

# Construction and Validation of Self-Assessment Inventory for English for Academic Purposes: A Case of Iranian Tertiary Students

*Goudarz Alibakhshi*

Allameh Tabataba'i University, alibakhshi@atu.ac.ir

*Received: 17/01/2014*

*Accepted: 26/03/2014*

## **Abstract**

Self-assessment helps learners to acquire self-confidence and gives them insightful feedbacks. There is no valid self-assessment inventory for learners of English for academic purposes (EAP). This study aimed at developing an academically appropriate instrument to measure tertiary students' English in academic situations. The study used both qualitative and quantitative methods in developing the self-assessment inventory (SAI). Sixty students were interviewed for the qualitative part of the study. For the quantitative phase, 600 students were selected through multistage random sampling. Qualitative data were analyzed by a constant comparative method. In addition, a questionnaire survey was used to assess the validity and reliability of SAI. A 41-item SAI was developed, the content validity, construct validity, and reliability of which were ensured by expert review, factor analysis, and internal consistency. This scale constitutes a validated self-assessment inventory of EAP.

**Keywords:** Self-Assessment Inventory; English for Academic Purposes (EAP); Language Assessment; L2 learners

## **1. Introduction**

The presents directions in teaching approaches which are believed to be learner-centered oriented and a great focus placed on the issues of authenticity and interactiveness (Bachman & Palmer, 1996) have contributed to increasing use of self-assessment in L2 learning and teaching process (Bachman, 2000; Hamayan, 1995). Great interest in involving the learners in all phases of the learning process and encouraging them to be autonomous and make decisions in the language classroom (Blanche & Merino, 1989) are the main causes of rapid growth in the use of self-assessment. To put in the words of Oscarson (1989), the use of self-assessment was assumed to be promising by many, especially for formative assessment purposes.

It is assumed that self-assessment evokes great sophistication in learner awareness, increases their self-confidence, helps them to acquire an evaluation perception which can cover the whole learning process; and see errors and mistakes

as something helpful and natural. It was also viewed to be potentially useful to teachers, providing information on learning and teaching styles, on educational areas which may need remediation and insightful feedback on teaching (Barbot, 1991).

As Alderson and Banerjee (2001) believe, the findings of a majority of the empirical studies on the effectiveness and applicability of self-assessment have yielded varying outcomes which have put test administrators and teachers doubtful about using self-assessment in the assessment and evaluation of learners' L2 ability. Such doubt is mainly because of concerns that learners may not be able to provide accurate judgments of their achievement and proficiency. As an instance, Blue (1988), although believing that self-assessment can play an important role in self-directed learning and that learners can have vital roles in the assessment of their language learning process, argues that learners cannot self-assess their progress in language without teachers' assistance. Viewing and analyzing self-assessment data gathered from students on a preseasonal English for academic purposes (EAP) program, he concludes that there is not a significant correlation between teachers' assessments of the students and their own self-assessments.

However, despite the abovementioned skepticism and concerns, investigation in self-assessment has not stopped, and since 1980s, research in this area has been concerned with the development of self-assessment instruments and their validation (e.g., Lewkowicz & Moon, 1985; Oscarson, 1984). For example, Hargan (1994) also introduces the application of a "do-it-yourself instrument" for placement purposes, and reports that such an instrument results in much the same placement levels as suggested by a traditional multiple-choice test. She believes that such an instrument might be effective in redressing the emphasis laid on grammar and stem the neglect of reading and writing skills in the classroom.

The review of the research methodology of the previous similar studies indicates that using correlation coefficient approaches for investigating the relationship between the self-assessment measures and one or more external measures of student performance has always been a main approach to validating self-assessment instruments to obtain concurrent validity statistics (e.g., Ross, 1998; Shameem, 1998). Multitrait multimethod (MTMM) designs, factor analysis (Bachman & Palmer, 1989) and a split-ballot technique (Heilenman, 1990) were the other frequently used approaches to validating self-assessment measures.

As the review of literature indicates, the self-assessment scales developed since 1980s did not involve L2 learners in designing and developing processes. Additionally, to the researcher's best of knowledge, in the context of the present study, no appropriate scale has been developed for L2 learners at tertiary level to assess EAP, despite the fact that tests of English for general purposes are different from tests of English for specific/academic purposes (Douglas, 2000; 2001; Douglas

& Selinker, 1992). The similar works of self-assessment are all directing attention to English for general academic purposes, whereas EAP and tests of EAP are of much concern to L2 learners at tertiary levels particularly those trying to enter universities in English speaking countries to accomplish their higher education degrees.

Due to the fact that L1 speakers of languages other than English wish to undertake their undergraduate or postgraduate education in an English speaking country or at a university in their own country where entire programs or courses are offered in English (Evans & Green, 2007; Turner, 2007), and because EAP is seen as a necessary adjunct to the academic success of individual students, it is of much importance to EAP teachers to use validated assessment instruments. As EAP learners may need a variety of language skills and subskills, it seems that teachers can not develop a self-assessment instrument without involving the learners. Moreover, the results of a couple of empirical studies verify that when learners utilize development-oriented self-assessment, there is increased autonomy and productivity, higher retention rates, higher motivation, and less frustration among learners (Gardner & McIntyre, 1991; Dickinson, 1987; Ellis, 1994; McNamara & Deane, 1995; O'Malley & Pierce, 1996; Oscarson, 1989; Peirce, Swain, & Hart, 1993; Rivers, 2001). Therefore, a self-assessment scale for learners of EAP is really of much significance. The main objective of the present is to develop a self-assessment inventory which involves language skills and components which EAP learners may need to meet their English academic needs.

## 2. Literature Review

A number of studies in the 1980s were concerned with the development of self-assessment instruments and their validation (e.g., Lewkowicz & Moon, 1985; Oscarso, 1997; Stefani, 1998; Taras, 2001, 2002). As a result, many approaches including pupil progress cards, learning diaries, log books, rating scales, and questionnaires were developed (Dearing, 1997; Falchikov, 1997). It was found that self-assessment helps learners to become autonomous and to be aware of their learning and reflect on their development (Boud, 2000; Freeman & Lewis, 1998).

Carton (1993) has discussed how self-assessment can become part of the learning process. He has described his use of questionnaires to encourage learners to reflect on their learning objective and preferred modes of learning. He has also presented an approach to monitoring learning that involves the learners in devising their own criteria; an approach that he believes helps learners to become more aware of their own cognitive processes.

Accordingly, Butler and Li (2001) have investigated the effectiveness of self-assessment among young EFL learners. They found some positive effects of self-assessment on the students' English performance as well as their confidence in learning English, though the effect sizes were small. The study also found that

teachers' and students' perception of the effectiveness of self-assessment are different depending on their teaching/learning contexts. A number of interesting insights were also discovered through interviews with teachers. The teachers were asked about the best way to utilize self-assessment as part of foreign language instruction in contexts wherein teacher-centered teaching has been traditionally valued.

McDonald and Boud (2003) found that when learners assess their own learning, their learning will be promoted to a high extent. Higgins, Hartley, and Skelton (2001) have mentioned that for the development of self-regulation in learners both teachers and learners' feedback on the learning process are needed and that this ability to give feedback must be promoted in both.

In the same vein, Alderson (2005) has investigated the importance of self-evaluation in the L2 class today, and stressed the advantages of having students keep a regular journal. Taking the methodological framework offered by the communicative approach to language teaching as a starting point, the dynamic interdependence of purpose, methodology and evaluation within the curriculum were studied. In this sense, formative or ongoing evaluation becomes one of the most practical assessment techniques for controlling our students' progress as well as the effectiveness of our teaching program. Self-evaluation has a number of additional advantages regarding both the students' affective implication in assessing their own learning processes and their participation in class management.

Some researchers have given some recommendations for doing a better self-assessment. Lejk and Wyvill (2001), for example, recommend a holistic approach rather than a category-based approach. Other studies, such as Blatchford (1997), have concluded that there is a significant association between self-assessments and attainments in both English/reading and mathematics. Taras (2001) reported that the active participation of learners and the teachers' experience will enhance the process of self-assessment.

The review of the related studies shows that self-assessment has been studied from several different perspectives. Fallows and Balasubramanyan (2001), for example, have reported that compulsory training combined with multiple ratings offer many benefits. In addition, motivation plays a significant role in the accuracy of self-assessment. AlFallay (2005) has concluded that those learners who have integrative motivation do more accurately in assessing themselves than those with instrumental motivation. Moreover, AlFallay has pointed out that language proficiency also influences the accuracy of self-assessment inducing that those with higher proficiency were more accurate than those with lower proficiency. He found that high proficiency learners to some extent underestimate their performances, whereas the lower proficiency ones often overestimate their performances.

Reliability of self-assessment has also been controversial among the related studies. For instance, in a comparison of a test of Dutch as an L2 for adult learners and a self-assessed version of the same test, Janssen-van-Dieten (1989) has found the self-assessed version was less reliable although earlier studies and her pilot studies had been more encouraging. Thomson (1996), in studying learners of Japanese as a foreign language, also felt very positive about using self-assessment despite finding considerable diversity in the accuracy of self-making.

Other studies have reported that self-assessment is reliable. Bachman and Palmer (1989), for example, found that members of a multilingual, multicultural group of adult learners of English as a foreign language in the US were able to reliably self-rate themselves for their communicative language abilities. Another example of success with self-assessment has been reported in Blanche's (1990) study wherein the ability of a group of adult learners of French as a foreign language to estimate their own speaking ability was investigated. He concludes that "the overall accuracy of the self-evaluation... is impressive" (Blanche 1990, p. 226). Variability in sample size, age of subjects, cultural/educational backgrounds, L2, the test format, the education context, and the criteria against which self-assessment is compared all affect reliability. What is comforting is that even when the results do not turn to have reliability, researchers maintain the value of self-assessment. One way to validate individual self-assessments is to have teachers randomly check some of the results. This would encourage learners to be honest and realistic in their self-assessment and would contribute to accreditation.

Xiao and Lucking (2008) examined the validity and reliability of student generated assessment scores. The findings indicated that the validity and reliability of student generated rating scores were extremely high. AlFallay (2005) has also investigated the role of some selected psychological and personality traits of learners of English as a foreign language for accuracy in their self- and peer-assessments. The study also showed that long periods of practice and sufficient feedback have a positive effect on the accuracy of self-assessment. He also maintained that students with low self-esteem are the most accurate in assessing their performance, whereas learners with instrumental motivation are the least accurate (AlFallay, 2005).

Diaska and Krekeler (2008) investigated the reliability of self-assessments of pronunciation skills and attempted to understand the causes of difficulties. In this study, 46 advanced learners of German assessed their own articulation of different speech sounds in comparison with the sounds produced by a native speaker. In 85% of all cases, the assessments of the raters and the self-assessments were identical. However, the learners only identified half of the number of speech sounds which the raters believed to be inaccurate. The study therefore concluded that even

experienced L2 learners seem to find it difficult to self-assess correctly their pronunciation skills.

Oscarson (1997) summed up the progress in the area of self-assessment by reminding us that the research in self-assessments is fairly new. He concludes that there are still many problems remained. For instance, learner goals and interpretations need to be adjusted with external necessities. Also, self-assessment is not self-explanatory; it must be introduced slowly, and learners need to be guided and supported in their use of the instruments.

### **3. Method**

This study used both qualitative and quantitative methods in constructing the self-assessment tool. Different phases of data collection and analysis are detailed as follows.

#### ***3.1 Phase I***

The initial phase of the development of the instrument consisted of three steps. First, the researchers conducted interviews and analyzed the contents of the interviews. Sixty postgraduate student volunteers were interviewed to explore the language skills and components which they may need in L2 use situations. The students signed up to be interviewed after the researcher requested volunteers for the research project through announcements in their own classrooms. Open-ended interviews were used to gather data on the participants' perceptions, experiences, and attitudes towards language skills and components which need to be assessed. The researcher interviewed five students majoring in Persian literature, social sciences, chemistry, mechanical engineering, civil engineering, electricity, biology, and geography from each field of study. Data saturation was reached with the 40<sup>th</sup> participant.

##### ***3.1.1 Ethical Considerations***

Before interviewing the students, the researcher informed the students of the purpose of the research and obtained written consent from each participant. The researchers also obtained the participants' permission to audiotape each interview for the purpose of content analysis and audit trail.

##### ***3.1.2 Step I: Interview Process***

Sixty students from different majors were interviewed. The interviews were conducted in an unstructured manner. Each interview began by asking the interviewees to express their experience of language use in their academic careers. The interviews began with the question: "Why do you think English language is needed for postgraduate students?" This was followed by questions designed to elicit specific language needs. The participants were also asked to describe what they used the English language for. The following are examples of prompts used to obtain further details from participants: "What are the main language skills you need

in L2 use situations?”, “can you tell me more about that?”, and “is there anything more you would like to tell me?” the interviews lasted between 30 and 60 min.

### **3.1.3 Step II: Content Analysis and Instrument Drafting**

The results of the interviews were immediately transcribed verbatim. The data were analyzed by a modified version of the constant comparative method (Glaser & Strauss, 1967; Lincoln & Guba, 1985). The constant comparative method consists of four steps: (1) comparing incidents applicable to each category, (2) integrating categories and their properties, (3) delimiting the construction, and (4) writing the construction (Glaser & Strauss, 1967). In the first step, meaningful descriptions of language needs were identified, coded, and compared; then, a preliminary classification into categories was performed. The analyzed data from each interview were constantly compared to each other in order to fit them into the most appropriate category. In the second stage, descriptions of language needs were compared to the preliminary version of the definitions of the various categories. The descriptions were resorted to increase the coherence of each category. Stage three was completed when the redundant categories had either been deleted or integrated into existing categories. The data collection and analysis were conducted until the data saturation was reached.

With the concepts from the review of literature on EAP, the needs analysis, and the interviews with the students, the researcher used the analysis of the student interview data to construct a self-assessment instrument. The instrument used a 5-point Likert scale: 1 (*very weak*), 2 (*weak*), 3 (*average*), 4 (*well*), and 5 (*very well*).

## **3.2 Phase II**

### **3.2.1 Step I: Assessing Content Validity**

In step 1, the 55-item self-assessment instrument was submitted to a panel of eight experts, including four experts from the field of applied linguistics and four instructors who had taught EAP to postgraduate students at Iranian universities in order to evaluate its clarity and relevance. A 5-point scale was used. The panel determined whether each item was appropriate, accurate, and representative: 1 = irrelevant and should be deleted, 2 = seemingly relevant but large-scale revision was required, 3 = relevant but in need of small adjustments, and 4 = relevant, clear and precise. The instrument was, then, revised accordingly.

### **3.2.2 Item Wording and Order Revisions**

For the sake of readability and clarity, revisions were made to eliminate awkward wording in some items. Then, the experts categorized the items into four main categories, each self-assessing one specific language skill. After revising the wording, content, and item order, the original 55 items became 47 items.



### 3.2.3 Step II: Assessing Face Validity

The language of the instrument was, then, reviewed for clarity by 20 students who were interviewed for the study. During the process of completing the survey, the students did not express any problems in understanding the item wording and meaning, demonstrating its face validity.

### 3.2.4 Step III: Data Analysis

A total of 600 postgraduate students of 30 majors (20 per major), studying at 10 different state universities in Iran, were selected through multistage sampling and given the tool. The participants were selected through convenience sampling. The students were also informed that they did not have to fill out the questionnaire, and that there would be no punishment for those who did not return the questionnaires. The questionnaires were anonymous, and the participants' consent was obtained on a tear-off form. The return rate was 85%. The data from the 510 returned questionnaires were analyzed using the SPSS (2000). Internal consistency was measured by Cronbach's alpha coefficient. Individual item characteristics were obtained by corrected average interitem correlation coefficients, interitem correlations, and item-to-total correlations. After deleting the items with an item-to-total correlation of less than .3, 41 items were left. The 41-item self-assessment inventory had a Cronbach's alpha of .9.

Table 1. *The 47 Items of the Self-Assessment Inventory (SAI)*

---

How would you assess your English ability in doing the following tasks? Give each item a rating from 1 to 5:

1 (*very weak*), 2 (*weak*), 3 (*average*), 4 (*well*), 5 (*very well*)

1. Understanding what university staff say in face to face conversation
  2. Understanding TV and radio programs related to general issues of universities
  3. Understanding TV and radio programs related to your own field of study
  - 4.\*Comprehending different types of lectures in nonacademic situations
  - 5.\*Understanding less frequently used idioms and expressions in teachers lectures
  6. Comprehending the main points in daily conversations
  7. Understanding teachers' speech without appealing for repetition
  8. Ability to retain the main points of teachers' lectures
  9. Comprehending the lectures by teachers of your own majors
  - 10.\*Comprehending lectures about nonspecific fields of the study
  11. Understanding general academic lectures in both American and British accent
  12. Repeating the main points of lectures to the students who may have problem in understanding
  13. Conveying our message to the discourse community members
  14. Contributing to all discussions in academic situations including teachers and university staff
  15. Presenting a short report of academic assignments
-



- 
16. Accurate use of structures while speaking
  17. Accurate pronunciation while speaking
  18. Conveying what you mean to your classmates
  19. Discussing the topics of your own fields with your classmates
  20. Presenting lectures in international conferences
  - 21.\*Answering all the questions raised in conferences
  22. Talking about technical issues without having a look at the notes
  - 23.\*Talking about general issues
  24. Explaining your demographic characteristics
  25. Talking about general issues of the universities
  - 26.\*Discussing the main social problems of your country with people
  27. Taking notes from the texts
  28. Taking notes from teachers' lectures
  29. Writing research papers
  30. Summarizing the texts
  31. Writing different academic letters
  32. Preparing research reports
  33. Writing research proposal
  34. Correct spelling and punctuation in writing
  35. Writing fluently and accurately
  36. Using appropriate diction
  37. Using appropriate cohesive devices
  38. Understanding the titles and headlines of famous newspapers
  39. Understanding scientific papers of your field of study
  40. Reading and understanding necessary forms such as application and registration ones
  41. Understating main ideas of texts easily
  42. Reading general texts due to high range of vocabulary
  43. Reading technical texts because of high range of vocabulary
  44. Understanding the erroneous parts of papers written by the other authors
  45. Translating into your native language
  46. Complete understanding of technical books and skimming skills
  47. Scanning

---

\*Deleted due to average interitem correlations less than .3

### **3.2.5 Construct Validity**

Using the SPSS, the researcher, then, examined the 41-item SAI, using exploratory factor analysis with principal component and varimax rotation. Four factors emerged, accounting for a total of 63.19% of the variance (see Table 2). Each factor is described in detail in the following parts:

Table 2. *The 41-item Version of the Self-Assessment Inventory (SAI):  
Exploratory Factor Analysis With Principal Component  
and Varimax Rotation*

Items	Factor 1	Factor 2	Factor 3	Factor 4
1	0.7	-0.3	0.12	0.12
2	0.64	0.2	0.25	0.15
3	0.8	0.12	0.23	-0.12
6	0.5	0.2	0.19	-0.35
7	0.64	0.38	-0.24	0.22
8	0.6	0.39	-0.32	0.28
9	0.82	0.11	0.22	0.38
11	0.64	-0.23	-0.37	0.33
12	0.6	-0.19	-0.2	0.23
13	-0.28	0.79	0.08	0.18
14	-0.11	0.79	-0.09	-0.31
15	0.21	0.79	0.23	0.12
16	0.13	0.67	0.14	0.16
17	-0.12	0.70	-0.13	-0.19
18	0.12	0.79	0.11	0.18
19	0.23	0.75	0.09	0.29
20	0.39	0.75	0.19	0.33
22	0.24	0.75	0.14	0.34
24	0.23	0.85	0.22	-0.22
25	0.18	.67	0.31	-0.28
27	-0.08	0.28	.83	-0.28
28	-0.09	0.11	.83	0.01
29	0.11	0.21	.79	0.04
30	0.03	0.23	.79	0.13
31	-0.02	-0.22	.79	-0.12
32	0.12	0.32	.73	0.12
33	0.13	0.33	.72	0.23
34	0.29	0.09	.69	0.19
35	0.14	0.22	.65	0.14
36	0.33	0.23	.55	0.13
37	0.28	0.18	.55	0.19

38	-0.22	0.24	-0.18	.80
39	-0.11	-0.13	-0.21	.85
40	-0.21	0.27	-0.28	.75
41	0.13	0.23	-0.32	.85
42	-0.12	-0.16	0.22	.75
43	-0.12	0.15	-0.18	.85
44	0.23	-0.23	0.33	.75
45	0.17	0.09	0.38	.65
46	0.34	-0.14	0.22	.65
47	0.33	-0.29	0.13	.85
Eigenvalue	17.81	3.738	2.14	1.46
Per cent of Total Variance Explained	44.5	9.34	5.65	3.63

## 4. Results

### 4.1 Factor One: Listening for Academic Purposes

This factor involves items which allow students to self-assess their listening comprehension ability in academic and nonacademic situations. The Eigenvalue of this factor is 17.81 and explains about 44.4% of the total variance. It consists of 9 items. The highest and lowest factor loading are .82, and .57, respectively. Item 9 had the highest factor loading (.82), followed by item 3 (.8). Understanding the main points in daily conversations has the third highest loading factor (.75). Items 3, 7, and 11 had the same factor loading (.64), followed by items 12 and 6 which had the lowest factor loading (.6).

### 4.2 Factor Two: Speaking for Academic Purposes

This factor gauges the learners' ability in using English to do speaking tasks in an academic situation. The initial Eigenvalue of this factor is 3.73 and explains about 9.34% of the total variance. It consists of 11 items. As the results of factor analysis in Table 2 show, item 24 had the highest loading factor (.85), followed by items 13, 14, 15, 18 (all .79), and 19, 20, 22 (all .75). Items 16 and 17 had the lowest loading factor (both .67).

### 4.3 Factor Three: Writing for Academic Purposes

The third dimension of this assessment scale deals with the items which measure the learners' ability in the use of academic English to do writing tasks in a classroom as well as in an academic situation. The initial Eigen value of this dimension was 2.14 and explained about 5.65% of the variance. Taking notes from the technical texts (item 27) and from the teachers' lectures (28) had the highest

factor loading (.83); items 29, 30, and 31 had the next highest loading factor (all .79). Items 36 and 37 had the lowest factor (each .55).

#### **4.4 Factor Four: Reading for Academic Purposes**

The last factor of this assessment scale includes the items which measure the tertiary level learners' reading skill in academic and target language use (TLU) situations. The initial Eigenvalue of this dimension was 1.46 and explained about 3.63% of the variance. It consists of 10 items. Items 47, 39, 41, and 43 had the highest factor loadings (all 0.85). The next item 38 has the factor loading of 0.80.

### **5. Discussion**

The conceptual and theoretical basis of the present study was provided by the works of assessment experts (e.g., Hargan, 1994; Lewkowicz & Moon, 1985; McNamara & Deane, 1995; Oscarson, 1984; Rivers, 2001) who believe in the effectiveness of self-assessment in educational settings. The SAI for academic English was constructed through a step-by-step development process. It can meaningfully evaluate L2 learners' EAP. The comparison of this instrument with those of the other researchers yields a number of important divergences that reflect differing teaching and testing procedures in EAP programs.

A key difference of the assessment in EAP programs lies in the definition of EAP and the learners' divergent needs in an EAP course. Due to these apparent divergences, an assessment inventory for general English cannot be used in an EAP context. Moreover, it seems that, due to particular needs of each individual learner, the standardized English tests may partially cover the contents of an EAP programs. This newly developed scale can be used by L2 learners at tertiary levels because its dimensions and items were constructed on the basis of a meticulous needs analysis. It also measures the domains of EAP in a TLU situations. The contents of this scale are all in line with Alderson's (2005) and Farhady's (2008) findings of need analysis research project. It consists of four main dimensions, each including some specific items.

The first dimension includes the items which invite L2 learners to assess their listening abilities in TLU situations. These items include L2 learners' reflections on their ability to understand the lectures by their teachers, the conversation between the learners and their classmates, the staff of universities, understanding the lectures at international conferences, comprehending radio and TV programs on general academic issues, and technical academic problems. It could be argued that that these listening tasks are the main ones needed by a tertiary level L2 learner in an academic TLU situation. The results of the present study are in line with Farhady (2008) and Alerson (2005).

The second dimension of this scale represents the main speaking tasks encountered by a university student in TLU situation. These tasks are ranked on the basis of their loading factor. These items invite L2 learners to self-assess their speaking ability including giving demographic explanation, contributing to academic discussions, arguing for or against the raised academic issues, presenting short research reports, giving a lecture at conferences, answering the raised questions at conferences, as well as accurate use of English structures and pronunciation while speaking English either in classroom in TLU situations. The items of this dimension are in line with the findings of the needs analysis project by Farhady (2008).

The third dimension of this instrument consists of the most frequently reported writing tasks needed by tertiary L2 learners in academic situations. As the results indicate, L2 learners need to self-assess their ability in writing in terms of their ability in note taking, writing research papers and research proposals, writing grammatically appropriate spelling and punctuation, writing research reports, and using appropriate diction in their writing. Therefore, it could be strongly mentioned that these items constitute the construct of the writing ability EAP learners. The learners' perceptions of their abilities in doing the writing tasks represented in Table 2 could be a good indicator of their writing ability in academic situations.

The last dimension of this scale contains the items which constitute the construct of reading tests for university students. The content analysis of the items of this dimension indicates skimming and scanning abilities, understanding technical texts, research papers, reading headlines of newspapers, reading general texts, understanding the erroneous parts of the other research papers, and the ability to translate from L2 into L1 are the main building blocks of reading comprehension construct. Therefore, in order to have a construct relevant reading test, tasks like the ones in this scale should be constructed.

In sum, due to the differences between tests of English for general purposes and EAP tests, general English self-assessment inventories cannot be used by tertiary level learners to self-assess their abilities in using English to accomplish their tasks. This is mainly because of the fact that if language tests do not include academic situations, they will certainly lack construct validity (Alibakhshi, Kiani, & Akbari, 2010). The inclusion of tasks representing what L2 learners need to know in assessment scales will certainly enhance the validity, authenticity, and generalization validity of the tests. Therefore, it could be strongly argued that the items of this scale will, to a great extent, assess the language components which L2 learners may need to know in an academic situation. Moreover, this scale can be used by L2 learners to assess their ability to English language skills and components and diagnose their problems.

### References

- Alderson, C.A. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. Continuum, New York, NY.
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part I). *Language Teaching*, 34, 213-236.
- AlFallay, I. (2005). The role of some selected psychological and personality traits of the rater in the accuracy of self- and peer-assessment. *System*, 32(3), 407-4.
- Alibakhshi, G., Kiani, R. G., & Akbari, R. (2010). On the authenticity of ESAP tests. *Journal of Asian ESP*, 6(2), 70-95.
- Bachman, L. F. (2000). Learner-directed assessment in ESL. In G. Ekbatani & H. Pierson (Eds.), *Learner-directed assessment in ESL* (pp. ix-xii). New Jersey: Lawrence Erlbaum Associates, Inc.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Barbot, M. J. (1991). New approaches to evaluation in self access learning (trans, from French). *Etudes de Linguistique Appliquee*, 79, 77-94.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning*, 39(3), 313-40.
- Blanche, P. (1990). Using standardized achievement and oral proficiency tests for self-assessment purposes: the DLIFLC study. *Language Testing*, 7(2), 202-29.
- Blatchford, P. (1997). Students' self-assessment of academic attainment: Accuracy and stability from 7 to 16 years and influence of domain and social comparison. *Educational Psychology*, 17, 354-360.
- Blue, G. M. (1988). Self-assessment: The limits of learner independence. *ELT Documents*, 131, 100-18.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Butler, Y., & Li, J. (2003). *The effect of self-assessment among young learners of English*. University of Pennsylvania.
- Calfee, R. C., & Hiebert, E. (1991). Classroom assessment of reading. In R. Barr, L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research, Volume 2* (pp. 281-309). White Plains, NY: Longman.
- Carton, F. (1993). Self-evaluation at the heart of learning. *Le Français dans le Monde*, 28-3.
- Dearing, R. (1997). *Higher education in the learning society*. London: HMSO.

- Dickinson, L. (1987). *Self-instruction in language learning*. London: Cambridge University Press.
- Dlaska, A., & Krekeler, C. (2008). *Self-assessment of pronunciation*. Oxford: Surrey University Press.
- Douglas, D., & Selinker, L. (1992). Analyzing oral proficiency test performance in general and specific-purpose contexts. *System*, 20(3), 317-28.
- Douglas, D. (2001). Language for Specific Purposes assessment criteria: where do they come from? *Language Testing*, 18(2), 171-85.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Evans, S., & Green, C. (2007). Why EGAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, 6, 3-17.
- Falchikov, N. (1997). *Why do lecturers involve students in assessment?* Paper presented at the 2<sup>nd</sup> North Umbria Assessment Conference, Encouraging Partnership in Assessing Learning, University of Northumbria, Newcastle.
- Fallows, S., & Balasubramanyan, C. (2001). Multiple approaches to assessment: reflections on the use of tutor, peer and self-assessment. *Teaching in Higher Education* 6, 229-24.
- Farhady, H. (2008). *Needs analysis*. Unpublished research project. SAMT: Tehran, Iran.
- Freeman, R., & Lewis, R. (1998). *Planning and Implementing Assessment*. Kogan Page, London.
- Fulcher, G. (1999). Assessment in English for Academic purposes: Putting content validity in its place. *Applied Linguistics*, 20(2), 221-36.
- Gardner, R. C., & MacIntyre, P. (1991). Motivational variables in second language acquisition. *Studies in Second Language Acquisition*, 13, 57-72.
- Glaser, B. G., Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine, Chicago.
- Hargan, N. (1994). Learner autonomy by remote control. *System*, 22(4), 455-62.
- Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback. *Teaching in Higher Education*, 6(2), 269-274.
- Janssen-van Dieten, A. (1989). The development of a test of Dutch as a foreign language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6(1), 30-46.



- Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: A comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education, 26*, 62-72.
- Lewkowicz, J. A., & Moon, J. (1985). Evaluation, a way of involving the learner. In J. C. Alderson (Ed.), *Lancaster Practical Papers in English Language Education* (Vol. 6; pp. 45-80). Oxford: Pergamon Press.
- Lincoln, Y. S., Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education, 10*, 209-220.
- McNamara, M., & Deane, D. (1995). Self-assessment activities toward autonomy in language learning. *TESOL Journal, 5*, 18-23.
- O'Malley, J. M., & Pierce, L. V. (1996). *Authentic assessment for English L2 learners*. Boston, MA: Addison-Wesley Publishing Company.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing, 6*, 1-13.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham, & D. Corson (Eds.), *Language testing and assessment* (Vol. 7; pp. 175-187). Dordrecht, The Netherlands: Kluwer Academic Publishers
- Peirce, B. M., Swain, M., & Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics, 14*, 25-42.
- Rivers, W. P. (2001) Autonomy at all costs: An ethnography of metacognitive self-assessment and self-management among experienced L2 learners. *The Modern Language Journal, 85*, 279-290.
- Ross, S. (1998). Self-assessment in second language testing: a meta-analysis of experiential factors. *Language Testing, 15*(1), 1-20.
- Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: the Wellington Indo-Fijians. *Language Testing, 15*(1), 86-108.
- Stefani, L. J. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education, 23*(4), 339-350.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: toward transparency for students and for tutors. *Assessment and Evaluation in Higher Education, 26*, 605-614.

- Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment and Evaluation in Higher Education*, 27, 501-510.
- Turner, J. (2004). Language as academic purpose. *Journal of English for Academic Purposes*, 3, 95-109.
- Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, 11(4), 186-193.