L2 SPEECH PRODUCTION IN THE CLASSROOM: THE EFFECTS OF TREATING SPEAKING IN ITS OWN RIGHT

Eduardo H. Diniz de FIGUEIREDO
(Alamance-Burlington School System, North Carolina)
Mailce Borges MOTA
(UFSC)

ABSTRACT: *In this study we investigate the effects of a 14-week instructional treatment, in which L2 speaking is addressed in terms of its specificities, on L2 learners' oral fluency. Twenty-four L2 learners at the upper-intermediate level of proficiency, assigned to an experimental group or to a control group, participated in the study. Speech performance was elicited by means of descriptive, narrative, and interview tasks on a pre- post-test basis. Results of statistical analyses show that there is trend toward gains in oral fluency after the treatment.*
KEYWORDS: *L2 speaking, fluency, oral performance*

RESUMO: *Este estudo investiga os efeitos de um tratamento instrucional com duração de 14 semanas sobre a fluência oral de aprendizes de L2. Vinte e quatro aprendizes, designados a um grupo experimental ou a um grupo controle, participaram do estudo. O desempenho oral foi eliciado através de atividades de descrição, narração e entrevistas, antes e depois do tratamento. Os resultados da análise estatística indicam que há uma tendência para ganhos na fluência oral após o tratamento.*
PALAVRAS-CHAVE: *expressão oral em L2, fluência, desempenho oral*

Introduction

Speaking has been regarded by many scholars (e.g. Levelt, 1989; Levelt, Roelofs & Meyer, 2000) as a highly demanding and complex cognitive skill that involves several different mechanisms. Levelt (1989) proposes a model of first language (L1) production that shows how speaking follows a series of processes, from the intention to speak to articulation of overt speech. In second language acquisition research (hereafter, SLA) scholars (e.g. De Bot, 1992; Poulisse & Bongaerts, 1994) have also given significant emphasis to the complexity of speaking in a second language (L2) and have highlighted the fact that,

given learners' incomplete knowledge of the L2, it might be an even more demanding skill than it is in the L1.

Interested in the complexity of this skill in L2, various researchers (e.g. Skehan, 1998; Bygate, 1998, 2001; Ejzemberg, 2000; Ellis, 2003; McCarthy, 1998; Fortkamp, 2000) have studied the *teaching* of L2 speaking. Some of them (e.g. Bygate, 1998, 2001; McCarthy, 1998; D'Ely & Mota, 2004) have defended the importance of the treatment and teaching of the speaking skill in its own right. In the present study, the term *speaking in its own right* (Bygate, 1998; McCarthy, 1998) refers to the specificities of the speaking skill when it is addressed pedagogically in the L2 classroom. These specificities are: (1)the acknowledgement of speaking as a genre that can be distinguished from writing, (2)the teaching, discussion and practice of different spoken genres in the L2 classroom (e.g., description, small talk, oral presentation, narrative), (3)the teaching, discussion and practice of different spoken registers, such as formal and informal, (4)the teaching and discussion of the concepts of pausing and hesitating, and the teaching of fillers (such as "you know", "you see" and "well", amongst others) as a means of diminishing silent time during speaking, and (5) the teaching of oral strategies that can be used before speaking (planning and rehearsing), while speaking (circumlocution, substitution of a word that is not recalled for another one which can be recalled, asking for clarification, active participation in speaking situation(s), monitoring performance), and after speaking (evaluating performance and planning for future tasks).

In the present study we investigate the effects of teaching speaking in its own right on the speech performance of L2 learners who were exposed to an instructional treatment consisting of teaching speaking according to the specificities outlined above. This group was pre- and post-tested so as to allow the analysis of the effectiveness of the treatment. The results were also compared to those obtained by a control group (i.e. a group that did not receive any treatment regarding speaking) which performed the same pre- and post-tests, so as to investigate whether any significant differences found in the results of the experimental group were a result of the treatment.

In what follows, we present the theoretical background to the study, the method of data collection and analysis, the discussion of results and the conclusions we draw from the findings.

1   Theoretical background

The study of L2 speech production has gained considerable attention in the past few decades, with scholars trying to understand, describe, explain and assess the processes involved in both L1 and L2

speech production. Mota, Xhafaj & Figueiredo (2005) presented a critical discussion of the literature on L2 speech production over the past sixteen years, with the objective of highlighting some of the most recent and important research that has been done in the area, both internationally and in Brazil. Mota, Xhafaj & Figueiredo (2005) found nine major areas of study of L2 speaking internationally (general aspects of the research about L2 speech production, models of L2 speech production, lexical access, fluency, teaching of L2 speaking, planning and repetition, assessing speaking, factors that affect L2 speech production, and the neurocognitive aspects of L2 speaking), and five in Brazil (treatment of speech errors, tools in the teaching of L2 speaking, classroom interaction as a fundamental part of L2 learning, pragmatics, and effects of different tasks upon L2 speech production). As stated in Mota, Xhafaj & Figueiredo's review, the research that has been done internationally differs significantly from that in Brazil. Much attention in Brazil is given to the teaching of L2 speaking and to the L2 classroom, whereas internationally, L2 speech production is mainly approached from a cognitive perspective, with more attention given to the learner, rather than to the teacher or to the classroom.

In the international scenario, a good number of empirical studies address issues of fluency, accuracy, and complexity in L2 speaking since these are viewed as important variables in the development of L2 speaking competence. The vast majority of the studies that investigate fluency mainly focus on determining what fluent speech is (e.g. Lennon, 1990; Riggenbach, 1991; Freed, 1995; Ejzemberg, 2000).

For example, Lennon (1990) investigated which variables are good indicators of fluency by classifying these into two groups: temporal variables (unprunned words per minute, prunned words per minute, total unfilled pause time, total filled pause time, mean length of runs, T-units followed by pauses, total pause time at T-unit boundaries, and mean pause time at T-unit boundaries), and dysfluency markers (repetitions, self-corrections, filled pauses and self-corrected words). Lennon analyzed speech samples from four German adult EFL students before and after residence abroad based on these variables. The data was also submitted to native-speaker teachers of English, who judged fluency based upon their experience and beliefs. Lennon's results indicated that speech rate and decrease in the number of filled pauses were the main indicators of fluency improvement. His claim was that this supports the existence of two main factors that determine fluent speech: a temporal factor and a dysfluency marker factor.

Lennon (1990) also discusses the two senses in which the term fluency is used. The "broad sense" (p. 389) encompasses the different components of oral competence, including "correctness, idiomaticness, relevance, appropriateness, pronunciation, lexical range and so on"

(p.389). Some researchers in the area of speech production seem to have this broad view of fluency but many others seem to agree that fluency is only one component of oral competence, which corresponds to Lennon's "narrow sense" (pp. 389-390) of fluency. When seen in this narrow sense, fluency is usually associated with temporal variables, such as speech rate (Ejzemberg, 2000; Riggenbach, 1991; Lennon, 1990; Fortkamp, 2000), pausing/hesitation phenomena (Riggenbach, 1991; Lennon, 1990; Fortkamp, 2000) and mean length of runs (Fortkamp, 2000).

Riggenbach (1991) claims that there is more to the term fluency that only temporal variables. In her study she analyzed speech samples from six nonnative speakers of English. Unlike Lennon (1990), however, she also analyzed speech in interaction and established a difference between fluency in monologues and fluency in conversations. Her study supports Lennon (1990) in that speech rate and number of filled pauses were good indicators of fluency. As for fluency in conversations, her study shows that fluency is related to the ability to initiate changes in topic (p. 439), make comments and responses (p. 439) and "show anticipation of end-of-turns as evident by *appropriate* latching and overlapping" (p. 439).

Ejzenberg (2000) collected speech samples from 46 Brazilian speakers of English at different levels of proficiency in order to investigate what differentiates the speech produced by fluent speakers as opposed to the speech produced by less-fluent speakers. The samples were analyzed both quantitatively and qualitatively. The results of the quantitative analysis suggested that the fluent participants spoke faster than the less fluent ones, which gives support to the importance of a temporal variable to the concept of fluency. The qualitative analysis, in turn, indicated that the fluent speakers organized their speech in a better manner and thus avoided interruptions. Once again, this is support for the idea that fluency encompasses more than only quantitative aspects.

Foster and Skehan (1996) investigated the impact of planning on L2 speaker's performance. The study involved 32 learners of English at pre-intermediate level of proficiency assigned to three different groups (no planning, planning without guidance, and planning with guidance), who had to perform three different tasks (personal information exchange task, narration, and decision-making task). Results showed that there was an effect for planning on learners' fluency and grammatical complexity on all tasks, and for grammatical accuracy only for the second group (planning without guidance). Foster and Skehan interpreted these results as evidence for a trade-off among fluency, accuracy, and complexity in speaking due to learners' limited attentional capacity.

Ortega (1999) also investigated the effects of planning on the performance of tasks. Her study consisted of 32 participants who had to retell a story under two different conditions: no planning and planning (for 10 minutes). Once again there was evidence for an effect of planning on performance for fluency and complexity and mixed results for accuracy. The researcher interpreted these results as evidence for a trade-off effect due to limited attentional capacity, giving support to Foster and Skehan (1996).

In Brazil, D'Ely (2006) investigated the impact of four types of pre-task planning on the L2 oral performance of 47 learners of English. Participant's speaking was assessed in terms of fluency, accuracy, complexity, and lexical density. Results of the statistical analyses conducted showed that the different planning conditions affected learners' performance in different ways. Thus, for instance, participants assigned to the planning plus repetition group (i.e., those who were required to plan performance and performed the oral task twice) obtained gains in lexical density, whereas those in the planning for repetition group (i.e., those who were instructed to plan with a view to performing the oral task again) showed gains in lexical density, accuracy and complexity. D'Ely's results also provide evidence for the trade-off among fluency, accuracy, complexity, and lexical density due to limitations in attentional resources during speaking and show that these limitations can be overcome by giving learners to opportunity to plan their performance. In a related study conducted with 50 Brazilian learners of English, Tavares (2008) investigated the role of working memory capacity in L2 oral performance under planning and no-planning conditions. Her results show that, in the no-planning condition, working memory capacity significantly correlates to accuracy and fluency whereas in the planning condition this capacity correlates significantly with fluency and complexity. In other words, in the planning condition, in which learners are encouraged to implement what they planned during performance, therefore maximizing the use of their attentional resources, high spans seem to be better able to manage and control attention thus showing improved fluency and complexity in their speech.

Other empirical studies have addressed the teaching of speaking strategies. Cohen, Weaver and Li (1998), for instance, conducted an experimental study seeking to investigate the role of speaking strategies in the teaching of L2 oral competence. They had 55 participants at intermediate level of proficiency, thirty-two of which comprised an experimental group which received treatment in the classroom, and twenty-three who were part of a comparison group. The treatment consisted of strategies-based speaking instruction during ten weeks, in which participants were instructed on strategies to use before, during,

and after speaking. All participants (in the comparison and experimental groups) had to complete three speaking tasks (self-description, story-retelling, city description) on a pre-posttest basis. The results of the study suggest that the use of strategies enhances speaking performance, as the experimental group outperformed the comparison group in the third task. The researchers also concluded that some specific strategies were linked to some specific tasks, while others (such as idioms, previewing lessons, attention to language form, word-for-word translation, avoidance of translation, and remembering words by image) seemed to be effective for all cases. Having reviewed the research found to be most relevant to the present research, we turn now to the review of a pedagogical proposal concerning the teaching of L2 speaking.

McCarthy (1998) analyses the features of speech in context, distinguishing the act of speaking from the act of writing, and thus giving speech a new perspective, with its own peculiarities. He describes the spoken genre, stating that "spoken language has its own grammar and lexicon" (p.47) and arguing that "the best data for a pedagogical theory of spoken language is everyday, informal talk" (p. 47). McCarthy argues that the spoken genre must be taken into consideration when the speaking skill is being targeted in the classroom and suggests that the first step in building teaching syllabuses and materials is to observe examples of real encounters by participants, thus focusing on real interactions and authentic language. In his proposal, the L2 class must include what he calls the structural, transactional, and generic features of spoken language.

Briefly put, the structural features are subdivided into transaction, exchange, and adjacency pair. Transaction is defined as a structure "to label stretches of talk identified by certain types of activity at their boundaries" (p. 51), such as "now then!", and "good!". McCarthy gives importance to the issue of the "teachability" (p. 51) of transactions, raising four questions: 1) if discourse features are automatically transferred from L1 behavior to L2; 2) if this transfer from L1 to L2 can occur effectively without lexico-grammatical input; 3) if the teacher/learner awareness of the spoken language is sufficient to address the possible problems raised by questions 1 and 2; and 4) if communication is adequate, satisfactory and satisfying without the features of transaction.

These questions are important for the development of a treatment that addresses speaking as proposed in the present investigation, in the sense that they are related to one specific feature of the structure of the spoken language. When addressing the first two questions, McCarthy defends that the fact that learners may pick up certain features of the spoken L2 should not be taken for granted; i.e. they must be addressed

in the L2 classroom. His claim is that it is important to "enable the appropriate contextual environments to be created in the classroom and in the teaching materials" (p. 52). That is why this specific feature of the spoken language (transactions) was addressed in the treatment designed for the purposes of the present study, in the study of discourse markers. McCarthy also gives considerable importance to answering the third question. He believes that the teacher must be aware of the spoken language as well as create the necessary conditions for his/her students to also become aware of them.

Exchange is defined by McCarthy as "the minimal structural unit of interaction, consisting of an initiation and a response (for example, a question and its answer, or a greeting and a return greeting)" (p.52). However, McCarthy also calls the attention to another feature of the exchange, the follow-up, which he considers important for the learner's repertoire. Briefly, the follow-up is the moment in which the person who initiates the exchange (e.g. asks the question) responds to the other person's answer. McCarthy believes that despite the formulaic nature of such responses, there must be contextual conditions for the occurrence of the follow-up. This is important because once again McCarthy is emphasizing the role of the contextual environment created by the teacher in order to give specific and adequate treatment to speaking, rather than the simple teaching of lexical and grammatical features of the language.

Finally, the adjacency pair consists of how participants "position themselves socially in relation to their interlocutor(s)" (p. 54). In other words, this specific feature involves the negotiation of meaning between two or more participants in a conversation, as well as the achievement of goals in the interaction, as it happens in openings to telephone conversations, congratulations-sequences, and seasonal greetings. McCarthy defends that there are a number of formulaic structures that may be taught to learners, thus enabling "fluent, natural and culturally and pragmatically appropriate adjacency pairs to be realised" (p. 55).

The transactional features of spoken language (also called interactional features) are also subdivided into three categories: turn-taking; discourse marking; and information staging. The first one of them (turn-taking) generates cultural problems, in the sense that pausing time may take longer depending on the culture. Besides, acceptance of forms of interruptions also varies amongst cultures. McCarthy (1998) also presents a third problem to this specific transactional feature, which is related to "the noises and verbalisations made by listeners to show understanding, continued interest, etc" (p. 58). This may be a problem to the teaching of turn-taking in L2, as the context of culture of the target language may differ from that of the learners. However, McCarthy defends the teaching of turn-taking in the

classroom, arguing that there must be a focus on the combination of lexical items and culturally acceptable behavior.

Discourse markers, defined as markers "widely used to signal many different functions in conversation" (p. 59), are also seen as highly relevant by McCarthy, due to their naturalness and high frequency in real conversations. Therefore, he defends their inclusion in the L2 classroom, once again emphasizing the importance of addressing both the lexical items and culturally acceptable behavior.

As for information staging, McCarthy states that it addresses the grammatical level of spoken language, rather than the lexical and cultural ones. Information staging refers to how the word-order for the clause in a language is manipulated in informal speech, so that the speaker can give more emphasis to specific entities or events of his/her speech.

Concerning the generic features of the spoken language, McCarthy (1998) emphasizes "the socially-rooted nature of genres and their recognisability for participants within 'discourse communities'" (p. 26). Thus, for him, a genre, be it written or spoken, encompasses the context of culture and context of situation (Eggins, 1993). Based on the idea that generic activities have a socio-psychological reality for language users (p. 33), McCarthy then proposes four dimensions in genre-oriented behavior: expectations, recollections, formulations, and instantiations.

Expectations have to do with resources used by speakers to signal the generic activity in which they are going to be engaged before or during the actual interaction. Signaling may happen with a specific utterance (such as "this is true", indicating truth in what one is about to say), a specific word (such as the topic shift marker "but", or the connector "also", indicating that there is additional information still to be given), or by other activities (such as laughing, smiling, or gesturing).

Recollections have to do with the previous participation of the individual in a certain social activity. For instance, when a speaker asks "have you heard the one about the donkey?", this indicates that the generic activity which is going to take place (or is already taking place) is that of a joke, and the fact that the speaker has already been exposed to this type of sentence engages him/her in the process of recollection.

Formulations occur while the social activity is taking place, with one or more speakers summing up where they believe discourse is at that specific moment. When a speaker asks, for example, "so this is what you mean by being silly", he/she is asking for clarification about a specific thing that is part of the discourse taking place, and he/she gives the other participant(s) of the social activity the chance to confirm his inference or reformulate it.

Finally, instantiations give the participant(s) in the act of discourse the possibility of activating a new set of goals in discourse. McCarthy states that "they enable transactional elements to proceed more efficiently (for example, signaling a change of mind in ordering food in a restaurant)" (p.37) and they may be interactionally-oriented (e.g. compliment/joking).

It is still important to emphasize the importance given by McCarthy (1998) to the teaching of the generic features of speaking, since they encompass a social involvement of the participants in the conversation. According to him, emphasis must be given to two problems. The first one is the fact that openings and closings may differ amongst cultures, a factor that may bring interesting discussions to the L2 classroom, and the second one is the fact that learners may have more difficulties in certain elements of a certain genre (such as the 'evaluation' function in narrative). Both of these problems were taken into consideration in the design of the treatment of speaking as proposed in this study.

Based on the concepts outlined above, McCarthy (1998) proposes an approach to speaking in which students learn *about* the spoken language and its specificities, through what he calls the "three Is" (Illustration, Interaction and Induction) as opposed to the "three Ps" (Presentation, Practice and Production) of the Communicative Approach.

In brief, "Illustration" is the exposure to real data, where the students have the opportunity to see how native speakers of the language speak in real contexts, taking into account not only what words and expressions they use to convey their ideas and feelings, but also how they interrupt one another, how they take turns, and how they react to certain situations. "Interaction" is the talk about language in which learners analyze the specific language items that they have recently been exposed to. This interaction is regarded as important in the sense that students discuss the real data presented with the aim of reaching conclusions from it. These discussions should address the generic, transactional and structural features of the target spoken language and should enable learners to become more aware of these features. Finally, "Induction" is the moment of reflection and conclusions about the L2 discourse. At this point, exposing and discussion have already occurred, thus learners have the opportunity of reflecting about the language to which they have been exposed and which they have recently discussed. This, according to McCarthy is the moment when learners become aware of the new concepts that have been discussed and/or practiced. It is important to note that the "three Is" is not meant to be a new methodology, according to McCarthy, but to be done "in tandem with the syllabus where the lexico-grammar and intonational components are discourse sensitive and not merely sentence-based

abstractions" (pp. 67-68). In the next section we state the specific research questions this study pursued and present the method we used to collect and analyze data, describing the treatment of L2 speaking as well.

3 Method

3.1 Objective and research question

The objective of the present study was to investigate whether exposing learners to an instructional treatment in which speaking is approached as a skill with specific characteristics – that is, in its own right, as defined earlier in this study – would result in gains in oral performance. The specific research question pursued was: Is the gain in *fluency* achieved by the learners in the experimental group significantly different from that achieved by the learners in the control group?

3.2 Participants

The study was conducted at Universidade Federal de Santa Catarina (UFSC) with 24 learners of English as a foreign language, assigned to either an experimental group or a control group. The experimental group consisted of 14 participants – 9 male and 5 female, ages ranging from 17 to 31 years - who were selected from various undergraduate and undergraduate courses at UFSC after being submitted to a proficiency test assessing their oral and written skills. In order to be selected to take part in the study, participants of the experimental group could not be taking English classes, should demonstrate proficiency at the upper-intermediate level, and should accept to receive the instructional treatment focusing on the speaking skills for an academic semester.

The control group consisted of 10 participants – 8 male and 2 female, ages ranging from 18 to 28 years - selected among the learners attending upper-intermediate classes at the Extracurricular English Program at UFSC. The participants in this group attended their English classes as usual and were required to take the pre- and post-tests.

Participants were required to sign a consent form. There were two different consent forms, one for the experimental group, and another for the control group. The one for the experimental group stated participants' consent in taking part in the differential treatment given to speaking in classes and in being tested and recorded for pre- and post-tests. The one for the control group only included their consent to participate in the study through the recording and testing, as there was no differential treatment given to the participants in this group.

3.3 Selection of participants for the experimental group

Over one-hundred volunteers answered our call for volunteers to participate in the study. The call was posted around the university campus and informed that those selected would attend a free English course focusing on the development of speaking skills. All volunteers were invited to participate in the selection process, which consisted of a written test and a subsequent oral test (only the ones who passed the written test were invited to take the oral test).

The written test consisted of a single composition in which students had to write about themselves, their field of study, their goals in the field, and how the English classes would help them achieve these goals. Assessment of their level was based on the cohesion and grammatical accuracy of their written texts, and whether or not they addressed all the items proposed. Only thirty-five candidates were chosen to proceed to the oral test. The oral test consisted of an interview related to the participant's written test. Interviews were recorded on ten Sony MC-60 mini-cassette tapes. The audio-recordings were made with the use of a Panasonic Voice Activated System RN-305 recorder. More than one recording was done in each one of the cassettes, as recordings were done individually at specific times.

Assessment of their oral level was based on perceptions of fluency, accuracy, and complexity. Three raters (two Brazilian and one American) helped in the assessment of volunteers' written and oral performance, so as to reach a final agreement on the candidates to be selected. The three raters were first-year MA students in the Graduate Program in English at UFSC and they all had at least one year of EFL teaching experience. They read the written exams and listened to the oral recordings (thirty-five in total), and then suggested which candidates they thought were the best ones for the course, based on their perceptions of cohesion, grammatical accuracy and completion of task on a scale from 0 to 5 (for the written test), and fluency, accuracy and complexity on a scale from 0 to 5 (for the oral test). Candidates to be selected for participation in the study were expected to be at the upper-intermediate level and this choice was based on the fact that at this point learners can already communicate in the target language. In addition, research has suggested that it is during this level that students in general have the highest improvement in their overall oral proficiency (e.g. Riggenbach, 1991). At the end of the selection process, 25 learners were invited to take place in the course but the final pool of participants whose data was entered for analysis consisted of 14 learners.

3.4 The teachers of the experimental and control groups

At the time of data collection, both the teacher of the control group and the teacher of the experimental group were current doctoral students in the Graduate Program in English at UFSC. Both of them had between five and ten years of teaching experience in this area (EFL) and held an MA Degree in English from UFSC. They were both female teachers in their early thirties.

The researchers contacted the two teachers prior to the beginning of the study. The teacher of the experimental group was invited to participate in the study, and promptly accepted. The teacher of the control group was asked to give permission for the study to be conducted with the students in her class, and permission was also promptly given. Prior to that, permission to conduct the study with the extracurricular courses at UFSC had already been granted by the coordination of the Language Department at UFSC.

3.5 The course book

The course book used by both groups was the same, as this was a main variable to be controlled in the present study. The course book chosen was Passages I (Richards & Sandy, 2000) and the first half (units 1 through 8) was used. This choice was based on the fact that this was the material used by the Extracurricular courses at UFSC for the upper-intermediate level (and, therefore, the material that the participants in the control group would have to use).

3.6 Instruments and procedures

The present study consisted of a pre-test and a pot-test of participants' speaking performance. Each test had three speaking tasks: picture description, personal narrative, and interview. The pre-tests were recorded on 24 Sony MC-60 mini-cassette tapes, one for each participant (i.e. each student's recording of the description, narrative and interview tasks was on one single tape). The post-tests were recorded on 24 Sony MC-60 mini-cassettes, one for each participant. The audio-recordings were made with the use of a Panasonic Voice Activated System RN-305 recorder.

3.6.1 The pre-test

The pre-test, which aimed at assessing the speaking performance of participants in both groups prior to the treatment given to the experimental group, occurred in the first week of classes. Participants were individually invited to leave the classroom and were taken to a different room, in which they were instructed on what to do. The reason why this was done individually was because part of the test was an oral interview, in which there was interaction between the participant and

the researcher. Due to this individual aspect of the test, two days were necessary for it to be concluded with all participants.

The test consisted of three different parts, being two monologues and one dialogue. The two monologues used (description and narrative) were chosen due to their wide use in empirical studies in speaking (e.g. Foster & Skehan, 1996; Fortkamp, 2000; Prebianca, 2004; Yuan & Ellis, 2003). The dialogue was chosen mainly because of the dialogical aspect of the treatment, which consisted of many different interactive activities.

The first part of the pre-test was the picture description. Each participant was asked to describe an advertising picture for a brand of clothes, showing many young people on the streets wearing clothes made by this brand and striking different poses. The picture was taken from an American magazine (Esquire, February 2005), and participants were asked to describe it as fully as they could, including every possible detail. The participants were given as long as they thought was necessary to observe the picture before actually speaking, and they were allowed to ask questions about the task, if necessary. Once they were ready to begin, they would indicate it, and recording would start. Once they began the description of the picture, they could not ask any more questions or interact in any manner with the researcher, until the completion of the task.

The second part of the pre-test consisted of a narrative. Participants were asked to narrate a real story that had happened to them based on the picture they had just described, or at least on part of it. The main objective of narrating a real story was the fact that students would be more involved in the story itself, as they had taken place in it, as well as the fact that they would be telling a story to someone who had never heard it before. The story could be based on the picture as a whole or on one specific part of the picture that called their attention. The story should also be told with as much detail as possible. Once again, the participants were given time to remember a specific story and ask any questions to clarify doubts, but after the beginning of the narrative, no interaction was allowed.

Finally, the participants were interviewed on one specific topic (festivals) by the researcher. The topic of the interview was also based on the picture, and the interview focused on the participants' opinions and attitudes towards the chosen topic, as well as their past experiences regarding this topic. As this was an interactive task, interruption was allowed (i.e. participants were allowed to orally interrupt the interviewer at any moment of the interview).

The participants were not given a time limit for producing any of the speech samples (picture description, narrative and interview), as the objective was for them to talk as much as they thought was possible or necessary at each moment of the pre-test. The whole pre-test lasted an

average of ten minutes per student, including instructions, clarification of doubts and performance.

3.6.2 The post-test

The post-test was carried out in a similar way to the pre-test. It happened on the days assigned for the groups to have their oral tests for the end of the semester, so it took place 28 classes after the pre-test, which was the time in which treatment occurred. A similar picture, also taken from a commercial advertisement, was selected for the description and as a basis for the follow-up activities. This time, the picture advertised a brand of musical products, showing several different images, each representing a famous band. The selection of a picture different from the one used in the pre-test was based on the fact that the treatment focused on the development of participants' overall oral proficiency, rather than the improvement of their speaking performance in one specific task. Moreover, different scholars (see Bygate, 2001b; Ellis, 2003; Ortega, 1999; and Foster and Skehan, 1996) have shown research that gives support to an effect of task repetition over results of post-tests, which means that if the same task that was given in the pre-test had been given in the post-test, gains might have occurred due to task repetition, rather than an effect of treatment.

One difference between the pre-test and post-test was that the latter was carried out on one single day. The students in the experimental group were asked to come at appointed times so that all the participants could do the post-test. As for the control group, participants were invited to stay a little longer in order to finish the test. They had been previously consulted on that and had agreed to stay as long as necessary. Once again, the activity lasted an average of ten minutes for each participant (in both groups).

For the narrative, once again participants were asked to narrate a real story that had happened to them based on the picture they had just described, or at least on part of it. The interview was also based on the picture used for the description, and this time the topic was a musical event that the participants had already been to.

3.6.3 Questionnaire

Participants in both groups were required to answer a questionnaire addressing their language learning and personal backgrounds.

3.7 The treatment

In this section, we present a brief description of the instructional treatment focusing on the speaking skill which was given to the learners

in the experimental group[1]. The treatment emphasized five different specificities of the spoken language: 1) the spoken genre, as opposed to the written genre; 2) different spoken genres (description, small talk, oral presentation, narrative); 3) register in speaking; 4) pausing and fillers; and 5) oral strategies. The treatment lasted an academic semester and consisted of 28 meetings of 1h30m each. Meetings took place twice a week in a classroom equipped with DVD and CD players, a TV set, an overhead projector, and power-point facilities.

The first three classes of the treatment were designed to introduce each student to the teacher, to the other students, to the course book material that was going to be used in class, and to the treatment that was going to be given to them. The treatment itself only began after those three classes. From then on, during each class, the teacher of the experimental group focused on a specific activity addressing one of these aspects, either introducing it, discussing it or practicing it with the students. Between thirty and forty-five minutes of classroom time was devoted to this activity, and the remaining time was used for activities in the course book (or based on the course book). Therefore, at least one-third of every class was addressed to activities dealing with the treatment of speaking in its own right.

The way in which each of the above aspects included in the treatment was introduced, discussed and practiced in class was either based on different authors (e.g. McCarthy & O'Keefe, 2004; Ur, 1981; Oxenden and Latham-Koenig, 2005; O'Dell, 1997) or created collectively by the researchers and the teacher of the group. Two main factors were taken into consideration for the preparation of the activities: 1) the use of authentic material for illustration[2]; and 2) McCarthy's proposal of the 3 Is. Table 1 displays the topic of each class or sequence of classes:

Table 1: Organization of treatment

Topic	Number of classes
Introduction to the treatment	3
Spoken language X written language	1
Register	2
Spoken genres	10

---

[1] Due to space limitations, only a brief overview of the treatment participants in the experimental group received will be given here. A full description of the treatment can be found in Figueiredo (2007).

[2] Some of the materials used for illustration, mainly the videos, were not necessarily illustrations of real oral interactions (e.g. the "Friends" videos, and the movie "Dead Poets' Society), but they were still considered by the researcher as relevant for the purposes of illustrating authentic language, in the sense that they were written for an audience of native-speakers of English. Therefore, it was assumed that the language used in these materials illustrates, up to a certain point, the way in which some oral interactions may happen in real life.

First Review	2
Pausing and fillers	2
Speaking strategies	6
Second Review	2
Total number of classes	28

3.8 Data Analysis

The data collected was analyzed quantitatively so as to reach the scores for fluency. Fluency, for the purposes of this study, was addressed as a temporal variable so as to reflect continuous L2 speech (Ejzenberg, 2000; Freed, 1995; Lennon, 1990; Fortkamp, 2000). It was assessed through unprunned speech rate, which, according to Fortkamp (2000), reflects the relationship of articulation to silence. Unprunned speech rate was determined by dividing the total number of spoken words, including immediate repetitions, by the total time of speech (in seconds) and multiplying the results by sixty, so as to reach the number of words spoken per minute.

The data was analyzed statistically through the use of independent sample and paired-sample t-tests, so as to check the statistical differences within the groups and between the groups in terms of fluency (as assessed by unprunned speech rate) both in the pre- and post-tests. The gain obtained by each group (score for post-test minus score for pre-test) was also analyzed. The independent-sample t-tests were necessary to contrast groups in the pre- and post-tests, and to contrast the gains obtained by the groups. The paired-sample t-tests were important to compare results of pre- and post-test within the same group, thus reaching a t-number. The data was also submitted to skewedness and kurtosis evaluation, so as to check whether it was possible to use parametric tests, and to Levene's test of equality of variances[3], so as to see whether both groups had similar levels before the beginning of the semester. The alpha level was set at 0.05.

4. Results

The fluency scores of both groups (for each one of the tasks in the pre- and post-tests) were submitted to descriptive statistical analyses (descriptive statistics and frequencies) to check for skewedness and/or kurtosis. No major problems of skewedness or kurtosis were found, although some of the normal curves plotted showed a slight tendency to be positively or negatively skewed. Hence, it was possible to submit the

---

[3] Levene's test of equality of variances is a statistical measure used to examine whether variances across samples are equal. If the test shows that there are significant differences between samples, different statistical tools (that do not assume equality of variances) are used for the analyses of data.

results of the pre- and post-tests to parametrical statistical tests when comparing the mean scores of both groups. Equality of variances was determined with the use of Levene's test. In all cases, the assumption of equal variances was satisfied for the pre-test, which means that the two groups had an approximately equal variance on the dependent variable (fluency).

4.1 Analysis of results within the same group

The analysis of results within the same group is necessary to investigate whether there were significant differences between the results obtained in the pre-test and the results obtained in the post-test by each group in each one of the tasks. Although the issue of significance in differences of results is extremely important here, increase and decrease in results are reported even when they were not significant, as this is an important issue for the explanation of the comparison between the groups and the comparison between the gains achieved by each group. Paired-sample t-tests were used in this analysis. The performance of each group was analyzed separately. The results, together with the t-values, are presented next.

4.1.1 Control group

Table 2 displays the t-values and increase/decrease of means achieved by the control group in the pre- and post-test. There were no significant differences between the mean results obtained by the control group in the description tasks of the pre- and post-tests. There was a decrease in the mean scores for fluency (i.e. the score for the post-test was smaller than that for the pre-test) but this difference was not significant. The T-value was 1.508 for fluency. Similarly, differences in the mean scores of the pre- and post-tests in the interview tasks were not significant. Once again, there was a decrease in the mean fluency score (t = 2.098) but this difference was not significant. For the narrative scores, again, the fluency mean score decreased (t = 1.878) but the difference was not significant. In sum, there were no major changes in the overall speech proficiency of the control group after the semester (see table 1 for t-values and increase/decrease of means achieved by the control group).

Table 2
T-values for control group

	Fluency
Description	Non-significant Decrease (t = 1.508)

Narrative	Non-significant Decrease (t = 1.878)
Interview	Non-significant Decrease (t = 2.098)

4.1.2 Experimental group

As can be seen in Table 3, the experimental group, which received the instructional treatment on speaking, demonstrated gains in fluency, between the pre- and post-tests, in two tasks. In the description task, there was a non-significant increase in the mean scores for fluency (t = -.612). The narrative task also showed non-significant increases in fluency (t = -.371). The interview task, however, shows a significant decrease in mean scores for fluency (p<.05; t = 4.389).

Table 3
T-values for experimental group

	Fluency
Description	Non-significant Increase (t = -.612)
Narrative	Non-significant Increase (t = -.371)
Interview	**Significant Decrease** (t = 4.389)

The results presented in Table 3 indicate that, overall, there were no major changes in the fluency of this group between the pre- and post-tests, similarly to what happened to the control group. However, despite the significant decrease in the fluency scores of the interview tasks, this group shows a trend towards improvement in fluency in the description and narrative tasks, a pattern that was not obtained in the control group.

4.2 Analysis of results between groups

The contrast between the mean scores of the groups before and after the treatment is one of the most important factors to establish whether the treatment caused any differences between the two groups.

The analysis between the mean scores of the groups was done with the use of independent-sample t-tests. Table 4 summarizes these results.

Table 4
Comparison between the results of the groups in the post-test

	Fluency
Description	**Significant difference**
Narrative	**Significant difference**
Interview	Non-significant difference

The description task was the one that presented the highest number of significant differences between mean scores in the pre-test and post-test. Significant differences ($p < .05$) were found between the mean scores of the groups in the post-test results for fluency ($t = 2.161$), suggesting that the results achieved by the experimental group were significantly better than those achieved by the control group. These results, however, have to be interpreted with caution.

The increase in the mean score of the experimental group was not significant when compared to the results of the pre-test. However, it is important to recall that the control group had a decrease in the fluency mean scores in the description task; it was this decrease together with the slight increase of mean score obtained by the experimental group that caused the significant difference between the experimental and control groups. In other words, in this case, the significant difference in the mean scores obtained by both groups was not only caused by the gains achieved by the experimental group, but also by a decrease in the results of the control group, which suggests that the treatment was not the main cause for this difference.

The results for the narrative tasks also show a significant difference between the mean scores for fluency achieved by the two groups in the post-test ($p < .05$; $t = 2.299$), similarly to what occurred to the results of the description tasks. However, once again this difference may be explained by the fact that there was a decrease in the mean scores obtained by the control group, whereas the experimental group presented a small increase in these scores. Once again, this may suggest that the difference between the scores achieved by the two groups was not necessarily caused by the treatment. Finally, the results for the interview show that there were no significant differences between the two groups.

In general terms, it may be argued that this comparative analysis between the two groups suggests a *trend* towards a positive effect of the differential treatment over the results of the post-test. In other words, the fact that the experimental group achieved results in fluency

that were significantly higher than those achieved by the control group in the descriptive and narrative tasks may be considered an indication of the positive effects of focusing on specific characteristics of L2 speaking.

4.3 Analysis of questionnaires

Learners in both groups answered a questionnaire about their language learning and personal backgrounds. The majority of learners in both groups had already studied English for over five years, and they had similar thoughts and feelings towards their English learning background - that is, in the majority of cases, students showed a positive attitude towards learning English as a Foreign Language. One participant in the experimental group and two in the control group stated that they did not enjoy learning English so much.   The questionnaire given to the experimental group did not seem revealing in terms of students' feelings and attitudes regarding the treatment to which they were submitted. The majority of students declared that they enjoyed the differential course and learned much with it, although one of them said that she found the methodology used "confusing". Two other students also said that they expected to use the book more in class, and one other student stated that she expected more grammar in class, but none of them pointed these factors as major difficulties or problems. Finally, it is important to state the students in both groups were also asked to answer questions related to the oral tests. This time, opinions varied from student to student (i.e. some students preferred the narrative task, whilst others showed preferences for the interview or description), and no clear overall preference could be inferred.

5. Discussion

A differential treatment given to a specific skill in the L2 classroom brings the expectation that this specific skill will be enhanced in a way, or up to a certain extent. In the specific case of this study, it is necessary to consider the significant differences between the results obtained by the groups in the post-tests as a *possible indicator of an effect of the treatment over the results*, but this fact, as pointed out earlier, does not put an end to the discussion over the effectiveness of the treatment. It may be argued that when there was a significant difference between the mean scores achieved by the groups, these differences were not caused by the treatment alone. Three main factors might account for the lack of stronger significant differences in fluency between the two groups. First, it is necessary to consider that the length of the treatment – 28 meetings – might have been too short for more significant changes in the learners' fluency to occur. In addition, it was the first time that the learners sin the experimental group were submitted to an approach of speaking in this manner. It would then be

an over-expectation to believe that these students would acquire all that was discussed and practiced in class regarding the treatment of speaking in its own right in such a short period of time, especially when other aspects (in the course book) also had to be approached in class. Further studies concerning this specific matter (the treatment of speaking in its own right) should take length of treatment in consideration.

The second factor is related to the material used in the treatment. To the best of our knowledge, there is no specific material available that addresses L2 speaking as a skill in its own right, as pointed out by various scholars in the area (Bygate, 2001; D'Ely and Mota, 2004; Burns, 1998). In the case of the present study, we adopted a well-known course book aimed at the development of the four main skills and designed extra material to address the specificities of speaking. The quality of this extra material as well as its effectiveness were only assumed. Ideally, this extra material should have been piloted prior to the treatment. Further research should consider the material an intervening variable.

Finally, the third factor is related to the difference between the nature of the treatment and the nature of the activities and variables used to test the students in the pre- and post-tests. The treatment consisted of several activities that were dialogical in nature, that is, that demanded interaction between the students. However, two out of the three activities in the pre- and post-tests were monological (description and narrative). This may have influenced the results of both the pre- and post-tests in the description and narrative activities. The interview task, on the other hand, was dialogical, and yet the treatment did not yield significant differences between the groups. This could be due to the fact that learners are not always prepared to answer questions on the topics presented during interviews, as the topics may not necessarily interest them. This, in turn, influences the interaction that happens between the interviewer and the interviewee, as the former needs to try to get the latter involved in the dialogue, since s/he is the one who is expected to speak more. Furthermore, in the case of the present study, learners might have found the interview task more threatening than the descriptive and narrative tasks, given that the interlocutor held a more powerful position in the interaction. In the he present study, learners were, after all, in a testing situation.

As regards the variables used to assess fluency, unprunned speech rate may not be a good variable to determine fluency in a dialogical situation. Performing well in a dialogue involves more than speaking fast. Other factors such as turn-taking and the initiation of topic changes may be better measures of fluent behavior in dialogues. These aspects were not addressed in the present study, mainly due to its quantitative

approach to data analysis and further research should take them into consideration.

6. Concluding remarks

The present study tried to investigate the treatment of L2 speaking as a skill with specific characteristics and not as a general means to learn the L2, as it is generally treated by teachers. The results of the investigation suggest that the effectiveness of the treatment of speaking in its own right in the EFL classroom is not as linear as it might be expected; that is, having such a treatment does not necessarily imply that learners' oral proficiency will improve significantly, at least when oral proficiency is addressed in the way that it was in the present study – in terms of fluency. Therefore, there are two main conclusions concerning the study itself. The first one is that a course in which the treatment of the speaking skill in its own right is implemented demands a complex pedagogical process that will imply changes in the course material to be used in the classroom, in the knowledge and perspectives of the teacher toward this skill, and consequently, in the way the skill is presented, discussed, practiced, and evaluated in class.

The other important conclusion of the study is that treating only some of the specificities of speaking in the L2 classroom does not necessarily give the treatment of this skill the complexity that it deserves. Issues of continuity, constant evaluation of progress through the use of different evaluation tools (see Luoma, 2004), and length of discussions and units, are also important for the overall success of the treatment.

References

BURNS, A. Teaching speaking. *Annual Review of Applied Linguistics*, 18:102-123, 1998.

BYGATE, M. Theoretical Perspectives on speaking. *Annual review of Applied Linguistics*, 18: 20-42, 1998.

BYGATE, M. Speaking. In R. Carter and D. Nunan (Eds.). *The Cambridge guide to the teaching English to speakers of other languages*. Cambridge: Cambridge University Press, p. 14-20, 2001.

BYGATE, M. Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.). *Researching pedagogic tasks: Second language learning and testing*. London: Longman, p. 37-53, 2001b.

COHEN, A. D.; WEAVER, S. F.; LI, T. The impact of strategies-based instruction on speaking a foreign language. In A. Cohen. *Strategies in*

learning and using a second language. New York: Longman, p. 107-156, 1998.

D'ELY, R.C.F. A focus on learners' metacognitive processes: the impact of strategic planning, repetition, strategic planning plus repetition, and strategic planning for repetition on L2 oral performance. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, 2006.

D'ELY, R.C.F.; MOTA, M. B. A teoria, o livro didático e o professor: uma análise da implementação de tarefas orais em LE. *Linguagem e Ensino*, 7: 65-98, 2004.

DE BOT, K. A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13: 1-24, 1992.

EGGINS, S. *An introduction to systemic functional linguistics*. London: Pinter, 1993.

EJZENBERG, R. The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.) *Perspectives on fluency*. Ann Arbor: The University of Michigan Press, 287-314, 2000.

ELLIS, R. *Task-based language learning and teaching*. Oxford: Oxford University Press, 2003.

FIGUEIREDO, E. H. D. de. *Treating speaking in its own right: an exploratory study*. Unpublished MA Thesis, Universidade Federal de Santa Catarina, 2007.

FORTKAMP, M. B. M. *Working memory capacity and L2 speech production: an exploratory study*. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, 2000

FOSTER, P.; SKEHAN, P. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18: 299-323, 1996.

FREED, B. F. What makes us think that students who study abroad become fluent? In B. Freed (Ed.). *Second language acquisition in a study abroad context*. Philadelphia: John Benjamins, p. 123-148, 1995.

LENNON, P. Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40: 387-417, 1990.

LEVELT, W. J. M. *Speaking: from intention to articulation*. Cambridge, MA: MIT, 1989.

LEVELT, W. J. M.; ROELOFS, A.; MEYER, A. S. A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22: 1-38, 2000.

LUOMA, S. *Assessing speaking*. Cambridge: Cambridge University Press, 2004.

MCCARTHY, M. *Spoken language and applied linguistics*. Cambridge: Cambridge University Press, 1998.

MCCARTHY, M.; O´KEEFE, A. Research in the teaching of speaking. *Annual Review of Applied Linguistics*, 24: 24-63, 2004.

MOTA, M. B.; XHAFAJ, D.; FIGUEIREDO, E. H. D. A literatura sobre produção oral em L2: uma apreciação crítica. *Anais do IV SENALE*.

Pelotas, UCPEL/ Programa de Pós-Graduação em Letras, p. 132-133, 2005.

O'DELL, F. *English Panorama 1 Students' book: a course for advanced learners*. Cambridge University Press, 1997.

ORTEGA, L. Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition,* 21: 109-148, 1999.

OXENDEN, C.; LATHAM-KOENIG, C. *New English file*. Oxford: Oxford University Press, 2005.

POULISSE, N.; BONGAERTS, T. First language use in second language production. *Applied Linguistics,* 15: 36-57, 1994.

PREBIANCA, G. V. V. *Communication strategies and L2 speech production.* Unpublished MA thesis, Universidade Federal de Santa Catarina, 2004.

RICHARDS, J.; SANDY, C. *Passages.* Cambridge: Cambridge University Press, 2000.

RIGGENBACH, H. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes*, 14: 423-441, 1991.

SKEHAN, P. *A cognitive approach to language learning*. Oxford: Oxford University Press, 1998.

TAVARES, M. da G. G. Pre-task planning, working memory capacity, and L2 speech performance. Unpublished doctoral dissertation, Universidade Federal de Santa Catarina, 2008.

UR, P. *Discussions that work: task-centered fluency practice*. Cambridge: Cambridge University Press, 1981.

YUAN, F.; ELLIS, R. The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, 24: 1-27, 2003.