



course or some aspect of the curriculum is having an effect on participants. In the context of second language programmes, a major aspect of programme evaluation concerns, of course, the extent to which the curriculum has the desired effect on students' second language skills. Bilingual immersion programmes, for example, are often evaluated to determine whether they are successful in promoting high levels of skill in the second language (cf. Lambert & Tucker, 1972; Swain & Lapkin, 1982). These programmes are also evaluated to see whether they interfere with or promote development in knowledge in non-language course content (e.g. mathematics) and in first language skills.

It is also meaningful to ask, however, whether programmes have an effect on those cognitive and affective individual difference variables that are thought to influence second language learning. By and large not too much research has been conducted on the effects of language programmes on cognitive variables, largely because it is generally believed that characteristics such as intelligence and language aptitude are relatively stable and not susceptible to modification. Cognitive language learning strategies in all probability are susceptible to change and development, though little research seems to have been directed to this question — quite likely because the language learning strategy research area is relatively new.

Somewhat more research has been concerned with the effects of language programmes on the affective dimension, or at least one aspect of it, viz., attitudes and motivation. One reason for this might be that it has come to be accepted that language programmes can influence students' attitudes and motivation. This is sometimes explicitly stated in curriculum guidelines. As early as 1974, the Gillin report prepared for the Ontario (Canada) Ministry of Education stated that:

Since motivation is one of the key factors in the success of any educational endeavour, the Committee hopes to promote this by introducing French to younger children and by presenting them with programs suitable to their age and interest level. In this way, favourable attitudes towards French can be created and maintained not only during their years of formal study of the language, but throughout their lives. (Ontario Ministry of Education, 1974: 3).

Curriculum guidelines often promote such goals, and even extend them to include an appreciation of the other language groups and other groups in general. For example, the following goals were articulated by the Ontario Ministry of Education in 1980:

Programs in French as a second language support these goals by assisting students:

- to develop communication skills;
- to begin to understand the structure and functioning of language;
- to pursue the mastery of a complex system of knowledge and skills;
- to acquire a sensitivity and exactness in the use of language;

— to gain an appreciation of the French presence in Canadian life and in the world;

— to develop sensitivity to culture and to people.

The ability to communicate in French can provide students with an additional tool to further their education and experience. [emphasis added] (Ontario Ministry of Education, 1980: 3).

As a consequence of such guidelines, it is reasonable to expect programme evaluations to include attitudinal components as well as linguistic and academic ones. Though not as common, some evaluations of immersion programmes, for example, have included attitudinal variables in their assessments (cf. Lambert & Tucker, 1972; Swain & Lapkin, 1982).

### A Theoretical Model Applicable to Programme Evaluation

One way of approaching the problem of programme evaluation as it relates to the affective dimension is by considering it from a particular theoretical perspective. One theoretical model that seems relevant in this case is the socio-educational model of second language acquisition (Gardner & Smythe, 1975; Gardner, 1985a). This model has four components, the sociocultural milieu, individual differences, language acquisition contexts and outcomes. A schematic representation of the model is presented in Figure 1; this is an integration of two versions of the model presented in Gardner (1985a). In this model, it is proposed that second language acquisition takes place in a particular socio-educational context and that this context must be considered when investigating factors associated with language learning. This seems particularly important when the question at issue is the effect of introducing some new aspect into the curriculum and investigating its effects on the affective characteristics of students, or simply on determining the affective implications of an existing programme.

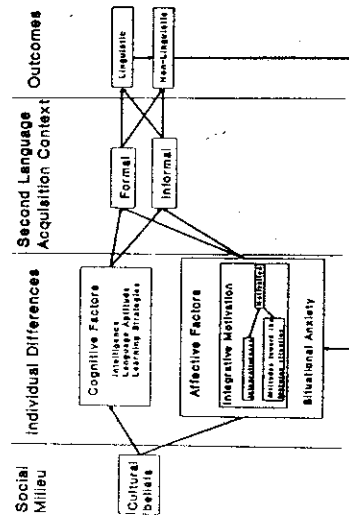


Figure 1 Schematic representation of the Socio-Educational Model

The student in any language course brings with him or her a host of pre-conceptions, inclinations and perceptions from the home background, and these can be either changed or reinforced in the classroom. Moreover, the language class itself exists in a social setting which has its own beliefs about the purpose and utility of the class and the outcomes to be expected. In the model all of these are subsumed under the heading of Cultural Beliefs, and these are considered to influence not only the development of both cognitive and affective individual differences but also their respective roles in the language learning process. This influence is depicted by the solid arrows linking Cultural Beliefs to the two classes of individual difference variables, Cognitive and Affective Factors.

Cognitive factors refer to those intellectual abilities that influence how quickly and easily individuals acquire second language material. Intelligence is viewed to play a role in that it would account for some individual differences in reasoning skills and in the degree to which lessons are understood and processed. Language aptitude refers specifically to those verbal skills, knowledge and abilities that relate directly to the language learning context. Intelligence would be assessed by means of any one of a number of valid indices of intellectual functioning. Language aptitude is measured by tests such as the Modern Language Aptitude Test (Carroll & Sapon, 1959), the Elementary Form of the Modern Language Aptitude Test (Carroll & Sapon, 1967) and the Pimsleur Language Aptitude Battery (1966). They assess cognitive abilities such as phonetic coding, grammatical sensitivity, memory ability and inductive language learning ability (Carroll, 1974). Some language learning strategies (see Oxford, 1986) could similarly be seen as instances of cognitive factors. Strategies refer to techniques that some individuals use to help them acquire and retain second language material and skill, and some of these strategies involve intellectual or linguistic aids. Oxford & Nyikos (1989) conducted a factor analysis of items describing various techniques used to facilitate language learning and identified five dimensions of language learning strategies. Two of them, *Formal Rule-Related Practice Strategies* and *Resourceful Independent Strategies*, would appear to reflect clear instances of cognitive-based strategies.

Affective factors refer to those non-cognitive variables that also have been shown to influence second language acquisition. To date, there are two major classes of these factors, motivation and situational anxiety. (As indicated above, one could consider a third class, viz., language learning strategies that focus on motivational aspects of learning, but at present it seems more parsimonious to consider them under the general category of motivation.) Motivation is viewed as comprising three components, motivational intensity, desire to learn the language, and attitudes towards learning the language. In Gardner's (1985a) socio-educational model it is argued that the truly motivated individual must display all three qualities, and that any one in and of itself does not really reflect motivation.

It is possible that there may be many motivational forces operating, but one that has been investigated quite substantially is the integrative motive. This refers to a constellation of attitudes and motivational characteristics subsumed under three categories, Motivation, Integrativeness, and Attitudes toward the Learning

Situation. It is assumed that Motivation is supported and facilitated by Integrativeness and Attitudes toward the Learning Situation. Integrativeness is reflected in such attitudes as an open willingness to interact with members of the other community (i.e. an integrative orientation to language study), an interest in foreign languages in general, and attitudes toward the second language community. Attitudes toward the Learning Situation are reflected in attitudes toward the teacher and toward the course.

The other class of affective variables is situational anxiety. In some formulations, situational anxiety is seen as having motivational components; however, most research to date has tended to consider it as an element in and of itself. Situational anxiety refers to such things as feelings of anxiety in the language class, or when called upon to use the second language in communicational settings. There is even research that suggests that such anxieties may have different components. Horwitz, Horwitz & Cope (1986), for example, posit three components of Foreign Language Class Anxiety: communication apprehension, test anxiety, and evaluation apprehension. As indicated above, it is not clear at present where situational anxiety belongs with respect to motivation in general and the integrative motive in particular. Some research (Gardner, Lalonde & Pierson, 1983; Lalonde & Gardner, 1984) has considered it to be caused by motivation (i.e. high levels of motivation cause low levels of anxiety), while other studies (e.g. Gardner, Smythe, Clément & Gijksman, 1976) view it as a separate correlate of proficiency. In any event, it is clear that indices of situational anxiety tend to correlate negatively with elements of the integrative motive (Gardner & Smythe, 1981; Trylong, 1987) even though they sometimes also share variance in common with other factors (Clément, Gardner & Smythe, 1977a; Clément & Kruidenier, 1985; Gardner, Smythe & Lalonde, 1984).

In the theoretical model depicted in Figure 1, both cognitive and affective variables are shown as having an effect on both formal and informal language acquisition contexts. Formal language acquisition contexts refer to those situations in which individuals receive direct language instruction. Individuals may or may not be in a formal context on a voluntary basis, but they are there for the express purpose of learning language material. Such contexts have instructors and some type of curriculum. The major characteristic of informal contexts, on the other hand, is that they are voluntary. Some instruction may be involved but by and large an informal language learning context provides opportunities to learn language material largely through exposure. Examples of informal contexts would be listening to a radio broadcast in the other language, reading novels, magazines or newspapers in the other language, speaking and listening to individuals in the other language, etc.

It is assumed that both cognitive and affective factors play direct roles in formal language acquisition contexts. Intelligence, language aptitude, language learning strategies and motivation would all aid in the acquisition of material taught, while situational anxiety would be expected to impede the learning of the material. Only affective factors would be expected to play a direct role in the informal contexts. Individuals with high levels of motivation would be expected to volunteer to participate, while those with high levels of anxiety would avoid

such contexts (Ely, 1986). Cognitive factors would play an indirect role (indicated by the broken line in Figure 1) since they would be implicated only if the individual chose to participate. Once in the situation, however, individual differences in intelligence, language aptitude and strategy use would play a clear role in the amount of material acquired.

Both formal and informal contexts could have both linguistic and non-linguistic outcomes. Linguistic outcomes refer to any acquisition of knowledge, skill or facility in the language, while non-linguistic outcomes refer to any other result of the experience, varying from a desire or willingness to use some aspect of the language itself to general feelings of self-esteem. All of the affective variables that are viewed as factors involved in the learning of the language could also be non-linguistic outcomes of language training. Thus, the model is dynamic, not static. The affective variables influence behaviour in both formal and informal contexts and are in turn influenced by the contexts and the achievement that is developed in these contexts.

This model has clear implications for research concerned with programme evaluation focusing on the affective dimension. It provides an overview of factors to consider — the sociocultural context, cognitive and affective factors that could interact with programme changes, and the potential role of both formal and informal language learning contexts. Moreover, it highlights the fact that language learning has both linguistic and non-linguistic outcomes. It also focuses attention on a number of affective variables that can influence language learning and in turn be influenced by language learning. The next section briefly considers tests that have been constructed to measure some important affective variables. These measures might be of use to researchers wishing to conduct programme evaluation studies and deal with affective variables associated with second language learning.

### Measures of Some Affective Variables

In our research, we developed the Attitude/Motivation Test Battery (Gardner & Smythe, 1981) to measure the major affective variables represented in this model. Because of the very important role played by the milieu in which language learning takes place, the test battery itself is rather dynamic in that the items often are modified or adapted to suit the situation under investigation. Some items simply are not appropriate to some settings, some ages of students, some programmes, and the like. The structure is, however, quite stable. The test battery generally consists of 9 or 10 subtests assessing the four components discussed above. As used with students learning French in the Canadian context, these would correspond to the following:

#### Motivation

- (1) Motivational Intensity
- (2) Desire to Learn French
- (3) Attitudes toward Learning French

#### Integrativeness

- (1) Integrative Orientation
- (2) Attitudes toward French Canadians
- (3) Interest in Foreign Languages

#### Attitudes toward the Learning Situation

- (1) Evaluation of the French Course
- (2) Evaluation of the French Teacher

#### Situational Anxiety

- (1) French Classroom Anxiety
- (2) French Use Anxiety

Other subtests are sometimes included in the battery. Two examples of such additions are the measures of Instrumental Orientation, and with younger students, Parental Encouragement to learn French. These two variables, however, are not considered major affective variables in much of our research. The preceding model and tests have developed from our work on the role of affective factors in second language learning. Initially our concern was with determining the extent to which these variables were related to achievement in the second language. Numerous studies demonstrated reasonable associations between scores on these measures and achievement in the second language (see, for example, Gardner & Lambert, 1959; 1972; Gardner & Smythe, 1975, 1981; Gardner, Smythe, Clément & Glikzman, 1976). Much of this research is reviewed by Gardner (1985a).

Other research has focused on other generalisations deriving from the model. For example, it was hypothesised that to the extent that integrative motivation plays a direct role in the formal learning context, it would be expected that integratively motivated students would volunteer more in class, be correct more frequently, and enjoy classes more than students not so motivated. This was confirmed by Glikzman (1976) and supported by Naiman, Fröhlich, Stern & Todesco (1978) (see also Glikzman, Gardner & Smythe, 1982), who assessed characteristics of the integrative motive using the subtests of the Attitude/Motivation Test Battery. In a similar vein, since integrative motivation was expected to play a direct role in informal contexts because it would determine who would or would not participate in them, it was predicted that participation in voluntary excursion programmes to the other community and the willingness to make contact with members of the other community would vary as a function of scores on the Attitude/Motivation Test Battery. This expectation similarly was supported (Clément, Gardner & Smythe, 1977b; Desrochers & Gardner, 1981). Finally, if integrative motivation is important for language acquisition because it orients the individual to seek out opportunities to learn the language, we reasoned that the tendency to enrol in subsequent classes would similarly be related to elements of the integrative motive as assessed by the Attitude/Motivation Test Battery. This too was supported by research (see Clément, Smythe & Gardner, 1978; Gardner & Smythe, 1975).

The socio-educational model of second language acquisition emphasises the importance of both formal and informal language acquisition contexts, but the research associated with it has seldom explicitly considered the nature of the contexts in which language study takes place. In fact, in most studies, context effects have been eliminated by standardising the data within classes. The focus has thus been on the relation of attitudes and motivation to achievement relative to everyone in a particular classroom setting. This was necessary quite simply because it was recognised that differences among curricula, programmes and teachers could play a role both in terms of language achievement and the affective variables themselves. For example, students who are in a programme stressing grammar knowledge might perform better on measures of grammar than those in a programme stressing audio-lingual components; those with a demanding teacher might perform better on tests of achievement than those with a more lenient one. Those with the demanding teacher may evidence higher levels of motivational intensity than those with a more friendly teacher, and they might have less favourable attitudes towards learning the language. By standardising within class, these class differences are eliminated, and attention is focused directly on differences in attitudes, motivation and achievement within the class itself and how they relate to achievement in the language, participation in class, or continuing on with language study.

Rather than playing down the importance of course-related factors, this approach recognises their unique contribution, but considers this as unwanted variance where attention is directed to the relation of attitudes and motivation to second language achievement (also see Gardner, 1985a: 78-80). These factors, however, become sources of important variance when the focus is on programme evaluation.

The theoretical model thus could easily be expanded to consider the explicit role played by different types of formal (and possibly even informal) contexts. This would require identifying subcategories of the formal context, and then studying whether they lead to different levels of linguistic and non-linguistic outcomes. It is even possible that the relationships of cognitive or affective factors with outcomes could vary as a function of the nature of the formal context. A basis of such classification of the formal context might be in terms of the type of programme, the nature of the curriculum, or characteristics of the teacher. It may even be that a factor such as the time of day in which the language is taught or the age of the students might reflect other meaningful ways of characterising the language acquisition context. Systematic research is required to determine what contextual factors are important, and to some extent studies evaluating particular programme innovations are one way of identifying them.

There are some studies that have focused on programme evaluation of the affective dimension, and, although they were not conducted with the socio-educational model in mind, its relevance is quite obvious. In the following section, attention will be directed towards some of these studies and relevant issues that derive from this research. Research evaluating the effects of specific second language programmes on students' attitudes and motivation can be grouped into two categories, first, there are those studies that focus on the effects primarily

of language training (regular and intensive language training courses); second, there are those directed towards the effects of contact with the other community (excursion programmes).

#### Programme Evaluation Involving Regular Second Language Courses

Research has indicated that students' attitudes and motivation towards studying a second language may be affected by specific programme variables. For example, Massey (1986) found that for Grade 6 and 7 children the principal factor for change in students' attitudes and motivation in the rural and suburban school was the students' attitude towards what happened day by day in class. Similarly, Burstall (1975) found that early achievement in French affected later attitudes towards learning French and later achievement in French (as expected in the socio-educational model). These studies would suggest that the acquisition of second-language skills and the development of attitudes towards second language learning during later years may be powerfully influenced by the learner's initial and formative experience of success or failure in the language-learning situation.

Some research has been conducted that compares students in a particular language programme with students in some other setting. Since subjects in this type of research are tested only once, after they have experienced the programme, this type of study is often referred to as a post-test only design. Other studies test the same students on more than one occasion, sometimes comparing the student with others taking some other type of programme, but also tested at the same times. These types of studies make use of what is often referred to as a pre/post design. If contrasting groups are not included, these are sometimes referred to as longitudinal studies, but since they involve two or more testings of the same subjects, they represent only a special case of the pre/post design, hence the latter distinction is not made here.

#### Comparisons of groups using a post-test only design

The post-test only design has been used in some studies. For example, Smythe Stennett & Feenstra (1972) found that students in a traditionally oriented French course had higher grades in French and more positive attitudes towards French than students in an audio-lingual French course. In like manner, Riestra & Johnson (1964) compared 63 grade five students who had been studying Spanish as a second language for two years with 63 students who had not studied Spanish, but who were matched in terms of sex, age and intelligence. They found that those students who had studied Spanish had significantly more favourable attitudes towards Spanish-speaking people than those who had not studied Spanish. The two groups did not differ in attitudes towards non-Spanish-speaking people. Finally, Halpern *et al.* (1976) investigated the effects of instruction time on attitudes towards French Canadians and towards learning French. They contrasted samples of students in regular programmes with those in

extended programmes and found no effects for students in grades 1 and 2 but some effects for students in grades 5 through 7.

Stennett & Earl (1982) also used a post-test only design in their investigation of instruction time and its effects on achievement and affective factors. Rather than comparing different classes selected for differential treatment, they capitalised on a change being made throughout the entire system over a three-year period. In the first year, instruction in grades 7 and 8 took place for 20 minutes each day. In the second year, students in grade 7 received instruction in French for 20 minutes per day, while those in grade 8 had 40 minutes per day. In the third year, students in both grades studied French for 40 minutes per day. Samples of 200-300 grade 8 students were tested on a number of French achievement measures and three attitude tests at the end of their school year. In addition to demonstrating the expected improvement on French achievement, the results demonstrated a significant increase in favourable attitudes towards French Canadians and a significant decrease in French class anxiety over the three-year period. No significant effects were evident on the measure of attitudes towards learning French.

A basic assumption underlying this type of approach is that the groups being compared were equivalent before the particular second-language learning experience was applied to some of the groups. If two groups are formed on the basis of random assignment, this is a reasonable assumption. If not, there is really no way of concluding unequivocally that the programme is in fact responsible for the differences. Even matching subjects in the groups, as was done in the Riestra & Johnson (1964) study, is no guarantee. It requires that the matching variables are the only ones that could account for any *a priori* difference in the dependent variables — and this is a very large assumption. Often, however, the post-test only type of design is the only one available to the researcher, and it seems reasonable to argue that it is better to conduct programme evaluation in this way rather than not at all. As long as its limitations are recognised, some useful information can be gained from this approach. Moreover, in cases where the pre-testing of affective characteristics may sensitise students to the affective nature of a programme intervention, the post-test only design with random assignment of subjects to conditions might be the preferred approach.

#### Comparison of groups using a pre/post design

Another way of comparing the effectiveness of a programme on students is to test them on more than one occasion, and sometimes to compare their performance with other students receiving different programmes. This is often referred to as a pre/post design. Using this type of approach, Gardner, Ginsberg & Smythe (1976) investigated the effects of self-instruction versus traditional instruction in a first year university French programme. The two programmes were comparable in that they used the same tests, drills and other materials; they differed in that students in the traditional programme took a relatively conventional language course while students in the self-instruction programme organised much of their own work while following a fairly structured curriculum.

There was a general tendency for attitudes to be less favourable on the post-test than the pre-test for both groups. More importantly, however, three measures showed differential changes for the two groups. Students in the self-instruction programme showed little change from pre-test to post-test in need achievement. Attitudes towards learning French and the intention to continue French study, while students in the traditional programme demonstrated appreciable declines on all three attributes. Such results suggest that self-paced instruction has a relatively positive effect on these characteristics.

Some students in the self-instruction programme missed the first testing for a number of reasons. It was thus possible to conduct another analysis comparing post-test scores for three groups of students, those in the traditional programme, students from the self-instruction programme who did not answer the pre-test questionnaires, and those who did. In this analysis, those students from the self-instruction condition who also completed the pre-test had more favourable attitudes towards French Canadians than the other two groups. Such results were interpreted as suggesting that the pre-test might sensitise students in some programmes to demonstrate attitude change. Of course, it may also be that there were some unknown differences between the two self-instruction groups that were responsible for the apparent post-test differences. This is possible since these two groups were not formed on the basis of random sampling.

A pre/post design was also used by McInnis (1976) to investigate whether classroom innovations could promote attitude change. In this study, however the researcher designed the study explicitly to permit an assessment of testing effects. He studied the effects of a 'Francobus Programme' on the attitudes of students in grades 2 through 8. This programme involved actor-teachers who visited the classrooms in the morning to sing, dance and teach the children French. In the afternoon they performed plays, gave puppet presentations, and supervised group dancing and singing, all in French. One study assessed pre and post attitudes using different groups of students, and found more positive attitudes after the experience. A second study investigated students who were tested both before and after experiencing the Francobus Programme. No significant changes in attitudes were obtained. A third study using analyses similar to the first one again showed positive effects. Clearly, as demonstrated by this study different results are obtained depending upon how one designs the research. Pre testing can have an effect!

Studies have also examined the effect of the programme on the development of children's attitudes towards a second language. What is generally found is that children come into school with positive attitudes and these attitudes decrease with age. For example, Shapson & Kaufman (1978) found that primary student in Core French generally hold positive attitudes towards French with 75% of children in Grades 1-3 surveyed wanting to continue to learn French, and 81% of children feeling that the time they spend in French is about right or too little. Shapson and Kaufman, however, found that children's attitudes towards French became less favourable the longer they study it. Similar results have been reported by Jones (1950) and Burstall (1975). In a somewhat related manner Nisbet & Welsh (1972) found that students who studied French in elementary

school had less favourable attitudes towards it in high school than students who did not. They hypothesised that since all pupils started French afresh in high school, the stimulus of novelty was lost for those who already knew some French and its popularity suffered. In any event, the early experience of learning French apparently had an effect on subsequent attitudes towards learning French.

#### Programme Evaluation Involving Intensive Language Programmes

These programmes can be classified in many ways, but the major distinction made here is between relatively long-term (i.e. one academic term or longer) and short-term (i.e. two to six weeks) programmes.

Two long-term intensive programmes that are most consistently studied are early immersion, and late immersion. Often students in these programmes are compared with students in a traditional core programme. In early immersion, students receive all instruction in the second language for the first few years and gradually the first language is introduced. In the late immersion programme all instruction is conducted in the second language, beginning in a later grade, generally Grade 7. Prior to that, students follow a traditional curriculum with one session per day in the second language.

The majority of evaluations of these programmes conducted in Canada involve anglophone students in French immersion. For example, Cziko, Lambert & Gutter (1979) found that early immersion students saw themselves as more similar to French Canadians, especially French Canadians who are bilingual, than did students in other programmes. Similarly, Edwards (1976) found that early immersion children had more positive attitudes towards learning French than did the children in a Core French programme.

Lambert & Tucker (1972) have also identified the importance of an early immersion programme in fostering positive attitudes towards French-speaking people. By grade 5, students' feelings towards French people had become decidedly more favourable and they now thought of themselves as being both French and English-Canadian in personal make-up. However, results obtained from students who participated in a later immersion programme did not show similar trends. Measures of attitudes towards French Canadians were obtained from grade 10 students who participated in the initial year of the grade 7 immersion programme in Montreal, and who had since taken up to 40% of their curriculum in French each year. Basically, their attitudes were no different from those of students who had been in the regular English programme. In other words, no change in attitudes towards French Canadians could be detected as a result of participation in a late immersion programme. These results suggest that one of the advantages of an early immersion programme is in the area of attitude development and change (Swain, 1974). Blake, Lambert, Sidoti & Wolfe (1981) also found differences between the attitudes of immersion and control subjects. Immersion subjects were more likely than control subjects to see English- and French-speaking people as similar. They concluded that early bilingual and bicultural experiences were promoting more receptive and less ethnocentric attitudes at an earlier age.

Cohen (1975) studied three samples of Mexican American children registered in bilingual education programmes at the elementary school level and compared them with comparable groups of children in conventional English-only schooling. One group had three years of bilingual education beginning in grade 1, another had two years of bilingual education beginning in kindergarten, and the third had two years of bilingual education beginning in grade 1. Attitudes were assessed by means of pictorial questionnaires using frowning and happy faces to indicate unfavourable and favourable responses respectively. For the first group, the students in the bilingual education programme expressed significantly more favourable attitudes towards the Mexican culture and towards school than did the comparison group. No significant effects were obtained with respect to student attitudes for the other two groups, suggesting possibly that two years of such training is not enough, particularly for very young children.

Research on short-term intensive language programmes has tended to focus on summer school programmes. In an early study, Lambert *et al.* (1963) investigated changes that took place during a six-week programme for adults in Montreal. The results showed that both elementary and advanced students increased significantly in anomie (feelings of normlessness) from the beginning to the end of the course, but that only the elementary students increased in authoritarianism. No significant changes were obtained on a measure of attitudes towards people from France.

A complex set of attitude changes were demonstrated by Gardner, Smythe & Brunet (1977) in their investigation of a five-week residential summer French programme for high school students. On the negative side, students demonstrated increases in ethnocentrism, a decreased interest in foreign languages, and a reduced integrative orientation for learning French. On the positive side, they decreased significantly in French classroom anxiety, became more motivated to learn French, felt the course was easier than their initial expectations, and reported having greater opportunities to use French.

Similar results were also obtained by Hoeh & Spuck (1975) in their investigation of a three-week programme abroad. During this time high school students lived with French families and attended French schools for two weeks and toured for the third week. Their attitudes were assessed on the day preceding their departure to France and one day following their return. Both positive and negative changes were observed. Positive effects included the students perceiving themselves as more potent, French people as more sociable, and life in France generally better. Negative changes included the perception that French family life is less cohesive and the French school system less valuable. Generally positive changes are reported by Chiebek & Coltrinari (1977) in their investigation of a European summer language programme for Canadian high school students. They analysed comments made by participants and concluded that the students developed considerable ease and self-confidence in French, more favourable attitudes towards learning French, more favourable attitudes towards other countries, greater maturity, and a broader appreciation for their own way of life.

The important role played by the nature of the programme was also demonstrated by Clément (1979). He conducted two studies, both of which involved

testing before and after the programme. One study involved a sample of 26 Yukon high school students, 12 of whom participated in a two-week immersion programme (with some formal instruction) in Quebec, and the remaining 14 serving as a control group. The second study employed 23 Yukon high school students, 13 of whom lived with French families in Quebec for eight days and received no formal language instruction. The other 10 students served as a control group. Two dependent measures were investigated in each study: attitudes towards French Canadians, and French use anxiety. Those students participating in the immersion programme demonstrated slight positive changes in attitudes towards French Canadians and large decreases in French use anxiety, while the control group showed relatively little change. Analyses for the second study showed that those participating in the residence programmes evidenced large positive changes in attitudes, while the control group showed virtually no change.

As was noted with the regular language courses, research with immersion programmes also has methodological problems. Although it may be demonstrated that students in different programmes have different attitudes one cannot assume that it is the programme that determines the student's attitude. Except on rare occasions, subjects are seldom assigned to long-term or short-term immersion programmes on a random basis. Pre-selection factors may account for many differences obtained (e.g. students who enter early immersion may have more positive attitudes towards French and may have parents who are more supportive than students who enter other French programmes). Moreover, those studies using pre/post designs have similar problems. There may also be pre-selection factors operating, or the pre-testing may attune subjects to potential programme effects. Any design has potential pitfalls.

#### Programme Evaluation Involving Excursion Programmes

Some language programmes seek to develop an interest in the second language and an appreciation of the other culture by providing for brief excursion to the other community. Often, too, educators are interested in evaluating the effects such excursion programmes have on the students concerned. In terms of the socio-educational model discussed earlier, this type of research would be focusing on the informal language learning context. To date, research on this question has led only to tentative conclusions, partly because the results have been somewhat ambiguous.

In an early study, Leonard (1964) compared university students before and after an excursion and found no significant attitude change. Similar null results were obtained by Hanna & Smith (1979) in a study of a four-week bilingual exchange programme involving French and English Canadian students, though they report that 'the exchanges did appear to have a positive effect on the attitudes of a minority of students' (p. 52). Somewhat more positive findings have been obtained in some other studies, though even here the effects were far from extensive. For example, Gardner *et al.* (1974) found a significant change in attitudes towards French Canadians among grade 8 students after a four-day

excursion to Quebec City, but no significant changes in other measures. In a similar study, Clément, Gardner & Smythe (1977b) also found that students who took part in the excursion, in contrast with their colleagues who did not, had more favourable attitudes towards French Canadians and towards learning French after the excursion (even after partialling out any variation attributable to pre-excursion attitudes). They also found, however, that the extent to which subjects reported interacting with members of the other community had an effect. Thus when the excursion subjects were classified as high or low contact based on their verbal reports, the high contact group differed significantly from the other groups on interest in foreign languages, instrumental and integrative orientations, parental encouragement to learn French, motivational intensity to learn French, evaluation and perceived utility of the French course, and attitudes towards European French people. These contrasts again were made on the post-excursion scores, once the effects of pre-excursion attitudes were removed. The implication is, therefore, that the extent to which individuals interact with members of the other community plays an important role in how much excursion programmes influence the affective dimension.

Desrochers & Gardner (1981) examined the effects of contact defined either in terms of self-ratings, as in the study above, or in terms of peer ratings. With self-ratings of contact, the high contact group differed from both groups on integrative orientation and parental encouragement to learn French, and from the group who didn't participate in the excursion on attitudes towards French Canadians, attitudes towards learning French and desire to learn French. When contact was defined in terms of peer ratings, the high contact group had more favourable attitudes towards French Canadians than the other two groups, and more favourable attitudes towards learning French, and lower French use anxiety than those who did not take part in the excursion. As above, these contrasts were based on post-excursion scores with the corresponding pre-excursion scores serving as covariates.

A somewhat different form of bicultural excursion was investigated by Cziko & Lambert (1976), with comparable results. They investigated English-speaking students who attended a French school for a one-day period at eight different times throughout the year, and compared them with a control group who did not take part in the excursion. They also distinguished between low and high contact participants, based on verbal reports. For one measure, attitudes towards learning French, they found no change for either the high contact or control group from the beginning to the end of the study, though they did find a significant decrease for the low contact group.

The results of these studies suggest that there may be some effect on affective characteristics of students as a result of excursions to the other community, but generally speaking the results are not pronounced. Moreover, there is evidence that the extent to which individuals actually seek out contact with members of the other community can play an important role in any affective changes that take place. To some extent the measures of contact in these studies are problematic in that in all but one case they were based on self-report. But even in that case, the results were sufficiently similar to suggest that the self-reports were



valid. Moreover, a study conducted by Hofman & Zak (1969) obtained comparable results, and their measure of contact was based on observations made by counsellors in the camp they were studying. Subjects in this case were participants of a five-week summer camp in Israel. Hofman and Zak report that subjects with high contact showed significant increases in the favourability of attitudes towards Jewishness and Israel during the five-week period, while low contact subjects demonstrated less favourable changes in attitudes towards Jewishness.

There is nothing in this research to indicate whether the changes are long-term, and this is unfortunate. It may be that any changes that do occur are relatively short-lived, and, if this is the case, the value of such excursions would be compromised. Clearly, some valuable research could be conducted along these lines. It is recommended, however, that any future research consider the issues raised later in this paper. All of the research reviewed in this section used a pre-test/post-test type of design, and, although this seems like a very straightforward way to assess the effects of any programme, it is not without its problems.

#### Effect of Teacher Variables on Students' Attitudes

In addition to pedagogical variables being important in themselves, it is also interesting to study the interaction between these variables and student characteristics. Gayle (1984) observed 14 classroom teachers and classified them in terms of four teaching styles. Significant interaction effects were found between specific teaching styles and specific categories of students defined by specific affective variables such as attitudes towards French Canadians, attitudes towards learning French and desire to learn French. Pre-treatment scores on aural comprehension were used as a covariate to eliminate any prior differences attributable to French achievement. More specifically, when instructed with teaching style H (lateral language) students with relatively less favourable attitudes towards French Canadians and low desire to learn French scored higher on aural comprehension than students with more favourable attitudes. In addition, when taught by style H, students with relatively less favourable attitudes towards learning French did not score quite as high as the high attitude students although they scored appreciably higher than the low attitude students did when taught by style L (linear language). This study indicates that students' attitudes may interact with teaching styles to influence the degree of achievement attained.

It is also possible that teacher attitudes might play a role in second language acquisition. Burstall (1975) found that the pupils' level of achievement in French was consistently rated more highly in those classes where the French teacher's attitude towards the class was considered 'positive' than in those classes where the French teacher's attitude towards the class was considered 'non-committal' or 'negative'. Thus, it is not only the students' affective characteristics that can influence achievement, but also the teachers'.

These potential effects of the teacher should also be considered in any research concerned with programme evaluation since such teacher effects could moderate any results obtained. Such effects are key sources of variance that could influence results, particularly when the studies focus on only a few classrooms.

#### Methodological and Analytic Issues

As indicated in the previous review, programme evaluation is not a straightforward task, and it is made somewhat more complex when the focus of attention is the affective dimension. There are a number of factors that must be considered in planning such research in order to ensure that the results are as meaningful as possible. Some of these factors involve issues of measurement, while other involve design and analytic considerations. This section will consider the two measurement issues of reliability and validity, and indicate how they have been investigated with respect to the Attitude/Motivation Test Battery. It will then focus on four factors that should be considered when designing programme evaluation studies and analysing the results obtained.

One important measurement issue is that of the reliability of the measuring instruments. Several methods for assessing reliability are available. The three that are most applicable to the paper and pencil measurement of affective characteristics are internal consistency, test/retest, and alternate forms reliability. The most basic type of reliability is that of internal consistency. This refers to the extent to which the items of a particular scale provide similar assessments of the same attribute. This is reflected in the items having similar variances, and similar intercorrelations with each other, and is referred to as psychometric parallelism (see Lord & Novick, 1968). For self-report measures, this tends to be reflected in high values of Cronbach's coefficient alpha, or high values of the split-half reliability coefficient. Cronbach alpha internal consistency measures of reliability are often reported in research using the Attitude/Motivation Test Battery, and the coefficients are generally quite substantial (see also Gardner, 1985b).

Another index of reliability is test/retest reliability. This is assessed in terms of the correlation between scores obtained on the same test given twice, and is thus a measure of the stability of scores over time. It is meaningful only if it can be assumed that the attribute under investigation does not change between the first and second testings. Obviously in programme evaluation studies this might not be the case. In fact, it is reasonable to assume that test/retest reliability coefficients of very reliable tests might be quite low in a programme evaluation study if the programme is having varying effects on some subjects. Test/retest reliability coefficients have been reported for the Attitude/Motivation Test Battery, and these values are generally quite high (see Gardner, 1985b).

The third index of reliability is alternate forms reliability. It is assessed in terms of the correlation of two different forms of the same measure. Unfortunately there are not many alternate forms of measures appropriate to assessing the affective dimension in second language learning. Some research using the Attitude/Motivation Test Battery has used alternate forms by the simple expedient of dividing the original measure into two (Gardner, Smythe & Brunet, 1977; Naiman, Fröhlich, Stern & Todesco, 1978). Unfortunately measures based on only half the number of items are not as reliable as the full test.

The issue of reliability is particularly important for programme evaluation studies because the interest in such studies is in changes that result from the intervention of the 'new' programme. Regardless of whether or not the same dependent measure is administered twice, it is important to have relatively error-

free measuring devices when assessing change. One factor over which a researcher has obvious control and which greatly influences the reliability of a test is the number of items comprising that test. Researchers should thus attempt to construct tests having a number of items assessing the same attribute, particularly when measuring affective characteristics. Subtle wording differences between individual items may influence individuals' reactions to them at any particular time, and with a large number of related items, these idiosyncratic effects tend to balance themselves. As a result, the total score for a fairly long test can give a much more reliable index of the attribute in question than that provided by only a few items. This is reflected in all three types of reliability.

Because of the above considerations, single item measures of any attribute should be avoided if at all possible. Internal consistency and alternate forms reliability obviously cannot be computed with single item measures, and even though test/retest coefficients can be computed, it is likely that they would be low because of the idiosyncratic effects referred to above. Since single items tend to yield unreliable measures, it is important that researchers devote considerable time to developing a homogeneous set of items for any constructs they consider important. With poor reliability, it is unlikely that any valid results will be obtained.

In programme evaluation studies, there is a lot to be said for using alternate forms of measures if a pre-test/post-test design is to be used. With repeated testing, the second and any subsequent administration has lost much of the novelty effect that occurred the first time and therefore the testing conditions may not be the same. Also, students may attempt to recall their previous responses in an attempt to appear consistent rather than responding to the items as they currently feel. By using alternate forms, these carry-over effects are reduced somewhat. This problem diminishes, however, as the length of time between testing sessions increases. In fact, for long delays between testings, alternate forms may not be that necessary; however, for short-term programmes, as in the excursion studies, using alternate forms would be a preferable procedure.

The second important measurement issue, that of validity, can be considered from a number of points of view. There are several ways of assessing validity, and each of them involves somewhat different questions being asked of the data. Five different forms of validity could be considered that are appropriate to programme evaluation studies that focus on the affective dimension. One question, for example, would be, 'Do scores on the test in question correlate as one would expect with scores on another test that is administered at the same time?' If it was believed that the two tests were measuring the same attribute, one would expect them to be highly correlated. This is an example of concurrent validity. Another question might be, 'Do scores on the test of interest correlate with (or predict) scores on some other test administered at some later date?' This would be an instance of predictive validity. A third question relates to face validity. One might ask, 'Do the items all appear to be measuring the same attribute?' This is the simplest form of validity, and also the most problematic in the context of programme evaluation of the affective dimension. If the items of a test all appear to be measuring a common attribute, it is possible that subjects may recognise the purpose of the assessment and answer accordingly.

A fourth aspect of validity is content validity. The question here is, 'Given that there is a population of items measuring the attribute in question, are all the items in the test representative of this population?' This type of validity is that considered to be of major importance by researchers concerned with test construction. Much of their work is concerned with clearly delineating the population of interest and ensuring that it is appropriately represented. This type of validity underlies the notion of psychometric parallelism referred to earlier.

A final form of validity involves construct validity. In this case the question is, 'Does the measure in question adequately reflect the construct it is purported to measure?' For programme evaluation, the most relevant consideration may be the extent to which a test measures the objectives of instruction. This is sometimes referred to as curricular validity. For example, in a French course, an excursion programme to a French-speaking community would be expected to influence attitudes towards the French people more than would a course primarily concerned with the subtleties of French grammar. The key here is the degree to which a programme changes the *relevant* attitudes, and the decision lies with the investigator. These sorts of decisions are not made lightly, however. They should be predicted before the study is initiated and should be based on a careful analysis of the objectives of the proposed programme and the processes that might account for any expected changes. The socio-educational model of second language acquisition described earlier in this article might well provide the basis for such an analysis.

In addition to the measurement issues of reliability and validity, there are a number of factors to be considered in the design of any programme evaluation study. Moreover, these issues must be resolved in light of the many practical considerations facing educational researchers. Campbell & Stanley (1963) list three 'true' experimental designs that can be applied to programme evaluation. The first is the classic experimental design with two groups. Both groups are tested prior to the start of the study, one group is then given the new treatment or programme and the other given no treatment (or the old programme), and then both groups are retested. There is a problem with this design in that the first testing may sensitise students to some of the content in the programme. If this sensitivity interacts with a given treatment, either affecting the new or the old one to a greater extent, the results of the experiment may be inaccurate. With this design there is no way of determining the extent of such an interaction. Furthermore, there are analytic problems with this sort of approach. The 'true effects' of the programme are assumed to be reflected in the interaction between Treatment and Time of Testing, though as Huck & McLean (1975) indicate this is not a powerful test. As an alternative, they suggest using the pre-test scores as a covariate.

The second design is the Solomon Four-Group design. As implied by the name, four groups are required as follows:

	<i>Pre-tested</i>	<i>Not Pre-tested</i>	
Treatment	1	3	TREATMENT EFFECT
No Treatment	2	4	TESTING EFFECT

This design is actually a combination of the classic experimental design above and the post-test only design described below. A  $2 \times 2$  ANOVA on the final scores estimates the effect of Treatment, the effect of Pre-testing, and the interaction between the two. The drawback of this design is the number of groups required in that two groups are tested merely to permit the assessment of the interaction. Moreover, it relies heavily on the random assignment of subjects to conditions.

The third design, the post-test only design referred to above, does not require the use of a pre-test which Campbell & Stanley (1963) argue is not at all necessary when proper randomisation is employed. The classic experiment uses the data from groups 1 and 2 in the above table, while the post-test only design uses groups 3 and 4. This design does not evaluate an interaction but does control for it if subjects are randomly assigned to conditions. This is especially useful since the use of affective measures is rather unusual in the classroom, and by pre-testing affective variables, a heightened sensitivity to the affective component of a programme would be likely to result. Unfortunately, in much educational research random assignment is often not possible, at which time the post-test only design has limited value.

A problem related to the validity of a study, especially with a post-test only design, involves the selection of students to be assigned to treatment conditions. In the interest of isolating the effects of a treatment, it is imperative that students be randomly assigned to conditions. Without random assignment, pre-selection factors are always an alternative explanation for significant differences between the groups. Administratively, random assignment can be done at the individual student level or at the class level. In any case, selection should not be done on the basis of an implicit theory of who would 'benefit' most or be 'harmed' least from a given change in a programme. It is not a useful study, for example, to assign only the brightest students to an experimental group and draw general conclusions about the effectiveness of a programme. For very similar reasons, teachers likewise should be randomly assigned to treatment conditions. As discussed earlier, teacher characteristics can influence students, thus these effects similarly should be randomised. It is not a sound practice to assign the experimental treatment only to those teachers who volunteer to take part.

The usefulness and interpretability of a study is also dependent to some extent on the statistical power of the study. Power refers to the probability of detecting an effect (say, of a programme), if indeed one exists. Power is influenced by the magnitude of the effect in the population, by the reliability of the measures, and by the 'sample size'. Often, however, it is this last factor over which the researcher has most control. In the literature a wide variety of sample sizes have been found, from over 1,000 (Gardner, Smythe, Clément & Glikzman, 1976) to N-of-1 studies (Backman, 1976). For large sample sizes, the power of statistics to declare differences as significant is great. Even the slightest effect, possibly with little practical significance, can reach statistical significance. The opposite is true for very small samples; a relatively strong effect may be lost because of one or two 'stray' data points.

When analysing data, the notion of sample size depends upon the level at

which randomisation occurs and the level at which treatments are applied. For example, if eight intact classes are randomly assigned to receive either a new or existing course, and those classes have 30 students each, the sample size is, in essence, eight. This is because assignment to treatment conditions was done in blocks. Moreover, if treatments are applied to the class as a block, sample size is still only eight as far as the appropriate statistical analysis is concerned. Any statistical comparison of the conditions should consider classes as the unit of analysis, not individual students. Thus, the comparisons would involve an analysis of variance for two treatments with Subjects nested in Classes nested in Treatments. Only if no Class effect is obtained is it meaningful to pool variability attributable to Subjects and Classes as an estimate of experimental error. Of course, if students are assigned to programmes randomly, and if programme effects are applied individually so that there are no confounding effects attributable to classes, then the unit of analysis would be students, and in our example, the sample size would be 240. For this reason, studies that compare only one experimental class with a control class are suspect since the proper sample size in the statistical sense is actually only one per treatment condition.

A major focus in this paper has been the assessment of changes in affective attributes, primarily attitudes and motivation, as a result of some programme intervention or alteration. There is one additional measurement issue that should be mentioned, however, and that is the possible effect that affective characteristics might have on other changes of interest. Assume, for example, that a researcher is concerned as to whether a particular programme change might facilitate the acquisition of vocabulary. The interest here is not on the affective dimension, but one might well ask whether the affective dimension could moderate the effects obtained. This might be expected from the perspective of the socio-educational model of second language acquisition discussed earlier.

This type of phenomenon involves what has been referred to in the literature as the aptitude-treatment interaction (Cronbach & Snow, 1977). This refers to the fact that some experimental treatment may have an effect on students at one level of an individual difference variable but not at another. For example, a treatment may have an effect on students with above average intelligence but not those with below average intelligence. In this example, intelligence is the 'aptitude' in the aptitude-treatment interaction. The aptitude could, however, be any individual difference variable such as motivation, attitudes toward French Canadians, or French classroom anxiety, etc.

It is meaningful, therefore, to ask whether some aspect of the affective dimension might moderate a treatment effect so much so that researchers involved in programme evaluation studies consider assessing relevant aspects of the affective dimension. Thus, if it was thought, for example, that a particular innovation might be more applicable to students high or low in language classroom anxiety, it would be meaningful to assess levels of classroom anxiety before the innovation is introduced. Subsequent analyses then would determine whether the effects differ for subjects varying in classroom anxiety. In this way, not only would a researcher be able to determine whether there is any aptitude-treatment interaction, but also perhaps provide some explanation for the effects obtained.

As such the study would be more informative than if the variable had not been studied. This added variable does, however, add a further complication already referred to when discussing pre-testing. That is, by pre-testing on the 'aptitude' variable, one may well be sensitising students to this aspect of the study.

### Summary and Conclusions

The purpose of this article was to consider the role of the affective dimension in programme evaluation studies in the domain of second language acquisition. It was proposed that the socio-educational model of second language acquisition provides a useful base from which to consider research in this area. It highlights major factors — the sociocultural milieu, cognitive and affective individual difference variables, language acquisition contexts, and language training outcomes — to be considered when contemplating programme interventions, as well as in the design of studies to evaluate them. It also focuses attention on the major affective variables that have been shown to be related to second language acquisition as well as being potential derivatives of second language training. Research associated with this model has provided reliable and valid measuring instruments that could facilitate the assessment of affective variables in any programme evaluation research concerned with this dimension.

To date, the socio-educational model has not been used expressly for this purpose. It has developed from research designed to explain individual differences in second language achievement. It could easily be extended for such a purpose, however. For example, a programme intervention designed to promote communicational skills could be addressed either to the formal learning context, or to the informal one. If the former, it might be expected that individuals with different levels of integrative motivation or language-related anxiety might respond differently in classes. The intervention may be more beneficial to those with high levels of integrative motivation, and/or low levels of second language anxiety than for the other students. If the intervention were directed towards informal learning contexts, then participation levels might be expected to differ for students with differing motivational and anxiety levels. Furthermore, one might expect that the intervention itself could raise some students' levels of integrative motivation and lower others, or cause changes in levels of language class or language use anxiety. Moreover, the effects of the programme on either affective variables or second language achievement might well differ from one social context to another. Clearly other examples could be proposed, and more research is needed to delineate clearly how the various components might be implicated in any programme evaluation situation, but the model would appear to be of potential use.

Some research was reviewed that focused on programme evaluation of the affective dimension. This review considered representative studies of regular 'core' second language programmes, short-term and long-term immersion programmes, and bicultural excursion programmes. The review demonstrated that effects appear to occur, but the interpretation of the results is not always that clear. A number of factors were discussed that could be responsible for the

ambiguities and uncertainties of interpretation. These included such things as effects due to non-random sampling or possible sensitising effects of pre-testing. In programme evaluation studies it is generally the case that the innovation being assessed is administered to entire classrooms, and the classrooms themselves are seldom, if ever, formed on the basis of random sampling. Under such conditions, unequivocal interpretation of effects is not possible. This is not to say that the programme evaluation studies should not be conducted. They do provide very useful information, and with a sufficient number of replications the potential ambiguities in interpretation decrease.

The final section in this paper dealt with issues that should be considered when conducting any programme evaluation, but they were described in terms of the affective dimension. Other issues could be added to the list, but those discussed seemed to be the major ones that researchers would be advised to keep in mind to ensure maximum information from programme evaluation studies. In the end, however, what is really needed is more research directed not only to evaluating particular programmes and programme interventions but in developing a schema that will serve to integrate the various studies. Perhaps the socio-educational model of second language acquisition will prove useful in this regard. Time and research will tell!

### Note

1. This research was supported by Grant No. 410-88-0158 from the Social Sciences and Humanities Research Council of Canada to the first author, on the topic 'Social factors in second language learning and ethnic relations'; and by Doctoral fellowships 453-90-0177 and 453-90-0315 from SSHRC to the second and third authors. We would like to thank V. Galbraith for her assistance in the preparation of this manuscript.

### References

- Backman, N. (1976) Two measures of affective factors as they relate to progress in adult second-language learning. *Working Papers on Bilingualism* 10, 100-22.
- Blake, L., Lambert, W. E., Sidoti, N. and Wolfe, D. (1981). Students' views of inter-group tensions in Quebec: the effects of language immersion experience. *Canadian Journal of Behavioural Science* 13, 144-60.
- Burstall, C. (1975) French in the primary school: the British experiment. *Canadian Modern Language Review* 31, 388-402.
- Campbell, D. T. and Stanley, J. C. (1963) *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Carroll, J. B. (1962) The prediction of success in intensive language training. In R. Glaser (ed.) *Training Research in Education*. Pittsburgh Press.
- (1967) *The Foreign Language Attainments of Language Majors in the Senior Year: A Survey Conducted in US Colleges and Universities*. Cambridge, MA: Harvard University Press.
- (1974) Aptitude in second language learning. *Proceedings of the Fifth Symposium of the Canadian Association of Applied Linguistics*, 8-23.
- Carroll, J. B. and Sapon, S. M. (1959) *Modern Language Aptitude Test (MLAT)*. New York: Psychological Corporation.

- (1967) *Modern Language Aptitude Test, Elementary Form*. New York: Psychological Corporation.
- Chiebek, A. and Coltrinar, H. (1977) Summer immersion programs abroad. *Canadian Modern Language Review* 33, 348–53.
- Clément, R. (1979) Immersion and residence programs: their effect on attitude and anxiety. *Interchange* 9, 52–8.
- Clément, R., Gardner, R. C. and Smythe, P. C. (1977a) Motivational variables in second language acquisition: A study of francophones learning English. *Canadian Journal of Behavioural Science* 9, 123–33.
- (1977b) Inter-ethnic contact: Attitudinal consequences. *Canadian Journal of Behavioural Science* 9, 205–15.
- Clément, R. and Kruidenier, B. G. (1985) Aptitude, attitude and motivation in second language proficiency: A test of Clément's model. *Journal of Language and Social Psychology* 4, 21–37.
- Clément, R., Smythe, P. C. and Gardner, R. C. (1978) Persistence in second language study: Motivational considerations. *Canadian Modern Language Review* 34, 688–94.
- Cohen, A. D. (1975) *A Sociolinguistic Approach to Bilingual Education*. Rowley, MA: Newbury House.
- Cronbach, L. J. and Snow, R. E. (1977) *Aptitude and Instructional Methods: A Handbook for Research on Interactions*. New York: Irvington.
- Cziko, G. A. and Lambert, W. E. (1976) A French-English school exchange program: Feasibility and effects on attitudes and motivation. *Canadian Modern Language Review* 32, 236–42.
- Cziko, G. A., Lambert, W. E. and Gutter, R. (1979) French immersion programs and students' social attitudes: A multidimensional investigation. *Working Papers on Bilingualism* 19, 13–29.
- Deshrochers, A. and Gardner, R. C. (1981). *Second Language Acquisition: An Investigation of a Bicultural Excursion Experience*. Quebec: International Center for Research on Bilingualism.
- Edwards, H. P. (1976) Evaluation of the French immersion programme offered by the Ottawa Roman Catholic Separate School Board. *Canadian Modern Language Review* 33, 137–50.
- Ely, C. M. (1986) Language learning motivation: A descriptive and causal analysis. *Modern Language Journal* 70, 28–35.
- Gardner, R. C. (1985a) *Social Psychology and Second Language Learning: The Role of Attitudes and Motivation*. London: Edward Arnold.
- (1985b) *The Attitude/Motivation Test Battery: Technical Report*. Department of Psychology, University of Western Ontario.
- Gardner, R. C., Ginsberg, R. C. and Smythe, P. C. (1976) Attitude and motivation in second language learning: course related changes. *Canadian Modern Language Review* 32, 243–66.
- Gardner, R. C., Kirby, D. M., Smythe, P. C., Dumas, G., Zelman, M. and Bramwell, J. R. (1974) Bicultural excursion programs: Their effects on students' stereotypes, attitudes and motivation. *Alberta Journal of Educational Research* 20, 270–7.
- Gardner, R. C., Lalonde, R. and Pierson, R. (1983) The socio-educational model of second language acquisition: an investigation using LISREL causal modeling. *Journal of Language and Social Psychology* 2, 51–65.
- Gardner, R. C. and Lambert, W. E. (1959) Motivational variables in second language acquisition. *Canadian Journal of Psychology* 13, 266–72.
- (1965) Language aptitude, intelligence and second language achievement. *Journal of Educational Psychology* 56, 191–9.
- (1972) *Attitudes and Motivation in Second Language Learning*. Rowley, MA: Newbury House.
- Gardner, R. C. and Smythe, P. C. (1975) *Second Language Acquisition: A Social Psychological Approach*. Research Bulletin No. 332, Department of Psychology, University of Western Ontario.

- (1981) On the development of the Attitude/Motivation Test Battery. *Canadian Modern Language Review* 37, 510–25.
- Gardner, R. C., Smythe, P. C. and Bruner, G. R. (1977) Intensive second language study: effects on attitudes, motivation and French achievement. *Language Learning* 27, 243–61.
- Gardner, R. C., Smythe, P. C., Clément, R. and Glikzman, L. (1976) Second-language learning: A social psychological perspective. *Canadian Modern Language Review* 32, 198–213.
- Gardner, R. C., Smythe, P. C. and Lalonde, R. N. (1984) *The Nature and Replicability of Factors in Second Language Acquisition*. Research Bulletin No. 605, Department of Psychology, University of Western Ontario.
- Gayle, C. M. H. (1984) Effective second-language teaching styles. *Canadian Modern Language Review* 40, 525–41.
- Glikzman, L. (1976) *Second Language Acquisition: The Effects of Student Attitudes on Classroom Behaviour*. Unpublished M.A. thesis, University of Western Ontario.
- Glikzman, L., Gardner, R. C. and Smythe, P. C. (1982). The role of the integrative motive on students' participation in the French classroom. *Canadian Modern Language Review* 38, 625–47.
- Halpern, G., MacNab, G. L., Kirby, D. M., Tuong, T. T., Martin, J. C., Hendelman, T. and Tourigny, R. (1976) *Alternative School Programs for French Language Learning*. Toronto: Ontario Ministry of Education.
- Hanna, G. and Smith, A. H. (1979) Evaluating summer bilingual exchanges: A progress report. *Working Papers on Bilingualism* 19, 29–58.
- Hoeh, J. A. and Spuck, D. W. (1975) Effects of a three phase acculturation process on language skill development and social and personal attitudes of high-school French students. *Foreign Language Annals* 8, 221–6.
- Hofman, J. E. and Zak, I. (1969) Interpersonal contact and attitude change in a cross-cultural situation. *Journal of Social Psychology* 78, 165–71.
- Horwitz, E. K., Horwitz, M. B. and Cope, J. (1986) Foreign language classroom anxiety. *Modern Language Journal* 70, 125–32.
- Huck, S. W. and McLean, R. A. (1975) Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin* 82, 511–18.
- Jones, W. R. (1950) Attitude towards Welsh as a second language: a preliminary investigation. *British Journal of Educational Psychology* 19, 44–52.
- Lalonde, R. N. and Gardner, R. C. (1984) Investigating a causal model of second language acquisition: Where does personality fit? *Canadian Journal of Behavioural Science* 16, 224–37.
- Lambert, W. E. (1963a) Psychological approaches to the study of language Part I: On learning, thinking and human abilities. *Modern Language Journal* 14, 51–62.
- (1963b) Psychological approaches to the study of language Part II. On second language learning and bilingualism. *Modern Language Journal* 14, 114–21.
- Lambert, W. E., Gardner, R. C., Barik, H. C. and Tunstall, K. (1963) Attitudinal and cognitive aspects of intensive study of a second language. *Journal of Abnormal and Social Psychology* 60, 44–51.
- Lambert, W. E. and Tucker, G. R. (1972) *Bilingual Education of Children: The St. Lambert Experiment*. Rowley, MA: Newbury House.
- Leonard, E. (1964) Attitude change in a college programme of foreign study and travel. *Educational Record* 173–81.
- Lord, F. M. and Novick, M. R. (1968) *Statistical Theories of Mental Test Scores*. Menlo Park, CA: Addison-Wesley.
- Massey, D. A. (1986) Variations in attitudes and motivation of adolescent learners of French as a second language. *Canadian Modern Language Review* 42, 607–18.
- McLinnis, C. E. (1976) Three studies of experimental French programs. *Canadian Modern Language Review* 33, 151–6.
- Naiman, N., Fröhlich, M., Stern, H. H. and Todesco, A. (1978). *The Good Language*

- Learner*. Research in Education Series No. 7. Toronto: Ontario Institute for Studies in Education.
- Nisbet, J. D. and Welsh, J. (1972) A local evaluation of primary school French. *Journal of Curriculum Studies* 4, 169-75.
- Ontario Ministry of Education (1974) *Report of the Ministerial Committee on the Teaching of French*. Toronto: Ministry of Education.
- (1980) *French, Core Programs 1980*. Toronto: Ministry of Education.
- Oxford, R. L. (1986) *Development of a New Survey and Taxonomy for Second Language Learning*. Paper presented at the Learning Strategy Symposium, New York.
- Oxford, R. L. and Nyikos, M. (1989) Variables affecting choice of language learning strategies by university students. *Modern Language Journal* 73, 291-300.
- Pimsleur, P. (1966) *Pimsleur Language Aptitude Battery*. New York: Harcourt Brace Jovanovich.
- Riestra, M. A. and Johnson, C. E. (1964) Changes in attitudes of elementary school pupils toward foreign-speaking peoples resulting from the study of a foreign language. *Journal of Experimental Education* 33, 65-72.
- Shapson, S. and Kaufman, D. (1978) Overview of elementary French programs in British Columbia: Issues and research. *Canadian Modern Language Review* 34, 586-603.
- Smythe, P. C., Stennett, R. G. and Feenstra, H. J. (1972) Attitude, aptitude and type of instructional programme in second language acquisition. *Canadian Journal of Behavioural Science* 4, 307-21.
- Stennett, R. G. and Earl, L. M. (1982) *Elementary French Programme Evaluation: The Effect of Successive Increases in Instructional Time*. Research Report. London, Ontario: The Board of Education for the City of London.
- Swain, M. (1974) *Early and Late French Immersion Programs in Canada: Research Findings*. Paper presented at the Federal-Provincial Conference on Bilingualism in Education.
- Swain, M. and Lapkin, S. (1982) *Evaluating Bilingual Education: A Canadian Case Study*. Clevedon, Avon: Multilingual Matters.
- Trylong, V. L. (1987) Aptitude, attitudes and anxiety: A study of their relationships to achievement in the foreign language classroom. Ph.D. dissertation, Purdue University.
- Wittenborn, J. R. and Larsen, R. P. (1944) A factorial study of achievement in college German. *Journal of Educational Psychology* 35, 39-48.