



Evaluation Criteria for English Listening and Speaking E-learning Courses

Yu-Chun Wang

(willie_wang@kuas.edu.tw)

National Kaohsiung University of Applied Sciences, Taiwan ROC

Chien-Tzu Chou

(lovewzkeita@gmail.com)

National Kaohsiung University of Applied Sciences, Taiwan ROC

Abstract

Some principles or criteria are provided to learners when they use English learning websites or CALL materials (Economides, 2003; Jamieson & Preiss, 2005; Johnson, Hornikb, & Salas, 2008; Liu, Liu, & Hwang, 2011; Wang & Chen, 2009; Yang & Chan, 2008). However, little research has been conducted to describe a set of evaluation criteria for English courses, especially on English listening and speaking. The main purpose of the study is trying to construct a multi-dimensional set of criteria for English teachers to evaluate the quality of e-learning English listening and speaking courses. These criteria can assist English teachers in designing effective English listening and speaking courses to improve students' English listening and speaking ability. The developmental research applied in this paper constructed and refined the evaluation criteria using literature review, procedure, experts' reviews and document analysis (George & Mallery, 2003). These evaluation guidelines were based on the aspects of a) information for e-learning course, b) English teaching, and c) listening and speaking teaching. In order to achieve this goal, the researcher used a four-stage procedure to refine and form the evaluation criteria. In the first stage, the 98 preliminary criteria were conducted based on general information of e-learning course, English teaching and teaching English listening and speaking related researches. The second stage focused on experts' opinions for the preliminary criterion through online Google Docs with five-Likert scale. The third stage was to conduct both experts' and learners' opinions according to the results of stage two. Last stage was to finalize the criteria based on quantitative and qualitative surveys. 90 items were finalized in the criteria for evaluating the English listening and speaking e-learning course.

1 Introduction

Learning English has become the norm not only in native English-speaking countries, but also in non-native English-speaking countries. English is today the common language in the global village. People can communicate with one another through English and achieve the main objectives of communication. Many methods are offered to language instructors, learners and trainees for language learning. For example, computer-assisted language learning (CALL) is a trend that allows stakeholders to gain certain advantages, because the design of CALL applications and curricula is grounded in pedagogical and second language acquisition theories, and offers learning opportunities for language learners (Chapelle & Mizuno, 1989; Garrett, 2009; Jamieson, Chapelle,

& Preiss, 2005). For language instructors who use this method to reach their teaching goals and for learners who adopt the method of self-learning through learning websites and applications, the use of CALL can be valuable in helping them achieve their objectives.

Based on computer-mediated communication theory, instructors can teach using interfaces and platforms such as Blackboard, WebCT, Skype, MSN and e-mail (Wang & Chen, 2009). Specifically, EFL learners in Taiwan need a suitable environment in which to learn English (Chen, 2009). Compared with a traditional face-to-face class, e-learning courses offers a more convenient approach for instructors because of its specific advantages in comparison with face-to-face courses. (Chen, 2009; Lantolf & Pavlenko, 1995). The characteristics of e-learning suggest that it can transfer the target materials through various electronic media to learners or even to higher education institutions (Johnson, Hornikb, & Salas, 2008; Rekkedal, 2008). However, language instructors have realized that even the high quality, power and availability of technology cannot guarantee the quality of learners' input and feedback on the teaching context (Robin, 2007). Course design and evaluation criteria are important and necessary elements for guiding teachers in creating a language course and assessing whether an e-learning course is suitable (Liu, Liu, & Hwang, 2011; Yang & Chan, 2008). Many criteria have been proposed for general e-learning courses and language learning websites (Economides, 2003; Jamieson et al., 2005; Liu et al., 2011; Rekkedal, 2008; Robin, 2007; Wang & Chen, 2009; Yang & Chan, 2008).

The Ministry of Education in Taiwan, in collaboration with the National Science and Technology Program, founded the e-Learning Quality Service Centre (eLQSC), which provides general integrated criteria for e-learning courseware certification (eLCC; see Reinders, 2009; Sung, Chang, & Yu, 2011). This evaluation criteria format is composed of four parts, including the e-courseware content, navigation, instructional design and instructional media (Reinders, 2009).

The American Council on the Teaching of Foreign Languages (ACTFL) provides the 5Cs, which are evaluation criteria for designing and evaluating language learning activities. The principles are based on Vygotsky's sociocultural theory, which asserts that human interaction through a target language is one of the elements of communication. Learning the target language means not only acquiring the features of language forms, but also gaining communicative competence. As Yang and Chan (2008) noted, language learning differs from other subjects because of the attributes of language acquisition.

In recent research, the trends and outcomes of general e-learning criteria have not yet met the objectives of e-learning language courses. However, some studies have established general evaluation criteria for English learning websites (Economides, 2003; Jamieson et al., 2005; Johnson et al., 2008; Liu et al., 2011; Wang & Chen, 2009; Yang & Chan, 2008). These studies have focused on criteria for English learning websites and instructors' perspectives. The opinions of learners should also be sought when evaluating a course, because the learners are the users of the course. Thus, the aim of the current study is to develop specific and valid criteria for teaching listening and speaking through e-learning courses that incorporate both experts' opinions and learners' perspectives.

2 The multiple aspects of constructing English listening and speaking e-learning course evaluation criteria

Developmental research is used not only for testing theories, but also for establishing processes, techniques and instruments through methodical analysis (George & Mallery, 2003). There are two types of developmental research: Type 1 emphasizes specific products or programs, and Type 2 focuses on the research process.

In Type 1 research, studies address product design and development as well as product evaluation. These studies may also involve constructing and verifying different design models and processes to facilitate their successful use. Type 2 studies may include a model construction phase, a model implementation phase and a model validation phase. Developmental research defines research problems, reviews related literature and follows research procedures (Santally, 2011).

Because they employ multiple research methods, such as evaluation, field observation, document analysis, in-depth interviews, expert reviews and surveys, developmental research studies

can be published in various formats. Based on a review of the developmental research literature, this research method was adopted in this study.

2.1 *E-learning courses*

Many courses have utilized the e-learning model, including medical, hospital, engineering, business training, intercultural competence and language learning courses. Learning is an activity or process that manifests as a change in a person's perceptions, attitudes, or cognitive or physical skills (Rekkedal, 2008). Through learning activities, instructors and learners interact with each other and can reach consensus on learning outcomes based on what they think and what they learn. Learning activities have been designed and applied in many courses, including language teaching, to provide students with more flexibility in their learning.

E-learning has been defined as an interactive form of learning in which the learning content is available online and automatic feedback is provided to learners with respect to their learning activities (Grifoll et al., 2010; Koochang, Riley, & Smith, 2009; Sung et al., 2011). This type of learning involves the use of educational technology, ranging from simply posting lecture notes on the Web to managing the learning process (Rekkedal, 2008). In e-learning, pedagogy and design, learner and instructor characteristics, technology use and peer interactions can contribute to the effectiveness of learning outcomes (Johnson et al., 2008; Wang & Chen, 2009).

With increased information sharing and processing, learners will be more likely to perceive value in the way the learning material applies in a course and the way the environment supports the sharing of information via peer interactions. Online communication with real people may or may not be included in this type of learning. The focus of e-learning is typically on the learning content rather than on communication between learners and instructors (Rekkedal, 2008). In this light, it would appear that a pedagogical model of Internet-based courses needs to be designed and organized using high-quality distance education to satisfy and support the needs of a wide variety of learners (Rekkedal, 2008).

2.2 *Evaluation of e-learning courses*

Evaluation refers to the process of investigating a program to judge its appropriateness for a given learning setting, identifying effective methods of implementation, assessing the program's degree of success and determining whether to continue its use or adjust its implementation for future use (Hubbard, 1988; Jamieson et al., 2005). The use of technology requires both pedagogical and technical skills, especially on the part of instructors, as well as a substantial investment of time and resources at the personal and institutional levels (Reinders, 2009).

The quality of education reflects the relationships among learning outcomes, course demands, goals, standards and requirements established by individuals, businesses, organizations, local communities and states (Chen, 2009; Chien & Chang, 2006; Grifoll et al., 2010; Santally, 2011). To maintain the quality of an e-learning course, evaluation criteria are necessary. In a study by Hay et al. (2008), the findings suggest that the measurement of learners' prior knowledge is key to the effective design of e-learning materials. Thus, e-learning materials should involve both instructors' opinions and learners' perspectives. Evaluation criteria have been provided for evaluating e-learning courses in various subjects (Chien & Chang, 2006; Johnson et al., 2008).

E-learning is a product of the people who conceive and implement it as well as of modern telecommunication systems. It has been shown to be an effective tool for bridging distance gaps on the Internet. In fact, e-learning is not a remote learning tool; it is a tool for overcoming the distance gap (Karlsudd & Tågerud, 2008). The distance gap is completely bridged for the parties involved in e-learning sessions within the framework of instructor-to-learner and learner-to-learner interactions. E-learning courses can convey all important instruction related to lectures, examinations and assignments. In recent years, the quality of e-learning courses has received increased attention. According to a public announcement by the Ministry of Education in Taiwan, the number of certified e-learning programs increased from 2009 to 2013. The eLCC evaluation criteria consist of

three aspects of quality measurement: personnel, course and system aspects. These three aspects were further developed into eight areas: (1) course description: a course can achieve the target goal and related information for learners; (2) learning motivation: the course provides activities for learners that can increase their learning motivation; (3) instructional design and media: the course utilizes multimedia to demonstrate the teaching/learning content; (4) instructional interaction: the course provides opportunities for learner-instructor interactions; (5) peer interaction: the course increases and offers learner-to-learner interaction; (6) learning evaluation: the course provides activities or assignments to evaluate learners' progress; (7) instructional management: the course provides surveys to indicate students' reflections on learning; and (8) assessment and evaluation: the course achieves the overall requirements of each category.

According to Sung et al. (2011), the reliability of quality assurance systems provides strong empirical support for the feasibility and benefits of providing certification for e-learning courses. The results of Chien and Chang's (2006) study showed that the eLCC evaluation criteria successfully increase awareness of e-related course quality among e-learners, developers, and vendors. Thus, the eLCC evaluation criteria have proven to be a valuable reference that can be transferred to other contexts (Chen, 2009). In sum, eLCC can provide a base model for the construction of a set of evaluation criteria.

2.3 Criteria for English teaching

The World Wide Web has opened a new world of opportunity to those who need or want to learn English. Because individuals' learning beliefs evolve through classroom interactions, language instructors' understanding of learners' beliefs is a key factor in learning outcomes (Fujiwara, 2012). For those who have little time, learning can occur at home using the Internet rather than by attending a regular class and studying with an instructor in a traditional class setting. Studies based on Krashen's second language acquisition (SLA) hypotheses have contributed to the development and assessment of language courses (Antrim, 2005; Ariza & Hancock, 2003; Ellis, 2005). Chapelle (2001) proposes three important aspects in CALL development: 1) findings and theory-based speculation about ideal conditions for SLA need to be taken into account; 2) a theory of evaluation needs to be articulated; and 3) criteria and theory need to be applied to CALL software and the tasks that teacher plan and learners carry out.

Chapelle and Jamieson (2008) suggested seven key imperatives in CALL: (1) making key linguistic characteristics salient (e.g. "The course presents different accents in the learning materials"); (2) offering modifications for linguistic input; (3) providing opportunities for comprehensible output (e.g. based on learners' current proficiency levels, they can choose the appropriate arrangement of lessons and tests to fit their needs); (4) providing opportunities for learners to notice their errors (e.g. "The course provides various types of exam questions"); (5) correcting learners' linguistic output; (6) supporting modified interactions between learners; and (7) using L2 learning tasks that involve the computer as a participant. Chapelle (2001) proposed a set of criteria for CALL evaluation that included language learning potential, meaning focus, learner fit, authenticity, positive influence and practicality. Some studies have adopted these evaluation criteria with positive results (Ahmada, Wana, & Jianga, 2011; Jamieson et al., 2005).

2.3 Criteria for teaching listening and speaking

In relation to the suggestions above, Ariza and Hancock (2003) proposed eight conditions for an optimal language learning environment: (1) learners have opportunities to interact with one another and to negotiate meaning; (2) learners interact with the target language through authentic audiences; (3) learners are involved in authentic tasks; (4) learners are exposed to and encouraged to produce varied and creative language; (5) learners have sufficient time and feedback during their learning processes; (6) learners are guided to mindfully attend to the learning process; (7) learners work in an atmosphere with a minimal level of stress/anxiety; and (8) learner autonomy is supported. Learners in speaking-related courses who have adequate knowledge and use of self-

regulated learning strategies appear to exhibit enhanced learning performance in spoken communication. Additionally, communication strategies can help learners to manage their speaking activities. Thus, teachers should be aware of the powerful role of self-regulated learning and should attempt to help learners become acquainted with these principles and strategies (Aregu, 2013). Learners must take initiative in their learning and control how knowledge is acquired during instruction. The design and implementation of an environment to encourage and enable this require specific tools for authoring and delivery, including explanations, tutoring and the intelligent diagnosis of learners' proficiency.

3 Method

To develop an effective and integrated set of evaluation criteria for English listening and speaking e-learning courses, the developmental research (George & Mallery, 2003) applied in this study constructed and refined the evaluation criteria through a literature review, expert reviews, and document analysis.

A four-stage research procedure was applied with the following steps. Stage 1 involved synthesizing and establishing a set of preliminary criteria from the reviewed literature. In Stage 2, experts were first asked to evaluate the preliminary criteria through a three-part questionnaire, and then these criteria were refined by categorizing items and ranking the importance of each item and the revised items. Stage 3 involved interviewing learners about their opinions of the quality and design of the e-learning courses and interviewing experts about specific items that needed expert validity. In Stage 4, the criteria for English listening and speaking e-learning courses were finalized based on the opinions of learners and experts.

These stages were necessary to ensure that the evaluation criteria were comprehensive and applicable to English listening and speaking e-learning courses. The following sections describe the details of the four-stage research procedure.

3.1 Stage 1 – Gathering a set of preliminary evaluation criteria

The language e-learning course evaluation requires specific components, such as course management, learning material, interaction between learners and instructors, and listening and speaking teaching theories. Based on a review of the related literature, the four constructs in the English listening and speaking e-learning course evaluation criteria are: 1) *information on the e-learning course*, 2) *the teaching of English*, 3) *the teaching of listening* and 4) *the teaching of speaking*. These criteria have been adopted in related studies, online resources and books. The first construct, *information on the e-learning course*, focuses on the information extracted from the Ministry of Education in Taiwan's eLCC, which refers to the information provided by the course. The second construct evaluates the overall messages relevant to *English teaching* approaches, content and materials. This construct refers to how instructors implement a language learning course. Based on the work of Chappelle and Jamieson (2008) on practical *listening and speaking teaching* approaches for CALL, the third and fourth constructs are related to multimedia aids to assist learners' in developing listening comprehension and speaking skills. The preliminary criteria included 98 items. In this set of evaluation criteria, some items were taken from previous studies and revised by the author, while other evaluation criteria were proposed by the author (see Table 1).

Table 1. Seven selected studies and response items

Selected Literature	Response Items
1. E-learning courseware certification (eLCC)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25
2. Multimedia CALL: Lessons to be learned from research on instructed SLA (Chapelle, 1998)	28, 33, 34, 37, 40, 47, 59, 66, 68, 70,71, 72, 81, 82, 86, 93
3. CALL environments: Research, practice, and critical issues (Egbert & Hanson-Smith, 1999)	31, 32, 39, 42, 58
4. Computer applications in second language acquisition (Chapelle, 2001)	30, 31, 35, 43, 50, 52, 54, 55, 56, 63, 66, 67, 75, 85, 90, 97, 98
5. Checklist: Evaluative criteria for computer-delivered language learning systems	36, 45, 48, 49, 62, 74, 87, 88, 95
6. Comprehensive evaluation criteria for English learning websites using expert validity surveys (Yang & Chan, 2008)	26, 27, 30, 41, 51, 53, 54, 61
7. Teaching listening & speaking, Latricia Trites, Ph.D., Academic Advisor, Fulbright Yilan Project 2008–2009	58, 60, 67, 72, 76, 77, 94
8. Created by the authors	29, 38, 44, 46, 57, 65, 69, 73, 78, 79, 80, 83, 84, 89, 91, 92, 96, 98,

3.2 Stage 2 – Conducting the expert survey

The goal of Stage 2 was to conduct expert evaluation surveys and gather opinions. In some small-sample studies, small numbers of panel experts are invited to judge the validity of the criteria developed for item acceptability. The experts must agree on the content validity of the representation of the items. The recommended number of experts in these studies ranges from three to 20 (Gable, 2001; Lynn, 1986; Rubio, Berg-Weger, Tebb, Lee, & Rauch, 2003)

Six panel experts who have been in-service professors either in the English teaching field or in the educational learning technology field for more than five years were recruited for this study. Three of these experts had experience using e-learning to teach English courses, and the others taught in traditional classrooms. All of these experts were from the Department of Applied Foreign Languages of a technology university in southern Taiwan.

Each expert had one week to complete the evaluation form, which consisted of three parts that asked the experts to provide responses for each evaluation criterion. The experts first considered the category to which each item should belong. Subsequently, the experts ranked the importance of each item on a five-point Likert scale, with 1 indicating “not very important” and 5 representing “very important.” In addition, room was provided for the experts to revise items through the online use of Google Docs (see Appendix A), if the item descriptions were not clear or inappropriate.

3.3 Stage 3 – Interviewing learners and experts

To ensure a comprehensive yet concise set of criteria, the researchers used SPSS to calculate the mean scores of each criterion. The average scores on the five-point Likert scale were categorized as low, medium, and high. Mean scores in the range of 1.0–2.4 were considered low, scores of 2.5–3.4 were considered medium, and scores of 3.5–5.0 were considered high (Su, 2005).

The researchers used SPSS to calculate the mean scores representing the importance of each item and the reliability of each category, and used factor analysis for construct validity to filter the items.

The purpose of Stage 3 was to verify the results from Stage 2 through interviews. Items with low mean scores were discussed with the experts per the interview protocol. The six experts provided comments on each low mean score item based on their teaching experiences.

Twelve English major graduate students who had English learning and teaching experience as well as e-learning course experience also participated in this study. These students were from the Department of Applied Foreign Languages of a technology university in southern Taiwan. The learners' perspectives and suggestions regarding English listening and speaking e-learning courses as well as their needs and preferences for instructional classes facilitated the clarification of the criteria. This objective was one of the main aims of the interviews.

3.4 Stage 4 – Finalizing the criteria

As noted previously, this study reviewed related literature and formed a set of preliminary criteria in Stage 1. The experts evaluated the set of preliminary criteria in Stage 2. Based on the expert survey results, the researcher analyzed the data using SPSS. The purpose of Stage 3 was to interview experts and learners. The researchers interviewed the participants, made audio recordings of the interviews and also took notes. Then, the researchers transcribed the notes for data analysis. Lastly, the goal of Stage 4 was to revise, finalize and synthesize the set of evaluation criteria based on the expert evaluation results and the expert and learner interviews from Stage 3. The researchers deleted and modified some items based on statistical analysis and the results of the interviews. Low-level mean scores and some items' classifications were discussed with the experts. For the factor analysis, the factor loading of each item was calculated. Finally, the reliability of each category was assessed.

4 Results and discussion

4.1 Stage 1 – Gathering a set of preliminary evaluation criteria

As noted previously, the first stage involved a review of related literature on e-learning course content, management, navigation, instructional design, instructional media and language teaching theory. Based on the literature, preliminary criteria, including 98 items, were collated and divided into four main categories: *information on e-learning courses* (25 items), *English teaching* (22 items), *teaching listening* (21 items), and *teaching speaking* (30 items).

4.2 Stage 2 – Conducting the expert survey

To ensure validity, the experts classified the items into three aspects in Stage 2. The expert opinions regarding the preliminary criteria were collected through an online Google Docs survey. The experts also provided their perspectives on each criterion.

Cronbach's alpha is an important tool for evaluating and assessing questionnaires. Researchers must ensure validity and accuracy in the interpretation of the data collected (Gadermann, Guhn, & Zumbo, 2012). Patanakul (2005) provided the following standards for Cronbach's alpha reliability: < .5 is unacceptable; > .5 is poor; > .6 is questionable; > .7 is acceptable; > .8 is good; and > .9 is excellent. The results show that the reliability of the items pertaining to information on the e-learning course, English teaching, teaching listening and teaching speaking were .930, .935, .948, and .965, respectively. The results also show that nine deleted items that were in the mean range of 1.0-2.4 required further discussion with the experts (see Table 2).

Table 2. Mean scores and standard deviations

Item number	Mean	SD	Item number	Mean	SD	Item number	Mean	SD
<i>Information on the e-learning course</i>								
1	4.40	.894	9	2.80	1.304	17	4.40	.548
2	4.60	.894	10	3.80	1.095	18	4.40	.548
3	4.20	1.304	11	4.00	.707	19	4.40	.894
4	4.60	.894	12	4.40	.894	20	4.40	.548
5	3.40	1.673	13	3.80	.837	21	2.40	1.949
6	3.80	1.789	14	3.60	1.517	22	2.40	1.949
7	3.80	1.643	15	3.80	.447	23	3.00	1.871
8	4.40	.548	16	4.00	1.414			
<i>English teaching</i>								
24	3.20	2.049	32	3.60	.894	40	4.40	.548
25	3.40	1.517	33	4.40	.894	41	4.00	.707
26	4.60	.548	34	3.20	2.049	42	4.00	.707
27	3.00	1.871	35	4.80	.447	43	4.20	.837
28	4.40	.894	36	4.60	.548	44	3.80	1.643
29	4.20	.837	37	4.40	.894	45	2.40	1.949
30	3.80	.837	38	4.80	.447	46	1.60	1.342
31	4.20	.837	39	4.80	.447	47	1.60	1.342
<i>Teaching listening</i>								
48	4.20	.837	55	4.80	.447	62	4.60	.548
49	4.20	.837	56	4.80	.447	63	4.80	.447
50	3.40	1.517	57	3.80	1.643	64	4.40	.548
51	4.40	.894	58	4.40	.548	65	4.80	.447
52	4.60	.548	59	4.60	.548	66	2.40	1.949
53	4.60	.548	60	4.00	1.732	67	3.00	1.871
54	3.60	1.517	61	4.60	.548	68	1.00	.000
<i>Teaching speaking</i>								
69	2.60	2.191	79	4.20	.837	89	5.00	.000
70	3.00	1.871	80	4.00	1.000	90	4.40	.548
71	4.80	.447	81	4.40	.894	91	4.00	.707
72	3.40	1.673	82	4.60	.894	92	4.20	.837
73	3.60	1.673	83	4.60	.894	93	3.80	1.789
74	3.80	1.643	84	4.60	.894	94	3.80	1.643
75	4.80	.447	85	4.60	.548	95	1.00	.000
76	4.20	.447	86	3.20	2.049	96	1.60	1.342
77	4.60	.548	87	3.80	1.643	97	3.40	2.191
78	4.40	.894	88	5.00	.000	98	3.40	2.191

Note: The items in boldface are the low-level mean scores that required further discussion with the experts.

4.3 Stage 3 – The opinions of experts and learners

The survey results were discussed with the six experts with respect to the nine items with low mean scores (mean \leq 2.40; see Table 3). Only nine items were considered unimportant or considered for deletion. The experts provided comments and suggestions on the nine items during the interviews. Discussions with these experts revealed three main reasons for the items with low mean scores. First, the items were considered optional; the choice to include this information depended on the decision of the instructors. Second, the items had low rankings because of their high degree of similarity; the redundant items could have been combined or removed. Finally, the item descriptions could be revised. For instance, items could have used examples that would have made them more specific and provided clues for readers. Ultimately, these nine items were deleted after interviews with the experts.

Table 3. The nine items with low mean scores

The low-level mean scores that were deleted after discussion with the experts
21. The course displays some outstanding work by learners on the e-learning platform
22. The instructor changes the schedule or sends an instant message through the online bulletin board
45. Consider how learners will respond
46. Teaching approach analyzes and synthesizes what is read by learners
47. Address a situation that is linguistically similar to the teaching content
66. Provide authentic audio files so that learners hear native speaker-like pronunciation of words, sentences and conversations
68. Learners extend their knowledge and draw conclusions according to what they hear
95. Choose communication strategies and perform systematic practice depending on learning style
96. Combine speaking with listening and reading

The expert surveys revealed that the categories of some items needed to be discussed. To confirm that the item classifications were suitable, 13 items were discussed (see Table 4). For the category of *information on the e-learning course*, the experts concluded that Items 4, 6, 7, 8, 9 and 23 were course-related. Based on the experts' recommendations, these six items remained under *information on the e-learning course*. However, it was determined that Items 16, 18, and 20 should be moved to the category of *English teaching*, because they were related to teaching materials, approaches, and assignments. In other words, the main concepts of these three items reflected decisions to be made by instructors. For the category of *English teaching*, most items were found to belong to this category, except for Item 28. The English teaching category concerned mainly language skills training. However, in regard to the aspect of reading and writing training, the item was not suitable. Thus, the experts recommended that Item 28 be moved to the category of *teaching listening*. For the category of *teaching speaking*, it was determined that Items 82, 85, and 90 should be moved to the category of *information on the e-learning course*. The experts suggested that the three items were applicable to every subject and were related to general information sharing.

The second goal of this stage was to gather opinions and suggestions from experienced e-learners to assist in forming the set of criteria. This step was included, because learners can offer insightful suggestions that experts may not have noticed.

The learners shared information that could be divided into two main categories: *the role of instructors* and *information about a course*. Most of the learners thought that the instructor should be consistent in arranging courses with regard to teaching activities, materials and teaching content as well as with respect to corrections. Moreover, instructors should adjust their courses based on the learning situations or outcomes of learners, and should assist them in a timely manner. However, some of the learners thought that the role of the instructor should be to assist students during their learning processes, because this type of course is learner-oriented. These students indicated that they did not particularly need the instructor.

With regard to information about the course, the learners first wanted to establish the main purposes or goals based on their expectations for the course. If courses could provide interactivity, well-adapted teaching strategies and practical materials, then these elements would increase their learning motivation.

Table 4. Items that were re-categorized after further discussion with experts

Original category	New category
<i>Information on e-learning course</i>	
4. The teaching materials and activities match the teaching and learning goals.	1
6. The instructor provides adequate teaching activities based on the course objective	1
7. The teaching content provides examples to assist learners' comprehension	1
8. The teaching materials provide supplementary materials such as Internet resources	1
9. The instructor provides at least 3/2 asynchronous and synchronous web-based courses	2
16. The instructor uses a cooperative learning strategy in class (e.g., learning strategies include group discussion, group project or peer review)	2
18. The assignments match the teaching goals and teaching materials	2
20. The assignments assist learners in organizing the key points and stimulate their critical thinking	1
23. The instructor holds a face-to-face panel discussion with learners during the last class to improve future classes	
	3
<i>English teaching</i>	
28. Learning materials are provided with different accents (e.g., British, American, Australian, Canadian)	
	1
<i>Teaching Speaking</i>	
82. Learners are allowed to practice the assigned segmental patterns through tasks (e.g., task that is separated into 1-3 parts is assigned; learners can proceed step by step to complete the task)	1
85. Clarification is sought	1
90. Clues are provided for learners to answer questions	

4.4 Stage 4 – Finalizing the criteria from the results of the learners' selections

Factor analysis allows relationships and patterns in the data to be clearly interpreted and understood (Yong & Pearce, 2013). Factor analysis is a statistical method that is used to examine the relationship between groups of observed variables through questions or items. Main factors are identified among the variables and placed in meaningful categories. The factor loading measures how the variables contribute to the factors; high loading scores are required (Hill & Hughes, 2007). Costello and Osborne (2005) indicated that a loading of .50 is sufficient to be considered strong. The factor loading of each item was higher than 0.5, indicating convergent validity. No item had a low factor loading, which means that this set of evaluation criteria had discriminant validity. In sum, this set of evaluation criteria possesses construct validity. To form a set of evaluation criteria, developmental research can be useful and significant because it involves testing theories and determining whether processes are beneficial to learners and instructors.

In sum, 89 items were finalized in the set of criteria to evaluate English listening and speaking e-learning courses. The final count of the items included 29 items in the category of *information on the e-learning course*, 18 items in the category of *English teaching*, 19 items in the category of *teaching listening*, and 23 items in the category of *teaching speaking* (see Appendix B).

5 Conclusion

The quality of e-learning courses plays an important role in the success of the courses (Puerto & Gamboa, 2009). Upon realizing this importance, the Taiwan Ministry of Education sought to ensure the quality of e-learning courses by implementing the eLCC in 2005. However, e-learning course evaluation criteria for English teaching have not been the focus of prior research. Thus, the

purpose of this study was to construct evaluation guidelines or principles for language instructors, particularly those teaching English listening and speaking.

A developmental research approach was adopted in this study. The set of criteria was developed based on related CALL theories and CALL evaluation studies. These evaluation criteria were reviewed by six in-service CALL experts, whose opinions lent credibility to the criteria. The researcher used SPSS for the data analysis. As the results show, the reliability of the evaluation criteria was high. Examining the mean scores led to a more concise set of criteria. In addition, factor analysis demonstrated that the set of evaluation criteria was valid.

This set of multidimensional guidelines for language learning courses, especially for those teaching English listening and speaking, represents a first step toward providing preliminary criteria. Based on the set of evaluation criteria, five suggestions are provided for future studies. First, because of time limitations, only six in-service CALL experts were consulted. More experts should thus be consulted to evaluate the criteria. However, the small sample size of panel experts was still within the recommended range (Gable, 2001; Lynn, 1986; Rubio et al., 2003). Second, the evaluation criteria must be updated regularly due to changes in insights and perspectives resulting from the research process (Chen, 2009; Liu et al., 2011). Third, this set of evaluation criteria pertains to listening and speaking courses, but reading- and writing-related criteria should also be considered. Fourth, this set of evaluation criteria can be applied to examine the practical aspects of specific courses. Finally, because the set of evaluation criteria was oriented toward teaching methods, listening and speaking texts and tests were not considered in this study. In future studies, these two factors could be considered, which would make this set of evaluation criteria more comprehensive.

References

- Ahmada, N. S. H. N., Wana, T. R., & Jianga, P. (2011). Immersive environment courseware evaluation. *Procedia Social and Behavioral Sciences*, *15*, 1667–1679.
- Antrim, N. M. (2005). A survey of SLA doctoral programs in Canada and the USA. *Second Language Research*, *21*(1), 72–94.
- Aregu, B. B. (2013). Enhancing self-regulated learning in teaching spoken communication: Does it affect speaking efficacy and performance? *Electronic Journal of Foreign Language Teaching*, *10*(1), 96–109.
- Ariza, E. N., & Hancock, S. (2003). Second language acquisition theories as a framework for creating distance learning courses. *International Review of Research in Open and Distance Learning*, *4*(2), 1–9.
- Chapelle, C., & Mizuno, S. (1989). Students' strategies with learner-controlled CALL. *CALICO*, *7*(2), 25–47.
- Chapelle, C. A. (2001). *Computer application in second language acquisition*. Cambridge: Cambridge University Press.
- Chapelle, C. A., & Jamieson, J. (2008). *Tips for teaching CALL*. White Plains, NY: Person Education, Inc.
- Chen, M.-P. (2009). An evaluation of the ELNP e-Learning quality assurance program: Perspectives of gap analysis and innovation diffusion. *Educational Technology & Society*, *12*(1), 18–33.
- Chien, M.-P., & Chang, K.-E. (2006, November). *An evaluation of compliance and effectiveness of ELNP e-courseware quality framework*. Paper presented at the The 5th WSEAS International Conference, Venice, Italy.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*, 1–9.
- Economides, G. E. A. A. (2003). An evaluation instrument for hypermedia courseware. *Educational Technology & Society*, *6*(2), 31–44.
- Ellis, R. (2005). Principles of instructed language learning. *System*, *33*, 209–224.
- Fujiwara, T. (2012). Beliefs about language learning of Thai students learning Chinese and Japanese: Relationships with past learning experiences and target language variations. *Electronic Journal of Foreign Language Teaching*, *9*(2), 170–182.
- Gable, C. T. B. R. K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement*, *9*, 201–215.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*(3), 1–13.
- Garrett, N. (2009). Computer-assisted language learning trends and issues revisited: Integrating innovation. *The Modern Language Journal*, *93*, 719–740.

- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Grifoll, J., Huertas, E., Prades, A., Sebastián Rodríguez, Rubin, Y., Mulder, F., & Ossiannilsson, E. (2010). *Quality assurance of e-learning*. Helsinki: European Association for Quality Assurance in Higher Education.
- Hay, D. B., Kehoe, C., Miquel, M. E., Hatzipanagos, S., Kinchin, I. M., Keevil, S. F., & Lygo-Baker, S. (2008). Measuring the quality of e-learning. *British Journal of Educational Technology*, 39(6), 1037–1056.
- Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the strengths and difficulties questionnaire. *School Psychology Quarterly*, 22(3), 380–406.
- Hubbard, P. (1988). An integrated framework for CALL courseware evaluation. *CALICO Journal*, 6(2), 51–72.
- Jamieson, J., Chapelle, C. A., & Preiss, S. (2005). CALL evaluation by developers, a teacher, and students. *CALICO Journal*, 23(1), 1–46.
- Johnson, R. D., Hornikb, S., & Salas, E. (2008). An empirical examination of factors contributing to the creation of successful e-learning environments. *International Journal of Human-Computer Studies*, 66(5), 356–369.
- Karlsson, P., & Tägerud, Y. (2008). Bridging the gap – Taking the distance out of e-learning. *The Electronic Journal of E-Learning*, 6(1), 43–52.
- Koohang, A., Riley, L., & Smith, T. (2009). E-learning and constructivism: From theory to application. *Interdisciplinary Journal of E-Learning and Learning Objects*, 5, 91–109.
- Lantolf, J. P., & Pavlenko, A. (1995). Sociocultural theory and second language acquisition. *Annual Review of Applied Linguistics*, 15, 108–124.
- Liu, G.-Z., Liu, Z.-H., & Hwang, G.-J. (2011). Developing multi-dimensional evaluation criteria for English learning websites with university students and professors. *Computers & Education*, 56, 65–79.
- Lynn, M. J. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382–385.
- Patanakul, D. M. P. (2005). Standardized project management may increase development projects success. *International Journal of Project Management*, 23, 181–192.
- Puerto, F. G. d., & Gamboa, E. (2009). The evaluation of computer-mediated technology by second language teachers: Collaboration and interaction in CALL. *Educational media international*, 46(2), 137–152.
- Reinders, H. (2009). Teaching (with) technology: The scope and practice of teacher education for technology. *Prospect*, 24(3), 15–23.
- Rekkedal, T. (2008). Internet based e-learning, pedagogy and support systems. *Journal of the Australasian Law Teachers Association*, 1(1), 239–246.
- Robin, R. (2007). Commentary: Learner-based listening and technological authenticity. *Language Learning & Technology*, 11(1), 109–115.
- Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94–104.
- Santally, M. I. (2011). Quality assurance and open educational resources in online courseware development and delivery. *International Journal of Instructional Technology and Distance Learning*, 8(5), 81–86.
- Su, M.-h. M. (2005). A study of EFL technological and vocational college students' language learning strategies and their self-perceived English proficiency. *Electronic Journal of Foreign Language Teaching*, 2(1), 44–56.
- Sung, Y.-T., Chang, K.-E., & Yu, W.-C. (2011). Evaluating the reliability and impact of a quality assurance system for E-learning courseware. *Computers & Education*, 57, 1615–1627.
- Wang, Y., & Chen, N.-S. (2009). Criteria for evaluating synchronous learning management systems: Arguments from the distance language classroom. *Computer Assisted Language Learning*, 22(1), 1–18.
- Yang, Y.-T. C., & Chan, C.-Y. (2008). Comprehensive evaluation criteria for English learning websites using expert validity surveys. *Computers & Education*, 51, 403–422.
- Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, 9(2), 79–94.

Appendices

Appendix A

The interface the experts used to evaluate the criteria



*必填

1. The course provides the main purpose of the course *

It should belongs to

- E-learning for general course
- The content of English teaching
- The criteria of listening teachin
- The criteria of speaking teaching

Importance *

1 2 3 4 5

Strongly Disagree Strongly Agree

Suggestions and comments

2. The course demonstrates the evaluation criteria. (e.g. the course explain the rate of tests, assignment and participation.) *

It should belongs to

- E-learning for general course
- The Content of English teaching
- The Criteria of listening teaching
- The Criteria of speaking teaching

Importance *

1 2 3 4 5

Strongly Disagree Strongly Agree

Suggestions and comments

3. The course specifies all learning activities (e.g. The teaching outline shows the date that learners will be going to do; The assignment due date, mid-term date.) *

It should belongs to

Appendix B

The refined and finalized criteria for the English listening and speaking e-learning course

1) Information on E-learning Course

1. The course provides a target
2. The course demonstrates the evaluation criteria (e.g., the course explains the criteria for tests, assignments, and participation)
3. The course specifies all learning activities (e.g., the teaching syllabus shows the dates for tasks that the learners will complete, assignments, and the mid-term)
4. The teaching materials and activities match the teaching and learning goals
5. The teaching materials and the interface show the progress of each teaching theme
6. The instructor provides adequate teaching activities based on the course objectives
7. The teaching content provides examples to assist learners' comprehension
8. The teaching materials provide supplementary materials, such as Internet resources
9. The instructor provides at least 3/2 asynchronous and synchronous web-based courses
10. Stakeholders (both teachers and learners) can actively communicate in asynchronous or synchronous web-based courses
11. The instructor in an asynchronous web-based course replies to learners' questions on the forum within two days
12. The instructor offers fixed "online office hours"
13. The course provides an online tutor
14. The instructor provides 1/5 face-to-face classes per semester
15. More than 50% of learners respond to each discussion issue on the forum
16. The instructor uses a cooperative learning strategy in class (e.g., group discussion, group projects, or peer review)
17. The course requires the provision of learner information (e.g., learners' self-introductions, personal profiles)
18. The assignments match the teaching goals and teaching materials
19. The course provides online tests, self-evaluation, feedback, or response options
20. The assignments assist learners in organizing the key points and stimulate critical thinking
21. The instructor holds a face-to-face panel discussion with learners during the last class to improve future classes
22. The course provides the instructor's profile, including his/her professional background and contact information
23. The instructor asks learners to state their main ideas and allows them to demonstrate their comprehension in tasks
24. Learners are provided with opportunities to interact with different types of learning content
25. The organization of lessons and tests according to the degree of difficulty can be established based on learners' current progress
26. Learners are allowed to practice the assigned segmental patterns through tasks (e.g., a task that is separated into 1-3 parts is assigned; learners proceed step by step to complete the task)
27. Clarification is sought
28. Clues are provided for learners to answer questions
29. The instructor provides activities that increase learners' learning motivation (e.g., the instructor chooses well-known videos to include in the teaching material)

2) English Teaching

30. Learners' personal needs are met
31. The topic is related to daily experiences
32. The task occurs in the context of situational learning (e.g., the aviation industry)
33. Learners are provided with learning tips on listening and speaking and with explanations of common mistakes
34. The materials are varied in terms of learners' gender and age, dialect, accent, topic, speed, level, and genre
35. Learners are asked for their main ideas and are allowed to demonstrate their comprehension in a task
36. Learners respond physically to the instructor's commands (e.g., when an instruction is given in the target language, learners can respond to the command)
37. Various tasks are taught (e.g., jigsaw, minute mysteries, no overt response activities, short/long response activities)
38. Explanations are provided for why the use of communication strategies can be helpful
39. The instructor provides logically sequenced introductions to grammar references and explanations
40. Vocabulary is organized into well-defined thematic units
41. The teaching content is relevant and consistent
42. Learners are provided with opportunities to interact with different types of learning content
43. Appropriate feedback is provided, and correction relies on learners' own performance
44. The teaching content and tests are correlated

45. Topics are discussed extensively and supported by opinions and hypotheses
46. The benefits of using a listening strategy are explained
47. Cultural appropriateness is ensured (e.g., proper words and acceptable terms)

3) Teaching Listening

48. Choices of listening strategies depend on learners' learning styles (e.g., spatial, auditory, linguistic, kinesthetic, mathematical, interpersonal, and intrapersonal)
49. The listening comprehension activities are designed to guide learners to answer questions related to the activities
50. The content is combined with a series of pictures that facilitate learners' listening comprehension
51. Text, images, and sounds are appropriately coordinated to the situation (e.g., dining language is presented with suitable conversation, relevant pictures, and movies)
52. The intonation of the audio is appropriate
53. The situation in the audio is authentic
54. Different levels of listening materials are provided
55. The form of listeners' responses is carefully considered
56. The listening comprehension activity can increase learners' interaction
57. Who, what, and why questions about the content of the audio are discussed
58. Both bottom-up and top-down activities are included in listening activities
59. The language and content of the audio are used in authentic situations
60. Background information is provided about the content of the audio
61. Learners transfer some of the information they hear (e.g., draw a picture, write key words)
62. Learners outline or take notes about what they hear
63. Clear speaking instructions and examples are provided (e.g., a sound track is provided)
64. Intrinsically motivational approaches for listening are used
65. Written aids are provided for audio (e.g., all content is presented after listening)
66. Learning materials are provided with different accents (e.g., British, American, Australian, Canadian)

4) Teaching Speaking

67. Informal conversations are based on discourse that is genuine, improvised, or spontaneous
68. Learners answer questions about the message that they hear
69. Learners translate the message into their native language or repeat it verbatim
70. Learners simulate what they hear
71. Learners provide feedback on the effectiveness of the strategy used
72. Learners are able to talk about whatever they want
73. The focus is on fluency and accuracy
74. Intrinsically motivational techniques for speaking are used
75. The model or models for English pronunciation are selected in advance
76. Intelligibility is fostered during spontaneous speech
77. Pronunciation is taught one-on-one so the learner can practice repetition of the sounds
78. Learners are provided opportunities to use sounds in phrases and sentences (e.g., access the whole text selection at once instead of in pieces or word by word; use phrases and sentences, long words, or formulaic sequences; perform activities that provide opportunities for learners to communicate meaningfully)
79. Learners are provided with explicit instruction/explanation on how to pronounce new words and phrases and the rules for using them
80. Learners are exposed to a variety of spoken text types
81. Both transactional and interpersonal speaking activities are designed
82. The speaking activities are personalized
83. Activities progress from mechanical drills to open-ended responses
84. Activities require the integration of linguistic, sociolinguistic, and pragmatic skills
85. Speaking activities are contextualized
86. Various speaking activities are used (e.g., extemporaneous speaking, information gap and jigsaw activities, guided conversations and interview activities, role play activities)
87. Problem-solving activities are provided (e.g., take a large picture and cut it into pieces of equal size, depending on the number of participants; participants try to make the picture complete)
88. A variety of examples are provided to inform learners that there are other correct options
89. Learners are encouraged to overcome their initial reluctance to speak