CONTEXT OF LEARNING AND SECOND LANGUAGE FLUENCY IN FRENCH: Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs
Author(s): Barbara F. Freed, Norman Segalowitz and Dan P. Dewey
Source: *Studies in Second Language Acquisition*, Vol. 26, No. 2, SPECIAL ISSUE: Learning Context and its Effects on Second Language Acquisition (June 2004), pp. 275-301
Published by: Cambridge University Press
Stable URL: https://www.jstor.org/stable/44486772
Accessed: 04-03-2020 14:33 UTC

SSLA, **26**, 275–301. Printed in the United States of America.
DOI: 10.1017/S0272263104062060

CONTEXT OF LEARNING AND SECOND LANGUAGE FLUENCY IN FRENCH

Comparing Regular Classroom, Study Abroad, and Intensive Domestic Immersion Programs

Barbara F. Freed
Carnegie Mellon University

Norman Segalowitz
Concordia University

Dan P. Dewey
University of Pittsburgh

We compared the acquisition of various dimensions of fluency by 28 students of French studying in three different learning contexts: formal language classrooms in an at home (AH) institution, an intensive summer immersion (IM) program, and a study abroad (SA) setting. For the purpose of oral data collection, students participated in oral

Appreciation is expressed to a number of organizations and individuals without whom this project would not have been brought to completion. Contributions of funding, technical support and expertise, or both are acknowledged from: ACTFL (Elvira Spender and Robert Vicars); Carnegie Mellon University (The Office of the President and Bonnie Youngs in the Department of Modern Languages); The Natural Sciences and Engineering Research Council of Canada; Concordia University (International Initiatives Research Program, Office of the Dean of the Faculty of Arts and Science); Concordia University (Randall Halter, Eric Buisson, Christine Brassard, Eowyn Crisfield, Nilmini de Silva, Sarah Frenkiel, and Heather Wilcox); Middlebury College (Clara Yu, Michael Katz, Beth Karnes, Kara Gennarelli, Paula Schwartz, Carol Rifelj, Jean-Claude Redonnet, Guy Spielmann, Beverly Keim, Anna Sun, and Alex Chapin). The authors also thank Nicole Lazar for her helpful statistical advice.

Address correspondence to: Barbara F. Freed, Department of Modern Languages, 160 Baker Hall, Carnegie Mellon University, Pittsburgh, PA 15213; e-mail: bf0u+@andrew.cmu.edu. Norman Segalowitz, Psychology Department, Concordia University, 7141 Sherbrooke Street West, Montreal QC, Canada H4B 1R6; e-mail: segalow@vax2.concordia.ca. Dan P. Dewey, Department of Instruction and Learning, University of Pittsburgh, W. W. Posvar Hall, Pittsburgh, PA 15260; e-mail: ddewey@pitt.edu.

interviews (similar to the Oral Proficiency Interview) at the beginning and the end of the semester and provided information regarding language use and interactions. Analyses included comparisons of gain scores as a function of the learning context and as a function of the time reported using French outside of class. The main findings that reached statistical significance include: (a) The IM group made significant gains in oral performance in terms of the total number of words spoken, in length of the longest turn, in rate of speech, and in speech fluidity based on a composite of fluidity measures. When compared to the AH group, the SA group made statistically significant gains only in terms of speech fluidity but fewer gains than the IM group. The AH group made no significant gains. (b) The IM students reported that they spoke and wrote French significantly more hours per week than the other two groups. The SA group reported using English more than French (although the difference was not statistically significant) and reported using significantly more English in out-of-class activities than the IM group. (c) Multiple regression analyses revealed that reported hours per week spent writing outside of class was significantly associated with oral fluidity gains.

Traditionally, teachers as well as their students, not to mention the lay public, have subscribed to the belief that second language (L2) learning in an "immersion" context is preferable to learning that is limited to the formal language classroom setting in an at home institutional context (AH).[1] The assumption regarding the supposed superiority of one particular immersion context—that of study abroad (SA)—has been the inspiration for recent research focusing exclusively on learning that takes place in this setting. Within the last decade or so, a small but interesting literature has emerged that explores various aspects of language learning abroad and that offers a series of contradictory, sometimes surprising, and occasionally provocative findings about language gain for students who study abroad, as well as about the nature of the immersion experience itself. Freed (1995a, 1998) and Coleman (1995, 1998) reviewed the majority of these studies from both American and European perspectives. This work has also engendered numerous questions about the specific linguistic benefits that accrue to students who have been abroad as compared to their peers whose learning takes place in a domestic setting (Freed, 1995a). Questions have also been asked as to what the nature of the SA context actually is and how the purported "immersion" in the native speech community abroad compares to the intensity of the language learning experience in an intensive domestic immersion (IM) program. There have now been a number of empirical studies that explore the language gains made by students who have spent time abroad, several of which compare the SA context with the AH context (DeKeyser, 1991; Huebner, 1995; Lafford, 1995; Lapkin, Hart, & Swain, 1995, among others) as well as a hand-

ful of studies that address learning by students in IM programs (Breiner-Sanders, Richter, & Chi, 1999; Liskin-Gasparro, 1998; McKee, 1983). However, to date there have been no carefully controlled studies comparing differences in language in these three primary learning contexts for American students: (a) the AH context, with its focus on formal classroom-based learning, (b) the SA context that combines classroom-based learning with a myriad of diverse options for interaction in the native speech community, and (c) the IM context that includes but goes beyond the classroom experience. Consequently, the profession continues to rely on a series of individual studies that investigate isolated aspects of L2 learning abroad, or the SA as compared to the formal classroom learning (i.e., AH context) and an even smaller number of qualitative studies that explore learning in IM settings. These studies are augmented by countless powerful anecdotal reports that extol the advantages of the SA context.

The work reported here is, to our knowledge, the first to systematically compare certain aspects of language use by adult (i.e., college age) students who spend time in any of these major language-learning settings. In undertaking this study, we assumed the challenge of providing a more enriched description of the individual learning contexts as well as identifying salient features of student language use that might distinguish one group of learners from the others. In particular, we were motivated by a desire to identify the characteristics of learner language that presumably have most contributed to the development of popular perceptions regarding superior language gain in the SA context. For this reason, we decided to build on prior work conducted on the topic of *fluency*—that ever-popular term used in a multitude of often contrasting ways by the public as well as language teaching professionals.

FLUENCY

The notion of fluency, applied to both first language (L1) and L2 use, is one of those topics about which there are many eagerly and enthusiastically expressed opinions. For example, an informal survey, in which first-year undergraduate students in a course on language and society were asked to define the term "fluency," yielded the following responses: "speaking quickly and smoothly," "speaking without saying *um*, without hesitations," "being bilingual," "speaking perfectly," "the ability to make jokes in a language," and "talking easily." Interestingly enough, these spontaneous student definitions corresponded to responses offered by six educated adult, native speaker (NS) judges in some of our prior work (Freed, 1995b) when asked which qualities of speech they believed influenced their decisions when evaluating student speech for fluency. Several of the judges used the term "articulate" to characterize the speech of those whom they rated as more fluent. They also evoked "lack of hesitation" and "less tendency to stumble over phrases" along with "faster rate of speech" as qualities that they rated highly. Most of the

judges also identified "better (or more complex) grammar" and "richness of vocabulary" as qualities that contributed to their decisions in evaluating speakers as more fluent. Higher evaluations seemed to encompass several other speaker attributes, including "ease" or "confidence in speech" and "comfort in the ability to converse." In addition to providing their subjective responses, the judges were also asked to review a list of eight possible components of fluency and to identify those they considered most important in determining which students they considered most fluent. Five of the six judges identified "rate of speech" and "smoother speech with fewer false starts" as the most important factors. Four of the six specified "better grammar" and "vocabulary" and "fewer pauses/hesitancies" as major factors on which they based their decisions. Half of them selected better accent as an important speech quality that contributed to their judgments.

In the professional literature, a brief survey of the term reveals that explorations of the notion of fluency span a continuum ranging from studies of its psychological manifestations and reflections of underlying speech planning and thinking processes (Bialystok, 1990; Chafe, 1980; Dewaele, 2002; Ochs, 1979; Segalowitz, 2000; Segalowitz & Segalowitz, 1993) to studies of speech production, hesitation phenomena, and temporal dimensions of speech (Dechert, 1980; Dechert & Raupach, 1987; Deese, 1980; Goldman-Eisler, 1961, 1968; Griffiths, 1991; Grosjean, 1980; Kowal & O'Connell, 1980; Möhle, 1984; Raupach, 1980, 1983) as well as to applied and sociolinguistic analyses that focus on the total communicative context and variables such as setting, participant, topic, and task, each of which has been shown to exert an influence on both the expression and perception of general fluency (Ejzenberg, 1992; Jefferson, 1979; Schegloff, 1981; Schegloff, Jefferson, & Sacks, 1977).

Definitions of NS fluency have often been based on a series of illustrative examples of the so-called fluent NSs. Prominent among these is Fillmore's (1979) evocative description of fluency that included not only "the ability to talk at length with few pauses" (p. 97) but also references to semantic density, sociolinguistic appropriateness, and creativity in language use. These insightful and engaging examples seem to capture a general language ability, particularly that of the NS.

By contrast, terms such as *communicative competence* or *language proficiency* (constructs with which fluency is sometimes compared) are not used when discussing NS fluency but are occasionally invoked in discussions of nonnative fluency. With respect to nonnative fluency, these three terms, each of which is subject to multiple and overlapping definitions, are frequently interchanged. With respect to L2 speakers, fluency is often used as a loose cover term to represent, as Lennon (1990) suggested, "the highest point on a scale that measures spoken command of a foreign language" (p. 389).

This broad use of the term "fluent," sometimes equated with global language ability, is familiar to all of us. It occurs in expressions such as "She spoke French fluently after spending a year in France." Such a nontechnical use of the term may be contrasted with the more restricted sense of fluency

that represents one of several identifiable components of language ability or proficiency, each of which may be evaluated independently. In the L2 learning literature, as Schmidt pointed out (1992, p. 358), a distinction is sometimes made between fluency and accuracy. It might be said, for example, that "David knows French grammar perfectly, but he doesn't speak the language fluently." Such statements echo Sajavaara's (1987, p. 62) observation that good linguistic or communicative competence is not always realized in fluent speech. By contrast, we have also heard statements such as "Sheldon speaks French fluently, but he makes many mistakes."

Within the field of SLA, several studies have explored various aspects of fluency. These include investigations that attempt to identify specific measures of fluency in speech and closely related efforts to identify characteristics that affect listener evaluations of fluency (see, e.g., Butler-Wall, 1986; Cucchiarini, Strik, & Boves, 2000; Ejzenberg, 1992; Foster & Skehan, 1996; Freed, So, & Lazar, 2003; Lennon, 1990; Möhle, 1984; Olynyk, d'Anglejan, & Sankoff, 1990; Raupach, 1984; Riggenbach, 1989, 1991; Temple, 1992; Towell, Hawkins, & Bazergui, 1996; Wennerstrom, 2000). Sometimes, the features studied can be viewed as representing particular dimensions of fluency. For example, as we do later in this article, it is possible to distinguish temporal phenomena (aspects concerned with the rate of production) from hesitancy phenomena (aspects concerned with the fluidity or smoothness of speech). We might, from logical considerations alone, view these as independent dimensions underlying fluency, although whether in fact temporal and hesitation phenomena are indeed independent of each other is an empirical question (e.g., Cucchiarini et al. reported high correlations between rate of speech and number of pauses in speech).

The general picture that emerges from this literature is that fluency is a complex phenomenon that encompasses a multitude of linguistic, psycholinguistic, and sociolinguistic features. Despite a cluster of agreed-on components of fluent language use, there appear to be considerable individual differences in both the expression and perception of fluency. Fluency functions as a relatively loose cover term, with both global and restricted interpretations that vary from context to context, speaker to speaker, and listener to listener.

LEARNING CONTEXTS

Although context of learning itself is not subject to an unlimited range of definitions, it might include settings as diverse as (a) exclusively technological academic contexts, (b) uninstructed learning settings for those residing in a native speech community, as well as (c) formal academic (in country or "at home") language classrooms, (d) IM settings that integrate formal classroom (content or language oriented) and out-of-class learning opportunities, and (e) SA contexts with potentially unlimited opportunities for use of the target lan-

guage. The present study focuses on the acquisition of oral fluency in the latter three contexts, each of which has been the subject of prior research. See, for example, for SA contexts: Brecht, Davidson, and Ginsberg, 1991, 1995; Coleman, 1995, 1998; DeKeyser, 1991; Freed, 1995b, 1998; Huebner, 1995; Lafford, 1995; Pellegrino, 1998; and Wilkinson, 1998; for IM contexts, see: Breiner-Sanders et al., 1999; Dewey, 2002; Liskin-Gasparro, 1998; and McKee, 1983; and for AH contexts, see, among others: Allwright and Bailey, 1991; Chaudron, 1987; and Schachter and Gass, 1996. For discussion of these comparisons, see Collentine and Freed (this issue). For a more general discussion of frameworks for thinking about learning context with respect to individual difference considerations and social factors, see Ellis (1994, pp. 222–229).

RESEARCH QUESTIONS

This study addresses the following questions:

1. Are there salient differences in the acquisition of oral fluency by students who have studied abroad, when compared to students whose learning takes place in IM programs or the regular AH language classroom?
2. Do time-on-task factors (e.g., instructional time, out-of-class time spent interacting orally with NSs or using the language within the literate domain) vary in each of these contexts?
3. To what extent are the measured differences in oral fluency associated with these time-on-task features?

METHOD

Participants

This study was conducted in the summer and fall of 2000 at a small liberal arts college on the East coast. The initial sample tested consisted of 41 students. Of these, 13 students failed to complete all of the test instruments or questionnaires, and therefore their data were not analyzed. The final sample, therefore, consisted of 28 students (18 female, 10 male), $M = 21.29$ years (AH: $M = 18.38$, $SD = .74$, median = 18.5, range = 17–19; SA: $M = 19.88$, $SD = .35$, median = 20, range = 19–20; IM: $M = 23.75$, $SD = 7.90$, median = 20.5, range = 18–48). All the participants were NSs of English and enrolled in one of three programs conducted by the same institution, and none of them had ever spent time abroad before the beginning of this study. The academic majors of the participants spanned the social sciences, humanities, and science curriculum. Of the total group, only six students were French majors. Depending on their prior study of French and the context in which they were living and studying, students were enrolled in 1–5 French courses during the study period. The number of participants in each of the three contexts were: AH, $n = 8$

(3 female, 5 male); IM, $n = 12$ (8 female, 4 male); and SA, $n = 8$ (7 female, 1 male). The student volunteers for this study were recruited either by the researchers or by the teachers. To encourage participation, we offered, in each of the three contexts, remuneration through a lottery consisting of three prizes of between $50 and $150.

Learning Contexts

Formal Classroom At Home Context. Classroom learning was inherently similar to instruction at other American institutions. That is, the fall semester lasted for approximately 12 weeks, and French language courses met for 2–4 hours a week, depending on the level of instruction. As a rule, students took one or two courses per semester and spent a weekly average of 3–4 hours in French class. The students in this study took courses in French language, literature, theater, history, and civilization. These students had 2–4 years of prior instruction ($M = 3.07$, $SD = 1.35$) calculated on the basis of one high school year being equal to one college semester. Of these, only one student listed French as a study major; the others listed international relations or left this question unspecified.

As on most college campuses, students tended to live in regular dormitories with other speakers of English. However, some of the learners of French, particularly majors, would sometimes gather in a special dormitory, built for use by the academic-year and summer French language program, where occasional French-related events were organized. Even here, however, French was not the lingua franca, and students spoke mostly English in the course of a normal day. Additionally, many students who did not speak French also gathered or resided there.

The curriculum emphasized a communicative approach to language learning. Although such an emphasis is avowedly de rigueur in most contemporary language classes, intensive classroom observation to fully document this statement was not possible. However, discussion with faculty and students convinced us that classroom instruction focused on providing students with an active, communicative command of the French language that would equip them to interact in the French-speaking world. As a general rule, instruction took place exclusively in French, but in courses that focused on the acquisition of content knowledge (cinema or art courses), instructors tended to use English more frequently. In lower level courses, speaking was the focus of instruction, but as students advanced in the French curriculum there was an extensive emphasis on reading and writing.

Extracurricular options in the AH context were not extensive and were consistent with what is found on most college campuses. They included a French table three times a week: twice for lunch and once for dinner. French movies were offered twice during the semester, and other activities (e.g., parties and performances) occurred on an irregular basis.

Intensive Domestic Immersion. The second context of learning in which some students were enrolled was an IM program, situated in the same locale as the academic-year program but operated during 7 weeks in the summer. Students enrolled in the IM program, as in the other two contexts, had completed 2–4 years of prior language instruction ($M = 2.26$ years, $SD = 1.78$), calculated on the basis of one high-school year being equal to one college semester. These students were enrolled in 1–3 language courses, depending on level and spent an average of 3–4 hours a day in formal classroom study or a weekly average of 17.5 class hours. The IM classes were characterized by a communicative language focus and instruction exclusively in French. In addition to classroom instruction, students were expected to engage in numerous learning activities outside of the classroom, and extensive efforts were made to integrate in- and out-of-classroom learning. Distinct from expectations in the AH context, students signed and were required to respect a strict "French only" language pledge (including e-mail and telephone use). Of the 12 students in this group, two listed French as their study major and the others listed English (2), history (3), biology (1), anthropology (1), and one was undecided.

Complementary to the seriousness of the summer IM program were the abundant extracurricular activities organized for students. These included daily opportunities to use French through participation on a soccer team, in a French School choir, and in painting classes. Additionally, there was a regular cultural calendar with weekly musical performances, films, and a cabaret offered on a regular basis. Enhancing these activities were frequent trips (e.g., to Montreal), parties, and cultural events to promote awareness of a diverse Francophone culture and the development of French language skills.

Almost all IM students lived in dormitories with other French School students, and they were generally not allowed to live in dormitories used by learners of other languages. Adherence to the language pledge was enhanced by the isolation of the natural environment and temptations to leave campus to engage in non-French activities.

Study Abroad Context. Students who elected to spend the semester studying in Paris were also enrolled in a 12-week course of instruction. They also had an average of 2–4 years of prior instruction ($M = 3.14$ years, $SD = 0.65$), again calculated on the basis of one high-school year being equal to one college semester. Unlike the students who remained at the home institution, the SA students took 2–5 French courses. As a result, these students spent 2–5 hours per day in formal classes (depending on the number of courses they were taking), with a weekly average of 16.4 class hours per week. Thus, their classroom-based exposure to French was on the same order of magnitude as that of the IM group. Instructional methods in the Paris-based program, in which all courses were taught in French, were described by students as being "very similar to those at the home institution" with an emphasis on communicative use of language. In addition to courses organized and sponsored by the home institution, students with sufficiently strong French skills were also

able to take courses in the social sciences at the Institut d'Etudes Politiques de Paris or at the University of Paris 3 in areas such as sociology, economics, art history, comparative and French literature, and film. Students who took advantage of this opportunity reported that these courses were generally "more demanding" than those offered at the French Study Center in Paris.

Of the eight students in this group, three listed French as their study major, and the others listed English (1), economics (1), or "other" (3). The SA students lived either with French families (3), in student residences known as *foyers* (2), in apartments by themselves (1), or in other similar arrangements (2). Sharing accommodations with other English-speaking students was not permitted. In principle, students pledged to speak only French, but this pledge was not strictly enforced outside the French Study Abroad Center.

Not surprisingly, opportunities to use French outside of the classroom abounded. In addition to activities in the native speech community, initiated and pursued by individual students, (voluntary) group trips to museums and historical sites were organized, and tickets to cultural performances were made available to students at very low rates.

Procedure

Recorded Interviews. Data collection consisted of a 15- to 30-minute interview (similar to the Oral Proficiency Interview or OPI; Breiner-Sanders, Lowe, Miles, & Swender, 2000) administered and tape-recorded at the beginning and end of the fall semester or the summer IM session.[2] The OPI ratings themselves were not considered a measure of gain nor used as a basis of comparisons in this study. However, the recorded oral data from the OPI provided the speech corpus that was used as the basis of analyses of features of oral fluency. For purposes of analysis, two 1-minute segments were extracted from each student's pretest and posttest oral interviews yielding a total of 4 minutes of each student's speech, 2 minutes at each testing time. The decision to use 2 minutes of student speech at each testing time was based on our desire to increase the quantity of each student's speech samples from the two 45-second samples used in some of our prior research (Freed, 1995b). These extracts occurred at the seventh and twelfth minute of both the pretest and posttest oral interviews. All of these interview segments were subsequently digitized to allow the acoustic measurements required for computing the fluency variables.[3]

Out-of-Class Contact. In addition to the oral data, students were asked to provide detailed information on a specially designed Language Contact Profile (LCP). The LCP, slightly modified for use in the three contexts, was an expanded and refined version of the instrument used by Freed (1990, 1995b); see Freed, Dewey, Segalowitz, and Halter, this issue. The LCP was administered on two separate occasions, at the beginning and again at the end of the semester or

the summer session. Although data collected via the LCP should not be equated with the potential wealth of data provided by frequently recorded student journals, the student information gathered via the LCP did help us to gain insight into specific qualities of each of the three learning contexts. Based on data in the LCP, we developed a clearer sense of how students spent their time, with whom they interacted, and in what language. One might wonder if students would be tempted to exaggerate their use of French or underestimate their use of English. However, there was no advantage to them in doing so, and based on the information they supplied, it strikes us that the low frequencies reported for the use of French suggest they were relatively honest in stating the extent to which they actually used English and had contact with NSs of French.

Data from the LCP enriched our understanding of the extent to which students took advantage of the various curricular and extracurricular activities and opportunities that were available to them. The LCP contained items asking how often (i.e., how many times per week) and for how long (i.e., how many hours per week) students interacted with NSs of French. Students were also asked to estimate what percentages of their speaking time was spent interacting with NSs of French in French, NSs of French in English, NSs of English in English (including e-mail and phone calls to friends and family), or with NSs of other languages in French and in English. Finally, students estimated the amount of time (hours per week) they participated in specific activities of a noninteractive nature, such as time spent watching French television or movies, reading French novels, and listening to French radio.

Data Transcription and Analysis

Following Freed (1995b), all of the digitized speech was transcribed into text by native and near-native speakers of French. Two other NSs of French confirmed these transcriptions. A formal reliability check was not deemed necessary because only minor discrepancies were found. The data were analyzed in a variety of ways, including looking at gain scores as a function of the learning context and as a function of time spent using French during the study period. A NS of French conducted the analyses using written transcripts. Based on the two 1-minute samples of each student's speech (2 minutes pretest and 2 minutes posttest oral interview), analyses were performed to measure various dimensions of oral performance and fluency.

MEASURES OF ORAL FLUENCY

For the purpose of this study, we adopted a construct of oral fluency based primarily on measures of temporal and hesitation phenomena of the type used previously by other SLA researchers interested in the construct of fluency (Lennon, 1990; Möhle, 1984; Olynyk et al., 1990; Raupach, 1984, 1987; Riggenbach, 1989; Sajavaara, 1987). We operationalized the notion of fluency in a very similar manner but used a different terminology to describe the six quantifiable

measures listed, all involving temporal or hesitation phenomena or both related to speech fluidity.

Speech Rate (Rate)

Speech rate was computed as words per minute. The following were counted as words: all French words and non-French proper nouns if also appropriate for use in French. Word counts excluded all false starts, repetitions, partial repetitions, and items contained in repairs (only words in the repaired construction were counted).

Hesitation-Free Speech Runs (Hesit-free)

This measure of fluency reflected the mean number of words spoken without dysfluencies consisting of silent hesitations of 400 ms or more (Freed, 1995b). These pauses were identified as follows. First, NSs listened to the recorded extracts and marked on the written transcripts where they thought they heard a dysfluent pause. Next, the durations of the identified silent gaps were measured with SoundEdit 16 software. These pauses tended to be 400 ms or longer, and so the 400-ms criterion was selected as the defining criterion of a dysfluent silent pause, consistent with Riggenbach (1991, p. 426), who suggested that pauses shorter than 400 ms are within the range of normal or fluent speech and do not reflect dysfluency. By using the criteria of both pause length and NS perceptions of dysfluency, we had a greater possibility of identifying students whose speech was less fluent. Cucchiarini et al. (2000) found that although a 200-ms criterion did distinguish more fluent from less fluent bilinguals, the R^2 change was only about .01 after speech rate was entered into the analysis. Freed and Riggenbach both reported using a criterion of 400 ms or greater for unfilled or silent pauses; see also Towell et al. (1996, p. 91) for a discussion of ways to define unfilled pauses.

Filler-Free Speech Runs (Filler-free)

This measure of fluency reflected the mean number of words spoken without filled dysfluent pauses (regardless of how long the pause was). A filled dysfluent pause was defined as an interruption in the speech flow by a French (*euh*) or a non-French filler item (e.g., *ah*, *uh*, laughing, and insertion of English words such as *like* or *I mean*).

Fluent Runs (Fluent-run)

This measure of fluency reflected the number of words in the longest run of speech without silent or filled pause dysfluencies.

Repetition-Free Speech Runs (Repeat-free)

This measure of fluency reflected the mean number of words in a run of speech not containing repeats, partial repeats, or false starts intended as self-corrections as illustrated by the following examples: repetition with no change—*um, j'ai besoin de va au, un, um, j'ai besoin de va, va, au* ... "um, I need to go to the, unh, um, I need to go, go, to the ..." (= 1 repeat); *Est-ce que tu peux, um, me, uh, est-ce que tu peux* "Can you, um, me, uh, can you" (= 1 repeat); partial repeats: *Je pense maintenant que, uh, je, j'étudie le medici, le* "I think now that, uh, I, I am studying medici(ne), the ..." (= 2 partial repeats: *je, medici*); *je voulais pa, passer* "I wanted to pa ..., pass" (= 1 partial repeat: *pa*); *le can, canaux* "the can ..., canals" (= 1 partial repeat); False starts—*je besoin, je pense* "I need, I think" (= 1 false start).

Grammatical-Repair-Free Speech Runs (Repair-free)

This measure of fluency reflected the number of words in the longest run of speech without grammatical repairs. For example: correcting grammar as part of a repeat: *UN journée, une journée* "a day" (= 1 self-correction grammatical repair).

In addition to these six measures, we used three other nonfluidity measures of general oral performance. These were:

Total Words Spoken (Total words)

This measure reflected the total number of words spoken in the 2-minute interview speech sample, where words were counted as described for Rate.

Duration of Speaking Time (Duration)

This measure of fluency reflected the duration of the student's speaking time in the 2-minute interview extract, measured in seconds of student speech (maximum possible = 120 s).

Longest Turn (Turn)

This measure of fluency reflected the student's ability, measured as the longest turn (in words), to hold the floor in the L2 during the 2-minute interview extract.

Finally, change scores were computed by partialling out pretest from the corresponding posttest scores. The resulting residualized score served as a gain score for the variable concerned.

RESULTS

The alpha level for significance was set at .05 for all analyses. The criterion adopted for identifying outliers in the data was three standard deviations from the mean. Where an ANOVA was conducted, only higher order significant interaction effects are reported if they explained lower order interactions and main effects. Where appropriate, tests were corrected using the adaptive False Discovery Rate procedure recommended by Benjamini and Hochberg (2000).[4] For correlations and *t*-tests, one-tailed tests of significance are indicated if there was a clear, a priori hypothesis concerning the direction of correlation or difference. Standard errors of means are shown in parenthesis following reports of means in the text and tables. The interpretation of the results appears in the "Discussion" section.

Gains in Oral Performance and Fluency and Their Relationship to Learning Context

The first set of analyses examined the measures of oral performance. Table 1 shows the means and standard errors for the nine oral measures, separately for each learning context. The first three were general measures of oral performance: the total number of words spoken during the 2-minute interview extract (Total words); the duration of the student's speaking time in seconds, within those 2 minutes (Duration); and the length (number of words) of the longest turn taken by the student in that extract (Turn).

The next six variables were measures of oral fluency based on hesitation and temporal phenomena, as previously discussed. These included: speech rate in words per minute (Rate); fluidity or absence of hesitations expressed as the mean run length (in words) with no silent hesitations or silences of 400 ms or more (Hesit-free); the mean run length with no filled pause dysfluencies (Filler-free); the number of words as the longest run containing no dysfluencies (Fluent-run); the mean run length spoken without repetitions (Repeat-free); and the mean run length spoken without grammatical repairs or self-correction and false starts (Repair-free).

With the exception of two categories (Filler-free and Fluent-run), the students in the three learning contexts did not differ significantly from each other on any of the pretest fluency measures or general oral performance in one-way ANOVAs conducted for this purpose. In the cases of Filler-free, $F(2, 25) = 6.542$, $p = .005$, $MSE = 13.998$, and Fluent-run, $F(2, 25) = 5.355$, $p = .012$, $MSE = 62.675$, the IM group scored significantly lower than the AH group on pretest Filler-free (Scheffé $p = .004$) as well as on pretest Fluent-run (Scheffé $p = .014$) but not significantly lower than the SA groups on these measures. There was no obvious explanation for these patterns or for why they were confined to these two particular variables.

As shown in Table 1, the measures of oral fluency and general oral performance on which students made significant gains (corrected for False Discov-

Table 1. Oral performance and fluency scores and pretest-posttest difference in the AH, IM, and SA contexts

Variables		Pretest		Posttest		*p* level of difference	eta^2
		M	(*SE*)	*M*	(*SE*)		
AH (*n* = 8)	Total words	144.13	(8.45)	139.63	(20.13)		
	Duration	86.13	(0.39)	84.88	(6.47)		
	Turn	38.75	(7.34)	55.25	(10.40)		
	Rate	100.47	(5.04)	95.93	(8.54)		
	Hesit-free	4.99	(0.51)	4.66	(0.81)		
	Filler-free	11.42	(1.26)	11.08	(3.24)		
	Repeat-free	25.68	(7.91)	13.85	(3.65)		
	Repair-free	130.66	(24.47)	78.85	(20.24)	.046	.46
	Fluent-run	28.00	(3.00)	23.63	(4.36)		
IM (*n* = 12)	Total words	127.67	(13.15)	178.33	(7.26)	.001*	.64
	Duration	90.33	(3.25)	92.00	(4.66)		
	Turn	45.75	(7.38)	66.50	(6.49)	.013*	.44
	Rate	83.16	(6.68)	(119.69)	(7.61)	<.001*	.71
	Hesit-free	4.87	(0.79)	7.44	(1.18)		
	Filler-free	5.25	(1.27)	6.89	(0.96)		
	Repeat-free	20.16	(3.70)	43.48	(9.92)	.030	.36
	Repair-free	94.02	(23.09)	94.32	(14.11)		
	Fluent-run	16.50	(2.03)	22.08	(1.87)	.034	.35
SA (*n* = 8)	Total words	156.63	(16.74)	148.75	(16.22)		
	Duration	87.50	(5.74)	78.63	(5.90)		
	Turn	41.00	(6.89)	51.00	(10.05)		
	Rate	106.78	(8.88)	113.33	(8.96)		
	Hesit-free	6.22	(0.98)	6.07	(0.68)		
	Filler-free	7.59	(0.93)	15.32	(6.02)		
	Repeat-free	51.91	(18.05)	36.78	(9.95)		
	Repair-free	80.07	(19.82)	87.94	(20.97)		
	Fluent-run	23.62	(3.05)	28.75	(4.11)		

*$p \leq .05$ after adjustment for False Discovery Rate separately within a given context of learning.

ery Rate) were Total words, Turn, and Rate, and only students in the IM context made gains on these. Moreover, when not corrected for False Discovery Rate, the IM data also yielded significant gains on Repeat-free and Fluent-run ($p < .05$ in each case). Table 1 also shows the eta-squared values associated with these gains (values ranged from .44 to .71 of the total variance accounted for). It should be noted that, whereas only the IM group registered a significant gain on Fluent-run, this gain may, in part, reflect their lower pretest scores on this variable rather than higher posttest performance. The AH group performed worse at posttest than on pretest on the measure Repair-free (marginally significant only, $p = .046$; not significant when corrected for False Discovery Rate). Overall, for the AH group, there did not seem to be a strong trend toward either improved or worse performance over time; on six of the nine variables about half (i.e., three to five students out of eight), the AH participants improved or did not change from pretest to posttest performance, and on three variables six students performed worse. By contrast, the IM group

made gains on all six fluency variables, and the SA group registered gains on four of the six fluency variables although not all these were statistically significant. Additionally, if one considers the five oral fluency variables other than Rate, 16 of the 28 participants made gains on three or more of these, although individually these did not come out to be statistically significant. Because these five measures have in common that they reflect aspects of the fluidity of speech (i.e., they all reflect absence of interruptions in the flow of speech), we accordingly computed a Fluidity gain score for each participant based on the average percentage gain across these five measures. As shown in Figure 1, the IM and SA groups made gains in Fluidity that were significantly above zero, whereas the AH group did not. The IM group made the greatest gain and the AH the least (i.e., no gain at all).

Table 2 presents the intercorrelations among the nine measures of general oral performance and oral fluency at pretest and posttest. At pretest, Rate correlated significantly with all the other oral variables except Repair-free. Six variables correlated significantly with at least two other oral variables, and three variables—Total words, Rate, and Fluent-run—each correlated with at least four other oral variables, which indicates that these three may provide a stronger reflection of oral skill than some of the others. Total words also correlated very highly with Rate, possibly reflecting the fact that within the 2-minute sample, faster speakers were able to produce more words. On the other hand, it should be noted that the mean duration of pretest speaking

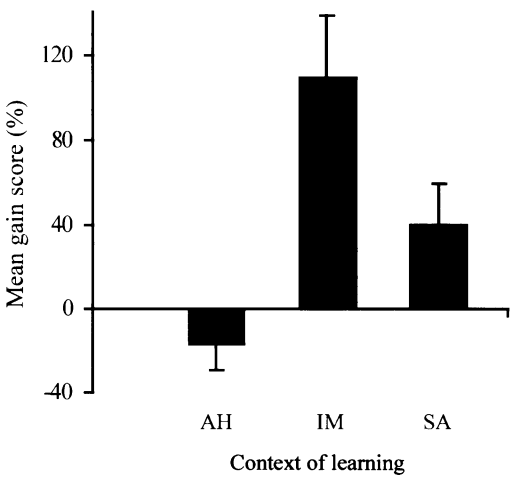

Figure 1. Mean percentage pretest to posttest gain score on oral fluidity (and standard errors), based on percentage gains collapsed over five oral fluency variables—Hesit-free, Filler-free, Repeat-free, Repair-free, and Fluent-run—as a function of context of learning—formal classroom, "at home" (AH), intensive domestic immersion (IM), and study abroad (SA).

Table 2. Correlations among the French L2 oral performance and fluency scores

Variables	Duration	Turn	Rate	Hesit-free	Filler-free	Fluent-run	Repeat-free	Repair-free
Pretest								
Total words	.622*	.635*	.896*	.670*	.476*	.643*	.375	.269
Duration	—	.769*	.223	.125	.032	.119	.005	−.063
Turn		—	.354	.275	.253	.337	.093	.089
Rate			—	.757*	.578*	.757*	.477*	.364
Hesit-free				—	.183	.471*	.109	.367
Filler-free					—	.667*	.246	.390
Fluent-run						—	.370	.207
Repeat-free							—	−.122
Repair-free								—
Posttest								
Total words	.597*	.791*	.725*	.589*	.401*	.656*	.283	−.321
Duration	—	.772*	−.090	−.053	.053	.203	.108	−.234
Turn		—	.293	.224	.242	.390*	.312	−.125
Rate			—	.755*	.411*	.618*	.266	−.212
Hesit-free				—	.341	.494*	.039	−.251
Filler-free					—	.720*	−.108	−.168
Fluent-run						—	.008	−.461*
Repeat-free							—	.422*
Repair-free								—

Note. All the tests were two-tailed. $N = 28$.
*$p \leq .05$ after controlling for the False Discovery Rate separately within the set of pretest and the set of posttest correlations.

time ranged from 85 to 92 seconds out of a maximum possible of 120 (the rest of the time being silences or short prompts spoken by the interviewer), and so it could be argued that Total words reflected more than simply speech Rate.

Finally, age appears not to have been a factor in any of these results nor in those reported below. All participants were in the age range of 17–24 years except for three students aged 28, 29, and 46. No outlier effects were ever detected that could be attributed to these three individuals.

Measures of Out-of-Class Language Contact and Learning Context

Table 3 shows the number of hours per week students in the three learning contexts reported on the LCP using either French or English. The IM group reported significantly more contact hours in French than the other two groups in speaking (Speak), in writing (Write), and in the global level of contact or the sum of speaking, writing, reading, and listening (Total). The SA group reported significantly more English language out-of-class contact activities than the IM group on all measures except listening. The SA group also reported more English than French out-of-class contact activity on the measures of speaking, writing, and global out-of-class, but none of these reached statistical signifi-

Table 3. Mean number of hours (and SE) per week in French and English out-of-class contact (OCC) during the semester reported by students

	AH (n = 8)		IM (n = 12)		SA (n = 8)	
	M	(*SE*)	*M*	(*SE*)	*M*	(*SE*)
French						
Total	10.85	(1.62)**	79.41	(16.84)*	26.38	(7.93)
Productive language activities						
Speak	3.35	(0.69)**	60.44	(17.34)*	10.38	(2.96)
Write	3.00	(0.22)*	14.23	(3.09)*	3.50	(1.32)
Receptive language activities						
Read	3.75	(0.64)	10.65	(2.65)	8.34	(3.29)
Listen	1.13	(0.48)	6.05	(2.21)	4.13	(1.16)
English						
Total	—		6.06	(1.82)***	31.06	(5.75)
Productive language activities						
Speak	—		2.13	(0.76)***	22.14	(5.68)
Write	—		1.26	(0.40)***	5.12	(0.72)
Receptive language activities						
Read	—		1.01	(0.26)***	3.06	(0.52)

Note. Students in the AH context were not asked about English language out-of-class language contact. Asterisks indicate significant differences between adjacent group means.
*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

cance. The IM group reported significantly more French out-of-class contact activity than English on all measures.

General Oral Performance Gains and Fluency Gains and Out-of-Class Contact Activity

Gains in general oral performance and in oral fluency were also examined for correlations with amount of out-of-class contact, on the basis of the number of hours per week students reported (on the LCP) for various extracurricular situations using French. These included time reported speaking, writing, reading, and listening in French (Speak, Write, Read, Listen), and total number of hours spent using French—that is, the sum of these four activities (Total).

The association between general oral performance gains and fluency gains with out-of-class contact was examined with respect to gains in Total words, Turn, Rate, and Fluidity gain, as these were the only measures that yielded significant gains in the earlier analyses. Multiple regression analyses were conducted with gain as the criterion variable and French out-of-class contact activities as the predictor variables. Gain was computed by partialling out pretest scores from the posttest scores and using the residuals as the gain measures.

Because of the relatively small sample size overall ($N = 28$), the multiple regression analyses were conducted in stages. In the first stage, the out-of-class contact variables were combined in two different ways to form two higher order categories. In one set, the categories were productive language out-of-class contact activities (computed as Speak plus Write) versus receptive language out-of-class contact activities (Read plus Listen). In the other set, the categories were Literate language activities (Read plus Write) versus Oral (Speak plus Listen). If, in a given analysis, either of the two variables significantly predicted fluency gain, then a second stage of analysis used the component variables as predictors (e.g., Read and Write in the case of a significant result for Literate).

Only the analysis with Fluidity gain as the dependent measure and with Literate and Oral entered as predictors yielded unambiguous, significant results. These two variables accounted for a significant proportion of the variance ($R^2 = .283$) of Fluidity gain, with only Literate yielding a significant beta coefficient. In the follow-up analysis, the literacy-based variables Read and Write were entered. Only Write yielded a significant beta coefficient, indicating that out-of-class writing activities, but not reading activities, were significantly associated with oral fluidity gains (see Table 4).

Although none of the other analyses yielded clear results, there were highly suggestive patterns. The analysis with Rate gain as the dependent measure and Productive and Receptive as predictors yielded a significant result indicating that productive language activities were a significant predictor, $R^2 = .283$, $F = 4.13$, $p < .05$. The follow-up analysis indicated that the productive language variable Write, but not Speak, was a marginally significant predictor,

Table 4. Multiple regression: Gains in fluidity aspects of L2 oral fluency as a function of reported L2 out-of-class contact (OCC) time

Variable	Parameter estimate	β	t	R^2	F
Fluidity gain: literate-oral analysis					
Literate OCC	3.227	.465	2.293*	.286	4.142*
Oral OCC	0.284	.129	0.637		
Fluidity gain: writing-reading analysis					
Writing OCC	4.421	.423	2.237*	.263	3.922*
Reading OCC	2.392	.201	1.062		

Note. $df = 2$. $N = 28$.
*$p < .05$.

$\beta = .460$, $t = 2.055$, $p = .052$, which suggested that writing activities predicted gains in speech rate, consistent with the previous result. The analysis with Fluidity gain as the dependent measure and Productive and Receptive as predictors yielded a significant result, $R^2 = .265$, $F = 3.79$, $p < .05$, which indicated that receptive language activities were a significant predictor, $\beta = .393$, $t = 2.087$, $p = .049$, but the follow-up analyses with the receptive language variables Read and Listen entered did not yield any significant results, $p > .15$. Finally, the analyses with general oral gain measures Total words and Turn did not yield significant patterns with the out-of-class language contact predictors.

It was not possible to further investigate these significant relationships as a function of learning context because the sample sizes were too small. Table 3 shows that the IM group reported engaging in out-of-class writing activities more than four times the amount for either of the other two groups. Moreover, the SA group reported spending more time writing in English outside of class than in French, whereas the IM group reported the reverse, reporting out-of-class writing activities in French to be more than 10 times the time spent writing in English. These patterns are consistent with previously reported findings, which indicated that the IM group made the greatest gains in oral fluidity and that writing activities were significantly associated with fluidity gains.

DISCUSSION

The findings yielded both negative and positive responses to the three questions that guided our study. In summary, we may first state that, whereas context-based differences in language learning opportunities promote salient differences in the acquisition of oral fluency, students who have studied abroad are not necessarily the ones who demonstrate greater gains in oral fluency when compared to the other two groups. Second, we are able to confirm, at

least for this study, that the actual amount of time spent by students using the L2 (i.e., out-of-class, time-on-task factors) varied significantly from one context to another. Third, the data demonstrated that time spent using the L2 in certain out-of-class activities was associated with differences in oral fluency gains.

Our principal interest was the relationship between context of learning and those aspects of language use that promote the impression that fluency is better acquired in one context than in another. We can confidently state that, for this population (as well as for others reported elsewhere), contextual differences do promote perceptible, identifiable, and significant differences in the acquisition of oral fluency. To the extent that fluency is defined by hesitation and temporal phenomena that are both perceived by listeners and supported by empirical fact as faster, smoother uninterrupted speech, one context of learning in particular appeared to facilitate the acquisition of such abilities. As previously reported, students in the IM context made significant gains on several of the nine measured variables. They produced significantly more words, delivered at faster rate than did students in either of the other two contexts. Additionally, students in the IM context also produced longer turns of speech (i.e., they held the floor longer) as well as more fluid speech runs, not interrupted by disruptive self-repetitions or punctuated by dysfluencies such as silent pauses (longer than those used by NSs) and filled pauses that produce the impression of stumbling and stalling for time. In comparison, students in the SA context demonstrated greater gains on several of the variables than the AH students but fewer than those in the IM context. At the same time, by the end of the semester, the speech of the AH students was actually characterized by more grammatical repairs (but marginally significant only). This is most likely consistent with the nature of classroom instruction—even that which is communicatively oriented. Students in the SA context made no significant improvement along these dimensions.

One perhaps unanticipated finding was that the students in the IM, rather than in the SA, context were those who made the greatest gains in oral fluency, despite the similar number of hours of classroom-based instruction for the two groups. Such a finding reinforces the importance of questions posed by scholars who emphasize the necessity of characterizing what it is that students do in one context compared to another and how much time they actually spend using the L2, despite assumptions about presumed immersion in a native speech community (e.g., Brecht & Robinson, 1995; Firth & Wagner, 1998; Miller & Ginsberg, 1995; Wilkinson, 1998). When this question was addressed by measuring the actual amount of time that students reported engaging in the use of French, we found that students in the IM contexts reported devoting significantly more time to using French in out-of-class activities compared not only to the students in the AH context but also to those in the SA context. In fact, students in the SA context reported using more English in out-of-class contact than they did using French. Such a finding may surprise, if not dis-

tress, the organizers of such programs, but it is consistent with other recent reports concerning the paucity of L2 use by students in an SA setting (Frank, 1997; Wilkinson).

Another possible explanation for the superior performance of the IM students is that the distribution of their hours of learning was over a 7-week period, not 12 as for the others. Although there is not much research on the relative merits of massed versus distributed classroom learning, Collins, Halter, Lightbown, and Spada (1999) reported that a group of French-speaking, 12- to 13-year-old learners in an intensive, massed, English immersion program outperformed comparable children in a distributed program with similar hours of learning over a longer period of time. Collins et al. were not able to clearly attribute the apparent benefits of massed over distributed learning in this context to any one particular factor or set of factors; nevertheless, their demonstration of an effect is important, and the learning they observed may have involved factors also at play in our study. Future research should address the role played by massed and distributed learning in determining learning-context effects.

In our efforts to determine the extent to which measured differences in oral fluency were associated with features of time-on-task (in particular, out-of-class activities in either the oral or the literate domain), we measured the relationship between reported use of French and gains in fluency. Not only did IM students report using French generally more (speaking, reading, writing, and listening combined) than the other two groups, but they also reported speaking and writing French out-of-class significantly more hours per week than did the SA or AH groups. It has been shown that hours per week spent writing French significantly predicted speaking rate gains, whereas the other uses of French did not. This unexpected finding regarding the importance of writing for fluidity of spoken language seemed to be driven by the IM students who engaged in a considerable amount of out-of-class writing activities in French. These students were given many research papers to write—assignments that involved gathering information from different sources, piecing together ideas, and then producing a polished report. It is possible that these students benefited from an effect that Swain (1993, 2000) has articulated as the basis of her output hypothesis—that the act of producing output forces the learner to process language more deeply. This would certainly be true of report writing. Additionally, sustained activity in looking up information in a variety of sources and accessing and practicing them in report writing could be expected to lead to automatization of important linguistic constructions, possibly including ready-made lexical chunks, through frequent encounters with those constructions and chunks (Gatbonton & Segalowitz, 1988; Segalowitz & Gatbonton, 1995; Skehan, 1998), which result in greater certainty about how to use them and in greater fluidity in their production. Such experience might thus be reflected in speech that is less hesitant and freer of self-generated interruptions.

Additionally, it is well known that individual differences exert a major influence on L2 learning and that, in IM contexts, individual differences appear to

be even more prominent (DeKeyser, 1991; Dewey, 2002; Freed, 1995b, among others). Therefore, the question remains as to whether individual differences make some students more capable of engaging in and benefiting from such activities. Although such questions cannot be answered by this study, future reports of research that investigate pretest and posttest cognitive differences among the students in this study and a similar study will be reported elsewhere (Segalowitz & Freed, this issue; Segalowitz, Freed, & Dewey, in preparation).

We alluded in the introduction to this article to a number of assumptions about the advantages of the SA context that are based, in large part, on years' worth of anecdotal reports about what students gain from being abroad and to what these gains may be attributed. Prominent among these beliefs is the assumption that students in the SA context are immersed in the native speech community. Beyond assumptions, there is also some accumulated evidence that supports these beliefs, most prominent among these are the multi-institution, multiyear studies conducted by Brecht and his colleagues (Brecht & Davidson, 1991; Brecht et al., 1991, 1995; Brecht & Robinson, 1995) of the acquisition of Russian by SA students. However, more recently, work by Frank (1997), Pellegrino (1998), and Wilkinson (1998), among others, has shed light on the fact that this may not always be the case. In our own data, we have observed that students in the SA context do not necessarily use French more than those in an AH context and certainly not more than those in an IM context. In fact, the students in our SA sample reported using more English than French, on a weekly basis, while living in Paris. By contrast, those in the IM context reported using significantly more French than English throughout the 7 weeks of their IM experience.

DIRECTIONS FOR FUTURE RESEARCH

The present study yields results that are, in part, consistent with prior work in this area and, in part, unanticipated. As such, it lays the groundwork for future work that may further elucidate our findings. To begin with, a larger sample size (or students in all three learning contexts) would yield more robust results. We would also hope that future research would explore, in greater depth, students' out-of-classroom language use, particularly for those in the IM and SA contexts. More extensive and frequent observation by the researchers, videotaping of student language use, regularly completed student journals, or other ethnographic procedures will enhance our understanding of what actually occurs in these settings. We would also urge future researchers to collect detailed information about L1 and L2 contact activities by means of questionnaires such as the LCP (Freed, Dewey, et al., this issue) and to strive to gather such data on numerous occasions throughout the study period.

Procedures for collecting oral data are problematic, and assuring comparable speech corpora for students at all levels is very challenging. The OPI is

a useful instrument for this purpose, but the use of the student language provided by the OPI might be improved if OPI testers were asked to identify portions of the interview that they considered level checks (efforts to confirm student proficiency ratings) as opposed to probes (attempts to challenge students linguistically to determine if they are able to demonstrate higher levels of proficiency) so that extracts always occur at precisely comparable spots in the interview.

Finally, notions of fluency are both intriguing and multifaceted. Given the numerous interpretations of this term as well as multiple interpretations as to what promotes greater oral fluency, it is not possible to prescribe specific procedures for the measurement of L2 fluency. However, we urge future researchers to investigate carefully both features of a context that might promote increased fluency as well as the interaction of various modes of language use that might complement each other. Additionally, we believe that a more fine-grained analysis of various hesitation and temporal phenomena would yield an increasing number of significant results that would inform both researchers and teachers as to what types of language gain might be best anticipated in one context as opposed to another. For example, it is known that L2 learners may, among other things, lengthen final vowels or make use of connectors or transitions such as *and* or *but* that serve as fillers rather than as links and thereby mask what are really hesitations and pauses. Future research might find it profitable to examine hesitation phenomena at this finer level of detail (see also Clark & Fox Tree, 2002, for an interesting perspective on filled-pause phenomena that merits examination in an L2 acquisition context).

We cannot retroactively respond to each of our own recommendations for the future, but we believe that our findings are suggestive of directions for future research of this type that will inform both researchers and practitioners in the general field of SLA.

CONCLUSION

Although certain limitations of this study might preclude the wide generalizability of our findings, our results suggest caution in subscribing to the prevailing belief that students who spend time in an SA context necessarily become more fluent in their L2 than do those whose learning takes place in other contexts. Admittedly, definitions of fluency vary widely but there seems to be common agreement on certain aspects of language use that promotes recognition of what sounds like fluent speech. Our data demonstrate that salient features of fluency associated with rate, quantity, and smoothness of speech—that is, speech devoid of dysfluent and awkward silent and filled pauses and intrusive repetitions—are differentially acquired by students in one context of learning as compared to another. These aspects of oral fluency use stood out above all else as characterizing the speech of students whose

language learning had not been limited to the AH setting. In our study, such fluency was acquired most easily by students in the IM context. However, when compared to students in the AH context, those in the SA context also demonstrated significant gains in several features of oral fluidity.

These findings do not lead us to suggest that the IM context is inherently superior to the SA context. Nor do we wish to imply that the SA learning context is less rich and less potentially beneficial to learning than has always been assumed. Rather, our findings move us one step closer to understanding some of what takes place in various learning contexts and to the realization that it is not the context per se that promotes various types of learning but rather, as some have always believed (R. Brecht, personal communication, March 19, 2002; G. R. Tucker, personal communication, April 17, 1994), the nature of the interactions, the quality of the experiences, and the efforts made to use the L2 that render one context superior to another with respect to language gain.

NOTES

1. As described by Collentine and Freed in the introduction to this issue, the term *immersion* is subject to numerous interpretations; these will not be reviewed again here. For our purposes, immersion is understood to be those settings in which L2 learners are surrounded by the target language on a potentially full-time basis. Their personal decisions as to how often to use the L2, with whom, for what purpose, and on what occasion vary according to their personal inclinations and these decisions, to be described within the context of this article, are likely to influence their ultimate language acquisition.

2. The OPI is based on an interview used by the various governmental agencies in the United States (the CIA, the Defense Language Institute, the Foreign Service Institute) that are interested in a candidate's ability to use language for a variety of functional purposes (e.g., basic survival, work-related goals, and social interaction). It was adapted in the mid 1980s for use in academic settings by one of the major L2 professional organizations in the United States—the American Council on the Teaching of Foreign Languages (ACTFL). The focus and goal of the OPI is to ascertain what students are able to do with language in a curriculum-free setting rather than on the mastery of the content of any particular course in any particular institution. The OPI ratings used in academic settings span a continuum that ranges from novice-low (those with no functional ability) to superior (those whose L2 skills might include functional categories such as the ability to hypothesize and to tailor their language to fit a variety of audiences from formal to casual). The OPI rating is one global holistic score that includes a range of linguistic features including text type, function, and content. The OPI scale is ordinal and does not have interval or ratio properties, and hence OPI ratings are not suitable for most parametric statistical analyses. The OPI is administered by testers who are trained and certified by ACTFL. Although no two interviews are identical, there is a precise structure to the interview that includes a series of level checks for establishing the minimal level of a student's ability and then a series of probes whose goal it is to push students to reach their highest level of ability. Each OPI (above the lowest level) includes a role play whose purpose is to oblige the students to assume a role different from the conversational interaction that characterizes the rest of the OPI. ACTFL maintains that trained and certified testers assure high levels of face validity in the ratings they award.

3. The seventh and twelfth minutes of the interviews were selected in an effort to establish as much parity as possible for students with respect to their comfort in the interview. Thus, the initial moments and the conclusion of the interview were avoided, as was the role play that is an inherent part of the standard OPI at most levels of proficiency. As one anonymous *SSLA* reviewer pointed out, the OPI procedure alternates level checks (used for the purpose of confirming where students are most secure in their use of the L2) and probes (intended to determine whether or not students can be pushed to higher levels of performance or whether their linguistic control declines or language "breakdown" occurs). It might be expected that students would be more fluent during level checks than during probes. To our knowledge no one has investigated whether the objective flu-

ency measures of the type used in our study vary systematically between level checks and probes, and we did not determine whether the two 1-minute samples obtained from each student coincided exactly with level-check or probe phases of the interview (presumably this varied unsystematically from one student to the next because level checks and probes are not specifically timed to occur at particular moments). It might be of considerable interest in a future study to examine the distribution of objective fluency indicators as a function of the different components of the OPI.

4. The False Discovery Rate procedure is used to protect against Type I errors and may be preferred over the better-known Bonferroni procedure because it is far less conservative and has greater power (Benjamini & Hochberg, 1995, 2000).

REFERENCES

Allwright, D., & Bailey, K. M. (1991). *Focus on the language classroom*. New York: Cambridge University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, B57*, 289–300.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the False Discovery Rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics, 25*, 60–83.

Bialystok, E. (1990). *Communication strategies: A psychological analysis of second-language use*. Oxford: Blackwell.

Brecht, R., & Davidson, D. (1991, March). *Language acquisition gains in study abroad: Program assessment and modification*. Paper presented at the National Foreign Language Center (NFLC) Conference on Language Testing, Washington, DC.

Brecht, R., Davidson, D., & Ginsberg, R. (1991). The empirical study of proficiency gain in study abroad environments of American students of Russian. In D. Davidson (Ed.), *American contributions to the 7th International Congress of MAPRIAL* (pp. 123–152). Washington, DC: American Council of Teachers of Russian.

Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predicting and measuring language gains in study abroad settings. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37–66). Amsterdam: Benjamins.

Brecht, R., & Robinson, J. L. (1995). On the value of formal instruction in study abroad: Student reactions in context. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 317–334). Amsterdam: Benjamins.

Breiner-Sanders, K., Richter, J., & Chi, T. R. (1999, November). *Total language immersion programs: Outcomes and assessments—the Middlebury experience*. Paper presented at the annual meeting of the American Council on the Teaching of Foreign Languages (ACTFL), Dallas, TX.

Breiner-Sanders, K., Lowe, P., Miles, J., & Swender, E. (2000) ACTFL proficiency guidelines: Speaking—revised 1999. *Foreign Language Annals, 33*, 13–18.

Butler-Wall, B. (1986). *The frequency and function of dysfluencies in native and nonnative conversational discourse*. Unpublished doctoral dissertation, University of California, Los Angeles.

Chaudron, C. (1987). *Second language classrooms: Research on teaching and learning*. New York: Cambridge University Press.

Chafe, W. L. (1980). Some reasons for hesitating. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 169–180). The Hague: Mouton.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*, 73–111.

Coleman, J. (1995). The current state of knowledge concerning student residence abroad. In G. Parker & A. Rouxeville (Eds.), *The year abroad* (pp. 17–42). London: AFLS/CILT.

Coleman, J. (1998). Language learning and study abroad: The European perspective. *Frontiers, 4*, 167–203.

Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly, 33*, 655–680.

Cucchiarini, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America, 107*, 989–999.

Dechert, H. W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 271–285). The Hague: Mouton.

Dechert, H. W. & Raupach, M. (1987). *Psycholinguistic models of production*. Norwood, NJ: Ablex.

Deese, J. (1980). Pauses, prosody, and the demands of production in language. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 69–84). The Hague: Mouton.

DeKeyser, R. (1991). Foreign language development during a semester abroad. In B. F. Freed (Ed.), *Foreign language acquisition and the classroom* (pp. 104–118). Lexington, MA: D. C. Heath.

Dewaele, J.-M. (2002). Individual differences in L2 fluency: The effect of neurobiological correlates. In V. Cook (Ed.), *Portraits of the L2 user* (pp. 221–249). Clevedon, UK: Multilingual Matters.

Dewey, D. (2002). *The effects of study context and environment on the acquisition of reading by students of Japanese as a second language during study-abroad and intensive domestic immersion.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Ejzenberg, R. (1992) *Understanding nonnative oral fluency: The role of task structure and discourse variability.* Unpublished doctoral dissertation, State University of New York, Albany.

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore (Ed.), *Individual differences in language ability and language behavior* (pp. 85–101). San Diego, CA: Academic Press.

Firth, A., & Wagner, J. (1998). SLA property: No trespassing. *Modern Language Journal, 82,* 91–94.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18,* 299–324.

Frank, V. (1997, March). *Potential negative effects of homestay.* Paper presented at the Middle Atlantic Conference of the American Association for the Advancement of Slavic Studies, Albany, NY.

Freed, B. F. (1990). Language learning in a study abroad context: The effects of interactive and noninteractive out-of-class contact on grammatical achievement and oral proficiency. In J. E. Alatis (Ed.), *Georgetown University Round Table on Language and Linguistics: Linguistics, language teaching and language acquisition—The interdependence of theory, practice, and research* (pp. 459–477). Washington, DC: Georgetown University Press.

Freed, B. F. (Ed.). (1995a). *Second language acquisition in a study abroad context.* Amsterdam: Benjamins.

Freed, B. F. (1995b). What makes us think that students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Amsterdam: Benjamins.

Freed, B. F. (1998). An overview of issues and research in language learning in a study abroad setting. *Frontiers, 4,* 31–60.

Freed, B. F., So, S., & Lazar, N. (2003). Language learning abroad: How do gains in written fluency compare with oral fluency in French as a second language? *Association of Departments of Foreign Languages Bulletin, 34,* 34–40.

Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly, 22,* 473–492.

Goldman-Eisler, F. (1961). The distribution of pause duration in speech. *Language and Speech, 4,* 232–237.

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech.* San Diego, CA: Academic Press.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale and review of selected studies. *Applied Linguistics, 12,* 345–364.

Grosjean, F. (1980). Linguistic structures and performance structures: Studies in pause distribution. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 91–106). The Hague: Mouton.

Huebner, T. (1995). The effects of overseas language programs: Report on a case study of an intensive Japanese course. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 171–193). Amsterdam: Benjamins.

Jefferson, G. (1979). Error correction as an interactional resource. *Language and Society, 3,* 181–199.

Kowal, S. & O'Connell, D. C. (1980). Pausological research at Saint Louis University. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 61–68). The Hague: Mouton.

Lafford, B. (1995). Getting into, through, and out of a situation: A comparison of communicative strategies used by students studying Spanish abroad and "at home." In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 97–121). Amsterdam: Benjamins.

Lapkin, S., Hart, D., & Swain, M. (1995). A Canadian interprovincial exchange: Evaluating the linguistic impact of a three-month stay in Quebec. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 67–94). Amsterdam: Benjamins.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40,* 387–417.

Liskin-Gasparro, J. (1998). Linguistic development in an immersion context: How advanced learners of Spanish perceive SLA. *Modern Language Journal, 82,* 159–175.

McKee, E. (1983). *The effects of intensive language instruction on student performance in beginning college French.* (Report no. FL013910). Washington, DC: Center for Applied Linguistics (ERIC Document Reproduction Service no. ED233601).

Miller, L., & Ginsberg, R. (1995). Folklinguistic theories of language learning. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 293–315). Amsterdam: Benjamins.

Möhle, D. (1984). A comparison of the second language speech of different native speakers. In H. Dechert, D. Möhle, & M. Raupach. (Eds.), *Second language productions* (pp. 26–49). Tübingen: Gunter Narr.

Ochs, E. (1979). Planned and unplanned discourse. In T. Givón (Ed.), *Discourse and syntax* (pp. 51–80). San Diego, CA: Academic Press.

Olynyk, M., d'Anglejan, A., & Sankoff, D. (1990). A quantitative and qualitative analysis of speech markers in the native and second language speech of bilinguals. In R. Scarcella, R. Andersen, & S. Krashen (Eds.), *Developing communicative competence in a second language* (pp. 139–155). Rowley, MA: Newbury House.

Pellegrino. V. (1998). Student perspectives on language learning in a study abroad context. *Frontiers, 4*, 91–120.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 49–60). The Hague: Mouton.

Raupach, M. (1983). Analysis and evaluation of communicative strategies. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 199–201). London: Longman.

Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Tübingen: Gunter Narr.

Raupach, M. (1987). Procedural knowledge in advanced learners of a foreign language. In J. Coleman & R. Towell (Eds.), *The advanced language learner* (pp. 123–155). London: CILT.

Riggenbach, H. (1989). *Nonnative fluency in dialogue versus monologue speech: A microanalytic approach*. Unpublished doctoral dissertation, University of California, Los Angeles.

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14*, 423–441.

Sajavaara, K. (1987). Second language speech production: Factors affecting fluency. In H. W. Dechert & M. Raupach (Eds.), *Psycholinguistic models of production* (pp. 45–65). Norwood, NJ: Ablex.

Schachter, J., & Gass. S. (Eds.). (1996). *Second language classroom research: Issues and opportunities*. Mahwah, NJ: Erlbaum.

Schegloff, E. A. (1981). Discourse as an interactional achievement: Some uses of *uh huh* and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Washington, DC: Georgetown University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language, 53*, 361–382.

Schmidt, R. (1992). Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357–385.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor: The University of Michigan Press.

Segalowitz, N., Freed, B. F., & Dewey, D. (in preparation). *Cognitive factors underlying oral fluency gains*.

Segalowitz, N., & Gatbonton, E. (1995). Automaticity and lexical skills in second language fluency: Implications for computer assisted language learning. *Computer Assisted Language Learning, 8*, 129–149.

Segalowitz, N., & Segalowitz, S. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics, 14*, 369–385.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Swain, M. (1993). The output hypothesis: Just speaking and writing aren't enough. *Canadian Modern Language Review, 50*, 158–164.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97–114). Oxford: Oxford University Press.

Temple, L. (1992). Dysfluencies in learner speech. *Australian Review of Applied Linguistics, 15*, 29–44.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*, 84–119.

Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102–127). Ann Arbor: The University of Michigan Press.

Wilkinson, S. (1998). Study abroad from the participants' perspective: A challenge to common beliefs. *Foreign Language Annals, 31*, 23–39.