ELABORATION AND VALIDATION OF AN ESL/EFL SOFTWARE EVALUATION INSTRUMENT

Vládia M. C. Borges

Universidade Federal do Ceará

Abstract:

This study aimed at elaborating and validating an ESL/EFL software evaluation instrument that encompassed the principles of Communicative Language Teaching and an interactive approach to computer use for language learning. Once the instrument was elaborated, it was tested for its internal consistency and inter-item and inter-rater reliability. The results of the Pearson Coefficient and the ICC Coefficient measures indicated high levels of inter-rater reliability for the group of 26 teachers. The Cronbach's Alpha Coefficients for the three programs indicated that the ESL/EFL software evaluation instrument had adequate levels of inter-item reliability. These results suggest that the ESL/EFL software evaluation instrument has high levels of internal consistency. These results indicate that the evaluation instrument has a high degree of reliability. The positive indicators of reliability obtained from the procedures used to assess the inter-rater reliability and the internal consistency and the face and content validity attributed to the instrument suggest that the ESL/EFL software evaluation

instrument is potentially valid for assessing the degree to which ESL/EFL software programs develop language skills according to the Communicative Language Teaching principles and an interactive approach to computer use in language learning.
Keywords: ESL/EFL software evaluation, Communicative Language Teaching, interactive approach to computer use for language learning.

Introduction

In recent years there has been a great deal of interest in using computers for language teaching and learning. With the advent of multimedia computing and the Internet, the role of computers in language instruction has become an important issue confronting large numbers of language teachers and researchers throughout the world.

During the four decades of CALL development, materials have gone from an emphasis on basic textual gap-filling tasks and simple programming exercises to interactive multimedia presentations with sound, animation and full-motion video. But this progress has not been purely linear and, "in terms of pedagogy, the 'new and improved' have not always replaced the 'old and tired'. Instead, many programs being produced today feature "little more than visually stimulating variations on the same gap-filling exercises used 40 years ago" (Beatty, 2003, p. 11).

There appears to be a substantial gap between what computer technology can do to support language learning and the way actual software programs provide for language learning. Therefore, there might be programs on the market which claim to be interactive but their design may lag behind current ESL pedagogy. Overall, changes in CALL tend to be governed more by advances in technology than by pedagogical insights (Beatty, 2003). Although a review of the

literature on ESL/EFL computer programs supports a pedagogical use of computer tools, the incorporation of modern theories of learning in the elaboration of ESL computer programs seems to be rather complex and difficult to achieve.

Consequently, it seems useful to identify meaningful ways of analyzing Computer Assisted Language Learning (CALL) material so that the characteristics associated with high quality interactive CALL material can be identified, articulated, and refined. There are many ways to evaluate and critique CALL programs, for in one sense their evaluation is analogous to the evaluation of a new textbook or other instructional resources. However, little attention has been given to whether particular programs effectively promote second language learning according to the principles of Communicative Language Teaching.

Although evaluating and critiquing CALL software programs can take different forms, there is an important point that cannot be forgotten: "the basic tenet of software evaluation is that "pedagogy must drive technology" (Burston, 2003, p. 35). This principle means identifying curricular needs first, then looking for software that meets these needs.

Scholfield (2000) presents three points to be considered when evaluating software materials: (a) the nature of the software, (b) the nature of the teacher/learner situation - the learners and their needs, uses etc, and (c) a rating system to judge the suitability of the software to one or both of the previously mentioned points, with due attention to relevant universal principles of good teaching/learning.

Plass (1998) suggests examining software based on the particular approach it uses or skill mastery it targets. He proposes steps to be followed when evaluating particular aspects of software:

1. Identify relevant skills, competencies, and domain knowledge.
2. Identify activities that cultivate and develop these skills, competencies, and knowledge.
3. Identify the cognitive processes involved in these activities.
4. Assess the level of support for the cognitive processes.
5. Using Plass's suggestions, a taxonomy of ESL/EFL software features could be based on the underlying pedagogy or principles of education, and would address how well the individual components of software programs are able to facilitate them.

Thorn (1995) has also outlined points to consider when evaluating interactive multimedia. He focused on six components of software that need to be addressed: (a) ease of navigation, (b) cognitive load, (c) information presentation, (d) media integration, (e) aesthetics, and (f) overall functionality. These components are a good starting point for evaluation, but do not address the questions presented by Chapelle (1997), Krashen (1982), or Laurillard (1998): how the software being evaluated addresses the students' linguistic needs.

With exceptional insight into the aspects of software selection, Healey and Johnson (1997) proposed a tool to assist educators with software selection. Healey and Johnson narrowed the focus of their tool to include six basic categories: (a) educational level, (b) academic goals/focus, (c) educational setting, (d) teacher interaction level, (e) computer hardware, and (f) money.

In addition to noting the lack of evaluative criteria which measure not only learning outcomes but also learning processes, Reeder, Heift, Roche, Tabyaninann, Schlickau, and Golz (2004) identified a number of shortcomings in current experimental evaluative practices:

- Problems of validity and generalizability in experimental evaluation designs, which includes difficulty in attributing outcomes validity to the treatment, and invalid reduction of complex learning processes.
- Failure to take educational goals into account.
- Inability to deal with new and multiple literacies.
- No reflection of the issues of intercultural communication and sociocultural variation among user groups.
- Lack of flexibility to include the interaction between emerging possibilities offered by new technologies and their effects on instructional paradigms.

A way of conducting evaluation studies suggested by Scholfield (2000) is introspective judgmental evaluation. This can be done individually, subjectively, globally and introspectively. However, to regard evaluation in a systematic way it is necessary at the very least to 'unpack' this approach. The teacher (or anyone else) acting alone as evaluator should break down the 'overall' or global judgment into parts.

This is where checklists come in, according to Scholfield (2000). Checklists are a type of introspective judgmental evaluation. They may be made by the teacher/ evaluator, or adopted from someone else. However, as pointed out by Burston (2003, p. 36), there is a problem with the existing checklists; they are extremely simplistic. A program can get good marks in many categories and still, intuitively, users may not be very impressed with it. One of the reasons for this situation may be that checklist items typically have the same relative weighting, whereas in any particular situation some feature of the software may be more important than others. Likewise, there is enormous variety in available software (stand alone usage, exploitation over the web; tutorial, collaborative, facilitative, etc.),

so it can be very difficult to know how predetermined checklist categories should be applied.

Perhaps a solution for the problems found in both experimental and introspective studies on software evaluations can be drawn from Chapelle's (1998) argument that some design features and evaluation criteria for multimedia CALL should be developed on the basis of hypotheses about ideal conditions for second language acquisition. For each of the hypotheses about SLA drawn from the interactionist perspective, Chapelle (1998) makes a corresponding suggestion for developing multimedia CALL. Each of the hypotheses comes primarily from the study of face-to-face oral communication between learners or between learners and proficient L2 speakers. As such, they attempt to describe multimedia characteristics with respect to the psycholinguistic responses that they might evoke from learners rather than exact replicas of conversations among learners.

While the majority of software developers as well as the evaluation systems agree on the significance of instructional considerations in the development and evaluation of educational courseware, there is no agreement among researchers and evaluators as to what criteria to use to assess this aspect of language software (Reeder at al., 2004).

If CALL software packages are to be properly evaluated and matched with learning needs, there must be a set of criteria to be taken into consideration in their evaluation. Because this research was concerned with criteria for the evaluation of CALL software programs designed for ESL/EFL learning, it was necessary to design an evaluation instrument that encompassed the principles of Communicative Language Teaching and an interactive approach to computer use for language learning. Therefore, this study aimed at examining the following questions:

1. To what extent does the ESL/EFL software evaluation instrument produce consistent results when administered under similar conditions? In other words, is the designed ESL/EFL software evaluation instrument reliable?

2. To what extent is the ESL/EFL software evaluation instrument valid to evaluate the potential of CALL programs to develop language skills according to the principles of Communicative Language Teaching principles and an interactive approach to computer use for language learning?

In order to answer these questions it was necessary to establish criteria for evaluating foreign or second language software. And the simplest approach to evaluate software was then to work from a checklist, since several already existed. However, none of them seemed to fit the goal of this research, because the available checklists did not integrate the systematic properties presented by Burston (2003) or the parameters suggested by Hubbard (1988) into their elaboration. These parameters and systematic properties were considered crucial by this researcher.

Burston (2003, p. 37) suggested a look at critical systematic properties as a viable alternative to software evaluation checklists. According to him, software programs must meet the first two of the following requirements and some combination of the last three: (a) pedagogical validity (software must conform to sound instructional methodology); (b) curriculum adaptability (it must fit, or be adaptable to, curricular needs); (d) efficiency (it must make something teachers are already doing easier to accomplish); (e) effectiveness (it must produce better results); (f) pedagogical innovation (it must allow teachers to do something new and different).

Still, the fundamental question here is not only whether a software program meets these conditions, but the extent to which and how it meets them. Determining the answer to this question

involves a close analysis of key program parameters. Hubbard (1988, pp. 60-63) established the following key parameters in CALL program analysis which were used as criteria in the development of the evaluation instrument designed for this study:

- The software's methodology (objectives, easy-to use-format, feedback to learners' responses).
- The software's approach to language instruction (linguistic assumptions, approach to language, support for a particular method of language teaching, and the platforms it is available for).
- The software's design (exercises geared toward or adjustable for learner variables, arrangement of exercises–according to the notional/functional approach, for example, integration to cultural and every-day life aspects, linguistic levels of exercises–discourse syntax, lexis, morphology).
- The software's procedures (types of activities the software offers, how the software presents these activities, and how much control it allows learners and/or instructors over the content of the lessons).

It is worth noting, however, that the parameters used by Hubbard (1992) are not geared toward any particular language learning approach. They are useful parameters for choosing software or for evaluating software appropriateness for different teaching/learning contexts.

In summary, a review of major publications on how to evaluate the use of software material for ESL/FL instruction/learning and the analysis of many evaluation checklists indicated that what was still needed was a comprehensive instrument based on what is known not only about the second language learning process but also about the best practices for language learning in order to assess the potential of CALL programs to develop language skills. Consequently, the

characteristics of the Communicative Approach to Language Teaching that promote language learning according to what is known about second language acquisition and the technological features that allow for an interactive pedagogical use of computers were taken into consideration in the elaboration of the ESL/EFL software evaluation instrument for this study.

This study, therefore, attempts to go beyond previous research which has investigated the advantages and disadvantages of using computers to develop specific language skills and to influence students' attitude toward language learning. Because Computer Assisted Language Learning (CALL) is an important aspect of many language-learning programs, it makes sense to be able to evaluate the types of programs which promote language development in second/foreign language learners.

Data gathered in this study was used to validate an instrument which is expected to help teachers select ESL/EFL software whose features fit the needs of their curriculum and students, and will inform researchers about the features in software programs that effectively develop language learning according to Communicative Language Teaching principles and to an interactive approach to computer use in language learning.

Theoretical Framework

The principles of Communicative Language Teaching (CLT) and an interactive approach to computer use for language learning served as theoretical background for the elaboration of the ESL/EFL software evaluation instrument. Thus, this section characterizes the main principles of CLT and presents research and theories that support an interactive approach for CALL.

Communicative Language Teaching

Communicative language teaching (CLT) has been an influential approach for more than three decades. The very term 'communicative' carries an obvious ring of truth: we 'learn to communicate by communicating' (Larsen-Freeman 1986, p. 131).

The Communicative Approach to Language Teaching is based on a theory of language as communication. The goal of language teaching is to develop what Hymes (1972) referred to as 'communicative competence'. Hymes's theory of communicative competence was a definition of what a speaker needs to know in order to be communicatively competent in a speech community.

Another linguistic theory favored in Communicative Language Teaching (CLT) is Halliday's functional account of language use. "Linguistics … is concerned … with the description of speech acts or texts, since only through the study of language in use are all the functions of language, and therefore all components of meaning, brought into focus" (Halliday 1970, p. 145).

Learning a second language was similarly viewed by proponents of Communicative Language Teaching, for example, Brumfit and Johnson, 1979; Savignon, 1983, as acquiring the linguistic means to perform different kinds of functions. Henry Widdowson is another theorist frequently cited for his views on the communicative nature of language. Widdowson (1978) presented a view of the relationship between linguistic systems and their communicative values in text and discourse. He focused on the communicative acts underlying the ability to use language for different purposes.

Expanding on the premise that language should be learned for communicative purposes, Canale and Swain (1980) identified four dimensions of communicative competence:

1. Grammatical competence–the domain of grammatical and lexical capacity.
2. Sociolinguistic competence–understanding of the social context in which communication takes place, including role relationships, the shared information of the participants, and the communicative purpose for their interaction.
3. Discourse competence–the interpretation of individual message elements in terms of their interconnectedness and of how meaning is represented in relationship to the entire discourse or text.
4. Strategic competence–the coping strategies that communicators employ to initiate, terminate, maintain, repair, and redirect communication.

In sum, Communicative Language Teaching has a rich and eclectic theoretical base, whose characteristics can be summarized (Richards & Rodgers, 1986) as:

1. Language is a system for the expression of meaning.
2. The primary function of language is for interaction and communication.
3. The structure of language reflects its functional and communicative uses.
4. The primary units of language are not merely its grammatical and structural features, but categories of functional and communicative meaning as exemplified in discourse.

Three key pedagogical principles that developed around CLT were: the presentation of language forms in context, the importance of genuine communication, and the need for learner-centered teaching. These were widely acknowledged but nevertheless open to interpretation, resulting in what Howatt (1984) described as weak

and strong versions of CLT. The former includes pre-communicative tasks (such as drills, cloze exercises, and controlled dialogue practice) along with communicative activities. Littlewood (1981), for example, described pre-communicative activities as a necessary stage between controlled and uncontrolled language use.

In strong versions of CLT the teacher is required to take a 'less dominant role' and the learners are encouraged to be 'more responsible managers of their own learning' (Larsen-Freeman 1986, p. 131). In order to encourage meaningful language use, many popular communicative activities involve 'elements of puzzle-solving, role play, or simulation' (Hadfield 1990). They encourage learners to do things with information such as: guessing, searching, matching, exchanging, collecting, sharing, combining, and arranging.

Studies based on the communicative approach also argue that social interaction is essential to language learning (e.g., Hall & Verplaetse, 2000; Lantolf, 2000; Long, 1983, 1996, 2005, 2006; Pica, 1994). Empirical evidence suggests that social interaction is a wellspring for negotiation of meaning, a communicative exchange that sustains and repairs conversations (Long 1983, 1996; Pica, 1994). Negotiation of meaning is a cognitive process that speakers use to better understand one another, that is, to increase the comprehensibility of language input. Furthermore, negotiation of meaning may result in modified interaction (Pica, 1994; Smith, 2004), which ostensibly optimizes second language acquisition (Gass, 1997). Modified interaction, as defined by Long (1983), is partly accomplished through the conversational repair moves of negotiation of meaning, including utterances such as clarification requests, comprehension checks, and incorporations in learners' speech.

An Interactive Approach to Computer Use for Language Learning

Communicative Language Teaching principles and an interactive approach to computer use in language learning are supported by Vygotsky's theories. Within Vygotskyan theory (1978), instruction is more than just didactic teaching, with a teacher explaining and demonstrating through language. Effective forms of teaching require learners to take an active role in the learning process. Scaffolded instruction does not mean teacher-initiated discourse and learner dependency. Higher order learning (problem-solving, evaluation, synthesis) requires the learner to be self-regulated, and to demonstrate initiative and independent thought. Studies by Barnes (1992) and Forman and Cazden (1985) suggest that students working together enjoy peer support and increased verbal exchange leading to higher levels of task involvement and problem solving behaviors. Social interaction and peer presence thus seem to be predictors of task related interaction and higher order thinking. If we accept that this is the case, how can verbal interaction be related to learning with computers?

The answer can be found in a sociocultural theory of learning in technology supported environments, because such a theory: endorses the fact that learning takes place in a social context; recognizes that language use is fundamental to learning; and acknowledges that learners need support and assistance to learn. All of these elements can be integrated in an interactive approach, which provides the basis for maximizing learning in technology-supported environments.

Research demonstrates the possibility of using a socio-constructivist theory in computer-based learning environments, and supports an interactive approach in the use of computers for language learning (Bonk & Cunningham, 1998; Jonassen, 1994; O'Malley, 1995; Schank & Cleary, 1995, Chong, 1998). Technology

can provide a socio-constructivist environment for relevant learning by creating whole, authentic, inherently interesting activities and setting up multiple representations of reality and actual experience for learners, thus enabling them to construct their own knowledge.

Research on computer based learning environments indicates that group work around computers offers opportunities for language use and enhanced learning outcomes. For example, group work with computers has been found to provide support for: relatively autonomous learning on the part of students (Collins & Berge, 1996, Jonassen, 1994; Ocker & Yaverbaum, 1999); increased collaboration and negotiation (Blake 2000; Fernandez-Garcia & Martinez-Arbelaiz, 2002; Kern, 1995; Kim, 1998; Pellettieri, 2000; Repman, 1993; Smith, 2003, and Warschauer, 1996); a higher quality of exploratory talk and cognitive discourse (Davis & Thiede, 2000; Irvine, 2000; Sengupta, 2001); greater problem solving competencies and higher order thinking (DeLoach & Greenlaw, 2002; Kamhi-Stein, 2000; MacKinley, 1999); development of writing skills and literary uses of language (Warschauer, 2004; Schultz, 1996).

In addition to increasing the comprehensibility of input, negotiation of meaning may also raise speakers' awareness of target language forms. Speakers may be alerted that their speech is inaccurate when interlocutors make the repair moves of negative feedback, such as the recasts and explicit corrections interlocutors make to inform speakers of grammatical inaccuracies (Gass, 1997; Long, 1996; Mackey, 1999; Spada, 1997). As a result, if the speaker recognizes the various types of negative feedback provided by interlocutors, the speaker may attempt to self-correct (Long, 1996).

Although it is widely agreed that computers can be used to facilitate interaction and negotiation of meaning, researchers still diverge as to the quality of interaction among second language

learners facilitated by the use of the technology. Some recent studies (Bohlke, 2003; Fitze, 2006; Lee, 2004; Simpson, 2005; Smith, 2003) attempted to investigate the differences in the quality of interaction between face-to-face and written electronic conferences. These studies suggest that in the written electronic setting, tasks involving negotiation of meaning tend to lead to incidental interactions in which students are asked to clarify and rearticulate what they have written in the target language. Also, as compared to face-to-face interaction, computer-mediated communication has an equalizing effect on the quantity and quality of participation across gender, socioeconomic status, and age, because participants feel less anxious or shy. Other studies (Iwasaki & Oliver, 2003; Jepson, 2005; Sauro, 2004; Warner, 2004) have compared learners' negotiated interactions in text and voice chat rooms. The studies suggest that although text chat is the more widely available and most studied form of chat, voice chat offers an environment in which learners are more apt to negotiate for meaning. Voice chats generated a number of repair moves, specifically negotiation of meaning-type repair moves. Because of the inherent absence of non-verbal communication and the focus that current voice chat technology places on pronunciation, voice chat may be an optimal environment for pronunciation work.

Besides facilitating language use, communicative, dialogic processes around computers can contribute to the development of higher order cognition. The potential benefits of discourse and learning are well documented in the literature. Learners' acquisition of new knowledge structures and cognitive strategies is facilitated by peer interaction where verbalization and dialogue are mediating forces. In groups, for example, students can learn from each other by giving and receiving help. By recognizing inconsistencies between their own and other people's perspectives, they can create mental

models of problems. By observing and participating in problem solving approaches that have been the product of joint effort, students increase their own repertoire of skills (Webb, Troper & Fall, 1995). In addition, exchanging ideas through verbal interaction promotes higher levels of thinking, such as question generation, explanation and elaboration (Webb Troper & Fall, 1995). Interpersonal discussion of ideas to resolve conflict and reach agreement is a further benefit of collaborative work with peers and computers (Pea, 1992). Overall, there is compelling evidence of the benefits of verbal interaction and communicative task-related talk in producing higher order learning within computer mediated environments.

Research conducted in classrooms indicates that judicious use of the computer has the potential to create conditions conducive to collaborative learning, and to sustain interactions leading to higher order learning (Light, 1993). Through talk, the process of representing one's thoughts for others, normally covert processes are made overt through language and dialogue.

Wegerif (1996) suggests that it is possible to plan for, and build exploratory talk within a teaching program using directive software. Exploratory talk can be achieved by changing the normally asymmetric patterns of interaction which characterize the classroom, resulting in predominantly teacher initiated discourse. This asymmetry needs to be balanced by student discussion, which can be achieved by encouraging students to engage with the software, and to discuss and evaluate their perceptions of working with a particular software package.

In summary, constructivists have found that communication technologies can realize constructivist ideals of learning (Bonk & Cunningham, 1998): active, collaborative construction of knowledge instead of knowledge transfer from one person to another (Jonassen,

1994; O'Malley, 1995), engagement in contextualized authentic tasks as opposed to abstract instruction, and less-controlled environments versus predetermined sequences of instruction where "conditions for shared understanding" are created and "alternative solutions and hypothesis building" (O'Malley, 1995, p. 289) are promoted through student interaction. Such learning environments encourage thoughtful reflection and "empower … learners … to assume ownership of their knowledge, rather than reproducing the teacher's" (Jonassen, 1994, p. 6).

However, various technologies differ in the way and extent to which they facilitate the realization of constructivist principles (Tella & Mononen-Aaltonen, 1998). Instructors need to identify the technologies and the implementation of those technologies, which best fulfill curricular goals (Chapelle, 1997; Tella & Mononen-Aaltonen, 1998).

Tasks should therefore be set to provide conditions for social collaboration. Pedagogical tasks using computer-based telecommunications should be carefully designed. Salaberry (1996) believes that the implementation of pedagogical tasks in computer mediated communication environments should be attentive to two important features of the design process: the nature of interaction among humans (communication paradigm) and the roles of the learner in such interaction (language learning goals). He proposes that a distinction be made between the concepts of interaction and communication to the effects of providing a better theoretical foundation for the pedagogical uses of internet environments. According to Salaberry (1996), the technical distinction between interaction (mutual or reciprocal action or influence) and communication (a process by which meanings are exchanged between individuals through a common system of symbols) can help

us keep in perspective the pedagogical value of computer mediated interaction for second language learning.

While many researchers agree that peer collaboration is conductive to learning, many debate the form that feedback needs to take (whether positive or negative) in order for second language learning acquisition to occur. Some researchers have maintained that positive evidence alone is sufficient for adult second language acquisition (e.g., Krashen, 1977, 1994). Others consider positive evidence as insufficient for second language learning to occur, and propose a role for both positive and negative evidence (e.g., Hatch, 1978; Long, 1983, 1996; White, 1987). Positive evidence tells the learner that linguistic features in the input are possible in the target language. In contrast to positive evidence, negative evidence provides information to learners about what is not possible in the target language (e.g., Lightbown & White, 1987; Long, 1996). Other researchers claim that recasts–reformulation of a learner's ill-formed utterance–can provide implicit negative feedback, positive evidence, and enhanced salience through the juxtaposition of the original ill-formed utterance and the target language recast form (Leeman, 2000; Saxton, Kulcsar, Marshall, & Rupra, 1998). In contrast to explicit correction and recasts, negotiation of form does not provide learners with the correct target language form. Instead, it indicates to learners that they have produced an error and that the error requires repair (Lyster, 1998 a, b).

Recent second-language acquisition research has developed a noticeable interest in the role that implicit negative feedback, such as recasts and negotiation, plays in second language development (Ayoun, 2001; Leeman, 2003; Long, Inagaki, & Ortega, 1998; Mackey & Philp, 1998; Morris, 2002; Muranoi, 2000). Findings suggest that implicit negative feedback facilitates learners' L2 development.

Because of the potential benefits of implicit negative feedback, research has attempted to examine whether it is available to learners in different interactional contexts (e.g., Braidi, 2002; Buckwalter, 2001; Chaudron, 1988; Doughty; Ellis, Basturkmen, & Loewen, 2001; Lyster, 1998 a, b; Lyster & Ranta, 1997; Mackey, Oliver, & Leeman, 2003; Moroishi, 2001; Morris, 2002; Ohta, 2000; Oliver, 1995, 2002; Panova & Lyster, 2002). These studies demonstrate that implicit negative feedback is frequently available and used by second language learners.

Learning around computers, therefore, may entail a new discourse role for teachers and students, as they engage in discussion, interaction, reflection and adaptation of ideas. It is clear that the quality of learning around computers is not entirely dependent upon the interface between learners and the technology. Instead, it is related to the whole social climate of the classroom and the opportunities created for interaction and exploratory talk between participants in the learning process.

As indicated, research provides evidence that the computer, when adequately used, can be a social facilitator in the sense that it provides opportunities for collaboration, group work and interaction which fosters cognitive change.

Method

Elaboration of the Instrument

The first step taken in the elaboration of the ESL/EFL software evaluation instrument consisted of a review of available software evaluation instruments designed by the following researchers: AI-Kahtani & Abalhassan (1999), Chapelle (1998), Cunningham (1995), Dudley-Marling & Owston (1988), Garrett et al. (1995), Healey &

Johnson (1997a), Hubbard (1987, 1988), Janello (1984), Kerr (2001), Mitra (2002), Murray and Barnes (1998), Odell (1986), Plass (1998), Poulsen (1990), Reeder et al. (2004), Reeves (1994 and 1997), Shueckler & Shuell (1989), Stieglitz (1997), Thorn (1995). The results of the review provided an understanding of the factors that should be considered in designing an effective ESL evaluation instrument.

The criteria used for the ESL/EFL software evaluation instrument were based on: (a) technological features associated with interactive CALL; (b) features of the Communicative Language Approach; and (c) technological features that allow for individualized instruction. The gathered criteria were grouped into three clusters encompassing desirable features of software programs for ESL/EFL learning: (a) Technological features; (b) Pedagogical features; and (c) Individualized-learning features. The following criteria were used in the elaboration of the questions in the "Critical Analysis" part of the ESL/EFL software evaluation instrument.

Technological Criteria

Although the quality of learning around computers is related to the social climate of the classroom and the opportunities created for interaction and exploratory talk between participants in the learning process, the interface between learners and the technology is highly correlated with how well users enjoy using a specific program. No matter how pedagogically appropriate a program might be, if students do not feel motivated to use it, it has very little chance to facilitate learning. Therefore, the analysis of software should start with an evaluation of the user interface. Criteria for the evaluation of specific technological features include:

- Ease of use - The perceived facility with which a user interacts with a multimedia program. The meaning of icons and symbols

should be easy to understand and remember. Options, choices, and menus have to be easily found. Instructions should be helpful.

- Navigation - Ability to move through the contents of an interactive program in an intentional manner. An important aspect of navigation is orientation, i.e., the degree to which a user feels that he/she knows where he/she is in the program and how to go to another part of it. A good approach to navigation is the WIMP (window-icons -mouse-pointing) interface.
- Screen design - The quality and design of: text, icons, graphics, color, and other visual aspects of interactive programs.
- Media integration - How well an interactive program combines different media to produce an effective whole. It also deals with whether or not the various media components are necessary to the function of the program.

Pedagogical Criteria

An analysis of the approach adopted by the programs is the most critical parameter of software evaluation, for it determines the pedagogical soundness and appropriateness of the program. This analysis primarily involves looking at the theoretical underpinnings of activities, judging how well they conform to the principles of Communicative Language Teaching, and determining how closely they are aligned with the program's objectives. The following criteria used for this type of analysis are based on the principles of Communicative Language Teaching.

- Theory of Language and of Learning: language is a system for the expression of meaning; its primary function is interaction and communication, and it is learned through tasks that are meaningful to the learner; the target linguistic system is learned best through the process of struggling to communicate; linguistic variation is a central concept in materials and methodology.

- Objectives: include functional skills (such as instrumental, interactional, and personal) as well as linguistic objectives (grammar and vocabulary for example); reflect the interest and needs of the learner; are made very clear to teachers and students.
- Syllabus: includes structures, functions, notions, themes, and tasks; ordering is guided by learner's needs and interests; sequencing is determined by consideration of content, function, or meaning.
- Activities: engage learners in communication, and involve processes such as information sharing, negotiation of meaning and interaction; allow unplanned and unpredictable responses; involve real communication; and develop communicative competence (i.e. the ability to use the linguistic system effectively).

Individualized Learning Criteria

Individualized instruction is a very strong premise of Communicative Language Teaching. Therefore, in order to evaluate the incorporation of Communicative Teaching principles into CALL, it is also important to analyze how technological features enable ESL/EFL software to adapt itself to fit different learning styles, needs and interests. Individualization of instruction has long been a major pursuit of CALL because it justifies having students practice ESL on the computer (Chapelle and Jamieson, 1986).

Individualization refers to the fact that the computer enables students to work alone and at their own pace. To provide an individualized learning environment, many developers have used a systems approach to design: a learning hierarchy is formulated, and a diagnostic mechanism is used so that either the computer program or the student can decide when the student needs to review (Dick & Carey, 1978; Tennyson, 1981). The difficulty, however, is in designing a diagnostic mechanism that will enable each student

to proceed along a tailor-made path. Although its potential has been demonstrated, individualization has not been achieved at a sophisticated level (Hart, 1981; Kearsley, Hunter, & Seidel, 1983). This traditional view of individualization in CALL is now seen in a new light. Some educators have proposed that students use the computer as a means of exploring and playing with material (such as the target language) through group work tasks and student-initiated exchanges. Individualization is directly related to the type of social environment students create for their own learning experiences.

The capability of collecting data and keeping records is a second aspect of individualized CALL. Data on any interaction that occurs between the student and the computer can be collected and subsequently analyzed. A good software program should be capable of keeping a permanent record of learner performance for the instructor, and of allowing the learner to carry on from where he/she has left off. This dimension also deals with immediate information on accuracy of response and /or a summary of total right and/or wrong answers.

The third aspect of individualized CALL instruction is embodied in answer judging. Answer judging occurs after students answer a question posed by the computer. The computer informs them whether it is right or wrong. Moreover, if the answer is wrong, the program should provide students with a meaningful explanation as to why the answer is wrong. So, this third distinguishable dimension of software programs refers to the type of feedback given to the students. Software programs should provide feedback that helps students judge when, and mainly why, their answer is wrong.

After the criteria were collected, the instrument was divided into two parts:

1. The "Descriptive Analysis", which describes the technical and pedagogical orientation of the software programs based on the documentation.
2. The "Critical Analysis", which assesses the extent to which software programs available on the market develop ESL/EFL skills according to the principles of the Communicative Approach and an interactive model of computer use for language learning.

The rating scale designed to answer all questions in the "Critical Analysis" part ranged from 1 (for low) to 4 (for high). Zero would be chosen when the feature was not present in the program. The higher the rating of a program, the more the program would be evaluated as having the potential to develop ESL/EFL skills according to the principles of Communicative Language Teaching and to an interactive approach to computer use for language learning.

A draft of the ESL/EFL software evaluation instrument was tested for its internal consistency in a Pilot Study. The Cronbach Alpha Coefficients for the ratings of the three programs analyzed in the pilot study indicated that the elaborated ESL/EFL software evaluation instrument had high levels of internal consistency. The data from the pilot study were used to construct a revised draft of the ESL/EFL software evaluation instrument with minor alterations. The validated ESL/EFL Software Evaluation Instrument is presented in Appendix A.

Validation of the Instrument

Study Participants

In order to test the internal consistency and the inter-rater reliability of the ESL/EFL software evaluation instrument, this instrument was used by ESL/EFL teachers in the evaluation of ESL/

EFL software programs. The selection of participants for this study was based on the following criteria:

1. Status of ESL and EFL teachers, since most software programs are designed for both ESL and EFL learners.
2. Teaching experiences that encompassed most contexts in which ESL/EFL is taught (middle, and high schools, university, and adult education).
3. ESL/EFL certified teachers.

Invitational letters were sent to fifteen American ESL teachers and to twenty Brazilian EFL teachers, known to this researcher, who met the above-specified criteria. Thirty teachers agreed to voluntarily participate in the study. Twenty-six teachers returned their evaluation results–15 Brazilian EFL teachers and 11 American ESL teachers.

Materials

For the reliability tests, three ESL/EFL software programs were selected–ELLIS, New Dynamic English, and Side-by-Side. The criteria for the selection of these three programs were based on: 1) programs designed for ESL/EFL young adult or adult learners; 2) multi-leveled programs (designed to take learners from one to another level of language development; 3) multi-skills programs (designed to teach listening, speaking, reading and writing skills in English as a second/foreign language; and 4) availability of enough demonstration CD-Roms that could be distributed to all participants in the study.

Procedures

The ESL/EFL software evaluation instrument was used by the ESL/EFL teachers to evaluate the selected ESL/EFL software programs. Each teacher received a sample CD-Rom of each of the

three programs which included sample lessons of different levels of the program, and an overview of the whole program. Since this study was also investigating how clearly and objectively the instrument could be used by teachers with different backgrounds and experiences to identify criteria of the Communicative Approach of Language Teaching and an interactive model of computer use for language learning used in the design of the programs, teachers were not given any training on how to use the instrument, nor were they assigned any specific time or amount of time to analyze the programs. However, all teachers were asked to go through at least one complete lesson in each level of the software program. All teachers were also encouraged to ask about and/or take notes on doubts or questions while using the instrument to analyze the programs.

Method of Data Collection and Analysis

The data used to test the reliability of the ESL/EFL software evaluation instrument came from the 26 participant-teachers' ratings for the items in the ESL/EFL software evaluation instrument for each of the three selected programs.

In order to test the inter-rater reliability of the ESL/EFL software evaluation instrument, the 26 teachers' ratings for the items in the instrument for each program's evaluation were entered into the Statistical Package for the Social Sciences program (SPSSP, Version 12.0). The following statistical tests were performed to test the inter-rater reliability:

1. The Pearson Coefficients were calculated to establish the degree of agreement among the raters for each program.
2. The Intra-Class Coefficients were calculated to determine the degree of correlation among the ratings of the raters for each

program. In the computation of the Intra-Class Coefficients (ICC), a *two-way mixed model* was used because the raters were seen as a fixed effect (not a random sample of all possible raters) and the items as a random effect. The ICC in this case is interpreted as not being generalizable beyond the given raters. The *Consistency* type was used because the interest was to measure raters' relative ratings; that is, if raters' ratings were consistent as long as their relative ratings were similar. The *Average Measure Reliability* gives the reliability of the mean of the ratings of all raters because the research design involved averaging multiple ratings for each item. The *Average Measure Reliability* is a Cronbach's Alpha Coefficient.

In order to evaluate the internal consistency of the ESL/EFL software evaluation instrument, the following statistical tests were run in the SPSS program, version 12.0, in this order:

1. Descriptive statistics were used to explore the dependent variables–the ratings for each item in the evaluation instrument for each program. The mean and the standard deviation for each item's ratings for each program were calculated.
2. The Cronbach Alpha Coefficients were calculated for the ratings for all the items of the ESL/EFL software evaluation instrument for each program.
3. Factor analysis was done with the ratings for the items of the ESL/EFL software evaluation instrument for the program with the highest Cronbach Alpha Coefficient.
4. The Cronbach Alpha Coefficients were calculated for the items loaded in each factor.
5. The Cronbach Alpha Coefficients were calculated for the items all together.

The ESL/EFL software evaluation instrument was also submitted to two experts' analysis for face and content validity. The criteria used

for the elaboration of the ESL/EFL software evaluation instrument were also presented to these experts.

Results

The results of the reliability and validation analyses of the ESL/EFL software evaluation instrument are presented in four sections: the first section describes the results for face validity and content validity; section two presents the results of inter-rater reliability measures; the third section describes the results of internal consistency measures; and the last section summarizes the results of the reliability and validity analysis.

Results for Face Validity and Content Validity

The ESL/EFL software evaluation instrument was submitted for the analysis to two experts in ESL/EFL pedagogy. These experts determined that 'on the face of it', the ESL/EFL software evaluation instrument seemed appropriate and valid.

Then, these experts were given the criteria that guided the elaboration of the ESL/EFL software evaluation instrument. They examined the evaluation instrument using these criteria, and both agreed that the items in the evaluation instrument comprised the set of criteria associated with the principles of the Communicative Approach to language teaching and an interactive model of computer use for language learning.

Results of Inter-Rater Reliability Measures

The ESL/EFL evaluation instrument was tested for its inter-rater reliability. Two statistical procedures–Pearson Coefficient and Intra-Class Coefficient (ICC)–were used to measure the inter-rater reliability among the 26 raters' ratings for the 60 items of the

ESL/EFL software evaluation instrument in the evaluation of the software programs–"ELLIS", "New Dynamic English", and "Side-by-Side. Although both procedures assess the homogeneity of the ratings among raters, the Pearson Coefficient establishes the degree of agreement on the ordering of the ratings among the raters only, while the Intra-Class Coefficient determines the degree of agreement not only of the ordering of the ratings but also of the range of the ratings among the raters. Therefore, the Pearson Coefficient is an index of the degree of correlation among raters, while the Intra-Class Coefficient is an index of the degree of correlation among the ratings of the raters.

Pearson-Coefficients

The degree of agreement among the 26 raters for the program "New Dynamic English" ranged from r = .729 to r = .995. For the program "ELLIS", the degree of agreement ranged from r = .749 to r = .938. The Pearson Coefficients for the program "Side-by-Side" ranged from r = .769 to r = .981. These results indicate a significant linear correlation among the raters.

Intra-Class Coefficients

In the computation of the Intra-Class Coefficients (ICC), a two-way mixed model was used because the raters were seen as a fixed effect (not a random sample of all possible raters) and the items as a random effect. The ICC is interpreted as not being generalizable beyond the given raters. The Consistency type was used because the interest was to measure raters' relative ratings; that is, if raters' ratings were consistent as long as their relative ratings were similar. The Average Measure Reliability gives the reliability of the mean of the ratings of all raters because the research design involved averaging

multiple ratings for each item. The Average Measure Reliability is a Cronbach's Alpha Coefficient. The Intra-Class Coefficients for the raters' ratings for the 60 items in the ESL/EFL software evaluation instrument are shown in Table 1.

Table 1: *Intra-Class Coefficients for Raters' Ratings for "New Dynamic English"*,
"*ELLIS*", and "*Side-by-Side*"

Programs	Intraclass Correlation/ Cronbach's Alpha	95% Confidence Interval	
		Lower Bound	Upper Bound
DynamicEnglish	.990	.986	.994
ELLIS	.994	.991	.996
Side-by-Side	.994	.992	.996

For the program "Dynamic English," the ICC correlation was .990, at 95% of confidence level, lower bound = .986, upper bound = .994. For the program "ELLIS", the ICC correlation was .994, lower bound = .991, upper bound = .996. For the program "Side-by-Side", the ICC was .994, lower bound = .992, upper bound = .996. These results indicate a strong correlation among the 26 raters' ratings for the three programs.

The results of the Pearson Coefficient and the ICC Coefficient measures indicate a high degree of inter-rater reliability. These results suggest that the study participants were an appropriate sample, and that despite their different backgrounds (American and Brazilian, ESL and EFL) there was a high level of agreement among the participants.

Results of Internal Consistency Measures

Internal consistency focuses on the degree to which the individual items are correlated with each other. The statistical measures used in this study to measure the level of internal consistency for the ESL/EFL software evaluation instrument were the Cronbach's Alpha Coefficients for inter item reliability and Factor Analysis followed by re-calculation of the Cronbach's Alpha Coefficients.

Cronbach- Alpha Coefficients for Inter-Item Reliability

The following table shows the Cronbach's Alphas for the three programs used to test the internal consistency of the instrument. The data used to compute the Cronbach's Alpha coefficients came from the results of the analysis of the three programs by 26 ESL/EFL teachers who used the instrument to evaluate the programs.

The Cronbach's alpha reliability tests applied to the analysis of the three specified programs indicated the following Cronbach's Alphas: .721 for ELLIS; .916 for New Dynamic English, and .866 for Side-by-Side. The results indicate adequate levels of inter item reliability among the items of the ESL/EFL software evaluation instrument (see Table 2).

Table 2: Cronbach's Alpha Coefficients for the programs "Ellis", "New Dynamic English", and Side-by-Side.

Programs	Cronbach's Alpha Coefficients
Ellis	.721
New Dynamic English	.916
Side-by-Side	.866

Factor Analysis

A Principal Component Analysis was conducted to determine what, if any, underlying structure existed for measures on the 60 items of the ESL/EFL software evaluation instrument. The data used for the factor analysis were derived from the results of the evaluations of the 26 ESL/EFL participant teachers who used the instrument to evaluate the program "New Dynamic English". The decision to use the data gathered from the evaluations of this program was based on the fact that "New Dynamic English" had the highest Cronbach's Alpha Coefficient (.916) when compared to the other two programs' Cronbach's coefficients (.721 for "ELLIS", and .866 for "Side-by-Side").

The 60 items were analyzed using a principal components analysis extraction method and varimax rotation with Kaizer normalization resulting in 15 components. The first three components accounted for 62.65% of the total variance in the original items. The first component–Factor 1–accounted for 36.95% of the total variance, and consisted of 28 of the 60 items. The second component–Factor 2–accounted for 14.93%, and consisted of 16 items. The third component–Factor 3–accounted for 10.76% of the total variance, and included 11 items. The other factors were not kept because, besides having few items loaded in them, the items loaded in these other factors were also loaded in one of the first three factors with higher path coefficients.

The 5 remaining items (out of the original 60 ones) had path coefficients < .4, and were eliminated:

- Item 1.6–"The sound adds to the understanding of the teaching point," loaded in Factor 8 (-399), was eliminated based on the analysis of that item being either too obvious (anyone could presuppose that the sound should add to the understanding of the teaching point), or not clearly understood (what was really

meant by adding to the understanding of the teaching point? what would that look like?).

- Item 1.12 - "Graphics organize information into a coherent structure," loaded in Factor 5 at .-393, was considered redundant; items 1.11–"Graphics aid understanding" - and 1.10–"Graphics make information attractive" - covered the same criteria. If 'graphics aid understanding' and if they 'make information attractive', it is very likely that they do so because they 'organize the information into a coherent structure'.

- Item 1.16–"The program allows learners to move through its contents on they own will" - loaded in Factor 2 at .-314, was eliminated since its content was considered to be repeated in Item 3.48, also loaded in Factor 2–"The program allows learner to go through its content at their own pace and rhythm." 'On their own will' can be understood as 'at their own pace and rhythm'. Besides, item 3.48 had a higher path coefficient (.643) than item 1.16 (.-314).

- Item 2.27 - "The program gives teachers a clear orientation of how to use its different components" - loaded low in Factor 1 at .346 - was eliminated because in the 'Descriptive Analysis' part of the ESL/EFL software evaluation instrument, one of the items asks for the same type of information (cf. item 2.14 in the Descriptive Analysis, Appendix B).

- Item 2.37 - "The program provides challenging activities" - loaded in Factor 1 at .388 - was eliminated because it was considered very subjective (different people have different concepts of challenge).

The path coefficients of the 55 items loaded in the first three components are reported in Appendix B.

After the 5 items were dropped from the ESL/EFL software evaluation instrument, and the remaining items were grouped into the three factors, the Cronbach Alpha Coefficients were calculated

for the items kept within each of the three major factors, and for all the 55 items together. The correlation was high, with the Cronbach's Alpha Coefficients for the items in the three factors and for all the items together reported in Table 3.

Table 3: *Cronbach's Alpha Coefficients for Each Factor*

Factor number	Cronbach's Alpha
1 (n = 28)	.96
2 (n =16)	.93
3 (n = 11)	.92
Total (n = 55)	.94

The inter-item reliability tests for each factor resulted in higher alpha values and thus a more robust set of items - .96 for Factor 1, .93 for Factor 2, and .92. The Cronbach's Alpha coefficient for the 55 items in the analysis of "New Dynamic English" was .94. This indicates that the inter-item reliability of the original 60 items, obtained from the data of the analysis of the program "New Dynamic English", which was already significant (Cronbach's Alpha = .916), turned out to be even higher after the five items were eliminated (Items 1.6, 1.12, 1,16, 2.27, and 2.37).

After deciding which components (factors) to retain and where to place items that were either not heavily loaded (<.4) in any component or loaded in more than one component, it was necessary to examine the items that comprised each component and try to interpret these factors in terms of their underlying parameters in order to be able to name the factors. The data from this study suggested that there were three key parameters that could be measured as represented in the three factors identified in the factor analysis.

The first parameter–Factor 1–that needed to be measured when evaluating ESL/EFL programs was thus called "instructional attributes." Clearly, when evaluating any educational material the first concern should be the potential effectiveness of its instructional attributes. These instructional attributes include: the theory of language learning and teaching that underlies the program (Items 2.17, 2.18, 2.19, 2.20, 2.21, 2.22 and 2.28), how content is presented and sequenced (Items 2.29, 2.30, 2.31, 2.32, 2.33, 2.34 and 2.35), the type and quality of language-learning activities (Items 2.36, 2.38, 2.39, 2.40 and 2.49), the type of feedback provided (Items 3.53, 3.54, 3.55 and 3.60), and how the program facilitates learning (Items 1.4, 1.7, 1.9, 1.10 and 1.13).

The second parameter–Factor 2–to be measured in the evaluation of ESL/EFL software programs was called "media attributes" since the items in this component were directly related to the technological features of software programs. These media attributes include: the degree and type of individualization of instruction the program provides (Items 3.41, 3.42, 3.43, 3.44. 3.45, 3.46, 3.47 and 3.48), how the program helps learners move through its content and sequence (Items 3.51, 3.52), how the media technology differentiates feedback (Items 3.50, 3.56, 3.57, 3.58 and 3.59), and how the integration of different types of media facilitates learning (Item 1.11).

The third parameter to be used in the evaluation of software programs was named "user-friendly attributes"–the items in Factor 3 relate to how the program is perceived as attractive, easy to follow, and/or motivating to learners and teachers. Although this is a more subjective parameter, it is also important because learning heavily depends on how motivated students and teachers feel about a specific course or program. The user-friendly attributes include: how the use of the program is made easy to learners (Items 1.1, 1.2, 1.3, 1.14, 1.15

and 2.26), how attractive the media technology makes the program to be (Items 1.5, 1.8 and 2.23) and how the program motivates learning (Items 2.24 and 2.25).

In summary, after dropping the 5 items with factor weight below .4 from the set of 60 original items and re-grouping the remaining 55 items into three factors, the alpha coefficients had higher values. This indicates that the revised draft of the ESL/EFL software evaluation instrument has a more robust set of items and higher levels of internal consistency.

Discussion

Analysis of data from the face and the content validity measures, the inter-rater coefficient measures, and the internal consistency measures suggest these answers to the questions addressed in this study:

1. *To what extent does the ESL/EFL software evaluation instrument produce consistent results when administered under similar conditions? In other words, is the designed ESL/EFL software evaluation instrument reliable?* The results of the Pearson Coefficient and the ICC Coefficient measures indicate high levels of inter-rater reliability for this group of 26 teachers. The Cronbach's Alpha Coefficients for the three programs indicated that the ESL/EFL software evaluation instrument had adequate levels of inter-item reliability. The factor analysis resulted in a set of 55 items with even higher alpha values and thus a more robust set of items. The results suggest that the ESL/EFL software evaluation instrument has high levels of internal consistency, and they also suggest inter-rater reliability. All these results assure the evaluation instrument a high degree of reliability.

2. To what extent is the ESL/EFL software evaluation instrument valid to evaluate the potential of CALL programs to develop language skills according to the Communicative Language Teaching principles and an interactive approach to computer

> use in language learning? The positive indicators of reliability obtained from the procedures used to assess the inter-rater reliability and the internal consistency and the face and content validity attributed to this instrument suggest that the ESL/EFL software evaluation instrument is potentially a valid instrument to assess the degree to which ESL/EFL software programs develop language skills according to the Communicative Language Teaching principles and an interactive approach to computer use in language learning.

In sum, the results of the reliability measures suggest that it is possible to use the ESL/EFL software evaluation instrument to measure the degree to which ESL/EFL software programs incorporate the features of Communicative Language Teaching and an interactive approach to computer use for language learning into their design. The 55 items in the ESL/EFL software evaluation instrument, clustered around three components, represent specific and distinguishing attributes of software programs that can be used as yardsticks for measuring the incorporation of the Media, User-Friendly, and Instructional Attributes into the elaboration and design of software programs. The incorporation of these attributes can, in turn, account for the integration of the features of Communicative Language Teaching and an interactive approach to computer use for language learning into software programs.

In this study, the positive indicators of reliability obtained from the procedures used to assess the inter-rater reliability and the internal consistency of the instrument, and the face and content validity attributed to the instrument, suggest that the ESL/EFL software evaluation instrument is potentially a valid instrument to assess the degree to which ESL/EFL software programs develop language skills according to the principles of Communicative Language Teaching and an interactive approach to computer use for language learning.

However, several limitations have to be taken into account in the discussion of the results of the validation of the ESL/EFL software evaluation instrument.

The study participants were either Brazilian EFL teachers or American ESL teachers. These teachers do not represent the entire EFL/ESL teacher population, since teachers from other countries where English is taught either as second or a foreign language were not represented in this study's sample. Thus, the sample represents a portion of the whole population of ESL/EFL teachers.

The study design consisted of the evaluation of three ESL/EFL software programs using the ESL/EFL software evaluation instrument. The method consisted, therefore, of an introspective judgmental evaluation, which can arguably be done purely individually, subjectively, globally and introspectively (Scholfield, 2000). No experimental studies were undertaken to test if the way learners responded to the programs matched the expectations raised by the evaluation of the programs using the ESL/EFL software evaluation instrument.

Also, because participation was voluntary, the teachers who agreed to participate may have knowingly or unwittingly biased the study results. Although the items in the ESL/EFL software evaluation instrument were objectively based on a set of criteria, every introspective evaluation involves a certain degree of subjectivity.

Limitations of the availability of enough demonstration CD-Roms to be distributed among the participating teachers did not allow random assignment of software programs to the 26 raters for the items in the ESL/EFL software evaluation instrument. Additionally, the 26 teacher-raters evaluated the three assigned programs by analyzing the sample lessons included in the demo CDs, and not the whole software program. Since the analyzed lessons were included in

a demonstration CD-Rom (and the purpose of a demonstration CD is to advertise and sell a software program), it can be argued that the lessons might not be good representations of all the other lessons of a software program, and so, of the programs as a whole.

In order to measure the validity of the ESL/EFL software evaluation instrument, two reliability measures were pursued: inter-rater reliability and internal consistency, and content and face validity were verified. However, although the inter-rater and the internal consistency coefficients were high, and the two experts granted face and content validity to the instrument, other types of reliability and validity measures, such as predictive validity, were not undertaken. Also, since the data used for the reliability tests came from a sample of convenience, it is important to run the same tests with data from a larger and more diverse sample of the ESL/EFL teacher population.

The results of this investigation suggest specific areas for further research. Additional studies using the ESL/EFL software evaluation instrument might further confirm the reliability and validity of this instrument. Additional studies should consider other types of reliability and validity measures.

Future studies should also incorporate more diverse populations of ESL/EFL teachers, and include teachers from different countries. Besides, further studies should also use a larger sample of participants. Furthermore, the evaluation of ESL/EFL software programs should be done through the analysis of the whole programs, rather than just through the analysis of sample lessons in demonstration CD ROMs.

Also, other experts should be asked to evaluate the extent to which the items in the ESL/EFL software evaluation instrument comprise the set of criteria to be taken into account when assessing a software program's potential to develop language skills according to principles of Communicative Language Teaching and an interactive

approach to computer use for language learning. These additional studies would grant the ESL/EFL software evaluation instrument more reliability and validity.

Further studies should attempt to use the ESL/EFL software evaluation instrument to evaluate not only commercial software, but also software programs privately developed by universities and other language learning institutions. Also, future research should evaluate CALL programs available on the Internet

Finally, other methods of investigating ESL/EFL software programs should be pursued. It would be relevant to first analyze an ESL/EFL software program using the ESL/EFL software evaluation instrument, and then have a group of ESL/EFL learners use this software program in order to see if the way actual learners respond to the program corresponds to the expectations of the program to potentially develop language skills according to the principles of Communicative Language Teaching and an interactive approach of computer use for language learning.

References

Al-Kahtani, S., & Abalhassan, K. (1999). Software review: MacReader and Storyboard programs in ESL reading classrooms. Retrieved June 7, 2000, from: http://www.lerc.ritsumei.ac.jp/caUej/index.html

Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of the Passé Composé and the Imparfait. *Modern Language Journal, 85*, 226-243.

Barnes, D. (1992). The role of talk in learning. In K. Norman (Ed.). *Thinking voices: The national oracy project*. London: Croom Helm.

Beatty, K. (2003). *Teaching and researching computer-assisted language learning*. Great Britain: Pearson Education Limited.

Blake, R. (2000). Computer-mediated communication: A window on L2 Spanish interlanguage. *Language Learning & Technology, 4*, 120-136.

Bohlke, O. (2003). A comparison of student participation levels by group size and language stages during chatroom and face-to-face discussions in German. *CALICO Journal, 21*, 67-87.

Bonk, C., & Cunningham, D. (1998). Searching for learner-centered, constructivist, and sociocultural components of collaborative educational learning tools. In C. Bonk & K. King (Eds.), *Electronic collaborators* (pp. 25-50). Mahwah, NJ: Lawrence Erlbaum.

Braidi, S. (2002). Reexamining the role of recasts in native-speaker/nonnative-speaker interactions. *Language Learning, 52*, 1-42.

Brumfit, C. J. & Johnson, K. (1979). *The communicative approach to language teaching.* Oxford: Oxford University Press.

Buckwalter, P. (2001). Repair sequences in Spanish L2 dyadic discourse: A descriptive study. *Modern Language Journal, 85*, 380-397.

Burston, J. (2003). Software selection: A primer on sources and evaluation. *CALICO Journal 21*, 29-40.

Canale, M. & Swain, M. (1980). Theoretical Bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1-47.

Chapelle, C. A. (1997). CALL in the year 2000: Still in search of research paradigms? *Language Learning and Technology, 1,* 19-43.

Chapelle, C. A. (1998).Multimedia CALL: Lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2*, 22-34

Chapelle, C. & Jamieson, J. (1986). Computer assisted language learning as a predictor of success in acquiring English as a second language. TESOL Quarterly, 20, 27-46.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning.* Cambridge, England: Cambridge University Press.

Chong, S-M. (1998). Models of asynchronous computer conferencing for collaborative learning in large college classes. In C. J. Bonk & K. S. King (Eds.), *Electronic collaborators* (pp. 157-182). Mahwah, NJ: Lawrence Erlbaum.

Collins, M., & Berge, Z. (1996). *Facilitating interaction in computer mediated online courses.* Retrieved June 15, 2000, from http://www.emoderators.com/moderators/flcc.html

Cunningham, P. (1995). Evaluating callware for your classroom. *CAELL Journal*, *6*, 13-19.

Davis, B., & Thiede, R. (2000). Writing into change: Style shifting in asynchronous electronic discourse. In M. Warschauer & R. Kern (Eds.), *Network-based language teaching: Concepts and practice* (pp. 87-120). Cambridge, England: Cambridge University Press.

DeLoach, S. B., & Greenlaw, S. A. (2002). *Do electronic discussions create critical-thinking spillovers?* Manuscript submitted for publication.

Dick, W., & Carey, L. (1978). *The systematic design of instruction.* Glenview, IL: Scott, Foresman.

Dudley-Marling, C. & Owston, R.D. (1988). A field-testing approach to software evaluation. *Computers in the Schools*, *5*, 241-249.

Ellis, R., Basturkmen, H., & Lowen, S. (2001). Learner uptake in Communicative ESL lessons. *Language Learning, 51*, 281-318.

Fernández-García, M., & Martínez-Arbelaiz, A. (2002). Negotiation of meaning in non-native speaker-non-native speaker synchronous discussions. *CALICO Journal, 19*, 279-294.

Fitze, M. (2006) Discourse and participation in ESL face-to-face and written electronic conferences. *Language Learning & Technology***,** *10*, pp. 67-86.

Forman, E. A & Cazden, C. B. (1985). Exploring Vygotskian perspectives on education: The cognitive value of peer interaction. In J. V. Wertsch, (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (pp. 323-247). New York: Cambridge University Press.

Garrett, M., Nimri, M., Peterson, J., & Barlow, M. (1995). *Software evaluation: Project of Rice University.* Retrieved May 17, 2000 from: http://www.owinet.rice.edu/-Iing4l7/guide.html.

Gass, S. (1997). *Input, interaction, and the second language learner*. Rowley, MA: Newbury House.

Hadfield, J. 1990. *Intermediate Communication Games.* Surrey: Thomas Nelson and Sons.

Hall, J. K., & Verplaetse, L. S. (2000). *Second and foreign language learning through classroom interaction.* Mahwah, NJ: Erlbaum.

Halliday, M. K. (1970). *Spoken and written language*. Oxford: Oxford University Press.

Hart, R. (1981). Language study and the Plato system. *Studies in Language Learning, 3*, 1-24.

Hatch, E. (1978). Discourse analysis and second language acquisition. In E. Hatch (Ed.), *Second language acquisition: A book of readings* (pp. 401-435). Rowley, MA: Newburry House.

Healey, D., & Johnson, N. (1997a). A place to start in selecting software. *CAELL Journal,* 8(1). Retrieved in May 29, 2000 from:

http:/www.ucs.orst.edu/-healeyd/cj_software_selection.html

Howatt, A. P. R. (1984). A *History of English Language Teaching*. Oxford: Oxford University Press.

Hubbard, P. L. (1987) Language teaching approaches, the evaluation of CALL software and design implementation. In W. Flint Smith (Ed.). *Modern media in foreign language education: Theory and implementation* (pp. 227-254). Lincolnwood, IL: National Textbook Company.

Hubbard, P. L. (1988). An Integrated Framework for CALL courseware evaluation. *CALICO Journal*, *6*, 51-73.

Hubbard, P. L. (1992). A methodological framework for CALL courseware development. I Pennington, M. and Stevens, V. (Eds), *Computers in applied linguistics*. Clevedon: Multilingual Matters.

Hymes, D. (1972). Models of the interaction of language and social setting. *Journal of social issues*, *23*, 8-28.

Irvine, S. E. (2000). What are we talking about? The impact of computer-mediated communication on student learning. (ERIC Document Reproduction Service No. ED444494).

Iwasaki, J., & Oliver, R. (2003). Chatline interaction and negative feedback. *Australian Review of Applied Linguistics, 17,* 60-73.

Janello, Pam. (1984). Software evaluation for the teacher of the English Language Arts. *ERIC Digest.*

Jepson, K. (2005). Conversations and negotiated interaction in text and voice chat rooms. *Language Learning & Technology*, *9*, pp. 79-98.

Jonassen, D. H. (1994). *Technology as cognitive tools: Learners as designers.* IT Forum Paper #1. Retrieved May 30, 2000, from http://it.coe.uga.edu/itforum/paper1/paper1.html

Kamhi-Stein, L. (2000). Looking to the future of TESOL teacher education: Web-based bulletin board discussions in a methods course. *TESOL Quarterly, 34*, 423-455.

Kearsley, G., Hunter, B., & Seidel, R. J. (1983). Two decades of computer-based instruction projects: What have we learned? *T.H.E. Journal*, *10*, 90-95.

Kern, R. (1995). Restructuring classroom interaction with networked computers: Effects on quantity and quality of language production. *Modern Language Journal*, 79, 457-476.

Kerr, J. (2001). Software evaluation. Retrieved in October 15, 2001 from http://www.ed.brocku.ca/~jkerr/sftwreva.htm

Kim, Y. (1998). The effect of a network computer-mediated discussion on subsequent oral discussion in the ESL classroom. Unpublished doctoral dissertation, University of Texas, Austin.

Krashen, S (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.), *Viewpoints on English as a second language* (pp. 152-161). New York: Regents.

Krashen, S. (1982). *Principles and practice in second language acquisition.* Oxford: Pergamon.

Krashen, S. (1994). The Input Hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.

Lantolf, J. (Ed.). (2000). *Sociocultural theory and second language learning.* Oxford, England: Oxford University Press.

Larsen-Freeman, D. (1986). *Techniques and principles in language teaching*. Oxford: Oxford University Press.

Lee, L. (2004). Learners' perspectives on networked collaborative interaction with native speakers of Spanish in the US. *Language Learning & Technology, 8* , 83-100.

Leeman, J. (2000). Towards a new classification of input: An empirical study of the effect of recasts, negative evidence and enhanced salience on L2 development. Unpublished doctoral dissertation, Georgetown University, Washington DC.

Leeman, J. (2003). Recasts and second language development: beyond negative evidence. *Studies in Second Language Acquisition, 25*, 37-63.

Light, P. H. (1993). Collaborative learning with computers. In P. Scrimshaw (Ed.), *Language, classrooms and computers* (pp.155-159). London: Routledge.

Lightbown, P., & White, L. (1987). The influence of linguistics theories on language acquisition research. *Language Learning, 37*, 483-510.

Littlewood, W. (1981) *Communicative Language Teaching*. New York: Cambridge University Press.

Long, M. H (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition, 5*, 177-193.

Long, M. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of research on language acquisition. Vol. 2: Second language acquisition* (pp. 413-468). New York: Academic Press.

Long, M. H. (2005). The delivery of negative feedback in naturalistic, classroom, and artificial environments. Proceedings of the UNTELE conference, Compiegne, France, 2005, 24-26.

Long, M. H. (2006). *Problems in SLA*. Mahwah, NJ: Lawrence Erlbaum Associates.

Long, M., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *Modern Language Journal, 82,* 357-371.

Lyster, R. (1998a). Negotiation of form, recasts explicit correction in relation to error types and learning repairs in L2 classrooms. *Language Learning, 48,* 183-213.

Lyster, R. (1998b). Recast, repetition and ambiguity in L2 classroom discourse. *Studies in second language acquisition, 20,* 51-81.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19,* 37-66.

Mackey, A. (1999). Input, interaction and second language development: An empirical study of question formation in ESL. *Studies in Second Language Acquisition, 21,* 557-587.

Mackey, A., & Philp, J. (1998). Conversation interaction and second language development: Recasts, response and red herrings? *Modern Language Journal, 82*, 338-356.

Mackey, A., Oliver, R., & Leeman, J. (2003). Interactional input and the incorporation of feedback: An exploration of NS-NNS and NNS-NNS adult and child dyads. *Language Learning, 53*, 35-66.

MacKinley, K. (1999). *Planning to use e-mail to support the learning process?* Retrieved December 10, 1999 from: http://www.edfac.unimelb.edu.au/online-ed/mailouts/1999/Sept6.html

Mitra, A. (2002). Toward developing questionnaire items to measure effectiveness of computers in teaching. *Journal of Educational Computing Research*, *25*, 381-394.

Moroishi, M. (2001). Recasts and learner uptake in Japanese classroom discourse. In X. Bonch-Bruevich, W. J. Crawford, J. Hellermann, C.,Higgins, & H. Nguyen (Eds.), *The past, present and future of second language research. Selected proceedings of the 2000 Second Language Research Forum* (pp. 197-208). Somerville, MA: Cascadilla Press.

Morris, F. (2002). Negotiation and recasts in relation to error types and learner repair. *Foreign Language Annals, 35*, 395-404.

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction with a communicative task in EFL classrooms. *Language Learning, 50*, 617-673.

Murray, L. & Barnes, A. (1998). Beyond the "wow" factor–evaluating multimedia language learning software from a pedagogical viewpoint. *System*, *26*, 249-259.

Ocker, R. J., & Yaverbaum, G. J. (1999). Asynchronous computer-mediated communication versus face-to-face collaboration: Results on student

learning, quality, and satisfaction. *Group Decision and Negotiation, 8*, 427-440.

Odell, A. (1986). Evaluating CALL software. In G. Leech, & C. Candlin (Eds.). *Computers in English Language Teaching and Research: Selected papers form the 1984 Lancaster Symposium "Computers in English Language Education and Research*" (pp.60-77). London: Longman.

Ohta, A. S. (2000). Rethinking recasts: A learner-centered examination of corrective feedback in the Japanese language classroom. In J. K. Hall & L. S. Verplaetse (Eds.), Second and foreign language learning through classroom interaction (pp. 47-71). Mahwah, NJ: Erlbaum.

Oliver, R. (1995*).* Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition, 17*, 459-481.

Oliver, R. (2002). The patterns of negotiation for meaning in child interactions. *Modern Language Journal, 86*, 97-111.

O'Malley, C. (1995). Designing computer support for collaborative learning. In C. O'Malley (Ed.), *Computer supported collaborative learning* (pp. 283-297). New York: Springer-Verlag.

Panova, I., & Lyster, R. (2002). Patterns of corrective feedback and classroom uptake in an adult ESL classroom. *TESOL Quarterly, 36*, 573-595.

Pea, R. D. (1992). Augmenting the discourse of learning with computer-based learning environments. In de Corte, E., Linn, M., Mandl, H., and Verschaffel, L. (Eds.), *Computer-Based Learning Environments and Problem-Solving* (pp, 87-99). New York: Springer-Verlag.

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. In M. Warschauer & R. Kerns (Eds.), *Network-based language teaching: Concepts and practice* (pp. 59-86). Cambridge, England: Cambridge University Press.

Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, *44*, 493-527.

Plass, J. L. (1998). Design and evaluation of the user interface of foreign language multimedia. *Language Learning and Technology,* 2(I), 35-45.

Poulsen, E. (1990). Evaluation of CALL from a classroom perspective. *CAELL Journal*, *1*, 73-78.

Reeder, K., Heift, T, Roche, J., Tabyanian, S., Schlickau, S., & Golz, P. (2004). Evaluating New Media in Language Development. In Fotos, S., & Browne, C. (Eds.). *New Perspectives on CALL for Second Language Classrooms* (pp. 255-278). Mahwah, NJ: Lawrence Erlbaum.

Reeves, T.C. (1994). Systematic Evaluation Procedures for Interactive multimedia for Education and Training. In Reisman, S. (Ed.) *Multimedia Computing: Preparing for the 21st century (pp.78-101)*. Harrisburg, PA: Idea Group

Reeves, T.C. (1997). *Evaluation Tools*. Retrieved September 26, 2002 from http://mime1.marc.gatech.edu/MMTools/ evaluation.html

Repman, J. (1993). Collaborative, computer-based learning: Cognitive and affective outcomes. *Journal of Educational Computing Research*, *9*, 149-163.

Richards, J. C. & Rodgers, T. S. (1986). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.

Salaberry, M. R. (1996). A theoretical foundation for the development of pedagogical tasks in computer mediated communication. *Calico Journal, 14*, 5-34.

Sauro, S. (2004). Cyberdiscursive tug-of-war: Learner repositioning in a multimodal CMC environment. *Working Papers in Educational Linguistics, 19,* 55-72.

Savignon, S. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA.:Addison-Wesley.

Saxton, M., Kulcsar, B., Marshall, G. , & Rupra, M. (1998). Long term effects of corrective input and experimental approach. *Journal of Child Language*, 25, 701-721.

Scholfield, P. J. (2000). *Evaluation of CALL software: Work in progress*. Department of Language and Linguistics, University of Essex, UK. Retrieved December 17, 2004 from http://privatewww.essex.ac.uk/~scholp/calleval.htm.

Schueckler, L.M. & Shuell, T.J. (1989). A Comparison of Software Evaluation Forms and reviews. *Journal of Educational Computing Research*, *5*, 17-33.

Schultz, J. (1996). *Computers and collaborative writing in the foreign language curriculum*. Unpublished manuscript, University of California at Berkeley.

Sengupta, S. (2001). Exchanging ideas with peers in network-based classrooms: An aid or a pain? *Language Learning and Technology, 5*, 103-134

Simpson, J. (2005). Conversational floors in synchronous text-based CMC discourse. *Discourse Studies*, *7*, 337-361.

Smith, B. (2003). Computer-mediated negotiated interaction: An expanded model. *The Modern Language Journal, 87,* 38-54.

Smith, B. (2004). Computer-mediated negotiated interaction and lexical acquisition. *Studies in Second Language Acquisition, 26,* 365-398.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 29,* 1-15.

Stieglitz, Ezra. (1997). *Form for reviewing educational software*. Unpublished manuscript.

Tella, S., & Mononen-Aaltonen, M. (1998). *Developing dialogic communication culture in media education: Integrating dialogism and technology.* Helsinki: Media Education Publications. Retrieved February 10, 1999 from http://www.helsinki.fi/~tella/mep7.html

Tennyson, R. (1981). Use of adaptive information for advisement in learning concepts and rules using computer-assisted instruction. *American Educational Research Journal, 18*, 423-438.

Thorn, W. J. (1995). Points to consider when evaluating interactive multimedia. *The Internet TESL Journal, 2*. Retrieved May 17, 2000 from: http://www.aitech.ac.jp/-iteslj/Articles/ Thorn-EvalueConsider.html

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* . Cambridge, MA: Harvard University Press.

Warner, C. N. (2004). It's just a game, right? Types of play in foreign language CMC. *Language Learning & Technology*, *8*, pp. 69-87.

Warschauer, M. (1996). Computer-assisted language learning: an introduction. In S. Fotos (Ed.), *Multimedia Language Teaching* (pp. 3-20). Tokyo: Logos.

Warschauer, M. (2004). Technology and writing. In C. Davison and J. Cummins (Eds.), *Handbook of English languaget teaching* (pp. 176-199). Kluwer, Dordrecht, Netherlands.

Webb, N.M., Troper, J.D., & Fall, R. (1995). Constructive activity and learning in collaborative small groups. *Journal of Educational Psychology*, *87*, 406-42.

White, L. (1987). Against comprehensible input: The input hypothesis and the development of L2 competence. *Applied Linguistics, 8*, 95-110.

Widdowson, H. G. (1978). *Teaching language as communication*. Oxford: Oxford University Press.

APPENDIX A

ESL/EFL SOFTWARE EVALUATION INSTRUMENT

Title of the program: _____

_____

Publisher: _____

_____

DESCRIPTIVE ANALYSIS

Please place a checkmark next to the item that best describes the program and/or its features.

I. TECHNICAL FEATURES

1.1 Components of the program:

Number of CD's: Per level Total	Teacher's guide ___Yes ____No	Support materials ___ Yes ___No What type?

1.2 Platform

Mac_____ _____ Mhz _____ MB of RAM	Windows Mhz MB of RAM

1.3 Tools

___ Word processing ____ Online forums ____ Online collaboration	___ Speech recognition ___ Web browsing ____ Encyclopedia or compendia	___ E-mail ___ Others

1.4 Other technological resources integrated into the program:

___ dictionaries on the web	____ grammar on line	___ other web sites	____ none

1.5 Directions for use:

___ are on the screen	___ are in the documentation	___ can be skipped at option of user

II. PEDAGOGICAL ORIENTATION

2.6 Program's theory of language

_____ structural	___ functional	_____ interactional

2.7 Type of program

___ remediation ___ enrichment ___ tutoring	___ demonstration ___ assessment	_____ education game _____ collaborative projects	___ problem solving ___ drill and practice ___ simulation

2.8 Curriculum capability

Grade range	Proficiency level range	
_____ elementary _____ middle school _____ high school _____ adult learning	From _____ beginning _____ pre-intermediate _____ intermediate _____ high-intermediate	To _____ pre- intermediate _____ intermediate _____ high-intermediate _____ advanced

2.9 Method of language teaching claimed by the program

_____ The Direct Method _____ The Natural Approach _____ Total Physical Response	_____ The Audio-lingual Method _____ The Communicative Approach

2.10 Language skills developed in the program

___ speaking	___ listening	___ reading	___ writing

2.11 Type of activities offered by the program

_____ games _____ quizzes _____ others	_____ simulation _____ exploratory	_____ tutorial _____ text construction	_____ drill and practice _____ problem solving

2.12 Orientation of how to use program and its feature

_____ for the teacher	_____ on the screen
_____ for the student	_____ on hard copy documentation
_____ for both	_____ on both

2.13 Role of the teacher

___ instructor	___ facilitator	___ lab manager	___ evaluator

2.14 Learners interact

___ with one another	___ with the teacher	___ with neither

2.15 In case learners interact with one another, they can interact

_____ in all activities	_____ in some activities

2.16 In case learners interact with the teacher, they can interact

_____ at any time _____ at the end of the lesson	_____ at the end of the activities _____ there's no interaction

CRITICAL ANALYSIS

For each item, circle the numbers 1 (for low) to 4 (for high) to indicate your judgment of the degree to which the program possesses or demonstrates the feature identified in each item. Circle 0 if the feature is not present in the program.

Instructional Attributes					
Graphics and sound enhance learning.	0	(low) 1	2	3	(high) 4
1.2 The animation is effective in minimizing boredom by motivating learners.	0	(low) 1	2	3	(high) 4
Screen displays are uncluttered.	0	(low) 1	2	3	(high) 4
1.4 Graphics make information attractive.	0	(low) 1	2	3	(high) 4
1.5 Graphics help memorization of key information.	0	(low) 1	2	3	(high) 4
1.6 The program makes use of authentic texts and other realia.	0	(low) 1	2	3	(high) 4

1.7 The program integrates information about culture/daily situations into the presentations and activities.	0	(low) 1	2	3	(high) 4
1.8 Lessons present and practice language structures in meaningful communicative contexts.	0	(low) 1	2	3	(high) 4
1.9 Lessons present and practice vocabulary in meaningful communicative contexts.	0	(low) 1	2	3	(high) 4
1.10 Lessons develop the communicative skills the program aims to develop.	0	(low) 1	2	3	(high) 4
1.11 Lessons develop the level of language proficiency the program aims to develop.	0	(low) 1	2	3	(high) 4
1.12 The course syllabus reflects a communicative approach to language teaching/learning.	0	(low) 1	2	3	(high) 4
1.13 Content selection is determined by communicative skills and/or themes.	0	(low) 1	2	3	(high) 4
1.14 Content is sequenced from simple communicative functions, such as introducing oneself, greeting, etc, to complex ones, such as stating an opinion, disagreeing, etc.	0	(low) 1	2	3	(high) 4
1.15 Content is presented communicatively.	0	(low) 1	2	3	(high) 4
1.16 The program develops the content at appropriate levels of language proficiency.	0	(low) 1	2	3	(high) 4
1.17 The program content is educationally relevant and interesting for the learner.	0	(low) 1	2	3	(high) 4
1.18 The program content is appropriate for intended learners.	0	(low) 1	2	3	(high) 4
1.19 The program content is applicable to real life contexts.	0	(low) 1	2	3	(high) 4
1.20 The program allows learners to work together in communicative activities.	0	(low) 1	2	3	(high) 4
1.21 The activities allow unplanned and/or unpredictable responses.	0	(low) 1	2	3	(high) 4
1.22 The activities lend themselves to group discussions.	0	(low) 1	2	3	(high) 4
1.23 The activities aim at developing other competencies in addition to syntactical and lexical	0	(low) 1	2	3	(high) 4

1.24 The program allows the teacher to interact with students while they are doing an activity.	0	(low) 1	2	3	(high) 4
1.25 The program provides non-threatening feedback.	0	(low) 1	2	3	(high) 4
1.26 The program allows learners to repeat an activity after feedback is provided.	0	(low) 1	2	3	(high) 4
1.27 Activities allow for more than one correct response.	0	(low) 1	2	3	(high) 4
1.28 The program provides the students with feedback that would allow them to correct their mistakes.	0	(low) 1	2	3	(high) 4
Sub-total =					

Media Attributes					
2.29 Graphics aid understanding.	0	(low) 1	2	3	(high) 4
2.30 The program allows for different routes and choices for learning.	0	(low) 1	2	3	(high) 4
2.31 The program allows for review of old information.	0	(low) 1	2	3	(high) 4
2.32 The program allows branching to new information.	0	(low) 1	2	3	(high) 4
2.33 The program allows students to select activities according to their ages.	0	(low) 1	2	3	(high) 4
2.34 The program allows students to select activities according to their learning styles.	0	(low) 1	2	3	(high) 4
2.35 The program allows students to select activities according to their interests.	0	(low) 1	2	3	(high) 4
2.36 The program adapts to the responses given by the learners, branching to more or less complicated questions as appropriate.	0	(low) 1	2	3	(high) 4
2.37 The program allows learners to go through its content at their own pace and rhythm.	0	(low) 1	2	3	(high) 4
2.38 The program prevents learners from repeating exercises, therefore, minimizing guessing.	0	(low) 1	2	3	(high) 4

2.39 The program keeps records of learners' performance to allow them to continue activities from where they left off.	0	(low) 1	2	3	(high) 4
2.40 The program keeps track of students' scores.	0	(low) 1	2	3	(high) 4
2.41 The program provides feedback for both correct and incorrect answers. The program provides feedback for both correct and incorrect answers.	0	(low) 1	2	3	(high) 4
2.42 The program gives learners the chance to correct their errors.	0	(low) 1	2	3	(high) 4
2.43 The program effectively signals the mistakes before providing the right answers.	0	(low) 1	2	3	(high) 4
2.44 The program effectively specifies different types of errors, such as differences between a syntactic error and an incorrect word choice.	0	(low) 1	2	3	(high) 4
Sub-total =					
User-Friendly Attributes					
3.45 Menu items are understandable and descriptive.	0	(low) 1	2	3	(high) 4
3.46 The commands and instructions for the activities are clear and objective.	0	(low) 1	2	3	(high) 4
3.47 The program gives the learners effective clues to clarify their doubts about its use.	0	(low) 1	2	3	(high) 4
3.48 Each screen uses text and graphic/ animation to make a particular teaching point clear.	0	(low) 1	2	3	(high) 4
3.49 The program is effectively integrated with other technological resources (such as dictionaries on the web, grammar on line, etc) as the learner uses it.	0	(low) 1	2	3	(high) 4
3.50 Icons, buttons and menus allow learners to readily search for additional information while doing an activity.	0	(low) 1	2	3	(high) 4
3.51 Buttons, icons or menu items make Help or Hint-type options easily accessible.	0	(low) 1	2	3	(high) 4
3.52 The program arouses sensory and cognitive curiosity	0	(low) 1	2	3	(high) 4

3.53 The program maintains attention throughout the lesson.	0	(low) 1	2	3	(high) 4
3.54 The use of animation invites learners' reaction or input.	0	(low) 1	2	3	(high) 4
3.55 The program gives teachers a clear explanation of its purposes and methodological orientation.	0	(low) 1	2	3	(high) 4
Sub-total =					

Sub-total of ratings for Instructional Attributes ……
Sub-total of ratings for Media Attributes ….. ……
Sub-total of ratings for User-Friendly Attributes …… ……
Total rating ……

APPENDIX B

Factor Loadings

Factors	Path Coef.
Factor 1	
1.4 Graphics and sound enhance learning.	(.457)
1.7 The animation is effective in minimizing boredom by motivating learners.	(.822)
1.9 Screen displays are uncluttered.	(.638)
1.10 Graphics make information attractive.	(.612)
1.13 Graphics help memorization of key information	(.782)
2.17 The program makes use of authentic texts and other realia.	(.543)
2.18 The program integrates information about culture/daily situations into the presentations and activities.	(.628)
2.19 Lessons present and practice language structures in meaningful communicative contexts.	(.636)
2.20 Lessons present and practice vocabulary in meaningful communicative contexts.	(.636)
2.21 Lessons develop the communicative skills the program aims to develop.	(.471)
2.22 Lessons develop the level of language proficiency the program aims to develop.	(.582)
2.28 The course syllabus reflects a communicative approach to language teaching/learning.	(.605)
2.29 Content selection is determined by communicative skills and/or themes.	(.582)
2.30 Content is sequenced from simple communicative functions, such as introducing oneself, greeting, etc, to complex ones, such as stating an opinion, disagreeing, etc.	(.605)
2.31 Content is presented communicatively.	(.687)
2.32 The program develops the content at appropriate levels of language proficiency.	(.518)
2.33 The program content is educationally relevant and interesting for the learner.	(.433)
2.34 The program content is appropriate for intended learners.	(.-604)
2.35 The program content is applicable to real life contexts.	(.625)
2.36 The program allows learners to work together in communicative activities.	(.753)
2.38 The activities allow unplanned and/or unpredictable responses.	(.624)
2.39 The activities lend themselves to group discussions.	(.471)
2.40 The activities aim at developing other competencies in addition to syntactical and lexical.	(.747)
2.49 The program allows the teacher to interact with students while they are doing an activity.	(.876)
3.53 The program provides non-threatening feedback.	(.601)

3.54 The program allows learners to repeat an activity after feedback is provided.	(.818)
3.55 Activities allow for more than one correct response.	(.-548)
3.60 The program provides the students with feedback that would allow them to correct their mistakes.	(.579)
Factor 2	
1.11 Graphics aid understanding.	(.729)
3.41The program allows for different routes and choices for learning.	(.661)
3.42 The program allows for review of old information.	(.-565)
3.43 The program allows branching to new information.	(.632)
3.44 The program allows students to select activities according to their ages.	(.588)
3.45 The program allows students to select activities according to their learning styles.	(.508)
3.46 The program allows students to select activities according to their interests.	(.445)
3.47 The program adapts to the responses given by the learners, branching to more or less complicated questions as appropriate.	(.741)
3.48 The program allows learners to go through its content at their own pace and rhythm.	(.430)
3.50 The program prevents learners from repeating exercises, therefore, minimizing guessing.	(.611)
3.51 The program keeps records of learners' performance to allow them to continue activities from where they left off.	(.643)
3.52 The program keeps track of students' scores.	(.632)
3.56 The program provides feedback for both correct and incorrect answers.	(.508)
3.57 The program gives learners the chance to correct their errors.	(.605)
3.58 The program effectively signals the mistakes before providing the right answers.	(.634)
3.59 The program effectively specifies different types of errors, such as differences between a syntactic error and an incorrect word choice.	(.409)
Factor 3	
1.1 Menu items are understandable and descriptive.	(.707)
1.2 The commands and instructions for the activities are clear and objective.	(.859)
1.3 The program gives the learners effective clues to clarify their doubts about its use.	(.-606)
1.5 Each screen uses text and graphic/animation to make a particular teaching point clear.	(.519)
1.8 The program is effectively integrated with other technological resources (such as dictionaries on the web, grammar on line, etc) as the learner uses it.	(.647)
1.14 Icons, buttons and menus allow learners to readily search for additional information while doing an activity.	(.624)
1.15 Buttons, icons or menu items make Help or Hint-type options easily accessible.	(.-599)

2.23 The program arouses sensory and cognitive curiosity.	(.582)
2.24 The program maintains attention throughout the lesson.	(.683)
2.25 The use of animation invites learners' reaction or input.	(.459)
2.26 The program gives teachers a clear explanation of its purposes and methodological orientation.	(.472)