Interaction of week and condition for draft ratio total words to new words.
Pre-writing conferences are often used in second language writing classrooms as a way of getting groups of students to generate language and ideas about a topic that can be used in subsequent individual writing. Such pre-writing activities have long been conducted as face-to-face conferences (Reid, 1993), or in other words, proximate conferences. In proximate conferences, the students and the teacher discuss the writing topic while seated around a table as a whole-class discussion. More recently, computers have been used as a medium for group pre-writing conferences. In these computer-mediated conferences, or in other words, graphic conferences, students and the teacher can also discuss the writing topic as a whole class. However, they do so by typing messages that can be viewed by all other participants while each participant is seated at his or her own computer in a computer laboratory. This study compared proximate and graphic conferences as pre-writing activities. In this chapter, after the rationale and objectives of this study are summarized, and the participants are described, the conclusions reached are stated, and the implications for practice, theory, and further research are addressed.
Summary of Rationale and Research Objectives

The use of graphic pre-writing conferences as an alternative to proximate pre-writing conferences in ESL education has not been unmotivated. Research (e.g., Beauvois, 1992, 1998; Braine, 1997, 2001; Chun, 1994; Kelm, 1992; Kern, 1995; Slatin, 1998; Sullivan & Pratt, 1996; Warschauer, 1996; Warschauer et al., 1996) suggests that graphic conferences shape patterns of teacher and student participation and discourse in ways that may be beneficial to second language students. More specifically, research suggests that the graphic mode of communication serves a democratizing function: tipping the balance of participation patterns towards the students and away from the teacher; facilitating balanced participation among students; and enabling students to take control of the discourse, or in other words, to display interactive competence. Moreover, research (Warschauer, 1996) suggests that the discourse in graphic conferences is more lexically complex, or in other words, has a higher proportion of new words to total words than discourse in proximate conferences. However, much of the research that compares participation patterns and discourse in graphic and proximate pre-writing conferences in the second language
classroom has been anecdotal, and the empirical studies that have been conducted have remained unreplicated.

Research (e.g., Braine, 1997, 2001; Ghaleb, 1993; Sullivan & Pratt, 1996) comparing the relative benefit of proximate and graphic pre-writing conferences on students' first drafts has remained inconclusive. One of the problems with such studies is the questionable choice of holistic measures used to compare the quality of first drafts written after the two types of conferences. Holistic measures are a questionable way to evaluate first drafts because holistic measures award higher scores for features of writing that, according to process writing theory, are not ordinarily incorporated into first drafts.

Another, and perhaps more pressing problem with research into the relative benefit of proximate and graphic pre-writing conferences on students' first drafts is that the studies that have been conducted (e.g., Braine, 1997, 2001; Ghaleb, 1993; Sullivan & Pratt, 1996) offer no direct evidence to show that the words and ideas generated in either type of conference are transferred to the drafts. Since no evidence is offered to show that words and ideas are transferred from conferences to drafts, it is difficult to tell if the conferences have any influence on the drafts.
at all, much less a relatively superior or inferior benefit.

The rationale for my current study is derived from a need for further research. As mentioned, the majority of research that compares participation patterns and discourse in proximate and graphic pre-writing conferences has been anecdotal, and the empirical studies that do exist have remained unreplicated. Moreover, research comparing the benefit of proximate and graphic pre-writing conferences on first drafts has been inconclusive. As a result, additional empirical research in these areas using appropriate quantitative measures is warranted (Braine, 1997; Ortega, 1997; Warschauer, 1996).

To answer this need for further research, this study addressed several research questions. I wished to determine the following. First, is there a difference in textual features between the two types of conferences? In particular, do both types of conferences allow students to produce an equivalent amount of discourse in an equivalent amount of time and is there a difference in the length of t-units contributed by participants in the two types of conferences? Moreover, is the discourse produced in both types of conferences equivalent in terms of lexical range, or do students in one type of conference tend to repeat the
same words more than they do in the other type of conference?

Second, with regard to conference participation, is the amount of researcher participation equivalent in both types of conferences; is participation equally balanced among students, and among the students and the researcher combined in both types of conferences; and does the addition of the researcher's discourse to the students' discourse tend to create greater inequality of participation in one or both of the two types of conferences?

Third, are the textual features of drafts produced after the two types of conferences equivalent in terms of total length, t-unit length, and topical range?

Finally, are both types of conferences equivalent in terms of students' ability to generate on-topic discourse in conferences, and to use the words and ideas that were generated in the conferences in their drafts?

**Participants**

The participants in this study were high intermediate to advanced students of English as a second language. These students were placed in level 4 of the Intensive English Language Program (IELP) at Brock University based on their Michigan Test scores at the beginning of the term.
in which this study was conducted. Since these classes were at the same proficiency level in the program based on their Michigan Test scores, all students were of comparable overall language proficiency at the outset of the study. In addition, the two class groups were approximately balanced in terms of number of students, gender, and language and cultural background. Since the participants in this study were high intermediate to advanced ESL students, there is no good reason to generalize these results to students at other levels of language proficiency or native language students.

Conclusions

In this section, my conclusions with respect to each of the research questions stated in the section on summary of rationale and research objectives in this chapter will be discussed.

Research Question 1 - Conference Textual Features

The first research question addressed possible differences in textual features between the two types of conferences. In particular, I wanted to know if both types of conferences would allow students to produce an equivalent amount of discourse in an equivalent amount of time, and if there would be a difference in the length of t-units contributed by participants in the two types of
conferences. Moreover, I wanted to find out if the discourse produced in both types of conferences would be of equivalent lexical range, or if students in one type of conference would tend to repeat the same words more than they did in the other type of conference.

**Amount of discourse.** No research to date has compared the total amount of discourse generated by students in graphic conferences with the total amount of discourse generated by students in proximate conferences. My analysis revealed that given equivalent amounts of time, students participating in graphic conference were able to write an equivalent number of words, and an equivalent number of contributions¹ as they could speak in proximate conferences. In other words, no statistically significant difference was found in total numbers of words or contributions produced by students in the two types of conferences.

This result may seem surprising since students were speaking in proximate conferences and writing in graphic conferences. However, this result is understandable for two reasons. First, as will be discussed in the upcoming

---

¹ To improve ease of readability in this chapter, I refer to the quantity of discourse in conferences as the number or length of "contributions." However, in fact, conference discourse was broken up into t-units. As a result, the number and the length of contributions made by students in conferences really refer to the number and the length of t-units. For the definition of t-units used in this study, see the section on the effect of electronic conferences on writing in chapter 2, and the glossary of terms in Appendix A.
section on conference participation, the conference facilitator (i.e., myself) contributed more during proximate conferences, perhaps taking time away from students to contribute. Second, as was discussed in the review of the literature, students did not need to take temporally sequential turns during graphic conferences (Beauvois, 1992, 1998; Kelm, 1992). Since students did not need to wait for their turn to contribute, during graphic conferences students could contribute as much as they wanted to, whenever they wanted to.

**Length of T-units.** Although analysis revealed that students produced equivalent numbers of words and contributions in the two conference conditions, analysis also revealed that contributions in the proximate conferences were statistically significantly longer than contributions in the graphic conferences. As shall become clearer in the next section, this was likely the case because, compared to the proximate conferences, the graphic conferences contained many more occurrences of a category of language functions that were expressed with short t-units.

**Lexical range.** The third part of this research question about conference textual features enquired if the discourse produced in both types of conferences would be of
equivalent lexical range, or if students in one type of conference would tend to repeat the same words more than they did in the other type of conference. In confirmation of previous research discussed in the review of the literature (Warschauer, 1996), analysis revealed that the discourse in graphic conferences covered a greater lexical range to a statistically significant degree. To put it another way, discourse in the graphic conferences had a statistically significantly higher proportion of new words to total words than discourse in proximate conferences.

Since the graphic conferences had a greater lexical range, one may be tempted to conclude that, in graphic conferences, students presented more diverse information about the topic. However, there is some reason to doubt that the greater lexical range in graphic conferences indicates that students used a greater range of topic-related words. In fact, the greater lexical range is likely a result of more short, off-topic comments in the graphic conferences. I shall return to this point in my discussion of the last research question.

Research Question 2 - Conference Participation

The second research objective in this study was about participation in the two types of conferences. I wanted to know if the amount of researcher participation would be
equivalent in both types of conferences; if participation would be equally balanced among students, and among the students and the researcher combined in both types of conferences; and if the addition of the researcher’s discourse to the students’ discourse would tend to create greater inequality of participation in one or both of the two types of conferences.

**Researcher Participation.** Although I attempted to minimize my participation in both conditions, during the proximate conferences, I contributed more than I did during the graphic conferences. Scores for the percentage of words I contributed, the percentage of t-units I contributed, and the length of my t-units were all higher in the proximate conferences to a statistically significant degree.

As I discussed in the review of the literature, this result does not come as much of a surprise (Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996). Like the teacher in Sullivan and Pratt’s (1996) study, in an effort to get students to participate, I ended up dominating the proximate class discussions. I felt myself drawn into proximate discussions; any long moments of silence seemed too much to bear in silence. On several occasions, I felt it was absolutely necessary for me to
bring students back to the topic through relatively lengthy summaries. Moreover, it seemed necessary for me to referee the discussion by deciding whose turn was next. Although I tried to resist this role, the students seemed to expect, even demand that I take it upon myself.

On the other hand, during graphic conferences, students seemed to require less guidance from me. They seemed quite capable of eliciting comments from classmates, questioning for clarification where necessary, and generally running their own show. While I always felt myself central to the proximate conferences, during electronic discussions I often felt myself to be on the periphery: one voice amongst many.

Balanced participation in conferences. As discussed in the review of the literature, previous research had found that participation in graphic conferences is more balanced among students, and between the students and the group leader, than participation in proximate conferences (e.g., Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996).

In confirmation of these research findings, my analysis revealed that participation was statistically significantly more balanced in graphic conferences. Moreover, when the classes were considered separately,
analysis revealed that for class 4a, participation in the graphic conferences was considerably more balanced among students than participation in proximate conferences. On the other hand, examination of the scores for balanced participation for class 4 revealed that conference condition had almost no impact on the degree to which participation was balanced among students. Although it was not possible to test these findings statistically, examination of the means strongly support these conclusions. This suggests that conference condition was not the only factor influencing balanced participation in conferences.

As discussed, the groups in this study were intact class groups. As such, the students had already had the chance to get to know each other quite well before this study began, and as a result, the group dynamic was already established. A difference in established group dynamics might help to explain the observed difference in patterns of participation. For instance, group 4a may have had a dynamic in which group members often vie for control of the discussion. On the other hand, group 4 may have had a more egalitarian approach to the group communication process in which members were more concerned with encouraging balanced participation. If this were so, then group 4 would likely
have carried their egalitarian group dynamic through, regardless of the conference condition. In other words, in both conference conditions, students would likely have been concerned with regulating their conversations so that everyone had a chance to participate.

On the other hand, if group 4a had a dynamic in which the inclusion of quieter members was not a concern, it would have been easier for a small number of members to take over the proximate conferences since they could have occupied most of the discussion time without allowing other students to take a turn. On the other hand, as noted in the review of the literature, in the electronic conferences it would have been more difficult for a small number of students to dominate the discussion since turn taking was unnecessary (Beauvois, 1992, 1998; Kelm, 1992). I will return to this point in my discussion of implications for further research.

*The effect of the researcher on balanced participation.* In confirmation of previous findings that were discussed in the review of the literature (Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996), my analysis revealed that when the group leader's input (i.e., my input) was considered in addition to the student's input, participation patterns in proximate
conferences became more unbalanced to a statistically significant degree. On the other hand, the addition of my comments to the graphic conferences did not affect the extent to which participation in the conferences was balanced among participants to a statistically significant degree.

As mentioned, I attempted to contribute as little as possible in both types of conferences. However, despite my intentions, not only did I contribute more words, more t-units, and longer t-units in the proximate conferences, I contributed so much more discourse to the proximate discussions that my discourse created considerably more inequality in participation in the group as a whole.

The results of this aspect of my study confirm the view discussed in the review of the literature that graphic conferences tend to be more student centred, while proximate conferences are more teacher centred (e.g., Beauvois, 1992, 1998; Braine, 1997; Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996; Warschauer, 1996).

**Research Question 3 - Draft Textual Features**

This research question enquired if drafts produced after the two types of conferences would be of equivalent total length, t-unit length, and topical range. No research studies to date have examined these questions.
When the data were examined, no evidence was found to suggest that students produced drafts that were longer, had longer t-units, or covered a wider range of topics after one type of conference compared to the other. In other words, no statistically significant difference was found between the drafts written after the two types of conferences in number of words, number of t-units, t-unit length, or vector length.

In isolation, these results suggest that both types of conferences equally prepared students for the drafting task. For instance, if the results had indicated that drafts after one type of conference were consistently longer and covered a greater topical range than drafts following the other, one might conclude that the former type of conference better prepared students for writing. However, the results in this section can only indicate that if there was a difference in the degree to which conferences prepared students for the drafting task, students must have compensated for this by individually coming up with their own ideas after the conferences.

**Research Question 4 - The Relation of Conferences and Drafts**

The final research question was about the relation of conferences and drafts. I wanted to know if both types of
conferences were equivalent in terms of students' ability to generate on-topic discourse in conferences, and to use the words and ideas that were generated in the conferences in their drafts.

**On-topic discourse in conferences.** The question of interest in this section was to compare how on topic discourse was in the two types of conferences. In my review of the literature, I found no studies that addressed this question. My analysis revealed modest evidence to suggest that discourse in the proximate conferences was, on average, more on topic than discourse in the graphic conferences.

For 2 out of 4 weeks, there was a statistically significant difference between the two types of conferences in the proportion of statements that were paraphrased or repeated in the drafts. During these 2 weeks, proximate conferences contained a statistically significantly higher proportion of statements that were repeated or paraphrased in the drafts.

Moreover, considering means only, for 3 out of 4 weeks, proximate conferences contained a higher proportion of statements that were repeated or paraphrased in the drafts, and a higher proportion of words that were widely applicable to the drafts. Taken together, these results
provide modest evidence to suggest that the proximate conferences contained a higher proportion of on-topic words and statements than graphic conferences.

I mentioned in my discussion of the first research question, in the section on lexical range, that there was some reason to doubt that the greater lexical range in graphic conferences meant that students had used a greater range of topic-related words. The results in this section provide evidence for this assertion.

Since the graphic conferences contained a larger proportion of off-topic statements, these off-topic statements may have caused the greater lexical range. To determine if this was indeed the case, I examined the off-topic graphic conference contributions. My examination revealed that compared with off-topic statements in proximate conferences (e.g., see the contributions presented of the end of Appendices F and O), off-topic contributions in graphic conferences (e.g., see the contributions presented of the end of Appendices E and N) more often performed interactive language functions such as greeting and leave-taking, expressing attitudes toward the communication, requesting clarification, or expressing agreement or disagreement.
In sum, since the graphic conferences contained a greater variety of interactive language functions, the graphic conferences likely had a greater lexical range because students used a greater variety of words that were not lexically related to the topic of discussion. On the other hand, proximate conferences had less lexical range because even off-topic comments contained words that were more restricted to the topic of discussion.

As I discussed in the review of the literature, this was likely the case because of the lack of paralinguistic cues in the graphic conferences. In graphic conferences, information that in proximate conferences can be communicated through paralinguistic cues, attitudes like boredom, curiosity, or agreement for instance, must be expressed graphically, in explicit language (Kiesler et al., 1984).

My interpretation of the significance of lexical range in proximate and graphic conferences casts doubt on naïve claims that graphic conferences are superior to proximate conferences as pre-writing activities because of the increased lexical range. Indeed, findings of increased lexical range in graphic conferences, such as were discussed in the review of the literature (e.g., Warschauer, 1996), must be interpreted carefully. In
particular, such findings should be accompanied by information about the sorts of words, or function of words, that constitute the increased lexical range in graphic conferences.

For example, I mentioned that the short contributions in graphic conferences, which I have characterized as off topic, performed interactive language functions such as initiating topics, requesting clarification, and expressing greetings and leave-takings. As a result, my results in this section, in combination with my finding that discussions in graphic conferences were more student centred, can be seen as providing further support for other researchers' findings which, as discussed in the review of the literature, suggest that second language students in graphic conferences display more interactive competence, or in other words, more control of the discussion (Chun, 1994; Kelm, 1992; Kern, 1995; Sullivan & Pratt, 1996; Slatin, 1998).

The relation of drafts to conferences. The question of interest in this section was to compare how related drafts were to the two types of conferences. No research studies to date have examined this question. My analysis revealed modest evidence to suggest that drafts written after the proximate conferences were, on average, more
related to the proximate conferences than drafts written after the graphic conferences were related to the graphic conferences.

For 1 out of 4 weeks, there was a statistically significant difference between the two types of drafts in the proportion of statements that were closely related to statements made in the conference. During this week, proximate drafts contained a statistically significantly higher proportion of statements that were closely related to statements made in the conference. Furthermore, for 2 out of 4 weeks, there was a statistically significant difference between the two types of drafts in the degree to which drafts relied on words that had been used in the conference. During these weeks, proximate drafts contained a statistically significantly lower proportion of words that had not been used in the conference.

Moreover, considering means only, for 3 out of 4 weeks, proximate drafts contained a higher proportion of statements that were closely related to statements made in the conference, and a lower proportion of words that had not been used in the conference. In other words, the graphic drafts had a lower proportion of statements that were related to the conference, and a higher proportion of words that had not been used in the conference. All in
all, these results provide modest evidence to suggest that
the drafts written after the proximate conferences were, on
average, more related to the proximate conferences than
drafts written after the graphic conferences were related
to the graphic conferences.

In sum, these results provide modest evidence to
suggest that proximate conferences better prepared students
for the drafting task. This may have occurred for two
possible reasons. First, since the two types of
conferences were equivalent in numbers of words and
contributions, and since there was a higher proportion of
off-topic statements in the graphic conferences that
expressed interactive language functions, it is reasonable
to conclude that the increased amount of interactive
discourse left students with less time to develop ideas to
use in their drafts, forcing students to come up with their
own ideas afterwards. The other possibility is that
students were less able to recall the ideas produced in the
graphic conferences.

Implications for Practice

My research revealed some features that differentiate
graphic conferences from proximate conferences.
Communication in graphic conferences was more student
centred, allowing students, rather than the teacher, to
produce a higher proportion of the discourse. Even though I found that students' statements in graphic conferences were, on average, less on topic, this was likely the case because students expressed more interactive language functions in these conferences. By expressing interactive language functions, students took more control of the discourse, or in other words, displayed more interactive competence. As was discussed in the review of the literature, practicing these sorts of language functions is considered to be an important part of learning to communicate in a second language (Chun, 1994; Kelm, 1992; Kern, 1995; Slatin, 1998; Sullivan & Pratt, 1996).

Moreover, in comparison to proximate conferences, graphic conferences supported patterns of participation that were, if not more balanced, at least as balanced as communication among students, and among students and the researcher combined.

I also found that my twin roles of instructor and researcher came into some conflict. Although, as researcher, I wanted to minimize my participation in both types of conferences, during graphic conferences, as instructor, I felt myself compelled to take a more active role. During proximate conferences, my research-motivated desire to minimize my participation was outweighed by the
responsibility I felt as an instructor to fulfill students' classroom expectations by presiding over the proximate conferences.

Overall, my research demonstrated that graphic conferences are a viable alternative to proximate conferences as a way of getting ESL students to generate language and ideas about a topic that they can use in subsequent writing. Even though discourse was written rather than spoken, students generated equivalent amounts of discourse in equivalent time periods. Moreover, students produced drafts of equivalent length and topical range following both types of conferences.

On the other hand, there was modest evidence to suggest that the drafts written after the proximate conferences were more related to the proximate conferences than drafts written after the graphic conferences were related to the graphic conferences. In other words, a higher proportion of the words and ideas used in drafts following proximate conferences could be traced back to the proximate conferences. In comparison, a lower proportion of the words and ideas used in drafts following graphic conferences could be traced back to the graphic conferences.
In sum, I found that the class groups did not make as much use of the ideas presented in graphic conferences when writing drafts. I mentioned that this could have occurred for two possible reasons. This may have occurred because students did not generate as many ideas in the graphic conferences in an equivalent period of time, due to the increased amount of time and discourse directed at comments that served an interactive function. If so, then teachers using graphic pre-writing conferences should allow students more time to develop ideas than they would for proximate pre-writing conferences.

If this occurred because students were not as able to recall the ideas presented in graphic conferences, there is a relatively straightforward way to mitigate this apparent disadvantage. As mentioned in the review of the literature, all computer-mediated communications have the characteristic of being permanent in the sense that communications can easily be logged and printed out. As a result, students can easily be provided with a copy of the transcript to which they can refer when writing. Others (e.g., Kern, 1995) have suggested providing students with graphic conference transcripts, and given the results of my study, there is some reason to believe that this would be a particularly supportive practice.
Implications for Theory

As I discussed in the introduction to this chapter, second language teachers have come to be enthusiastic about using graphic conferences in their writing classrooms because of several features identified by researchers. In other words, research has identified several relatively consistent aspects of graphic conferences that have generally come to form a theoretical base to justify and promote the use of graphic pre-writing conferences. For instance, as discussed in the review of the literature, research (e.g., Beauvois, 1992, 1998; Braine, 1997, 2001; Chun, 1994; Kelm, 1992; Kern, 1995; Slatin, 1998; Sullivan & Pratt, 1996; Warschauer, 1996; Warschauer et al., 1996) has come to show that the graphic mode of communication serves a democratizing function: tipping the balance of participation patterns towards the students and away from the teacher, facilitating balanced participation among students, and enabling students to take control of the discourse. As has already been discussed, the results of my study tend to provide further support for this theoretical position.

In light of this general consensus in the ESL literature regarding communication in graphic conferences, as discussed in the review of the literature, graphic
conferences have been tied to a social-constructivist theoretical orientation to pedagogy (Beauvois, 1998; Peterson, 1997; Warschauer, 1996; Warschauer et al., 1996). A social-constructivist orientation stresses a social and collaborative view of learning. Because communication in graphic conferences is more student centred and balanced among students, and because students in graphic conferences display more control over the discourse, it is not difficult to see why researchers and teachers interested in a social-constructivist approach to education have been quite enthusiastic about using graphic conferences in second language classrooms.

In contrast to the theory of communication in graphic conferences just discussed, no well-established theory regarding the benefit of graphic conferences to second language writing has been developed. This is partially because of the very small number of research studies in this area, and the conflicting results of the studies that have been done. For instance, as discussed in the review of the literature, while Ghaleb (1993) and Braine (1997; 2001) found that drafts written after graphic conferences were of higher quality than drafts written after proximate conferences, Sullivan and Pratt (1996) found the opposite.
Although I did not consider writing quality, the results of my research suggest that both graphic conferences and proximate conferences can be helpful to students' writing. However, since students tend to spend more time controlling and directing the discourse in graphic conferences, students should be allowed additional time so that they can develop an equivalent number of ideas as they would in proximate conferences.

Implications for Further Research

One important direction for further research comes from my finding that for one of the groups, participation in graphic conferences was much more balanced among students than participation in proximate conferences, while for the other group, participation was almost equally balanced regardless of the conference condition. This finding is not completely consistent with the findings of other researchers. As discussed in the review of the literature, previous research has generally concluded that participation in graphic conferences is more balanced among students than participation in proximate conferences (e.g., Beauvois, 1992, 1998; Chun, 1994; Kelm, 1992; Kern, 1995; Warschauer, 1996). No research to date has found that this difference in balanced participation between conference
conditions might be mediated by group characteristics. As a result, my finding warrants further investigation.

In addition, although I found moderate evidence to suggest that the drafts written after proximate conferences were more related to conferences than the drafts written after graphic conferences, the question still remains as to what caused this. I suggested two possibilities. This may have occurred because students were not as able to generate as many ideas in the graphic conferences in an equivalent period of time due to the increased amount of time and discourse spent on contributions that served an interactive function. On the other hand, this might have occurred because students were not as able to recall the ideas presented in graphic conferences. This uncertainty points to a need for further research.

Two more important areas for further research are an extension of my current focus. Both of them are new possibilities that derive from the measures I have introduced in this study that assess the semantic relation between conferences and drafts. I shall discuss each of these areas for further research in turn.

First, my current focus was, in part, on comparing the influence of the two types of pre-writing conferences on students writing. For instance, I compared the degree to
which the drafts repeated language and ideas generated in the conferences. However, measures of writing quality are also of crucial importance to studies of writing pedagogy because such measures allow researchers to compare the benefit of the two types of pre-writing conferences on students writing. Together, measures of the degree to which drafts were based on conferences and measures of draft quality would allow researcher to compare the influence and benefit of pre-writing conferences on drafts.

In fact, as I argued in the review of the literature, comparisons of the quality of drafts written following the two types of conferences, by themselves, do not allow researchers to establish that the conferences had the decisive influence on writing quality. However, by including measures of influence and benefit, researchers can more convincingly compare the benefit of the two types of conferences on drafts. For instance, by using both types of measures, researchers could establish that one type of conference or the other was more beneficial to students' drafts because the same ideas that were introduced in the conferences, and used in the drafts, were of higher quality.

Second, with the ability to compare the semantic relation between statements in conferences and drafts,
researchers can begin to investigate important questions regarding collaboration in pre-writing conferences. As I mentioned in the review of the literature, collaboration is important because it has generally been found to positively influence the number and quality of ideas generated by groups (Neuwirth et al., 1993; Pendergast & Hayne, 1999).

For instance, using measures of semantic relation, researchers can compare the two types of conferences to determine if students mostly wrote about the ideas that they introduced themselves in the conferences, or if they wrote about other students' ideas as well. Similarly, researchers can compare the two types of conferences to determine if students tended to mostly use the ideas introduced by the teacher or some other dominant member of the group. This sort of analysis can reveal if there is a difference between dominating the discussion in terms of taking the most turns, and in terms of being influential on other students' ideas.

**Conclusion**

Perhaps the greatest success of this exploratory study with ESL students has been to implement a new approach to investigating the writing process. The measures of semantic and lexical similarity introduced in this study provide a powerful way for researchers to track the
development of language and ideas throughout the writing process. Since the derivation of these measure is completely automated, in the future, researchers will be able to base conclusions on larger samples of texts than was previously feasible, increasing the power of statistical analysis and the external validity of research findings.
References


Appendix A

Glossary of Terms

N.B. Words that are set in boldface within some definitions are defined in this glossary.

**Absence** - refers to one side of the traditional binary opposition between communications in written versus spoken forms. Writing is considered to be absent because the reader and writer are not spatially co-present, nor communicating with each other at the same time.

**Asynchronous** - refers to communication in which interlocutors are not interacting with each other at the same time. For example, traditional writing and e-mail are ordinarily considered to be asynchronous forms of communication.

**Average cosine** - refers to an aspect of one of the LSA measures used in this study. An average cosine is the average cosine that a conference t-unit has in relation to many draft t-units. See conference mean average cosine.

**Average maximum cosine** - refers to an aspect of two of the LSA measures used in this study. An average maximum cosine is the maximum cosine that a student's conference t-
units or draft t-units obtained with, respectively, many
draft t-units or many conference t-units, averaged across
the total number of, respectively, the conference t-units or
draft t-units contributed by that student. See conference
average maximum cosine and draft average maximum cosine.

Bandwidth - refers, in this study, to the amount and
type of information that can be sent over a computer
network. Bandwidth also refers to the degree to which a
computer-mediated interaction approximates the linguistic
and paralinguistic information typically conveyed in face-
to-face interactions. For instance, graphic conferences are
considered to be relatively low-bandwidth because only
graphic linguistic and paralinguistic information (e.g.,
emoticons) can be conveyed. In contrast, videoconferencing
is considered to be a relatively high-bandwidth
communication because aural and visual linguistic and
paralinguistic information can be conveyed.

Class discussion - refers, in this study, to the whole-
class discussions that were based on one of the topics
presented in Appendix I. Subjects participated in whole-
class proximate conferences and whole-class graphic
conferences.

Communication environments - refers to the settings in
which the class discussions occurred. In this study, the
communication environments were, on the one hand, the proximate environment (i.e., proximate conferences), and on the other hand, the graphic environment (i.e., graphic conferences).

**Communicative competence** - refers, in a general way, to an interlocutor's ability to make him or herself understood through verbal and nonverbal communication, and to communicate appropriately within the context of the culture within which a language is used.

**Computer-mediated** - refers to the use of a computer to perform an activity. The computer may be seen as an intervening part of the performance of the task. For example, a computer-mediated discussion uses the computer as a communications device.

**Conference** - see class discussion.

**Conference average maximum cosine** - refers to one of the LSA measures used in this study. A conference average maximum cosine is the sum of the maximum cosines the conference t-units contributed by a student reached with any the draft t-units that followed that conference, divided by the total number of conference t-units contributed by that student.
Conference mean average cosine - refers to one of the LSA measures used in this study. A conference mean average cosine is the sum of the average cosines the conference t-units contributed by a student had with all the draft t-units that followed that conference, divided by the total number of conference t-units contributed by that student.

Content Words - refers to words of an open grammatical class, such as nouns and verbs, that carry a good deal more of the semantic meaning of a sentence than words of a closed grammatical class, such as prepositions, articles, conjunctions, and pronouns, etc.

Control of discussion (control of discourse) - see interactive competence.

Corpus Linguistics - refers to the application of computer technology to linguistic analysis, and the collection of computer programs that make such application possible. For instance, the computer application Wordsmith Tools allows users to perform several types of Corpus Linguistics analysis such as listing the words in a document, and finding key words in a document (see Scott, 1997; 1998).
Corpus (of texts) - refers to a collection of texts. Texts in a corpus usually share something in common.

Cosine - refers, in this study, to a way of measuring the semantic similarity between two words or documents in a semantic space as part of LSA. If the cosine between the vector representation of 2 words or documents in a semantic space is close to 1, the words or documents are semantically similar. If the cosine between the vector representation of 2 words or documents in a semantic space is close to 0, the words or documents are semantically dissimilar.

Cyberspace - refers to a virtual environment created by networked computers, such as the Internet, in which users can locate, access and send information to other computer users. The science fiction author, William Gibson (1984), originally coined the term in his novel, Neuromancer.

Discourse management - see interactive competence.

Document - refers, in LSA, to a unit of language greater than a single word, such as a phrase, a clause, a sentence, a t-unit, a paragraph, a section, a chapter, or an entire text.

Document space - refers, in LSA, to a type of semantic space that includes the information theoretic weighting
step. In such a space, content words contribute more than non-content words to determining the length and direction of a document vector.

Document vector - refers, in LSA to a vector in term space or document space that represents the meaning, length and semantic breadth of a document.

Draft average maximum cosine - refers to one of the LSA measures used in this study. A draft average maximum cosine is the sum of the maximum cosines the draft t-units contributed by a student reached with any the conference t-units that preceded it, divided by the total number of draft t-units contributed by that student.

Emoticon - refers to a graphic emotion icon. Emoticons are small graphics that are used to convey the emotional content of a written message in much the same way that a smile, a laugh, or a frown carries emotional content in face-to-face communications. For a couple of examples, the emoticons :-) and ☺ indicate a smile, whereas :-( and ☻ indicate a frown.

Electronic classroom - a classroom in which computer-mediated activities are carried out in order to support learning objectives.
Electronic communication - refers, in general, to a communication that is made possible through the use of an information technology such as a telephone, a fax machine, or a computer.

Electronic conference (discussion) - see graphic conference (discussion).

Face-to-face classroom - a classroom in which face-to-face activities are carried out in order to support learning objectives.

Face-to-face conference (discussion) - see proximate conference (discussion).

First drafts - refers to the drafts written by students immediately following the class discussions.

Form of communication - refers to the mode of communication. For instance, speech is one form of communication; writing is another; graphic conferences are yet a third. The types of linguistic and paralinguistic information that can be conveyed in a given medium differentiate forms of communication. For instance, within electronic communications, different forms of communication are differentiated by the bandwidth for the communication.
Function of communication - refers, in this study, to the purpose to which a given form of communication is habitually put. For instance, for Socrates, the functions of speaking and writing were derived from their forms: While speaking can impart or generate truth and wisdom through dialectical exchanges, writing is used to transmit information of wide historical or cultural significance through space and time. Function of communication also refers to the speech act performed by an utterance, such as requesting information, greeting, initiating topics, etc.

Graphic - typically refers to the orthographic, or written form of language. However, graphic communication, as used in this document, refers to written language, as well as icons, diagrams, and pictures.

Graphic conference (discussion) - refers, in this study, to pre-writing conferences in which participants discuss a writing topic by typing messages that can be viewed by all other participants while each participant is seated at his or her own computer in a computer laboratory.
Graphic synchronous conference - see graphic conference.

Higher-order communication - see interactive competence.

Hyponym - refers to a relationship between words. A hyponym is a word that represents a general category under which specific examples can be given. For a couple of examples, animal is a hyponym of dog or bear; building is a hyponym of school or house.

Impermanence - refers to the amount of time that a communication exists in one form or another (e.g., as a sound wave, as a written text, or as electronically stored data). Speech is typically considered to be relatively impermanent in comparison with writing and electronically stored data.

Independent-sample t test - refers to a t test used to compare the means of two sets of scores when the sets of scores were obtained from two different groups of subjects.

Information theoretic weighting - refers to a component of Latent Semantic Analysis in which a mathematical algorithm is used to reduce the importance of
some words and increase the importance of other words in defining a semantic context. The importance, or "weight," of words that occur across a wide variety of sentences is reduced, and the importance of words that appear distinctively in a small number of sentences is increased. For instance, since a word such as the occurs in a wide variety of contexts, the information theoretic weighting reduces the importance of the in defining the semantic context of any given sentence.

Interactive competence - refers to an interlocutor's ability to control the discourse through speech acts such as expanding on topics, taking turns, giving feedback, capturing attention, steering or avoiding topics, starting and ending conversations, requesting confirmation or clarification, and apologizing.

Interactive language functions - refers to statements that demonstrate interactive competence.

Latent Semantic Analysis (LSA) - refers to a corpus-based statistical model of word usage that allows researchers to make comparisons of semantic similarity between pieces of textual information.
Lexical complexity - refers to the amount of exact lexical repetition that a document contains, as commonly measured with type-token ratios. Documents that do not repeat the same words much, and thus have higher type-token ratios, are considered to be more lexically complex, or in other words, to have a greater lexical range than documents that do use a lot of exact lexical repetition.

Lexical overlap - refers, in general, to the number of words that two documents or two sentences have in common. Simple lexical overlap includes exact word repetition as well as variants of the same root word (i.e., the same word in noun, adjective, or adverb form (e.g., happiness, happy, and happily). Complex lexical overlap includes synonyms (e.g., happy and glad), hyponyms (e.g., school and building), and phrasal definitions (e.g., “classroom” and “the place where we have our classes”).

Lexical range - see lexical complexity.

Logarithmic entropy weighting - See information theoretic weighting.

Matched-sample t test - refers to a t test used to compare the means of two sets of scores when the two sets of scores were obtained from the same group or person (e.g., as
in a pretest, posttest situation) or when the sets of scores were obtained from two groups that produced the same number of scores and that were matched so that they were comparable.

**Maximum cosine** - refers to an aspect of two of the LSA measures used in this study. A maximum cosine is the maximum cosine that a conference t-unit or a draft t-unit obtained with, respectively, many draft t-units or many conference t-units. See conference average maximum cosine and draft average maximum cosine.

**Mean average cosine** - refers to an aspect of one of the LSA measures used in this study. A mean average cosine is the sum of the average cosines that a student's conference t-units had in relation to many draft t-units, divided by the total number of conference t-units contributed by that student. See conference mean average cosine.

**Networked classes** - see electronic classroom.

**Non-content Words** - refers to words of a closed grammatical class, such as prepositions, articles, conjunctions, pronouns, etc., that do not carry much semantic meaning compared with words of an open grammatical class, such as nouns and verbs.
Nonlexical cohesive devices - refers to words of a closed grammatical class, particularly pronouns, demonstrative pronouns, and the definite article, that serve to tie a statement to a proceeding or following statement. For instance, consider the sentences, "How could I live without Mary?" and, "She is such a great friend." The pronoun she in the second sentence refers to Mary in the first sentence, and thus serves to tie the two sentences together.

One-sample t test - see Matched-sample t test

One-tailed test - refers, in inferential statistics, to a way of testing the null hypothesis when a researcher has some reason to predict the direction of an outcome a priori. For instance, one might suppose that a researcher has some reason to suspect that an individual's IQ score will be statistically significantly higher, but not lower, than the mean of a sample distribution of IQ scores. In such a situation, the researcher would want to test the individual's IQ score only against the high end of the sample distribution. Thus, if the chosen probability cutoff level for a one-tailed test is 5% (e.g., alpha = 0.05), the null hypothesis will be rejected only if the individual's score falls within the top 5% of the sample distribution.
Paralinguistic (cues) - refers to communication that takes place in a form other than spoken or written language. The "aural" category of paralinguistic cues is conveyed through the pitch, rhythm, and intonation of the voice. The "visual" category of paralinguistic cues is transmitted through visual channels such as gesture or facial expression. The "graphical" category of paralinguistic cues is communicated through the use of emoticons.

Permanence - refers to the amount of time that a communication exists in one form or another (e.g., as a sound wave, as a written text, or as electronically stored data). Writing and electronically stored data are typically considered to be relatively permanent in comparison with speech.

Presence - refers to one side of the traditional binary opposition between communications in traditional written versus spoken forms. Speaking is considered to be present because the interlocutors are spatially co-present and communicating with each other at the same time.

Pre-writing - refers to the first stage of the writing process. During the pre-writing stage, the writer's main focus is on the inclusion of content. The writer begins to gather and organize ideas about the topic.
Pre-writing conference - refers to the class discussion upon which subsequent first drafts are based. In this study, pre-writing conferences were held as proximate conferences and as graphic conferences.

Process writing - refers to the process approach to writing pedagogy. Compositions are built up in stages with the creation of multiple drafts. Early pre-writing activities lead to the creation of a first draft. The first draft is then subjected to multiple revisions and editing before the final product emerges.

Propositional analysis - refers to a text analysis technique in which the propositions contained in texts are identified and categorized according to their semantic content.

Proximity (Proximate) - refers to spatial co-presence. Proximity is a component of the distinction between presence and absence.

Proximate conference (discussion) - refers to class discussions that take place in the oral/aural mode, in which participants engage in a whole-class discussion while being present together in the same room.
Semantic space - refers, in LSA, to the space in which documents can be compared semantically. A semantic space can be a term space or a document space.

Singular-value decomposition (SVD) - refers to a mathematical technique in which a high dimensional matrix is reduced to an equivalent smaller matrix from which the original high dimensional matrix can be approximated by linear combination.

Social-constructivist (pedagogy) - refers to shift in educational theory from an interest in cognitive and developmental models of learning to a social and collaborative view of learning.

Subsequent written drafts - see first drafts.

Strategic discourse management - see interactive competence.

Synchronous - refers to communication in which interlocutors are interacting with each other at the same time. For example, spoken face-to-face communications, or
communications on the telephone are ordinarily considered to be synchronous forms of communication.

**Term-document matrix** - refers, in LSA, to a matrix that contains information about the similarity among many related documents. Each column in this matrix represents each document in the corpus and each row represents each word in the corpus. Each cell in the matrix contains the number of times that each word occurred in each document.

**Term space** - refers, in LSA, to a type of semantic space formed without the information theoretic weighting step. In such a space, content words and non-content words contribute equally to determining the length and direction of a document vector.

**T-unit** - refers to a minimum terminable unit. In this study, a t-unit is defined in psycholinguistic terms as a complete thought. Structurally, in this study, a t-unit is defined as the longer of: the entirety of a participant's contribution, or an independent clause with all its dependent clauses attached.

**Two-sample t test** - see Independent-sample t test
Two-tailed test - refers, in inferential statistics, to a way of testing the null hypothesis when a researcher has no reason to predict the direction of an outcome a priori. A situation in which a researcher does not know if an individual's IQ score will be statistically significantly higher or statistically significantly lower than some sample distribution of IQ scores will provide a good example for explanatory purposes. In such a situation, the researcher must test the individual's score against both the high and low ends of the sample distribution. Thus, if the chosen probability cutoff level for a two-tailed test is 5% (e.g., alpha = 0.05), the null hypothesis will only be rejected if the individual's score falls within the top 2.5% or bottom 2.5% of the sample distribution.

Type I error - a term used in inferential statistics to refer to incorrectly rejecting a true null hypothesis. A type I error is also called a false positive. For example, if a researcher's statistical test reveals a statistically significant difference between, for instance, an experimental and a control condition, when no difference in fact exists in the real world, the researcher has made a type I error. Common causes of type I errors include pooling heterogeneous variances, and using overly liberal probability cutoff values.
Type II error - a term used in inferential statistics to refer to failing to reject a false null hypothesis. In other words, if a researcher's statistical test fails to reveal a statistically significant difference between, for instance, an experimental and a control condition, when a difference does in fact exist in the real world, the researcher has made a type II error. Type II errors can result from selecting inappropriate statistical tests, such as using an independent sample t test to test the improvement between pretest and posttest scores for a single group of subjects. Type II errors also commonly result from a combination of: small sample sizes, large variances within comparison groups, and overly conservative probability cutoff values.

Type-token ratio - a measure derived from Corpus Linguistics that refers to the ratio of unique words (i.e., types) to total words (i.e., tokens) in a document. Type-token ratios are commonly used as a measure of the lexical complexity, or lexical range of a document.

Writing process - see process writing.

Written conference (discussion) - see graphic conference.
Written electronic conference (discussion) - see graphic conference.

Written drafts - see first drafts.
Appendix B

Semantic Space Information

The Web site at the University of Colorado at Boulder (http://lsa.colorado.edu/) was used to carry out the latent semantic analysis used in this study. To compute cosines, one must first choose an appropriate semantic space within which to conduct the analysis. As mentioned, the semantic space used in this study was "General reading up to first year college," also referred to as tasaALL. The following is an extract taken from the Web site that explains the composition of the tasaALL semantic space.

This space uses a variety of texts, novels, newspaper articles, and other information, from the TASA (Touchstone Applied Science Associates, Inc.) corpus used to develop The Educator's Word Frequency Guide.

This first incarnation of TASA-based spaces breaks out by grade level -- there are spaces for 3rd, 6th, 9th and 12th grades plus one for 'college' level. These are cumulative spaces, i.e. the 6th grade space includes all the 3rd grade docs, the 9th grade space includes all the 6th and 3rd, etc.
The judgment for inclusion in a grade level space comes from a readability score (DRP-Degrees of Reading Power Scale) assigned by TASA to each sample. DRP scores in the TASA corpus range from about 30 to about 73. TASA studies determined what ranges of difficulty are being used in different grade levels, e.g. the texts used in 3rd grade classes range from 45-51 DRP units. For the LSA spaces, all documents less than or equal to the maximum DRP score for a grade level are included, e.g. the 3rd grade corpus includes all text samples that score <= 51 DRP units.

Following are the specifics for each space:

<table>
<thead>
<tr>
<th>name</th>
<th>grade</th>
<th>maxDRP</th>
<th>#docs</th>
<th>#terms</th>
<th>#dims</th>
</tr>
</thead>
<tbody>
<tr>
<td>tasa03</td>
<td>3</td>
<td>51</td>
<td>6,974</td>
<td>29,315</td>
<td>432</td>
</tr>
<tr>
<td>tasa06</td>
<td>6</td>
<td>59</td>
<td>17,949</td>
<td>55,105</td>
<td>412</td>
</tr>
<tr>
<td>tasa09</td>
<td>9</td>
<td>62</td>
<td>22,211</td>
<td>63,582</td>
<td>407</td>
</tr>
<tr>
<td>tasa12</td>
<td>12</td>
<td>67</td>
<td>28,882</td>
<td>76,132</td>
<td>412</td>
</tr>
<tr>
<td>tasaALL</td>
<td>college</td>
<td>73</td>
<td>37,651</td>
<td>92,409</td>
<td>419</td>
</tr>
</tbody>
</table>

The breakdown for samples by academic area (in tasaALL):
To analyze cosines of drafts to conferences, and conferences to drafts, within this semantic space, the matrix module was used. Punctuation was removed from the texts before they were submitted for analysis. A document-to-document analysis rather than a term-to-term analysis was used because the document-to-document analysis performs the logarithmic entropy weighting step of the LSA while the term-to-term analysis does not (Foltz, personal communication, May 1, 2002). The logarithmic entropy weighting step was desired because, as discussed in the review of the literature, it serves to increase the importance of less frequent terms and decrease the importance of more frequent terms in defining the vector of a document. In other words, a term-to-term analysis is
more a measure of lexical similarity than semantic similarity (Foltz, personal communication, May 1, 2002).

To compute the vector length of documents, the one-to-many module was used. Again, the "General reading up to first year college" semantic space was chosen and a document-to-document analysis was selected. In the case of vector length, the document-to-document analysis returns considerably shorter vectors for documents because the vectors for high-frequency words are considerably shortened. Entire texts were entered into the lower text box one at a time, and the checkbox for returning vector lengths was selected.
Appendix C

Excerpt of Graphic Conference Sorted by Average Maximum Cosine

The following excerpt comes from the graphic conference transcript and drafts created by class 4a during the second week of the study. The texts have been broken up into t-units, as defined by Vavra (2000) and the conference statements, on the left, are sorted in descending order of mean average cosine. On the right of each conference contribution is the draft t-unit that had the highest cosine with it. Few short conference contributions attained high cosines with draft t-units.

<table>
<thead>
<tr>
<th>Conference T-Unit</th>
<th>Draft T-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>the cross cultural marriage can be successful</td>
<td>benefit of cross cultural marriage</td>
</tr>
<tr>
<td>also it is very useful for children</td>
<td>therefore all the children's understanding is very useful for them</td>
</tr>
<tr>
<td>ok for example their children have possibility to hurt from other children because of the different appearance</td>
<td>but for the different appearance children receive hurt</td>
</tr>
<tr>
<td>you just have to communicate with your partner</td>
<td>and you can't communicate with your partner deeply</td>
</tr>
<tr>
<td>is everybody agree with the cross cultural marriages</td>
<td>disadvantage of the cross cultural marriages</td>
</tr>
<tr>
<td>cross cultural marriage has the same opportunity to be successful as marriages with same culture background</td>
<td>same cultural marriages and cross cultural marriages are same</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>what is necessary to make a cross cultural marriage succeed</td>
<td>finally couples of a cross cultural marriage have to be able to make compromises</td>
</tr>
<tr>
<td>because different culture is some interesting problem</td>
<td>but it is very difficult to adjust to new culture and customs because of different cultural background</td>
</tr>
<tr>
<td>what are you opinion about can a cross cultural marriage be as successful as a marriage of people from the same culture</td>
<td>if you want to marry which one will you choose cross cultural marriage or same culture marriage</td>
</tr>
<tr>
<td>maybe there's benefit from cross culture marriage</td>
<td>benefit of cross cultural marriage</td>
</tr>
<tr>
<td>to educate the children that do not shame yourself</td>
<td>the children never hate one's appearance and never receive bullying from other children</td>
</tr>
<tr>
<td>they must think about their children welfare</td>
<td>their children might be abused</td>
</tr>
<tr>
<td>love</td>
<td>it is likely to a feeling of love</td>
</tr>
<tr>
<td>children from cross cultural marriages have the advantage to domain at least two different languages</td>
<td>also their children have two kinds of culture and language</td>
</tr>
<tr>
<td>even if they love each other</td>
<td>but it is possible if both spouses really love each other</td>
</tr>
</tbody>
</table>
Appendix D

Excerpt of Proximate Conference Sorted by Average Maximum Cosine

The following excerpt comes from the proximate conference transcript and drafts created by class 4 during the second week of the study. The texts have been broken up into t-units, as defined by Vavra (2000), and the conference statements, on the left, are sorted in descending order of mean average cosine. On the right of each conference contribution is the draft t-unit that had the highest cosine with it. Several short conference contributions attained high cosines with draft t-units.

<table>
<thead>
<tr>
<th>Conference T-Unit</th>
<th>Draft T-Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>so it's problem</td>
<td>they have to solve the problem is a lot</td>
</tr>
<tr>
<td>bilingual children</td>
<td>however when they have children and children are growing up they can be bilingual</td>
</tr>
<tr>
<td>so do you think cross cultural marriage can be as successful as marriages of people from the same culture</td>
<td>cross cultural marriage can be as successful as a marriage of people from the same culture</td>
</tr>
<tr>
<td>so automatically their children can speak two languages</td>
<td>one thing is that their children could speak two languages</td>
</tr>
<tr>
<td>so it can be a problem</td>
<td>they have to solve the problem is a lot</td>
</tr>
<tr>
<td>education is most difficult for their children</td>
<td>where they going to do for their children's education</td>
</tr>
<tr>
<td>---------------------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>I think cross cultural marriage is very difficult</td>
<td>I think a cross cultural marriage is a very difficult dream</td>
</tr>
<tr>
<td>all different cultures</td>
<td>because if cultures are different languages are mostly different</td>
</tr>
<tr>
<td>so the culture is different</td>
<td>they had lived with totally different culture</td>
</tr>
<tr>
<td>even though children have the problems a lot of problem but their children are basically bilingual</td>
<td>the problem is their children's identity</td>
</tr>
<tr>
<td>maybe when they are young it's ok they can solve the problem</td>
<td>they have to solve the problem is a lot</td>
</tr>
<tr>
<td>a person who can speak another language the partner's language</td>
<td>and they have to speak the other language</td>
</tr>
<tr>
<td>but I think that this is important in people from the same culture too the same characteristic because even in the same culture people are totally different</td>
<td>either in people from same or different culture love is important though not the only important</td>
</tr>
<tr>
<td>I think one thing was mentioned about children would learn two languages</td>
<td>children might be confused when they learn two languages</td>
</tr>
<tr>
<td>because we spend a lot of money to learn another language</td>
<td>we are spending a lot of money and time to learn second languages</td>
</tr>
</tbody>
</table>
Appendix E

Excerpt of Graphic Conference Sorted by Mean Average Cosine

The following excerpt was taken from the graphic conference transcript created by class 4 during the first week of the study. The text has been broken up into t-units, as defined by Vavra (2000), and sorted in descending order of mean average cosine. The statements at the beginning of the transcript, which have higher mean average cosines, are both generally longer, and appear to be much more on topic than the statements at the end of the transcript. The statements at the end of the transcript tend to perform interactive language functions.

1. we are too technology equipped
2. we can use technology to save lives
3. now we have technology and the lowest problems
4. we must have more technological advances
5. technology can do something to make our health better
6. we should use technology only in emergency cases
7. if we use this technology to cure disease it is ok
8. but why we should copy human
9. I think if we use biotechnology without limit it might makes people all look same in the future
10. technology is made by people [24]
11. but we can not produce the spirits the souls
12. but they can cure us
13. just we do our best to reduce the problems
14. or do you think that we have lose our freedom because
    sometimes we use it very bad
15. technology cannot be stopped
16. if biotechnology copies a human the same as you and
    does the copied one has human right
17. we can produce a lot of food you know and no hunger
    any more in our planet
18. you means biotechnology can manipulate human
19. and you think about heaven which is no technology
20. so we must make limit how it can do
21. now we can cure poor people
22. anyway technology is very good
23. we have to advance
24. we can limit anything if we want

    --- TRANSCRIPT CONTINUES ---
136. what do you think about it
137. sorry spelling was wrong
138. thanks a lot
139. [18] I don't want to hear your joke anymore
140. the sun is yellow the sea is blue
141. everything is fine
142. yes I know
143. I'm very starving that's it
144. why
145. why [1]
146. my name is [19]
147. thank you [17]
148. my name is [18]
149. everybody is good students except me
150. I've seen
151. [19] wake up
152. be my guest
153. ok
154. babo [19]
155. sorry [19] wake up
156. yes [11]
157. yes [17]
158. everybody read the article
159. I'm so tired very miserable
160. hi [27] at last your in
161. hi again
162. me too
163. me too [18]
164. yes
Appendix F

Excerpt of Proximate Conference Sorted by Mean Average Cosine

The following excerpt was taken from the proximate conference transcript created by class 4a during the first week of the study. The text has been broken up into t-units, as defined by Vavra (2000), and sorted in descending order of mean average cosine. The statements at the end of the transcript, which have lower mean average cosines, are less coherently on topic than the statements at the beginning of the transcript; however, the words in most of these statements still seem lexically related to the topic.

1. in that way we can use manipulate new technology to improve our quality of life
2. and maybe if we use this technology it could be bad use because always there are more problems with the technology then the benefits
3. I think it's a good technology for the future to cure many diseases and to help people have better food
4. we have to adjust technology
5. so we have to carefully technology
6. I think we can limit the technology only to food and medicine
7. today we are starting to know about this technology
8. we can't stop the advances in technology
9. now if we add to this the new technology to create
   super warriors or super killers it could be situation
   worse than now
10. I think that every technology should be used for our
    human or for food like for genetic food
11. some of you might have some particular ideas for what
    you would say yes we can use this technology for and
    what we have to say no we can't use this technology
    for
12. so the technology can be used for good or it can be
    used for evil
13. and what effects could use of this technology have on
    society
14. and we think about that then genetic people when we
    make a clone a human
15. we have think a lot about this technology
16. and we can use the technology for bad things
17. and for that reason I not agree to clone people only
    use technology for the food
18. because with the ability that we can improve or get we
    also can get bad things of the human beings
143. that kind of baby grow up with lacking of the parents of the principals
144. and I am a single woman and I don't need a man to do that
145. so sometimes I not convenient
146. but he has a brain is different
147. in my opinion I think the proper sensitive is the general
148. they something writing and they discussion some topic and they make decision
149. I think everybody's different
150. no because I'm saying it is sexual instinct to survive instinct to
151. does anyone agree
152. of the atomic bomb
153. yes I think so
154. maybe for you that is your opinion
155. instinct
156. and where is the father
157. pardon me
158. yes
159. yes
Appendix G

Ethics Approval

Brock University

FROM: Robert Ogilvie, Chair
Senate Research Ethics Board (REB)

TO: Michael Fitz
Hedy McGarrell, Applied Language Studies

FILE: 99-158, Fitz

DATE: February 7, 2000

The Brock University Research Ethics Board has reviewed the research proposal:

Electronic and Face to Face Prewriting Discussions: Student Engagement and the Quality of First Drafts

The Research Ethics Board finds that your revised proposal conforms to the Brock University guidelines set out for ethical research.

**Approved as revised.

RO/dvo
Appendix H

Consent Form

February 7, 2000

Researcher: Michael Fitze
Supervisor: Hedy McGarrell

Name of Participant: (Please print)__________________________

I understand that this study in which I have agreed to partake will involve participating in class discussions led by the researcher named above (i.e., not my regular IELP writing teacher), and writing first drafts of papers based on those discussions. I understand that the purpose of this study is to investigate how face-to-face and computer-based class discussions can be used to help second language students improve their English writing skill.

I understand that as part of my regular classroom activities, and as part of my grade in my IELP writing class, I will participate in four (4) class discussions. Two (2) discussions will occur face-to-face and two (2) will be done using a computer. I know that I will be required to write a first draft of a paper based on each of these discussions (i.e., four (4) drafts). I realize that my regular writing teacher will be responsible for assigning my grade for these activities (i.e., not the researcher).

I understand that my participation in these regular classroom activities will account for part of my IELP writing class final mark, assigned by my regular teacher.

I understand that my participation in the face-to-face conferences will be videotaped and that my participation in
the computer-based conferences will be electronically transcribed and that I, and my teacher, will have access to these videotapes and written transcripts. I understand that I, or my teacher, may review these records to help my learning.

By agreeing to be a subject in this study, I understand that I am agreeing to allow what I express during the pre-writing conferences and in the papers I write to be used as data and analyzed in a research study. By agreeing to be a subject, I also agree to allow my most recent Michigan Test scores and writing class scores to be made available to the researcher and/or supervisor for research purposes. I understand that all my personal data will be kept strictly confidential and that all information will be coded so that my name cannot be connected to my answers. I understand that only the individuals named above (i.e. researcher and supervisor) will have access to my personal data.

I understand that I may withdraw my consent to participate in the study by refusing to allow my personal data to be analyzed for research purposes at any time and for any reason without any negative consequences. I also understand that withdrawing my permission for my personal data or participation to be analyzed will not affect my grade for these activities, as assigned by my regular teacher.

I understand that there will be no payment for my participation.

I understand that there is no obligation to answer any question/participate in any aspect of this project that I consider invasive, offensive, or inappropriate.
This study has been reviewed and approved by the Brock Research Ethics Board. (File # 99-158 Fitze)

If you have any concerns about your participation in this study, you may contact Michael Fitze at (905) 688-5550, ext. 4608 or Professor Hedy McGarrell at (905) 688-5550, ext. 3757.

Feedback about the use of the data collected will be available during the month of April 2000 in the Faculty of Education office at Brock University. A written explanation will be provided for you upon request.

Thank you for your help!

Please take one copy of this form with you for further reference.

I have fully explained the procedures of this study to the above volunteer.

Participant's Signature: __________________________

Date: __________________________

Researcher Signature: __________________________

Date: __________________________
Appendix I

Topics for Discussion

Week 1 - Modern Medicine/Biotechnology

Once we have mapped the human genome, we can use our knowledge to control the sort of babies we have. We will be able to purchase medical treatments that will ensure our babies are healthy, "normal" in appearance, and intelligent. Biotechnology can be used to detect a genetic predisposition to certain diseases, and perhaps, to manipulate the DNA into a more desirable sequence. Biotechnology can also be used for more conventional purposes such as increasing food production and supply. Other possibilities include growing custom organs for transplant patients. What possible effects on society could this technology have? Should limits be placed on the use of this technology?

Week 2 - Cross-Cultural Marriages

The invention of the airplane and communication technologies have revolutionized travel and made the world a smaller place. People are coming together in ways that were never previously possible. These new opportunities for travel and communication have accompanied an increase in marriages among people from different cultural backgrounds. How do you feel about cross-cultural
marriages? What may be the advantages and disadvantages of cross-cultural marriages for couples and for their children? Do you think cross-cultural marriages can work? Can a cross-cultural marriage be as successful as a marriage of people from the same culture? Under what conditions can and do these marriages work?

Week 3 - The Environment

We have heard much lately about the greenhouse effect. Carbon monoxide in the atmosphere is causing depletion of the layer of ozone that protects us from much of the harmful ultraviolet radiation from the sun. In addition, increased levels of carbon dioxide (as a result of industrial processes) are causing the world's temperature to increase, leading to more water in the atmosphere and more violent storms. Moreover, the population is increasing at an alarming rate. Can we feed, clothe, and provide medical care for all these people? Should governments and countries (continue to) take steps to control population growth? Are we headed for much more poverty and suffering in our world or will our technology solve our problems? What advice do you have for governments who face these problems?
Week 4 - Professional Sporting

Professional sports seem to have great popularity among people from many countries. Although the favoured game may differ, professional sporting events draw huge crowds. One may argue that the significance of professional sports is overrated, and that people would be better served either playing the sport themselves, or engaging in some other pastime. Others point out that professional athletes are outrageously overpaid and that the innocence of sporting has been lost; athletes no longer play for the love of the game, they play for the money.

What positive side do you see to professional sporting that might justify its popularity? Do you think that professional sports are overrated or do you see value in these people and events? Explain.
Appendix J

Summary of Procedures in Temporal Order

The data collection portion of this study took place over a 14-week term between January and April 2000. In this appendix, I will summarize the procedures in temporal order making reference to week numbers. A discussion of the rationale for the procedures can be found in the procedures section of chapter 3.

The first 5 weeks of the term:

- teachers selected 4 writing topics out of a possible 10
- teachers agreed that students' participation in the conferences would constitute a portion of their participation mark for their writing class
- teachers agreed that submitting drafts for this study would count towards the homework portion of students' marks
- teachers agreed that students would be required to develop the first drafts written for this study into final drafts, and that the final drafts would count towards their final mark in the writing course.
• I clarified to the teachers that they would attend, but not participate in the conferences, and that they would retain responsibility for assigning all marks.

• Teachers agreed that I would be allowed time during the eighth week of the term to take each group of students into the computer lab for a training session on the computer software used in the study.

The sixth and seventh weeks of the term:

• I visited with the students in both of the classes to describe the nature of the study and what was required of them.

• I informed students about how the activities to be carried out during this study would relate to their writing class marks.

• I told students that their regular classroom teacher would attend but not participate in the conferences.

• I let students know that the regular classroom teacher, rather than the researcher, would assign all marks.
I distributed the consent forms (i.e., see Appendix H), and students were given an opportunity to read the consent form and ask any questions about the study.

I collected the signed consent forms.

The eighth week of the term:

I took each class group into the computer lab for an introductory session on the computer.

Students learned how to log in and out of the system, and a short rehearsal conference was held so that students could practice sending and receiving messages.

The ninth, tenth, twelfth, and thirteenth weeks of the term:

Part 1 - The day before each of the conferences:

Each classroom teacher gave a written copy of the writing topic for the week to the students in each class, and reviewed the topic with the class to ensure that students understood the vocabulary and general meaning.
• the teacher reminded students that on the following day, they would be required, as a whole-class activity, to discuss the issue for approximately 20 minutes, after which they would have about 25 minutes to individually begin writing a first draft that would be due for homework on the following day

Part 2 - The day of each conference

• I attempted to keep my role identical during all the conferences
• at the beginning of each conference, I orally reminded students that they had 20 minutes to discuss the topic as a class, after which they were to individually write first drafts
• I reminded students that they would be required to complete the drafts for homework, and that the drafts would be collected in class on the following day
• I read aloud the question that students had been assigned on the previous day to orient students to the task
• I facilitated the discussions by summarizing students' contributions and questioning for clarification when
necessary, but as far as was possible, I attempted to restrain myself from offering new ideas or giving examples about the topic.

- after discussing the topic for 20 minutes, I orally called the discussion to a close, and asked students to individually and silently begin writing their first draft
Appendix K

Meaning and Calculation of the Gini Coefficient

The gini coefficient is a measure of inequality. Larger gini coefficients correspond to greater inequalities. Gini coefficients were first used to describe the distribution of wealth in human populations, but the underlying concept can be applied equally well to the distribution of words contributed by individuals in a conference.

The meaning of the gini coefficient can be illustrated graphically by plotting a Lorenz curve. For the purpose of this study, the vertical axis of the Lorenz curve is the cumulative proportion of total number of words contributed in the conference, and the horizontal axis is the cumulative proportion of the total number of individuals in the conference (i.e., see Figure 1, this appendix).

The first step is to rank individuals according to the number of words contributed (i.e., from least to most). If each student in the class contributed the same amount of words, each would contribute to the total number of words in the conference equally. As each successive student is added to the population, the cumulative proportion of the total number of words would increase linearly, resulting in the diagonal line (i.e., the line of equality in Figure 1,
this appendix). If there were inequality of numbers of words among individuals, however, the cumulative proportion of the total number of words would first grow very slowly. The curve would bow below the line of equality because the students who contributed the least would add little to the total number of words. Only when those who contributed the most were considered would the curve begin to bow up because these students contribute disproportionately to the overall number of words. A greater degree of bowing would result from a greater inequality of contributions among individuals in the conference.

Figure 1. Graphic illustration of the gini coefficient.
The gini coefficient is equivalent to the area between the diagonal "line of equality" and the Lorenz curve. It can be calculated from the following formula:

\[ G_1 = \left( \frac{2}{n^2 \bar{x}} \right) \sum_{i=1}^{n} \left( \left( i - \frac{n+1}{2} \right) x_i \right) \]

where:

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \]

The explanation of the gini coefficient offered in this appendix was referenced from Koide and Shumway (1998).
Appendix L

Conference Transcription Procedures

This appendix consists of three sections. In the first and second sections, the conventions followed when transcribing graphic and proximate conferences, respectively, are described. The third section contains a discussion of information about the t-unit analysis in relation to data transcription, with respect to both types of conferences.

Transcription of Graphic Conferences

For the most part, transcription of the graphic conferences was very straightforward. After each conference, the log file from the Wet-CT server was saved to local disk as a word-processing file. The saved file contained a record of all messages, with information about who sent each message, and information about the order in which messages were sent. To protect the anonymity of the participants, the names of the participants who sent the messages, as well as the names of participants used in the body of the messages, were converted to numerical codes.

Next, mistakes in spelling in the conference transcript were corrected. Spelling mistakes were corrected for three reasons. First, spelling mistakes were corrected to be consistent with transcription of the
proximate conferences. Since the proximate conferences were transcribed by the researcher and subjected to a computer spell-check, and thus contained no spelling mistakes, to be comparable, it was felt that graphic conferences should also be subjected to the spell-check.

Second, since Corpus Linguistics analysis was used to compare the words used in the conferences to the words used in the drafts, it was necessary to ensure that words used in conferences and drafts were spelled consistently. As a result, all drafts were also subjected to a spell-check. In sum, inconsistent spellings of the same words between conferences and drafts would have flawed the Corpus Linguistics exact word repetition analysis.

Third, since the LSA measures used in this study were based on comparison to a corpus of published texts, and as a result, LSA would not have been able to recognize misspelled words, the LSA analysis would have been flawed if spelling errors were allowed to stand. In addition, since the corpus of texts used in the LSA was primarily based on articles written in the United States, United States spelling conventions were observed.

Transcription of Proximate Conferences

Proximate conferences were videotaped, and the videotapes were transcribed. To do so, the name of the
participant speaking and his or her comment, in the order in which comments occurred, were first transcribed. Next, to protect the anonymity of the participants, the names of the participants who produced the messages, as well as the names of participants used in the body of the messages, were converted to numerical codes.

Paralinguistic aspects of communication were not recorded in the conference transcripts. For instance, nodding or shaking the head were not recorded as yes or no in the conference transcripts. Nor were hand gestures, or the appearance of students (e.g., appearing interested or bored) recorded. Such paralinguistic aspects of communication were not recorded for two reasons.

First, one of the problems with trying to include an explicit linguistic reference to, for instance, a nod, is the difficulty in interpreting the gesture. For instance, a nod could be meant to express "yes, I agree," or, "very interesting," or even, "I understand what you are saying." As a result, I felt it would be better not to attempt such interpretation. One might argue that this decision may have biased comparisons of total amount of discourse to favour more communication in graphic conferences. For example, for contributions such as agree, or yes, to be communicated at all in graphic conferences, they must be
expressed in explicit language. However, such verbal contributions also occurred, and were allowed to stand as t-units in the proximate conferences.

Second, and perhaps more importantly, this study focused on measuring the number of words and ideas expressed in explicit linguistic communication, and the transfer of those words and ideas from conferences to drafts. Qualitative comparison of group process and interaction in the two types of conferences was not an objective of this study. In other words, although this study compared the total amount of participation, and the degree to which participation was balanced among participants in the two types of conferences, the focus of this comparison was on explicit linguistic communication rather than on communication through paralinguistic cues. In fact, such paralinguistic communication was understood to be a consequence of participation in proximate conferences, which might have some impact on the amount of students' explicit verbal participation.

When single-word semi-verbal contributions such as "uh-huh" were made by the researcher to show that the researcher was listening, as opposed to expressing agreement, these contributions were not recorded in the transcript. "Fillers" such as "uh-huh" were omitted
because they were deemed to be roughly equivalent to a non-verbal gesture such as appearing interested. Moreover, recording such contributions would have tended to overestimate the total number of words and t-units and to underestimate the average length of t-units contributed by the researcher during proximate conferences.

Finally, the transcriptions of proximate conferences were checked for spelling. During this spell-check procedure, speech words such as "gonna" were converted to standard written forms (e.g., going to). This was done for the same reasons discussed in the section on graphic conference transcription. Inclusion of misspelled and non-standard words would have been inconsistent with the language used in the drafts, and would have flawed the Corpus Linguistics and LSA analyses. Moreover, in cases where the same word was repeated several times in a row, (i.e., when a participant stuttered) the repetitions were dropped. This was done because allowing such repetitions would have flawed the Corpus Linguistics lexical range analysis.

**T-Unit Analysis in Relation to Data Transcription**

Comments in conferences were broken into t-units based on the KISS model of t-units (Vavra, 2000). For a more detailed explanation of the definition of t-units used in
this study, see the section on the effect of electronic conferences on writing in chapter 2, and Appendix A of this study.

In graphic conferences, a contribution was defined as a comment made by a participant, up to one t-unit in length, concluded by the participant pressing the send key. Consecutive contributions by the same person in graphic conferences, even sentence fragments, were not combined into one contribution because, in this study, an orthographic fragment counted as a t-unit. Similarly, in proximate conferences, a contribution was defined as a single conversational turn taken by a participant up to one t-unit in length.

In both types of conferences, where contributions were equal to or shorter than a main clause with all its subordinate clauses attached, or in other words, a t-unit, no manipulation of the conference transcript was necessary. Where contributions were longer than a main clause with all its subordinate clauses attached, the main clauses were separated into consecutive contributions made by the same participant.
Appendix M

Example of Conference Text Transcription

This appendix contains an example of transcribed conference text. This text is taken from the proximate conference during the second week of the study. The first part of the appendix shows an excerpt of the original transcription, except the names of participants have been converted to numerical codes. In the second part, the text has been broken into t-units in preparation for LSA and Corpus Linguistics analysis.

Original Transcription

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>In Quebec now it's more common now to see marriage from different cultures.</td>
</tr>
<tr>
<td>30</td>
<td>In Quebec.</td>
</tr>
<tr>
<td>4</td>
<td>In Montreal, especially in Montreal.</td>
</tr>
<tr>
<td>30</td>
<td>Right, and what different cultures?</td>
</tr>
<tr>
<td>4</td>
<td>All.</td>
</tr>
<tr>
<td>30</td>
<td>All different cultures.</td>
</tr>
<tr>
<td>4</td>
<td>In Montreal we have lot of cultures there, but in Quebec City it is less common, I think. In Montreal it is common.</td>
</tr>
</tbody>
</table>
And what do you think about it, do you notice a
difference between people who are married from
different cultures and people from the same cultures?

I don't know any people.

Does anyone know anyone who has a cross-cultural
marriage?

My host parents here.

Between a married couple Tanzanian and Russia.

Russian and Tanzanian

In my church, one of couples, man is Korean and woman
is Canadian. Sometimes I feel they don't understand
each other sometimes. I think sometimes we can have
children, and I think the most difficult one is they
understand each other very hard their culture

So the culture is different, and it's hard for them to
understand each other.

Even though they have the problems a lot of problem
but their children are basically bilingual. Because
we spend a lot of money to learn another language, so
automatically their children can speak two languages.

It's an opportunity

But I think part of the problem is the decision to be
married, because obviously, different culture has
different things. They think different. But even in
the same culture it's the decision: what do you think, what I think, and if we can live together. More than the culture is the decision because in same culture most of marriages have difficulties. Most of them finish early.

30 Right, so even people who are married in the same culture have difficulties

Text Ready for Analysis

Student ID T-unit
4 in Quebec now it's more common now to see marriage from different cultures
30 in Quebec
4 in Montreal especially in Montreal
30 right and what different cultures
4 all
30 all different cultures
4 in Montreal we have lot of cultures there
4 but in Quebec City it is less common I think
4 in Montreal it is common
30 and what do you think about it
do you notice a difference between people who are married from different cultures and people from the same cultures

I don't know any people
does anyone know anyone who has a cross-cultural marriage

my host parents here

between a married couple Tanzanian and Russia

Russian and Tanzanian

in my church one of couples man is Korean and woman is Canadian

sometimes I feel they don't understand each other

sometimes

I think sometimes we can have children

and I think the most difficult one is they understand each other very hard their culture

so the culture is different

and it's hard for them to understand each other

even though they have the problems a lot of problem but their children are basically bilingual

because we spend a lot of money to learn another language

so automatically their children can speak two languages
it's an opportunity

but I think part of the problem is the decision to be married

because obviously different culture has different things

ey think different

but even in the same culture it's the decision what do you think what I think and if we can live together

more than the culture is the decision because in same culture most of marriages have difficulties

most of them finish early
	right so even people who are married in the same culture have difficulties
Appendix N

Excerpt of Graphic Conference Sorted by Average Maximum Cosine

The following excerpt was taken from the graphic conference transcript created by class 4a during the last week of the study. The text has been broken up into t-units as defined by Vavra (2000) and sorted in descending order of average maximum cosine. The statements at the beginning of the transcript, which have higher average maximum cosines, are more on topic than the statements at the end of the transcript. The statements at the end of the transcript tend to perform interactive language functions.

1. do athletes play for love of the game or for money
2. athletes give dreaming children hope which is good
3. they earn a lot of money
4. it is controlled by money
5. they receive a lot of money
6. sports is good for children
7. but some player enjoys only play
8. but I like to watch the game
9. today the majority of players are playing for money
10. what is positive about watching sports
11. some players have qualification for receiving much money not every
12. I think player play for money and for fun
13. yes I think it is good to children
14. it is a necessity to make money to support themselves
15. it is true I hope they can earn money as much as they can
16. I thinks that professional sports is overrated
17. professional player love their sports
18. other players want money
19. but when there is money it seems difficult to keep fair play
20. olympic game give to them kind of glory
21. I think sport is good for children
22. but when they start to earn money they play for money
23. do we put too much importance on winning a game
24. but you know many player got injure that game these days
25. but I think their money is not limited
26. ok entertainers give children dreams
27. I thinks professional sports is too exaggerated
28. I think sport player need to earn a lot of money because they can die by playing
29. some players just care about money
30. I believe sports are very worthy
31. do you think a famous baseball player would still play
   if the team only offered him half his usual pay

--- TRANSCRIPT CONTINUES ---

225. yes who is crazy
226. bye
227. lunch time
228. bye
229. bye
230. yes you are right
231. yes
232. yes
233. no that not true
234. no of course
235. no[30]
236. no [30]
237. yes [30]
238. you think so [14]
239. yeah explain [22]
240. bye everybody
241. congratulation [29]
242. check it out
243. who is crazy [21]
244. how about [15]
245. anyway [23]
246. it sound interesting
247. yes [21]
248. yes [21]
249. yes [16]
250. sorry [21]
251. [14]
252. hi [21]
253. hi [21]
Appendix O

Excerpt of Proximate Conference Sorted by Average Maximum Cosine

The following excerpt was taken from the proximate conference transcript created by class 4 during the last week of the study. The text has been broken up into t-units, as defined by Vavra (2000), and sorted in descending order of average maximum cosine. The statements at the end of the transcript, which have lower average maximum cosines, are less coherently on topic than the statements at the beginning of the transcript; however, the words in most of these statements still seem lexically related to the topic.

1. what is the value of watching sports
2. the really good professional players can give some dreams to children
3. it's the very best influence for children
4. they have to earn money
5. this is a way they earn money
6. and so they are satisfied by watching the professional player
7. so they have to get money
8. so if that team lose the game people say the same way
9. and only few children can think about money from sports.

10. so if a baseball team has a lot of money the result is very high strong.

11. so it seems like professional sports is good for the athletes.

12. athletes always try to win the game.

13. for example some athletes help other children poor children.

14. maybe children can dream I want to be a basketball player like Michael Jordan or somebody.

15. and I worry about the bad influence to the children.

16. they deserve to get this money for their playing.

17. so instead of studying they just pursue their life style of professional sports player.

18. or why should we watch professional sports.

19. to compete with other children and to be the best.

20. only few professionals can get a lot of money.

21. and I think it influences children.

22. but instead of watching it's better to go and play sports.

23. the role model can give some encourage the children by watching tv or by playing same sports like basketball of baseball.
24. when they root for a team and that team wins they get very excited

--- TRANSCRIPT CONTINUES ---

181. so what kinds of enjoyment do they get
182. people feel upset
183. maybe though they need to study a lot
184. the danger maybe
185. it's not just about that
186. so yes why
187. they are fully involved in effort
188. I would like to say about sammy sosa because he is from dominican republic a very poor country
189. right it's true
190. in the future they still have the dream
191. in canada it is hockey
192. right working together
193. anyone else
194. cooperation
195. cooperation
196. cooperation
197. yeah [19] you too
198. yeah
199. yeah

200. like a couch potato

201. I'm really disappointed

202. yes [17] say something

203. even sex scandals

204. it's a business

205. [17] has a comment

206. it's a business

207. please finish that thought

208. it's a business yeah

209. yes it's only my opinion

210. [18] [1] say something please

211. and he's going to have a heart attack

212. it's not conclusive in my opinion

213. determined

214. determined

215. persistent

216. persistent

217. yes

218. yes
Appendix P

Example of Researcher Participation During a Proximate Conference

In this appendix, I have included all the comments I made in the proximate conference with class 4 during the second week of the study in the temporal order in which my comments occurred. I chose this particular conference because it is representative of my proximate conference contributions throughout the study. Although it may not be immediately apparent because these comments are taken out of context, in most cases where my comments begin with so, (e.g., statements 1, 2, 4, 5, 8, 14, 15, 17, 22, 23, 28, & 30) I am not presenting a conclusion, but rather summarizing what I believe a student said.

1. so do you think cross cultural marriage can be as successful as marriages of people from the same culture

2. right so even people who are married in the same culture have difficulties

3. do you notice a difference between people who are married from different cultures and people from the same cultures
4. so you think respect and accepting the other culture is important
5. so the culture is different
6. more than if they're from the same culture
7. does anyone know anyone who has a cross cultural marriage
8. so you think that it might be easy at first for cross cultural couples
9. all different cultures
10. right and what different cultures
11. so far we've talked about the difficulties with cross cultural marriages
12. I think one thing was mentioned about children would learn two languages
13. are there any other advantages
14. so there might be a problem just trying to live together in the same country
15. so when they have children it becomes more difficult
16. do you think it's possible that the child might feel they belong both places
17. so they're confused about their identity
18. when you say global mind do you mean the couple's themselves the man and the woman or their children
19. can you think of any possible advantages
20. and it's hard for them to understand each other
21. do you think it might be easier for them to choose a third country
22. so you think it's a challenge
23. so that's difficult for the children
24. what kind of people would it work for
25. and what do you think about it
26. their children too
27. and you said that the child will feel like they don't belong either place
28. so maybe they would make good politicians
29. but then later it becomes more difficult
30. so for example if the father is korean and the mother is canadian am I korean or canadian and where I belong
31. you said the children miss their adolescence
32. I'm not sure I completely understand their siblings
33. what do you mean exactly
34. does everyone agree with that
35. oh children
36. what do you mean habits
37. does anyone want to add anything else
38. more independent
39. in quebec
40. russian and tanzanian
41. any final summary