

# **A Comparison of ESL Students' Ability in Four Language Modes: Oral Composing, Independent Writing, Story Retelling and Reading\***

**Victor Froese**

---

Thirty-nine students (ages 10-14) were drawn from classrooms containing ESL students in three schools in one Winnipeg school division and comparisons were made for these ethnolinguistic groups — Filipino, Vietnamese, and Chinese. For each of the four language modes — oral composing, independent writing, reading, and retelling — a number of language units were compared: number of words pro-

duced, number of t-units produced, average number of words per t-unit, and number of dependent clauses produced. The answers to the following four questions are discussed: 1) In terms of language units, how do the modes compare across language groups? 2) How do oral reading miscues compare? 3) How does reading comprehension and listening comprehension compare?

---

## **SCOPE & OBJECTIVES OF STUDY**

The purpose of the present study was to compare ecologically valid measurement techniques which took into account both the form and function of language, and which occurred within the context of a pragmatic or integrative language assessment paradigm. Such techniques should be suitable for formative and summative evaluation purposes and be applicable to multicultural settings. They should measure power over the living language (Loban 1976).

More specifically this study posed the following questions about subjects within the largest language and ethnocultural groups — Filipino, Vietnamese, Chinese — within the Manitoba multicultural milieu:

- a. In terms of common quantitative language units (mean words per t-unit, mean number of dependent clauses, and length of composition) how do the four modes — oral composing (the subject tells a "story" which the adult writes on paper), independent writing, reading and story retelling — compare?
- b. How do oral reading miscues compare in terms of type and number?
- c. How does reading comprehension compare to aural comprehension?

## THEORETICAL SIGNIFICANCE & PRACTICAL IMPORTANCE OF STUDY

The combination of recent waves of immigration and renewed (and more sophisticated) interest in language research — especially in composition and comprehension — has focussed attention on the need for contemporary multicultural research. In addition, the Canadian, British, and American view of language has been shifting towards a more holistic paradigm (sometimes referred to as the “communicative competence” model) of language teaching and assessment (Shuy 1981).

Within this milieu, it is not surprising to find the report *Testing, Assessment, Counselling, and Placement of Ethnic Minority Students* recommending “that a study be conducted to determine the validity and reliability of standardized tests and other means of assessment presently being used to evaluate ethnic minority immigrant students across the Province of Ontario” (Samuda & Crawford 1980). In provinces such as Manitoba where the mandated school curriculum is an “integrated” one (i.e. holistic) it becomes essential to develop assessment procedures with some ecological validity such as the present study proposes. Standardized tests may be criticized on the basis that they do not fit the current paradigm — they are suitable for a lock-step, skills oriented curriculum.

While a considerable body of research has examined the relationship among the language arts in the English-speaking population and even in French Immersion programs, Stotsky (1981), for example, has noted that:

There is apparently no research ... examining the writing of English-as-a-Second-Language students and the relationship between traits in their writing and their reading scores in English. Studies of the writing of ESL students of various language groups in relation to their level of proficiency in English in both speech and reading seems needed not only from the practical point of view of error analysis but from a theoretical perspective as well.

Further support for the proposed research comes from the field of text analysis and from comprehension studies. Kintsch and Yarbrough's (1982) work on disassociating macro and micro processes in text comprehension has important implications for assessment purposes. And a series of important studies from the Center for the Study of Reading at the University of Illinois suggest that:

Often such readers (those with nonmainstream cultural backgrounds) are those having the most difficulty in comprehension, and their difficulties understanding texts might be due to mismatched expectations arising from cultural differences (Bruce et al. 1981, p. 34)."

In summary, then, the paucity of cross-ethnic/language study has been duly noted along with a changing language paradigm which necessitates a different assessment strategy. Finally, the recent psycholinguistic work on text analysis and comprehension has provided us with new insights and measures which may more nearly adequately identify differences in language production and reception. This study, in an exploratory manner, seeks to apply these current findings to the multicultural setting.

This overview of the related research deals with four major areas: the validity of standardized testing in multicultural settings; the validity of using standardized measures with "integrated" curricula; the interrelationship of language measures in first-language instruction; and methods of second-language assessment.

## **MULTICULTURALISM & STANDARDIZED TESTING**

By normally acceptable measurement standards, results of norm-referenced tests must only be used with populations similar to the norming group. When testing instructions are manipulated (i.e. time is extended, items are omitted, and words are substituted) as reported by Samuda & Crawford (1980) in Ontario schools and when even Canadian norms are absent (for example, the SRA Achievement Test), then results from standardized tests become meaningless even for first-language subjects. What, then, can scores gathered under such conditions mean when applied to ESL students? Labov (1976) noted that norm-referenced standardized tests make normally intelligent minority students appear inept. Samuda and Crawford (1980) also documented that these tests do not distinguish between performance and capacity; they show cultural differences, not deficits. In short, while commercial standardized tests and similar provincially constructed measures are commonly used, their validity must be seriously questioned when used with minority ethnic/language groups.

## **INTEGRATED CURRICULA & ASSESSMENT**

With the "communicative competence" model, Hymes (1980) cogently argues that both linguistic competence and sociolinguistic competence must be considered in language learning. If both form and function must be taken into account in language learning, it should also be generalizable to language assessment. That is the logic used in this research. Shuy (1981) expresses it thus:

Research shows that good language learners move gradually toward acquiring the forms which reveal that function. They learn holisti-

cally, not by isolated skills ... we have developed a tradition of teaching reading, writing, and foreign language which goes in just the opposite direction — from surface to deep (structure), from form to function, from part to whole. (p. 106)

The Province of Manitoba states explicitly in its Language Arts Curriculum Guide (1982) that “integration is the aim of this guide”. If learning is set in a functional context, then assessment of that same learning, when possible, should also take place in its sociological context.

## INTERRELATIONSHIP OF LANGUAGE ARTS

Loban's (1976) longitudinal study following children from kindergarten through grade twelve has best documented the close relationship of the four language arts areas — listening, speaking, reading and writing. Of particular noteworthiness is the relationship of speaking to writing. Loban (1976) observed that “subjects tend to speak and write in units of virtually the same average length” and that “it is of special note that those superior in oral language in kindergarten and grade one *before they learn to read and write* are the very ones who excel in reading and writing by the time they are in grade six. Our data shows a positive relationship of success among the language arts” (p. 35).

The “unit” referred to by Loban was defined as “each independent clause with its modifiers” and was earlier named the “t-unit” (for minimal terminal unit) by Hunt (1965), and is presently the most widely used single index of syntactic growth and development (Neilsen 1977). Stotsky (1975, p. 47) pointed out that:

With the development of Hunt's units of analysis, reliable techniques for measuring syntactic growth from grade to grade and between ability groups have become available”.

In ESL research, Braun (1969) and Klassen (1976) have used these measures and have argued for their validity.

For the present study, t-unit analysis and related measures (mean words per t-unit, and mean number of dependent clauses) provide a convenient means of comparing the syntactic fluency of language in the various modes: speaking (oral composing), writing, retelling, and reading.

## ESL LANGUAGE ASSESSMENT

Discourse analysis, pragmatics, communicative competence and integrative testing are terms well known in the ESL field (Oller 1979). The essential unifying concept is that language must be assessed in context and that form and function must be accounted for. Often this type of analysis

is compared to discrete point testing, which is more akin to skills testing in the standardized test manner, and which has been argued to be inappropriate for current curricular paradigms.

Assessment in second-language instruction has been comprehensively dealt with in works by Lado (1964), Valette (1977), and Oller (1979) and many others. The present study differs from previous language measures primarily in that the units of analysis (t-units and derived measures) are common across modes of expression (i.e. writing, speaking, listening and reading) and that they are collected in an ecologically valid context.

## **RESEARCH PLAN & METHODS**

Thirty nine subjects (ages 10-14) for the study were drawn from classrooms containing ESL students in three schools in one Winnipeg school division which registers the majority of ethnic minority students. In this phase of the study 10-15 subjects from each of the three most dominant ethnic/language groups — Filipino, Vietnamese, Chinese — were used.

Two part-time graduate students were trained to collect the data using methods developed by Froese (1983), Goodman & Burke (1972), Hay (1984), King & Rentel (1981), and Loban (1976). To assure adequate reliability, two samples of independent writing, two samples of oral composing, two samples of retellings, and two samples of reading were collected and analyzed. The data representing 25 variables were collected in March and April of 1986.

In order to collect the oral composing samples, examiners used a common picture from the Interaction materials (Moffett & Wagner 1973) to elicit oral narratives which were written for the subject (original to student, carbon copy for study) while s/he observed the process.

To produce the independent writing samples again a common stimulus from the Interaction materials was used but only after an initial practice session.

Retellings were generated by the following procedures: First, narrative stories were written to a common story structure appropriate to this age group (McConaughy 1982). Then the story was read to the subject by one examiner and retold to the other examiner in an adjacent room. The retellings were tape-recorded and later transcribed for analysis (for linguistic units, story structure and aural comprehension).

Samples of reading were based on stories written to the same narrative schema used for the retelling stories but they were scored for "miscues" after procedures developed by Goodman & Burke (1972) but modified slightly for this study.

After the transcription and analysis of the language samples, the data were analyzed using the SAS (1984 Version 5.15) statistical package, using

the covariate — years in Canada — since obvious differences were noted across the language groups.

FINDINGS & CONCLUSIONS

Question #1:  
*In terms of language units, how do the modes compare across language groups?*

Table 1 presents the means for total words produced, number of t-units, words per t-unit, and the number of dependent clauses which has been adjusted to control for length of time that the student has been in Canada. It is clear that uncued recall of the story independently produced the least number of words; independent writing the most. Second and third in rank order are oral composing and retelling. Most of the clausal measures follow the same pattern in terms of magnitude.

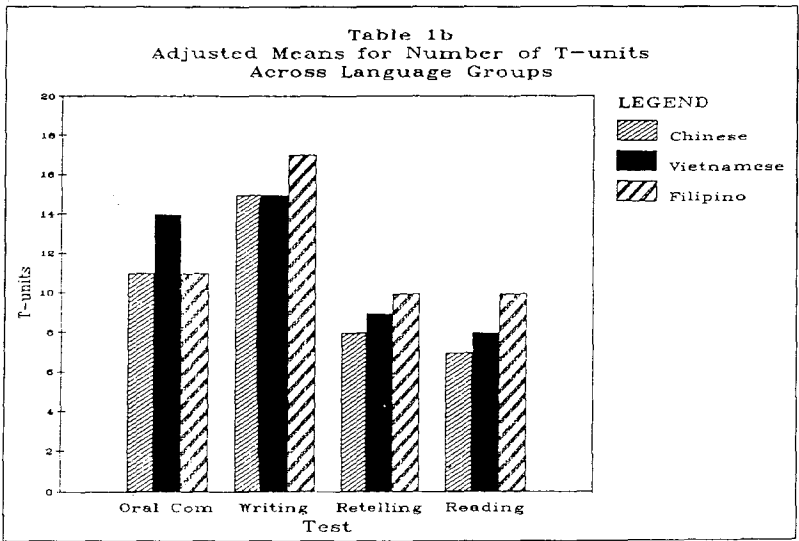
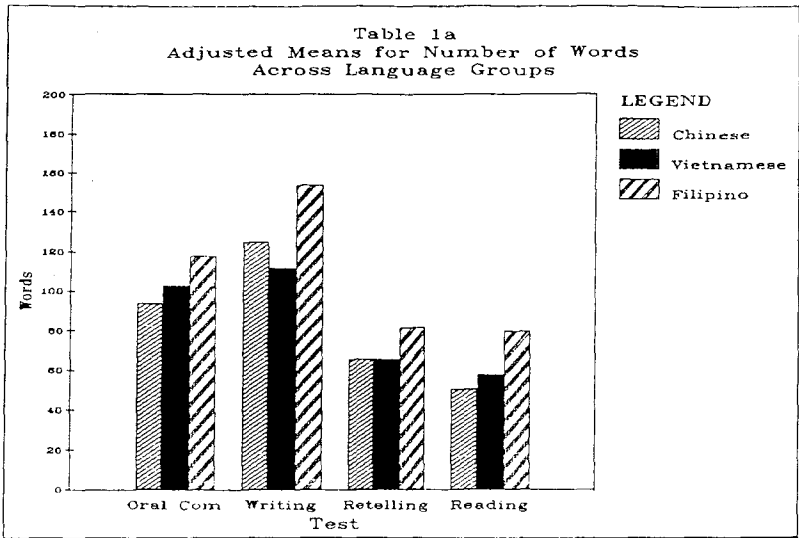
Table 1  
Adjusted Means for Number of Words, Number of T-units, Words per T-unit, and Dependent Clauses in Four Modes Across Language Groups

Mode	Language Group											
	Chinese				Vietnamese				Filipino			
	W	T	W/T	DC	W	T	W/T	DC	W	T	W/T	DC
Oral Com	94	11	8.18	1.57	103	14	7.93	0.89	118	11	10.44	2.75
Writing	125	15	8.48	2.20	112	15	8.05	1.27	154	17	9.40	3.69
Retelling	66	8	6.85	1.18	66	9	7.20	0.79	82	10	8.39	1.68
Reading	51	7	6.04	0.98	58	8	6.74	0.64	80	10	8.02	1.11

W = Number of words produced  
T = Number of t-units produced  
W/T = Number of words per t-unit  
DC = Number of dependent clauses

Statistical analysis for language groups across mode are presented in Table 2. The most obvious conclusion is that the number of dependent clauses is the most consistent measure which differentiates among the language groups — Chinese, Vietnamese, and Filipino — and does so

only in the oral composing and writing modes. On the other hand, the number of words per t-unit differentiates only in the oral composing mode.



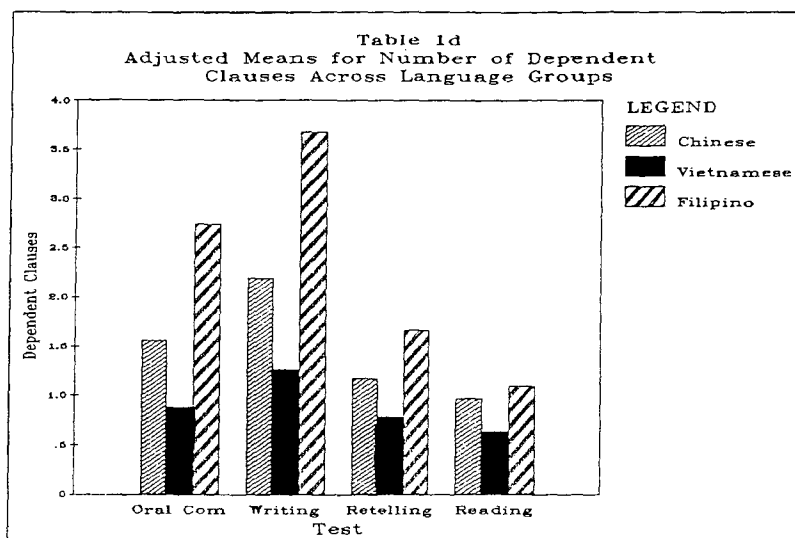
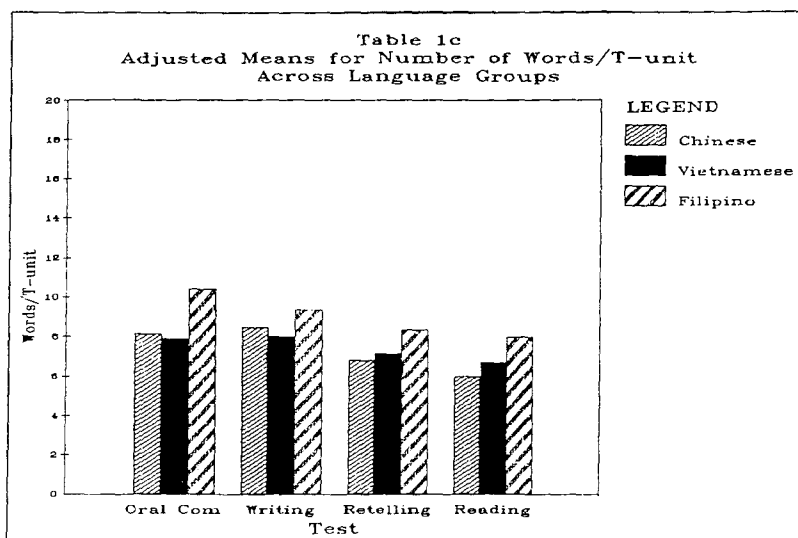




Table 2  
Analysis of Covariance for Four Modes Among Language Groups  
— Chinese, Vietnamese, Filipino

	Unit	F-ratio	Probability	Significance
Mode				
Oral	W	0.96	0.3929	NS
Composing	T	1.30	0.2846	NS
	W/T	4.85	0.0139	S
	DC	5.18	0.0107	S
Writing	W	1.68	0.2007	NS
	T	0.66	0.5237	NS
	W/T	1.24	0.3030	NS
	DC	3.85	0.0308	S
Reading	W	3.04	0.0608	NS
	T	2.17	0.1292	NS
	W/T	2.37	0.1080	NS
	DC	0.69	0.5091	NS
Retelling	W	0.95	0.3960	NS
	T	0.91	0.4111	NS
	W/T	1.24	0.3031	NS
	DC	2.01	0.1490	NS

$p < .05$

W = Number of words produced

T = Number of t-units produced

W/T = Number of words per t-unit

DC = Number of dependent clauses produced

Question #2:

*How do oral reading miscues compare?*

In terms of miscues, accuracy in all language groups was relatively high: Filipino 99%, Chinese 91%, and Vietnamese 87%. The ANCOVA results

presented in Table 3 indicate that there was a significant difference among the groups, and Duncan's post-hoc test indicated that the Filipino group was significantly different from the other two, but that the results for the Vietnamese and Chinese speaking students were not different.

Table 3  
Analysis of Covariance for Comprehension Measures Across  
Language Groups — Chinese, Vietnamese, Filipino —

	F-ratios	Probability	Significance
Comprehension Measures			
Miscues	4.52	0.0180	S
Reading			
Total Comp.	3.59	0.0381	S
Inferential	1.15	0.3295	NS
Cued	1.10	0.3439	NS
Uncued	5.68	0.0073	S
Retelling			
Total Retelling	1.46	0.2454	NS
Inferential	0.79	0.4606	NS
Cued	1.53	0.2312	NS
Uncued	1.82	0.1769	NS

$p = < .05$

### Question #3:

*How does reading comprehension and listening comprehension (i.e. retelling comprehension) compare?*

As might be predicted from the miscue analysis, the Filipino language group averaged 73% on total comprehension; the Chinese language group 49%, and the Vietnamese language group 57%. Table 3 presents the ANCOVA comparisons, and as may be seen, the difference among the language groups for reading comprehension is significant and the Duncan test indicates that the Filipino group is the superior one in comprehension — the other two groups do not differ from each other. A further analysis (also Table 3) indicates that uncued comprehension is the measure which best differentiates among language groups. It is also of importance to note

that the measures of inferential comprehension were NOT statistically different among the groups, although it might be considered low for all groups (i.e. Filipino 54%, Vietnamese 42%, Chinese 40%).

## CONCLUSIONS

These analyses must be considered interim in nature and hence the conclusions tentative. Also, it might be recognized that these results may not be generalizable beyond the group examined here. The intent of the study was to explore what seemed to be promising, valid, and reliable measures of language which might be more appropriate to ESL students than many of the measures presently employed. Hence the study is exploratory in nature; it is not intended to be definitive. Further, much of the more fine-grained analyses (and perhaps more revealing) have not been completed at this time and these might well temper the omnibus statistics presented here. Also, in order to control the varying length of time that students have been in Canada (Vietnamese=3.0 yrs; Chinese=3.9 years; Filipino=5.7 years), the covariance technique was used to adjust all means. While statistically acceptable, it may not be the best approach.

Turning to the findings, then, it becomes obvious that there are modal differences. That is, in rank order of production we find Writing producing the greatest quantity, Oral composing coming second, Retelling coming third, and recalling information for Reading coming last. This is not the expected order in first language learning and perhaps this fact could help teachers in organizing for instruction in a differentiated way for ESL students. The exact nature of these differences and their implication remain to be explored. The fact that the use of dependent clauses differentiates among language groups should be of interest; as should be the fact that words per t-unit were not differential (see also Crowhurst 1983). It is interesting, however, to compare the number of words per t-unit produced by these ESL students to the results obtained in the Manitoba Writing Assessment Program (1983) which represents a provincial average. The independent writing average for Grade 6 was 10.03 words per t-unit as compared to 9.40 for the Filipino group, 8.48 for the Chinese group, and 8.05 for the Vietnamese group. On the other hand, the number of dependent clauses was 9.23 for Manitoba, 3.69 for the Filipino group, 2.20 and 1.27 for the Chinese and Vietnamese groups respectively. Comparisons for dictation, retelling and reading are, of course, not available. Schewe (1986) working with fourth-graders found them to produce 10.7 words per t-unit for the high group, 7.4 for the average, and 7.6 for the low when writing independent compositions.

The accuracy of oral reading is perhaps surprising (i.e. 87-97%). But the nature of the miscues must first be analyzed to determine what the errors mean for these ESL students. Certainly the close relationship between miscues made and reading comprehension attests to the validity of these measures. In this study the rank ordering of miscue accuracy and comprehension accuracy among language groups was identical (i.e. Filipino, Vietnamese, Chinese in descending order).

The comparison of reading comprehension and retelling comprehension (i.e. listening) is unique in this study. While the retelling task confounds listening comprehension and general comprehension, it provides us with a measure that eliminates "decoding" in the usual sense. Table 3 indicates that total reading comprehension (cued/uncued) is significantly different among the language groups (i.e. the Filipino group is significantly higher than the other two groups by Duncan's test); but total retelling comprehension is not. Perhaps this suggests that students are able to understand oral language — the story read to them — but when asked to decode it themselves certain difficulties arise. Again, more fine-grained analyses are in progress and are necessary to understand the reasons for these results. Also of possible importance is the finding that students in the language groups do not differ in their ability to deal with inferential comprehension, although this must be tempered by the fact that the scores were relatively low in all groups (i.e. 62%, 38%, 36%).

In conclusion, then, the findings from this study must be considered tentative since the more fine-grained analyses are not yet completed. On the other hand, the data raise an interesting array of questions for the teacher and researcher. Perhaps, as Murphy suggests: "Enough research will tend to support your theory" or "Once you open a can of worms the only way to recan them is to use a larger can."

#### FOOTNOTE

- \* Support for this research was provided by the Department of the Secretary of State: Multiculturalism and by the University of Manitoba Research Board.

---

#### REFERENCES

- Baratz, J. (1977). *Requiring Performance Standards for Children: What is the State's Responsibility?* Educational Policy Research Institute of the Educational Testing Service.
- Braun, C. (1969). *A Transformational Analysis of Syntactic Development of Children from Varying Ethnolinguistic Communities*. Winnipeg: Manitoba Department of Youth & Education.
- Bruce, Bertram and Bolt Beranek and Newman Inc. (1981, July). *A New Point of View on Children's Stories. Reading Education Report No. 25*, Center for the Study of Reading. Champaign, Illinois: University of Illinois.

- Crowhurst, Marion. (1983). Syntactic complexity and writing quality: A review. *Canadian Journal of Education*, 8, 1-16.
- English Language Arts (1982): Middle Years*. Winnipeg, Manitoba: Province of Manitoba, Department of Education.
- Froese, Victor. (1983, November). A comparison of first-graders' ability in three modes of expression: Dictation, independent writing, and story retelling. Paper presented at the National Reading Conference, Austin, Texas.
- Goodman, Y. and Burke, C. (1972). *Reading Miscue Analysis: Handbook*. New York: Macmillan.
- Hay, T. A. (1984). A Study of Grade Two Conservers' and Non-conservers' Ability to Retell, to Tell, to Dictate, and to Write Stories. Unpublished Master's Thesis, University of Manitoba.
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. NCTE Research Report No. 3. Champaign, Illinois: National Council of Teachers of English.
- Hymes, Dell. (1980). On communicative competence. In A.K. Pugh and J. Swann (Eds.). *Language and Language Use*. London: Heineman Educational.
- King, Martha L. and Rentel, Victor, M. (1981). *How Children Learn to Write: A Longitudinal Study*. Columbus, Ohio: Ohio State University Research Foundation.
- Kintsch, W. and Yarbrough, J. C. (1972, December). Role of rhetorical structure in text comprehension. *Journal of Education Psychology*, 74, 823-834.
- Klassen, B. R. (1976). Sentence-Combining Exercises as an Aid to Expediting Syntactic Fluency in Learning English as a Second Language. Unpublished Doctoral Dissertation, University of Minnesota.
- Labov, W. (1976). Systematically misleading data from test questions. *Urban Review*, 9, 146-170.
- Loban, Walter, (1976). *Language Development: Kindergarten Through Grade Twelve*. Urbana: NCTE.
- Lado, Robert. (1964). *Language Testing: The Construction and Use of Foreign Language Tests*. New York: McGraw-Hill.
- Manitoba Writing Assessment Program 1982, Preliminary Report (1983)*. Winnipeg, Manitoba: Department of Education.
- McConaughy, S. H. (1982). Developmental changes in story comprehension and levels of questioning. *Language Arts*, 59, 580-589.
- Moffett, J. & Wagner, B. J. (1973). *Interaction: Student-Centered Language Arts and Reading*. Boston: Houghton Mifflin.
- Neilsen, L. (1977). The Effect of Headed Nominal Complexity and Vocabulary on Qualitative Judgements of Written Composition. Unpublished Master's Thesis, University of Minnesota.
- Oller, John W., Jr. (1979). *Language Tests at School*. London: Longman.
- Samuda, Ronald J. and Crawford, Douglas H. (1980). *Testing, Assessment, Counselling and Placement of Ethnic Minority Students: Current Methods in Ontario*. Ministry of Education, Ontario.
- SAS Institute Inc. (1984). *SAS user's guide* (Version 5.15). Cary, North Carolina.

- Schewe, A. C. (1986). Relating Reading and Writing: A Study of Fourth-Graders' Use of Story Schema. Unpublished M.Ed. Thesis, University of Manitoba.
- Shuy, Roger W. (1981, May). A holistic view of language. *Research in the Teaching of English*, 15, 101-111.
- Stotsky, Sandra. (1981, November). A review of research on the relationship between reading and writing: Directions for future research. Paper presented at National Council of Teachers of English Conference, Boston, Massachusetts.
- Stotsky, Sandra. (1975). Sentence-combining as a curricular activity: Its effect on written language development and reading comprehension. *Research in the Teaching of English*, 9, 30-71.
- Valette, Rebecca M. (1977). *Modern Language Testing*, Second Edition. New York: Harcourt Brace Javonovich.

### THE AUTHOR

Dr. Victor Froese is currently Head of the Department of Language Education at UBC; however, the work presented here is based on research conducted while at the University of Manitoba in Winnipeg. Dr. Froese has written extensively in journals such as *Language Arts*, *The Reading Teacher*, and *The Alberta Journal of Educational Research* and has co-authored and edited two books *Research in the Language Arts: Language and Schooling* and *An Experience-based Approach to Language & Reading*. He is also co-editor of *English Quarterly*, a journal of the Canadian Council of Teachers of English dedicated to scholarship, research, and theory.