

Testing the Surf: Criteria for Evaluating English Learning Websites

Chia-Ying Chan and Ya-Ting C. Yang

Institute of Education, National Cheng Kung University

u3692113@mail.ncku.edu.tw, yangyt@mail.ncku.edu.tw

ABSTRACT

Due to the importance of English learning and the significance of Computer Assisted Language Learning, more and more English learning websites and software have been developed to increase English learning opportunities. This study developed a set of evaluation criteria for English learning websites by conducting an expert validity survey. The finalized criteria have 48 items, including six categories—general information, general English learning, listening, speaking, reading, and writing. The ratings of the finalized criteria received Representativeness of 3.69, Importance of 3.57, Clarity of 3.79, Content Validity Index of .94, and Factorial Validity Index of .97.

1: INTRODUCTIONS

Along with the dramatic increase in websites, Internet World Stats [1] estimated that 1,023 million people use the Internet worldwide from all locations, representing 15.7% of the world's total population. However, the rapid proliferation of Internet resources has generated a large amount of useless information being developed in the world [2]. Similarly, the number of English learning websites is increasing to assist users in teaching/learning English, but this large number of English as Second Language (ESL) websites has made it extremely difficult for users to make the right choices. That is, it is doubtful that the quality of these websites has increased with their quantity. A good English website has to highlight the characteristics of learning and also meet learners' needs. Therefore, as applying technology or using websites on teaching and learning, one should consider the appropriateness of the media and the arrangement of the material for the students. Warschauer, Shetzer, and Meloni [3] mentioned that an unstructured website is a simple place for users to surf aimlessly with little direction. They emphasized that a good website not only provides users with a place to surf, but also helps users create some "waves" from surfing. With proper criteria, it will help students, educators, and other Internet users focus on valuable insights from the website and save users' time from searching unreliable information and websites [4], [2].

Thus, the quality of the websites is at present an important issue. One way to screen for good English learning websites is to develop a set of standard criteria that can be applied in evaluating the Internet

information sources. With these criteria, users can easily evaluate the quality and characteristics of the websites, and therefore decide whether the websites meet their needs. Specifically, teachers can evaluate English learning websites to screen them as teaching resources. With predetermined criteria, students are not likely to waste their time surfing unreliable or poorly designed sites. In addition, website designers can also take the evaluation criteria as a guideline or direction on how to build a reliable and well-designed English learning website.

Certain research studies have developed criteria to evaluate the information on general websites, such as the content, objectivity, currency, navigation, and authority [4]. These general criteria are the basis for creating a good website. However, a set of general criteria seems not enough to evaluate educational websites and their information. Clayton [5] developed a set of criteria for evaluating the quality of online courses. Each criterion was selected to identify specific course components, qualities, or procedures proven to be helpful to learners and/or instructors. A good learning website should further consider the aspects of learners' attributes, learning motivation, presentation of online resources, and interaction among users [6]. Learning websites generally aim to be teaching aids for instruction or self-learning materials for learners. Thus, the principles for developing the criteria should be based on learning theories and the characteristics of the content [7].

However, a review of the literature has shown that English learning is distinct from other subjects based on the attributes of language acquisition. Language learning is not only a subject matter, but also the acquisition of a culture, social rules, linguistic functions, and psychological reactions [8], [9]. In other words, a good speech is produced in the appropriate situation with the proper communicative manner. The meaning of a language is different when people say the same words in different social settings and with different psychological status. Because language learning is unique, the criteria to evaluate English learning websites may need to be designed specifically for English learning, and be distinctly different from other subjects. The critical points to evaluate English learning websites are emphasized in communication and situational settings [8] since the language generates different meanings in different situations with various people. However, from the review of research on evaluation criteria pertinent to the CALL environment, four reasons are synthesized to

explain the need for the development of evaluation criteria for English learning websites in this study. First, some studies focused on developing general guidance for evaluation, but were not specific enough to include the essential characteristics for English learning [10]. Second, many studies explored the conceptual ideas of evaluating English learning materials, but did not specifically provide detailed guidance for designing English learning websites [11]. Third, relatively few research studies have been conducted on the exploration of a complete set of criteria; that is, most research only focused on one or two of the language learning aspects, such as listening, speaking, or writing skills. Fourth, researchers in the social sciences study complex constructs for which valid and reliable measures are needed. In these circumstances, a content validity study should be conducted [12]. However, most of the developed criteria were only based on theoretical concepts, but not validated by the empirical study. In sum, the goal of this study is to develop a specific, complete, and validated set of evaluation criteria for English learning websites, so that teachers and web designers can use them to develop reliable and well designed websites and Learners can use them to acquire knowledge effectively and conveniently in the E-world.

2: METHOD

In order to achieve the above goal, the researcher conducted an expert validity survey to validate the preliminary criteria developed by the researchers according to a review of the literature.

2.1: SUBJECTS

A panel of experts, who were professors in the area of CALL programs from across Northern, Central, and Southern Taiwan, participated in this study. All selected participants had backgrounds in the educational technology and language teaching fields. The professors were from national and private universities, science and technology universities, and teachers colleges in Taiwan. Rubio et al. [12] stated that 6 to 20 participants is an adequate number of panel experts. In general, more experts will generate more information towards the measure. Thus, 17 experts were invited to validate the preliminary criteria in this study.

2.2: INSTRUMENT

The experts were asked to revise and rate each item of the preliminary criteria. There were four indexes for experts to use in evaluating each item: Factors, Representativeness, Importance, and Clarity (Table 1).

1. Factors: The researcher has assigned the item to a factor and the items are arranged by factors. The expert can either check the “yes, it belongs to factor #” box or “no, it should be assigned to factor #.” Experts can also leave other opinions in the box to further clarify their judgments.

2. Representativeness: Representativeness is used to determine whether the item or statement stands for the right factor. Experts can also provide a suggested revision or opinion under the rating. For example, a value of 1 indicates that the item is not representative of the corresponding factor. Also, experts can state reasons as to why a certain item might be more appropriate for another factor.
3. Importance: Importance represents whether the item is a crucial statement for the factor. Similarly, a rating of 1 indicates that the item is not important.
4. Clarity: Clarity represents whether the wording or sentences are clear enough for users to follow. Also, the experts can offer some recommendations about unclear statements. Through this step, the researcher can gather complete opinions towards the overall format.

2.3: DATA ANALYSIS

The purpose of conducting a content validity study is to analyze the measure to determine whether the expert validity survey in this study is valid. According to Rubio et al., three analyses need to be performed—Interrater Reliability (IR), Content Validity Index (CVI), and Factorial Validity Index (FVI).

First, Interrater Reliability (IR) is to determine how well the participants in the content validity study agree with each other based on their responses to both Representativeness and Clarity. Because the panel of experts consisted of 17 professors, Intraclass Correlation Coefficient (ICC) was adopted to compute the reliability among the raters in this study. Due to the purposes and participants in this study, the two-way mixed effected model with measures of consistency was selected for computing the statistics. The reasons involved in selecting this model included that the experts were not chosen at random to participate in the study. Furthermore, the purpose of IR is to measure whether the raters' scores are highly correlated. Barrett [13] stated that interrater/intraclass $r > .74$ is excellent, $.60 - .74$ is good, $.40 - .59$ is fair, and $< .40$ is poor. The IR for the evaluation criteria of English learning websites is $.72$, which means in good condition.

Second, Content Validity Index (CVI) is to determine if the item and the instrument as a whole are valid. This analysis is calculated based on the Representativeness responses of the measure. The CVI is calculated by counting the number of experts who rated the item as 3 or 4 and dividing that number by the total number of experts. The appropriate CVI should have a minimum of $.80$ [14]. The CVI for the measurement is calculated by estimating the average CVI across the items. The CVI for the evaluation criteria of English learning websites is $.90$.

The final analysis is the Factorial Validity Index (FVI), which will determine how each item and the instrument as a whole is associated with the appropriate Factors included in the content validity survey. The FVI is calculated by counting the number of experts who agreed with the item in the assigned factor and dividing

that number by the total number of experts. Each item has its FVI from individual expert's responses. Rubio et al. [12] recommended an FVI of at least .80. The FVI for the evaluation criteria of English learning websites is .97.

3: RESULTS AND DISCUSSIONS

3.1: CRITERIA FOR GENERAL INFORMATION

The criteria for general information and the basic rules to use in constructing websites can be used to evaluate the basic information provided on the English learning websites. The ratings of Representativeness, Importance, Clarity, CVI, and FVI are 3.51, 3.47, 3.72, .85, and .97 in the preliminary criteria and 3.60, 3.55, 3.70, .89, and .98 in the finalized criteria, respectively. Most of the ratings improved based on the experts' suggestions. Specifically, they showed an increase of 8% in Representativeness and Importance. The principles used to select the items were based on the ratings of Representativeness and Importance rather than on those of Clarity. This is the reason that the rating of Clarity decreased to .02 in this category of the finalized criteria.

In this category, four new items were added, seven less important items taken out, and six relevant items integrated. The four new items—the accuracy of the content, adaptive design, identification of copyright, and best resolution were added in this category. Correct and fluent wording and information on the website will lead users to easily acquire the resources they surfed for. Similarly, the adaptive design can successfully guide the target users to surf a series of information on the website. Gathering diverse resources online is easy; however, it is important to highlight the copyright of the documents to avoid improper or illegal application of the data. Besides, to specify the best resolution on the website can increase the readability for users surfing the webpage.

On the other side, the researcher took out seven relatively less representative and less important items (e.g. providing site map, search engines on the websites). The items about site map and search engines are less important because both ideas for evaluation depend on the sizes of the website. If the website is concise, it would not be necessary to contain the site map or search engine for the Intranet or the Internet.

Besides, the researcher integrated similar items into one item to make the criteria more concise. There were six items integrated into two items in this category. For example, four similar items, which included surfing the website easily, surfing in plain designs, suitable linkage from screen to screen, and classifying data by its characteristics are for evaluating whether the resources are placed in an organized format for users to browse efficiently. Thus, these items were combined into one item—*The design of the format is clear and easy to browse (i.e. there are explicit linking words, graphs, and website frameworks)*. As a result, there are a total of 13 items in this category.

3.2: CRITERIA FOR GENERAL ENGLISH LEARNING

The category of general English learning includes the comprehensive evaluation of criteria for the overall materials of English learning on the website. The ratings of Representativeness, Importance, Clarity, CVI, and FVI were 3.35, 3.27, 3.54, .83, and .88 in the preliminary criteria and 3.75, 3.51, 3.78, .95, and .97 in the finalized criteria, respectively. All of the ratings anticipatively increased in the finalized criteria as well as in the quality of evaluation criteria.

In this category, four new items were added, 11 items taken out, and six similar items integrated. Four new items were added: (a) the content design should be based on the learning topic; (b) progressive approaches should be applied to language learning skills; (c) comprehension tests should be included, and (d) the tests should be related to the topics. Users can effectively comprehend the learning materials when the content or information is designed based on topics. For example, the topic of Christmas accompanies Christmas pictures or presents relevant culture. Using progressive approaches for teaching language skills will gradually enhance users' language ability. That is, the reading section can contain pre-, during, and post-reading activities. The design of the comprehension tests can specifically focus on what learners acquired from a series of learning materials. Besides, the questions can guide learners to clarify the misconceptions and let them know what they perceived. However, including comprehension tests is not enough for explicitly examining learners' acquisition. The tests should be related to the learning topics.

In this category, 11 less important or representative items were taken out, such as topics based on daily experiences and designs for stimulating users' multi-sense organs. Because any topic is useful for learners in acquiring English, the item only emphasizing daily experience-based topics is not needed in every English learning website. The designs for stimulating users' multi-sense organs are too vague for explicitly evaluating the materials with this design. As evaluation criteria, more explicit descriptions will be better to guide users in evaluating the websites they surfed.

Six similar items were integrated into two items. For example, four items—the website should include enough text, pictures, video, and proper fitness for multimedia are used in evaluating the multimedia design on the website. The multimedia will be meaningful as they are designed to fit in with the content. Thus, the integrated item is *The content is accompanied by proper wording, videos, or images (e.g. the topic of dining is presented with conversation, pictures, and animation)*. The finalized criteria for general English learning turn out to be 13 items.

3.3: CRITERIA FOR LISTENING

Listening input will be acquired efficiently by users when it exists along with multimedia aids. For example, the visual context can enhance learners' comprehension, and the function of the accurate playback enables users simply to find the specific segment. Thus, listening resources, relevant intonation to a natural semantic situation, as well as appropriate multimedia applications are considered in this study. The ratings of Representativeness, Importance, Clarity, CVI, and FVI were 3.64, 3.48, 3.68, .92, and .96 in the preliminary criteria and 3.72, 3.59, 3.73, .97, and .96 in the finalized criteria, respectively. Most of the rating results increased from the preliminary criteria except that the FVI did not change (.96). Specifically, there was an increase of .11 in Importance. The quality of these criteria thus improved as well.

In this category, four less important or less representative items were taken out, three relevant items integrated, and no new ideas added. Concerning taking out items (e.g. users' attention spans, different levels of learning materials), although the learners' attention span for designing listening comprehension is desirable, it is hard to adaptively design for every learner on the website and the design should be based on the purposes of the website. Another item taken out concerned includes different levels of learning materials on the website. Because the criteria for general English learning have already contained the item of classifying materials according to different levels, this item was then taken out.

In the idea of integration (e.g. teaching aids of text, relevant pictures, and situational animation), these items are concerned with applying multimedia to assist in listening comprehension. Thus, they are integrated into one item as follows: *Multimedia-aided listening materials are provided to enhance students' comprehension (e.g. pictures, flash, or videos)*. In the end, the finalized criteria for listening include four items.

3.4: CRITERIA FOR SPEAKING

The focus of the speaking criteria is to provide appropriate examples and skills in discourse, as well as an oral device for users when communicating with each other. Speaking instruction is hard to implement well on a website due to the limited technology. In order to promote learners' oral skills, the criteria thus contain the example of the linkage of sound as well as online feedback information to use in evaluating the materials related to speaking. The ratings of Representativeness, Importance, Clarity, CVI, and FVI were 3.56, 3.53, 3.72, .89, and .96 in the preliminary criteria and 3.62, 3.55, 3.73, .92, and .96 in the finalized criteria, respectively. The finalized ratings were in the acceptable range.

In this category, five less representative and less important items were taken out and two new items added. The items taken out were good quality of voices over the Internet, examples for addressing a speech, and

explanation of colloquial skills on the website, etc. The first item, good quality of interactive discourse, should be attributed not only to the design of the website but also to the speed of the Internet. Therefore, this item is not directly relevant to the category of speaking. The second and third items are too specific and more appropriate for advanced oral training than are general English learning websites.

This category furthermore included two new items: observing other people's speaking tasks, and specifying needed hardware requirements. Learning will take place as learners construct knowledge by themselves. The design for users to observe other people's speaking tasks is the process of modeling. Learners will be aware of other people's merits or shortcomings in their performances; thus, learners might reconstruct their own knowledge about performing good speaking on similar topics or might discuss with peers about the shows. In the second new item, specifying the needed hardware requirements is important because speaking design online needs relevant programs to convey the voices well. Eventually, the finalized criteria for speaking contained seven items for evaluation.

3.5: CRITERIA FOR READING

The reading category includes evaluating the presentation of vocabulary, reading resources, and self-evaluation designs. The ratings of Representativeness, Importance, Clarity, CVI, and FVI were 3.47, 3.37, 3.78, .85, and .99 in the preliminary criteria and 3.69, 3.66, 3.86, .92, and 1.00 in the finalized criteria, respectively. The rating increased at least .22 in the criteria of both Representativeness and Importance.

In this category, six items were taken out, three items integrated, and no items added. For taking out (e.g. including the synonyms of new vocabulary, reading skills, and summary of the articles), first, the item of synonym is too specific to be an evaluation criterion in this study. In addition, the definition of synonym in this item is ambiguous because two kinds of synonym exist: morphology and semantics of the word. The experts have suggested that to state an example sentence for a word is more important than its synonym. Due to these reasons, although this item is important, it is not necessary. Second, the idea of including reading skills on the website has been taken out because a similar concept of presenting language skills is included in the criteria for general English learning. The importance of this item depends on the length or level of the writing to provide a summary of the articles. Thus, this item is not required for every situation and should be based on the purposes of the website.

For integration (e.g. including relevant resources, pictures, background designs for the reading), these three items include the relevant resources for readings promoting learners' comprehension. Thus, the researcher integrated these ideas into one item as follows: *Reading resources are provided to readers for*

better comprehension (e.g. multi-media aids, similar topics, and related links). As a result, the finalized reading criteria included five items.

3.6: CRITERIA FOR WRITING

Writing is an output process which contains the communication with relevant information or other learners to enhance one's inspiration for generating the production. The ratings of Representativeness, Importance, Clarity, CVI, and FVI were 3.52, 3.39, 3.77, .88, and .96 in the preliminary criteria and 3.81, 3.61, 3.90, .98, and .97 in the finalized criteria, respectively.

In this category, seven less representative and important items were taken out, and three relevant items were integrated. The items taken out were mostly not pertinent to the concept of writing, such as diverse fonts for typing and topic designs based on users' ages, and materials relevant to daily experiences. The flexible fonts for typing might only increase learners' interest in arranging different fonts, but not in their writing quality. Therefore, this idea would be more suitable as an option in writing courses. The second and third ideas for taking out were about the topic designs. According to the experts' suggestions, it is more important to design the topics based on the target learners, the purposes of the website, or learners' interests than their ages or daily experiences.

Besides this taking out of items, the researcher furthermore preliminary three similar items (e.g. relevant resources, vocabulary, and phrases of articles) into one criterion. Because the ideas were regarded as information relevant to the topics, the preliminary item was revised to: *Resources related to the writing topic are provided (e.g. vocabulary, phrase, or related concept)*. The finalized criteria for writing finally contained six items for use in evaluation.

3.7: FINALIZED CRITERIA

The ratings of the finalized criteria received Representativeness of 3.69, Importance of 3.57, Clarity of 3.79, CVI of .94, and FVI of .97. In the development of the evaluation criteria, to ameliorate the quality of each item in the individual categories, the same methods were followed: taking out less important items whose Representativeness or Importance might be below 3.50, revising unclear concepts of the items based on the experts' suggestions, and adding other items to the category which were recommended by the experts. Thus, the results of the finalized criteria mostly improved as well. A set of 48 criteria gathered in six different categories was listed as follows:

I. General information

1. The website includes the background information of the developer.
2. The website includes the contact information of the developer.
3. The information provided by the website coincides with the goals of the establishment.

4. The format design is clear and easy to browse (e.g. there are explicit links to words, graphs, and website frameworks).
5. There are no failed links (e.g. when the webpage is temporarily inaccessible, relevant information is provided for browsers).
6. When downloading files, the format and estimated duration is provided.
7. The information is objective without stereotypical or obvious discrimination.
8. The wording is accurate and fluent.
9. The website includes the date of the last revision, or informs users of how often the website is updated.
10. The website specifies the sources of the data.
11. The website states the copyright and limits of authority of the data.
12. The website is accessible, and employs adaptive designs (e.g. readability, consistent format).
13. The best resolution is specified.

II. General English learning

14. The website includes the background information of the developer.
15. The content is designed based on English learning topics (e.g. the topic of Christmas is accompanied with relevant cultural information).
16. Various learning methods and opportunities are provided (e.g. online tests, discussion forums, listening practice).
17. The learning content is rich and diverse (e.g. provides a wide variety of learning themes).
18. The content is accompanied by proper wording, videos, or images (e.g. the topic of dining is presented with conversation, pictures, and animation).
19. The learning activities are relevant (e.g. the topic of Disney cartoons: vocabulary → read story → evaluation).
20. The learning methods follow a gradual progress (e.g. the reading activities involve pre-, during- and post-reading activities).
21. The materials are graded (e.g. for novice learner or intermediate learner).
22. The learning materials are relevant to the objective of the website (e.g. advanced websites contain international news; introductory websites contain everyday events).
23. Frequently asked questions are answered.
24. Learning strategies are provided (e.g. when reading, the meaning can be inferred from the context).
25. Online dictionary or website links are provided.
26. To help learners understand their learning efficacy, the website includes comprehension tests.

III. Listening

27. The content of the listening comprehension is in an authentic context (e.g. conversation, news, story, or academic speeches).
28. In listening comprehension, the speech intonation is appropriate.
29. In listening comprehension, the pronunciation is recognized by most English speakers.
30. Multimedia-aided listening materials are provided (e.g. picture, flash, or video)

IV. Speaking

31. In connection to speaking, the website specifies the needed hardware requirements.
32. Authentic examples of sound links are adequately provided.
33. The strategy of appropriate usage of tone is provided.
34. Examples of interactive conversation are provided.
35. In the speech design, online learners can communicate in English.

- 36. Users can listen to the spoken word of other learners.
- 37. Online feedback is given based on the recorded work of the users (e.g. users' frequent mistakes in pronunciation.)

V. Reading

- 38. New vocabulary in an article is highlighted with special effects (e.g. different colors or fonts are used for new words).
- 39. Articles are categorized based on their characteristics (e.g. according to the difficulty or topic).
- 40. Through various interesting tasks, users can self-evaluate their reading ability. (e.g. cloze test, multiple choice, and crossword).
- 41. For new vocabulary, definitions and explanation are provided.
- 42. Reading resources are provided for readers (e.g. multi-media aids, similar topics, and related links)

VI. Writing

- 43. Readers are encouraged to communicate in English on the discussion board.
- 44. Users can discuss the composition with online advisers (e.g. through email or discussion forums).
- 45. A guided composition activity is provided (e.g. writing with pictures or filling in the dialogue of a comic).
- 46. Examples of various articles and writing styles are provided (e.g. exposition and description).

- 47. Resources related to the writing topic are provided (e.g. vocabulary, phrase, or related concepts).
- 48. Users can view work from other peers, famous writers, newspapers, and magazines

4: CONCLUSION

This study, serving as a cornerstone, hopes to stimulate further studies related to the evaluation of educational technology, in particular conception and multi-disciplinary approaches in the evaluation of language learning websites. Finally, teachers in practice can apply the comprehensive characteristics of English learning websites to build their own ones, to guide users to screen for reliable and well-designed websites from among the many websites, and to reorganize their teaching focuses and approaches to various language skills. It is hoped that students' English proficiency will increase with comprehensive guidance and sufficient technology following the trend of English learning at present.

Table 1 *Exemplification of Survey for Panel Experts*

Items	Factors	Representative				Importance			Clarity				
	1. General website evaluation; 2. English overview; 3. Listening; 4. Speaking; 5. Reading; 6. Writing	Not at all	Major	Minor	Good	Not at all	unimportant	importance	Very	Not at all	Major	Minor	Good
		1	2	3	4	1	2	3	4	1	2	3	4
1. <u>Item 1</u> Users can link to information about the sponsor and author	<input type="checkbox"/> yes, it belongs to factor 1 <input type="checkbox"/> no, it should belong to _____ Comments:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	Comments:	Comments:				Comments:			Comments:				

REFERENCES

[1] Internet World Stats. (2006). Internet usage statistics—The big picture. Retrieved July 20, 2006 from <http://www.internetworldstats.com/stats.htm>

[2] Wilkinson, G. L., Bennett, L.T., & Oliver, K.M. (1997). Evaluation criteria and indicators of quality for Internet resources. *Educational Technology*, 37(3), 52-59.

[3] Warschauer, M., Shetzer, H., & Meloni, C. (2000). *Internet for English teaching*. Alexandria, VA: TESOL Publications.

[4] Laura, G. (1999). Evaluating net evaluators. *Searcher*, 7(2), 57-66.

[5] Clayton, R. W. (2006). Instructional media & design. Criteria for Evaluating the Quality of Online Courses. Retrieved July 20, 2006 from <http://www.imd.macewan.ca/imd/content.php?contentid=36>

[6] Furner, J. M., & Daigle, D. (2004). The educational software/website effectiveness survey. *Int'l J of instructional Media*, 31(1), 61-77.

[7] Najjar, L. J. (1998). Principles of educational multimedia user interface design. *Human Factors*, 40(2), 311-323.

[8] Davies, G. (2001). New technologies and language learning: A suitable subject for research? In A. Chambers & G. Davies (Eds.), *ICT and language learning: A European perspective* (pp.13-27). Lisse: Swets & Zeitlinger.

[9] Gardner, H. (1985) *Social psychology and language learning: The role of attitudes and motivation*. London: Edward Arnold.

[10] Comer, P., & Geissler, C. (1998). A methodology for software evaluation. *SITE 98: Society for information technology & teacher education*, Washington, DC.

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

[12] Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E.S., & Ruch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104.

[13] Barrett, P. (2001). Assessing the reliability of rating data. Retrieved July 20, 2006 from <http://www.pbarrett.net/rater.pdf>

[14] Davis, L. (1992). Instrument review: Getting the most from your panel of experts. *Applied Nursing Research: ANR*, 5, 194-197.