Assessing Fluency of English learner in a speaking examination

Dr.Nabil Nasr Noaman   **Inst.Dr. Thamer Yousif Allawi**

College of Fine Arts/ Baghdad University   / Directorate of education in Anbar

nabilnasr71@yahoo.com

The Literature review

Abstract

Recently, there have been discussions on the evaluation of students performance in speaking since self-cultural issues involved in raising awareness about how to enable teachers to assess students fluency in speaking. The main focus of this study is how to investigate the student level and their speaking skill in terms of fluency. This study concentrate on personal interview as one of the most common means of assessing speaking skills used in official speaking tests locally and internationally. The study analyses the level of fluency aptitude of student in terms of their commination in an oral test by following speak fluency criteria such as the use of , er , long pauses , word per minute and the contraction .

Key Words: fluency, speaking, assessment,

الملخص

في الآونة الأخيرة ، كانت هناك مناقشات حول تقييم أداء الطلاب في التحدث منذ أن شاركت القضايا الثقافية الذاتية في زيادة الوعي حول كيفية تمكين المعلمين من تقييم طلاقة الطلاب في التحدث. التركيز الرئيسي لهذه الدراسة هو كيفية التحقيق من مستوى الطلاب ومهارات التحدث لديهم من حيث الطلاقة التحدث وتركز هذه الدراسة على المقابلة الشخصية باعتبارها واحدة من أكثر الوسائل شيوعًا لتقييم مهارات التحدث المستخدمة في اختبارات التحدث الرسمية محليًا ودوليًا. وتحلل الدراسة مستوى كفاءة الطلاب في الطلاقة من حيث بدءهم في الاختبار الشفوي باتباع معايير التحدث بطلاقة مثل استخدام ، إيه، وقفات طويلة ، و كم كلمة في الدقيقة

الكلمات المفتاحية: الطلاقة ، التحدث ، التقييم

Introduction

There is a number of assessment missions conversation often used by teachers to assess oral communication skills for learners. These tasks include an oral reply to vouchers questions, describe the visual prompts and respond to them, stories tell, and oral presentation. The most common assessing speaking skills choices among teachers is the face to face interview. Personal interview is considered one of the most common means of assessing speaking skills used in official speaking tests locally and internationally. It is a direct exchange of face-to-face between the learner and themes (interlocutors) where the previous performance is evaluated in the evaluation of the oral interview, "there is sometimes a resident is not involved in operative interaction, but listen and monitor and evaluate the capabilities of the learner". (Underhill, 1998; Weir, 1993 , 102)

Format

Audio recording of four pairs. Each group consists of one pair of students.

Aim of the Study

To investigate the student level and their speaking skill in terms of fluency.

Principles of Assessment

1.Validity

In science and statistics, validity is the extent to which a concept, conclusion or measurement is well-founded and corresponds accurately to the real world. The word "valid is derived from the Latin validus, meaning strong. The validity of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure". (Wikipedia, 2014)

Validity shows whether the test is actually assessing: what it claims to test. Validity also shows whether the test telling you what you want to know? If we intend to examine student's level of speaking or reading comprehension, but instead test intelligence or background knowledge, in that case the test is not valid. When we design a test it should be assessed for its validity as much as possible. The examiner should be recommended that it is not the test itself, that the use of the test for a particular purpose it should be valid; for example a test for doctoral students in literature will not be valid for undergraduate scholars, because the level of their knowledge is different. Test measures translation skills will not be valid if the requirement is speaking or teaching skills.

2. Types of Validity

A. Construct Validity: is used to affirm that the assessment actually evaluating what it is meant to assess.

B. Formative Validity: this kind of validity is about how well   assessing is Able to provide information to help and improve the program of study

C. Validity can be improved in the following ways
1.Aims and ideas should be clearly defined and operationalized.
2.Assessment measures should match the aims and ideas.
 3. If possible, we should compare the assessment used with other assessment

3. Reliability

"The reliability of a test depends on whether the test would produce the same results if given to the same student at a different time Reliability therefore depends on the performance of the candidate as well as the accuracy and consistency of the marking" In other words a student must be able to obtain the same mark if he takes the same exam with other examiner (Anderson, 1990, 88).

A. Types of Reliability

1- Test-Retest Reliability:
 This kind of reliability assesses the consistency of a test across time.

2- Inter-rater Reliability
This kind of reliability is assessed by two or more independent judges, and the score must be convergent.

3-Internal Consistency Reliability
"Internal consistency is typically a measure based on the correlations between different items on the same test (or the same subscale on a larger test). It measures whether several items that propose to measure the same general construct produce similar scores"(Wikipedia 2014).

Testing Speaking By Paired Format
    Verbal or spoken communication is a significant interpersonal skills. Other important skills for non-verbal _communication are istening _skills and clarification. According to Heaton 1975 and Madsen1983 they show that speaking is important skill and the most difficult to test**.** Saville and Hargreaves (1999) emphasize that paired format of the Cambridge Speaking Tests advantages are: "candidates are more

relaxed; they have the possibility of more varied patterns of interaction during the tests; and this format can lead to positive washback in the classroom by encouraging learners to interact together in preparation for the test". Norton on the other says, "highlights several potential problems: candidates may not necessarily perform 'better' if they are more relaxed; nervous candidates could make their partner feel more nervous; candidates who do not know each other may feel more anxious about interacting with a stranger".(Norton,2005, 76)

The Scoring Perspective

1-**The Holistic Rating Scale**
In this type of measure, it gives the resident a single sign of performance, based on a predetermined scale. An example of this type of standards is a measure Carroll (1980), where this measure provides a general description of each domain.
The Holistic Rating Scale (Carroll, 1980, 45)

Table 1 : Holistic Rating Scale

Band	
1	An expert, speaks of authority in a variety of subjects. You can start the topic and the expansion and development
2	A very good speaker who is not a native speaker. Effectively maintains his party for discussion. Starts, maintains, and separates as necessary.
3	A good speaker. The situation presents a clear and logical, and can develop a coherent and constructive dialogue, rather less flexibility and versatility of the performance of the band 8 but can respond to changes in key tone or subject. Some repetition frequency and scale due to the restriction of language, but it interacts effectively.
4	Professional spokesperson. Able to preserve the topic of dialogue, keep track of topic clues, and use and appreciate key situation makers. Stumble and hesitate sometimes, but otherwise reasonably fluent. Some mistakes and appropriate language, but they will not hinder the interchange of opinion. Some appear independently in the talk with the ability to get started.
5	Humble speaker. Although the essence of proper dialogue can be understood primarily, but there were notable shortcomings in the mastery of methods and style of language. Needs to be repeated or clarified and question them in kind. its Lacks initiative and flexibility . Often they have to talk to the interviewer intentionally. It adapts but not with great style or attention.
6	Marginal speaker. It can maintain dialogue but in a somewhat negative way, and will seldom take the initiative or direct the discussion. He has difficulty in pursuing English language at ordinary speed; It lacks fluency and

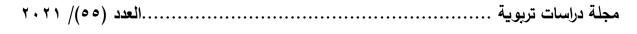|	
	accuracy in perhaps speaking. Therefore, the dialogue is not easy and flowing. However, it gives the impression that touches the essence of the dialogue, even if not fully proficient in it.
7	Very imperfect speaker. The issue of dialogue punctuated by lengthy hesitation and confusion. It holds only part of ordinary speech and is incapable to create a continuous and correct speech. Simple merit is just a comment on the substance of the debate, without creating a significant contribution to it.
8	Non-speaking. It is unable for understanding or speaking.

The Analytic Rating Scale

In this type of scale "
the developer first identifies the operations involved in responding to the task(s) and then attempts to create a marking scheme specifically to reflect these operations". (Barry O'Sullivan 2008,97) .criteria is presented for each aspect that is assessed.

Table 2: The Foreign Services Institute (FSI)

	Rating
Accent • Pronunciation is often incomprehensible. • Repeated gross mistakes and severe accent makes understanding hard and needs many repetition. • A "foreign accent" needs focused listening, and incorrect pronunciation leads to occasional confusion and obvious mistakes in grammar and vocabulary. • Distinguishing a "foreign accent" and accidental mispronunciation that does not intervene with understanding.. • There are no obvious verbal errors, but they will not be treated as a native speaker. • The original pronunciation, without trace of a foreign accent.	Rating
Grammar • The rules of Grammer are almost completely imprecise except for the inventory phases. • Constant faults that show control in too few main patterns and prevent frequent communication. • the large number of errors that some of the main patterns appear out of control and cause imitation and misunderstanding from time to time.	Rating

• Occasional mistakes that show incomplete control of some patterns, but there is no weakness which leads to misunderstanding.. • Few mistakes, without failure patterns. • There are a few negligible mistakes through the interaction.	
Vocabulary • Vocabulary is insufficient for even the simplest discussion. • Vocabulary restricted to essential areas of personal and existence (family, time, transportation, food, etc.). • The choice of words is sometimes imprecise, and restrictions on vocabulary avoid discussion at some steps of communication • Enough vocabulary to participate in the communication with some kind of circumlocutions. • A broad and accurate vocabulary suitable for dealing with the most difficult problems. • The vocabulary is accurate and broad, apparently like that of native speakers.	Rating
Fluency • Speech is interrupted and fragmented to the point that conversation is almost impossible. • Speech is very slow and erratic except for short sentences or routine. • Speech is often jerky and hesitant. Sentences might become incomplete. • Speech hesitation at times, with some difference resulting from reformulating words • . Speech is easy and smooth, but not original in speech and equality • Speech in all subjects easy and smooth like the original speaker.	Rating
Comprehension • Understands very little for the simplest kind of dialog. • Understands only slow, very simple speech on the most basic topics. Requires constant and rephrasing. • Understands precise and somewhat easy speech addressed to him / her with great repetition and paraphrasing. • Well understands his / her direct natural speech, but requires repetition and paraphrasing from time to time. • Able to understand every well in normal dialogue but low frequency or very low in the vernacular language elements, or	Rating

rapid speech is unclear or exceptionally .	

The CEFR (the Common European Framework of Reference) provides analytic scales divided in three levels :

" A	B	C
Basic User	*Independent User*	*Proficient User*

A1 A2	*B1 B2*	*C1 C2*
A2+	*B1+ B2+*	

The CEFR focuses on five criteria for spoken language use; Range, Accuracy, Fluency, Interaction and Coherence" (council of Europe, 2001)

Part two: Research design

A.Participant
1- The data that will be analyses is a pair format speaking test from the English Department of Suleyman Demirel University
2-. They are a four groups, each group contains 2 students
3- One examiner directed the test
4- The examiner asked the student to introduce themselves
5-The examiner asked each student individually about memory and then asked them to speak with each other about the topic he chose.
6- Each group takes different times to finishes the test. Group one finish the test at 10.18 minutes .group two 10.42 minutes, group three 9.46 minutes and group four 9.39 minutes. There are several different approaches to assessment show above .In this paper I use the levels from the CEFRR .this involves criteria referring and use of analytic scales. (appendix 1)

B.Method of Data Analysis
The study analyses the level of fluency aptitude of the students in terms of their commination in an oral test by following speak fluency criteria such as the use of, er, long pauses, word per minute and the contraction. According to: *the Common Reference Levels: qualitative aspects of spoken language use* fluency can be described in the following ways.

Table 3: Fluency criteria

A1	It can basically manage very short, isolated and pre-filled words, with a lot of pauses for expressions / expressions of less common words / to fix the connection
A2	Can make him / her understandable with a very short pronunciation, although pauses, false starts and paraphrasing are very clear
B1	It can still work understandably, although the pause in grammatical and lexical planning and reform is very evident, especially in the longer free production periods
B2	They can produce extensions of the language at a fairly equal pace; Although he may hesitate to look for patterns and expressions. There are little remarkably long downtime
C1	He can express himself easily and spontaneously without bothering about. Only conceptually difficult topic can hinder the natural flow and smooth language
C2	He can express it spontaneously prolonged flow slang normal, avoid or undo any difficulty so smoothly is hardly familiar with axes

1.Scoring scale:
 I split my scale range from 1-4 .One represents the inability to speak fluently and four represent the ability to speak fluently.

 1-Interjections or Er Use of sounds, syllables, and words that are independent of context of utterance like uh ,er , eam. I will concentrate on the using these kind of words in their speech.

Table 4: score for interjection

31	_	40 ER	1
21	_	30 ER	2
11	_	20 ER	3
1	_	10 ER	4

2- Words per minute: I will concentrate in this part on the word count that the each pair had used in their speech, as in the table below

table5: words per minute

120	-	190	words	1
191	-	250	words	2
251	-	300	words	3
301	-	400	words	4

3- Hesitation or **Long pauses:** Any non-tense break in the forward flow of speech, unacceptable within –word

Table 6: long pauses

15	-	25 Pauses	1
14	-	11 pauses	2
10	-	5 Pauses	3
4	-	1 Pauses	4

4- A Feature of Connected Speech**:** In this part, the researcher will concentrate on contraction in their answer.

Table 7: feature of connection

2	-	1	1
4	-	3	2
7	-	5	3
9	-	8	4

Table 8: score for each part

Fluency Feature	Score
ER	1- 4

Words per minute	1-4
Long pauses	1-4
A Feature of Connected Speech	1-4

In the table beneath, I will show a general overview of each part score, and the highest scores of the performance will be 16.

level	Rang
B2	13-16
B1+	09-12
B1	05-08
A1	01-04

As mentioned above, the rating scale will be related the Common European.
Table 9: rating scale

Data analysis:
Group one (Hasan and Hasan)
Table 10 : Hasan and Hasan

	Scoring		Comments
	Hasan (H)	**Hasan (HY)**	
ER	1	4	(H) used 31 er in his speech while (Hy)used 6 er
Words per minute	2	2	(H) used 215 words in his speech ,and he take 2minutes and 93 second and (Hy) used 218 words, and he take 3 minutes and 53 second
Long pauses	1	2	(H) had (23)long pauses in his speech ,while (Hy)had (12)long pauses in his speech.
A Feature of Connected Speech	3	3	(H) and (Hy) each one tries to make his speech more modern by using contraction
7		11	

Group (2) Muzeyyen and Şule

	Scoring		Comments
	Muzeyyen (M)	**Sule (S)**	
	1	3	(M) used 34 er in her speech while (S)used 17 er
Words per minute	2	1	(M) used 218 words in her speech and she take 3 minutes 10 second. (S) used 151 words and she take 2 minutes 18 second
Long pauses	2	2	(M) Had (12) long pauses in her speech, while (S) had (14) long pauses in her speech.
A Feature of Connected Speech	3	2	(M) is better than (S) in contraction
8		7	

Table 11:Muzeyyen and sule

Group (3) Hatice and Ayten (pembe)

Table 12: Hatice and Ayten

	Scoring		Comments
	Hatice (H)	**Ayten (P)**	
ER	2	1	(H) used 29 er in her speech while (P)used 47 er

Words per minute	**1**	**2**	(H) used 128 words in her speech and she take 3 minutes 25 second (P) used 191 words and she take 3 minutes 47 second
Long pauses	**2**	**3**	(H) Had (12) long pauses in her speech while (p) had (10) long pauses in her speech.
A Feature of Connected Speech	**1**	**3**	(P) is able to use contraction in her speech better than (H)
6		**9**	

Group (4) Sevda and Duygu

Table 13:Sevda and Duygu

	Scoring		Comments
	Sevda (S)	**Duygu (D)**	
ER	**2**	**1**	(S) used 23 er in her speech while (D)used 38 er
Words per minute	**1**	**1**	(S) Used 132 words in her speech and she take 1 minutes and 20 second while (D) used 139 words and take 3minutes and 08 second.
Long pauses	**3**	**3**	Each one had (9) long pauses
A Feature of Connected Speech	**1**	**1**	Two of them are not good in contraction
7		**6**	

(General Assessment of the groups)

Table 14: general assessment

NAME	ER	WOR PER MINUTES	LONG PAUSES	FEATURES OF CONNECTED SPEECH	TOTAL
Hasan (h)	1	2	1	3	7
Hasan (hy)	4	2	2	3	11
Muzeyyen (M)	1	2	2	4	8
Sule (S)	3	1	2	2	7
Hatice (h)	2	1	2	1	6
AYTEN (P)	1	2	3	3	9
Sevda (S)	2	1	3	1	7
Duygu (D)	1	1	3	1	6

The final assessment of the four participants will be as follows:

Group one (Hasan and Hasan)

table 15 :H and Hy score

Name	Score	level
Hasan(H)	7	B1
Hasan(Hy)	11	B1+

Group (2) Muzeyyen and Şule

Table 16:M and S score

Name	Score	level
Muzeyyen (M)	9	B1+
Sule (S)	8	B1

Group (3) Hatice and Ayten (pembe)

Table 17: H and P score

Name	Score	level
Hatic	6	B1
Ayten (Pembe)	9	B1+

Group (4) Sevda and Duygu

Table 18: S and D score

Name	Score	level
Sevda	7	B1
Duygu	6	B1

Discussion

The researcher listen to recorded between Hasan (H) and Hasan (Hy). The researcher felt Hasan (Hy) better than H, because he was able to communicate easily with the examiner and able to answer the question successfully, and he was helping his partner with words to his partners speech when his partner's cannot remember it. The researcher decided the student deserved B1+, While Hy was hesitating in his speech, he was able to express himself and I gave him B1. through the analysis I discovered that what I decided about them was correct.

Muzeyyen and sule

M was not able to express herself at the beginning ,and she was hesitating, she uses long pauses, but after a long pause, she was able to answer the question which means she takes time to arrange her thought. While Sule was not able to speak very well and she take a long pauses in her speech, even after that long pause, she was not able to answer well, Also when she was not able to give example she reaped what she said .she cannot connect word in a grammatical way, she used pronoun (you) instead of (I) . So The researcher expectation about their level that (M) take B1 and S will take A1. After the analysis my expectation was wrong (M) take B1+ and S take B1.

Hatic (H) and Ayten (p):

The examiner asked them about memory and from recorder, it was clear that(P) speaks better than (H), and she was able to speak in a logic way, she also supported her idea with an example, so I decided she deserve B2 .while (H) is not bad in English, but she uses Turkish language in her speech with her partner, she sometimes used the idea or the example that (p) gave it in her speech .so in my opinion, she will take B1. after the analysis, I discovered that I was wrong because (P) take B1+while(H) take B1 .

Sevda and Duyge

Sevda, starts to speak about the memory at the beginning of the test she is not able to speak in a good way. even she sometime she forgets words , but by the time she began to be more confident to speak .I think she deserves B1+.while (D) is more comfortable than(S) she was able to express herself from the very begging and she was able to speak well than her classmate and she also gave example to express her idea. I will give her B1. It was clear that I'm wrong because the analysis showed us that level of Sevda and Duyge are B1.

Conclusion

In conclusion, this paper has measured fluency for non-native speakers, from the English Department of Suleyman Demirel University in oral proficiency test. The examiner asks his students many questions .Through these questions we can measure their fluency level by the also the use of the features of connected speech and the number of their answer words they used it.

References

- Barry O'Sullivan, (2008). '*Notes on Assessing Speaking'* www.lrc.cornell.edu/events/past/**2008**-2009/papers08/osull1.pdf

- **Carroll, B. J.** (1980). *Testing Communicative Performance.* Oxford: Pergamon. In Barry O'Sullivan, (2008). '*Notes on Assessing Speaking'*
- Chinda, B. (2009). *Professional development in language testing and assessment: A casestudy of supporting change in assessment practice in in-service EFL teachers in Thailand* (Doctoral dissertation). University of Nottingham, UK. In Sinwongsuwat, Kemtong. (2012). "Rethinking Assessment of Thai EFL *Learners' Speaking Skills"*
- Council of European, (2001).*the Common European Framework of Reference for languages: earing teaching, assessment*. Cambridge university press
- Foot, M. C. 1999. *'Relaxing in pairs'*. ELT Journal 53/1: (36–41). J. Charles Alderson, Caroline Clapham, Dianne Wal,(1995) Language Test Construction and Evaluation, Cambridge
- Heaton,John Brian (1975) writing English Language Test, Longman. In Foot ,m.c (1999) Relaxing in pairs .ELT Journal
- Khamkhien, A. (2010). Teaching English speaking and English speaking tests in the Thai context: A reflection from Thai perspective. *English Language Teaching, 3*(1), 184-190. In Sinwongsuwat, Kemtong. (2012). "Rethinking Assessment of Thai EFL *Learners' Speaking Skills"* Language Testing in Asia.
- Madsen, H. S. 1983. Techniques in testing. New York: Oxford University Press In Foot,m.c(1999) Relaxing in pairs .ELT Journal
- Norton, Julie.(2005) 'The paired format in the Cambridge Speaking Tests'

- Saville, N. and P. Hargreaves. (1999). '*Assessing speaking in the revised FCE'*. ELT Journal 53/1: (42–51.) in Norton, Julie.(2005) 'The paired format in the Cambridge Speaking Tests' Language Testing in Asia.
- Sinwongsuwat, Kemtong. (2012). "Rethinking Assessment of Thai EFL *Learners' Speaking Skills"* Language Testing in Asia
- Sook, K. H. (2003). The types of speaking assessment tasks used by Korean Junior Secondary school English teachers. *Asian EFL Journal.* Retrieved from http://www.asian-efl-journal.com/dec_03_gl.pdf. In Sinwongsuwat, Kemtong. (2012). "Rethinking Assessment of Thai EFL *Learners' Speaking Skills"* Language Testing in Asia.
- http://en.wikipedia.org/wiki/Internal_consistency(2014)
- http://en.wikipedia.org/wiki/Validity_statistics (2014)