Research Article

Assessment of English Language Oral Production Using CALF: an Automatic Analytic Tool for Complexity, Accuracy, Lexis and Fluency

[1]Infanta Vincy R [2]Greta D'Souza
[1]CHRIST (Deemed to be University), Email: infanta.r@res.christuniversity.in
[2]CHRIST (Deemed to be University), Email:greta.dsouza@christuniversity.in

Abstract

This paper shows the evident lack of ICT based oral language production assessment tools and how the English language assessment in the State school education system is limited to testing the listening, reading and writing language skills. This paper projects four important factors which are proven to be effective indicators of an individual's oral language proficiency. This research paper explores the use of ICT in language assessment and explains the usage of CALF tool in depth which is an automatic analytic tool developed by Gavin Bui and Peter Skehan (2016) to analyse the complexity, accuracy and fluency of spoken language for research purposes. The study was conducted primarily to pilot the CALF tool and test its functionality in the context of interest for a research project. The CALF tool was piloted in a Government school in Tamil Nadu and the participants were learners of English as a second language from VI standard. A total of 15 students were present for all the sessions. The speech data of the learners were recorded, transcribed, coded and uploaded in the CALF system for obtaining the indicative measures of complexity, accuracy, lexical and fluency in English language production. The recorded speech data was analysed twice and the Intraclass Correlation Coefficient (ICC) was obtained using SPSS to establish the reliability of the CALF tool. The obtained ICC value of 0.955 proved an excellent level of conformity in the Complexity, Accuracy, Lexical and Fluency measures derived from the CALF system. The study most importantly exhibits the potential of the CALF tool to be added as a significant online language production assessment tool with further advancement research on the transcription and coding stages.

Keywords: CALF tool, ICT in language assessment, English Language Oral Production

Introduction

The English language production skills are one of the least assessed and focused skills in English language teaching spaces in the school education. "In a society where English proficiency guarantees economic and social upward mobility, much greater investment in research that informs policy and practice is necessary but does not seem to be forthcoming" (Ramanathan, 2016). The Draft National Educational Policy (2019) clearly expresses the desire to promote high accessibility and quality in English language education in India

in all government and non- government schools "to help break the current divide between the economic elite and the rest of the country, in addition to teaching Indian languages" with the focus on functionality and fluency (p. 82).

There's a huge dissimilarity in terms of English language proficiency between the different types of schools (NCERT, 2006, p.2) which are primarily categorized on the basis of two factors such as teachers' English language proficiency (TP) and English in the environment for language acquisition (EE). Ignoring these two factors, the state government of Tamil Nadu prescribes a common English language curriculum for students from the English and Tamil medium schools. There is no compromise done in terms of syllabus planning and textbook designing for the lack of both exposure to the target language and opportunities to use the language among the students from the Tamil medium schools. With the given conditions of limited teachers' proficiency and students' exposure to English language, it needs to be made sure that the amount to which the students are exposed to the target language and the amount to which they are made to produce the target language inside the classroom space is optimum.

A general review of the Tamil Nadu Government's English language curriculum, syllabus and textbooks reveal that the teaching and learning processes are inclined towards developing the English language receptive skills. Under productive skills, writing skills are given more importance when compared to the speaking skills. English language assessment models in school education also reflect the same patterns of testing where maximum focus is given on content knowledge, the grammar and the writing skills of the students. The speaking skills of the students remain ignored throughout the school education both in teaching and assessment. Most common reasons cited for the absence of English language oral or speech assessment in the English language curriculum is the lack of time and effective oral assessment tools. As a result, the speaking skills of the students to a large extent develop implicitly depending on the amount of language input they are exposed to inside and outside English language classrooms.

There is a need for establishing the most suitable English language production assessment tool for applied linguistics research purposes to facilitate the examination of language teaching approaches or methods. In the present curriculum of Tamil Nadu State Syllabus, the idea of testing the production skills is non-existent. It is the need of the hour to explore the available speaking assessment models and to find if they can be adapted and be simplified to later develop something that would suit the English language classroom requirements. To achieve this, the assessment models need to be researched and piloted in the respective contexts. It is also important to establish the reliability and validity of these assessment tools to look for adaptation ways. This paper presents the piloting of CALF: An Automatic Analytic Tool for Complexity, Accuracy, Lexis and Fluency developed by Gavin

Bui and Peter Skehan in the year 2016 in a Government school in Tamil Nadu to measure English language oral production.

English Language Oral Production

English language production is the production of spoken or written language. The researcher would be using the speech output as the testing variable. The oral language output from the students would be used to measure the progress in language learning. Complexity, Accuracy and Fluency (CAF) are language performance descriptors and are widely used in second language research in recent times as an alternative to standardized proficiency tests. CAF are "dimensions for describing language performance, most frequently used as dependent variables to assess variation with respect to independent variables such as acquisitional level or task features" (Pallotti, 2009, p. 590). The task performance was defined by the dimensions of complexity, accuracy and fluency based on the Task-based language teaching approach (TBLT). It is important to establish the relationship between the teaching approach and assessment model to explain the suitability of the assessment model and also to project the constructs of assessment in an integrated manner. In simple terms it is to show whether the learners are tested based on what they were taught or exposed to in the language syllabus. The language teaching approach that is dominantly present around the world and specifically in the State government school syllabus is task-based language teaching. Skehan (1996) proposed a framework for the implementation of task- based instruction which was designed in such a way that it is highly favourable towards achieving complexity, accuracy and fluency (CAF). He stated that CAF are the most relevant goals for task-based instruction.

Constructs

Complexity

Complexity is essentially "the extent to which the language produced in performing a task is elaborate and varied" (Ellis, 2003, as cited in Housen and Kuiken, 2009, p.461). L2 complexity can be subdivided into cognitive complexity and linguistic complexity. "Both types of complexity in essence refer to properties of language features (items, patterns, structures, rules) or (sub) systems (phonological, morphological, syntactic, lexical) thereof" (Housen and Kuiken, 2009, p.463). Linguistic complexity in TBLT refers to how differently and distinctly learners are able to phrase their language based on number and length of clauses (t- unit) and a range of grammatical structures. "A T-unit is an independent clause and any associated dependent clauses, that is, clauses that are attached to or embedded within it" (Larsen- Freeman, 2006, p. 597). The linguistic complexity can be further divided into grammatical complexity and lexical complexity. Grammatical complexity is the "average number of clauses per T-unit" (Larsen- Freeman, 2006, p. 597) and lexical complexity is "a sophisticated type-token ratio, word types per square root of two times the words" (Larsen- Freeman, 2006, p. 597).

Accuracy

Linguistic accuracy refers to how well the learners abide by the structural rules of the target language. Accuracy is "the ability to produce error-free speech" (Ellis, 2003, as cited in Housen and Kuiken, 2009, p.461). "It is the degree of conformity to certain norms" of a particular language (Pallotti, 2009, p. 592). Grammatical Accuracy is defined as "the proportion of error-free T-units to T-units" (Larsen- Freeman, 2006, p. 597). There are many kinds of grammatical errors, namely the subject-verb agreement, tense usage, article usage and punctuations. There are accordingly specific measures to gauge these different types of errors and the general measures to produce one value which is representative enough of all kinds of errors in language production (Bui & Skehan, 2018).

Fluency

Linguistic fluency refers to the ease of communication and the smooth delivery of the intended message. Fluency is "the extent to which the language produced in performing a task manifests pausing, hesitation, or reformulation" (Ellis, 2003, as cited in Housen and Kuiken, 2009, p.461). It is also defined as the "capacity to produce speech at normal rate and without interruption or as the production of language in real time without pausing or hesitation" (Pallotti, 2009, p. 591). Fluency is measured as the "average number of words per T-unit" (Larsen- Freeman, 2006, p. 597). Linguistic fluency can further be divided into cognitive fluency and uttered fluency. Cognitive fluency refers to the extent to which the learner has a clear understanding of the message to be delivered and the extent to which it is planned and structured in the learner's before the delivery of the speech. Utterance fluency is the actual language performance and that is the component that can be measured. The utterance fluency is gauged using the indices such as number of words per T-unit or minute, the frequency and length of speech, repetitions, reformulation, replacements and false starts.

CALF: An automatic analytic tool for complexity, accuracy, lexis and fluency

The CALF tool which is an automatic analytic tool for analyzing the complexity, accuracy and fluency of spoken and written language was developed by Gavin Bui and Peter Skehan (2016). The written or spoken language data need to be transcribed and coded in a specific way and then be updated on the CALF main system. On the system, the coded files will be processed and give an output of a range of measures of complexity, accuracy, lexis and fluency. It is important to note that the CALF measures are not accurate measurements of learners' oral language proficiency. These measures are only indicative of the English language proficiency based on the transcribed data. These measures can primarily be used for research purposes in applied linguistics and also can further be developed and automated for efficient gauging of English language production.

A detailed demonstration of the way in which the spoken data is transcribed

and coded is provided in the data analysis and interpretation section.

Participants

The participants in the study were sixth grade students from the government higher secondary school in Tamil Nadu, India. The number of participants was 15 and the medium of instruction is Tamil. In conversation with the students it was observed that the students understood what was being said to some extent when communicated slowly and with repetitions. When the students were involved in general everyday conversations they were hesitant to reply and communicated certain messages only through the utterance of words with no grammatical structure. The learners' vocabulary size was observed to be very limited and there were severe punctuation problems where groups of sentences were uttered without any pauses by the end of each sentence. Errors were committed in areas of subject-verb agreement and tense usage.

Procedure

The pilot study was conducted in a Government Higher Secondary Tamil Medium School in Coimbatore with the permission of the City Educational Officer. The researcher got the consent of both the teacher and the students to record the class and informed them how it would be used for research purposes. The pilot study happened over three weeks' time and the researcher got to interact with the students during the English language classes. Each period was for duration of 40 minutes. In the first two classes the researcher was present in the English language classes to get acquainted with the students and the teacher. The researcher even assisted the teacher in the transaction of a prose lesson in Term III. In the third class (in the absence of the English language teacher) the researcher individually developed conversations with the students to obtain a general idea of their proficiency level. The casual conversation was staged to also break the ice in order to create a favorable environment for language production in the latter classes.

In the next class the students were asked to assemble in the verandah where some of their classes usually happen and were asked to sit together. The change of setting was carried out to bring down their effective filter levels. The researcher sat along with them and introduced herself to the learners. To provide the students a little of content clarity and fluency they were guided with what topics can be covered in their self- introduction. The students were asked to talk about their city, family, friends, favourite food, favourite stars and hobbies. The students were asked to talk in random order and the entire class was recorded.

Procedure for CALF Analysis

The complexity, accuracy, lexis and fluency of the spoken data were measured using the CALF tool through the following four stages.

i) Recording

The researcher recorded the entire class sessions and the participants' oral speech. The students were given identification numbers and their audio

speech files were categorized and labeled with their roll number. A total of 15 audio files were segregated (1 each participant) and were stored in folders according to the participants roll numbers. Later the oral speech of each of the participants was transcribed into plain texts.

ii) Transcription

The next stage was the transcription of the spoken data using the CHAT (Codes for Human Analysis of Transcripts) (MacWhinney, 2000). The CALF system requires the spoken data to be transcribed in the CHAT format to process the input in an efficient manner. Brian MacWhinney (2000) developed a specific transcription format for transcribing child's talk which was one the two components of the CHILDES project which aimed to develop tools for analyzing talk. The CHAT system prescribes a set of coding features which facilitates the analysis of data using the CLAN software. The three main components of the CHAT format are the file headers, the main tiers and the dependent tiers. The headers give important information of the transcribed data namely the participants, the setting, the time and the details of the coder and the participants. The header files are followed by the four block tiers where the students' speech is transcribed into individual AS-unit tiers. The three tier block for single utterances is

*ID: <tab>- "The pruned line. Utterance transcribed into words without dysfluencies or pauses or any grammatical marking." (Bui & Skehan, 2016, p. 4)

%snd: <tab>- "The duration line. This line indicates the start and end of the utterance in that AS- unit." (Bui & Skehan, 2016, p. 4)

%ID: <tab> - "The main working line. This line includes all dysfluencies and pauses in the actual speech recorded, and syntactic marking." (Bui & Skehan, 2016, p. 4) The main working line is coded using the CHAT format.

The second line which is an automatically generated one will be explained in the next section.

iii) Coding

The main line is coded for fluency (including repairs, fillers, pseudo filled pauses, timing), complexity, accuracy and lexis following the user manual provided by Bui and Gavin (2016). The coded transcription (*ID, %snd, %ID in the AS-unit tiers) is then uploaded in the CLAN software to obtain the %mor. The second line in the four block tier is %mor and is generated automatically by the CLAN software when the transcription is run on CLAN with the command (see Appendix A). The %mor line produces the part-of-speech (POS) to each and every word in the transcription from the pruned line. Since it is automatically generated, the authors advise the POS tagging needs to be manually checked for ensuring accuracy in the result.

iv) Output

The final stage in the analysis is to drop the completed CLAN (txtin.cha) file in the CALF system. The system produces a range of results under Complexity, Accuracy, Lexis and Fluency. The output from the CALF tool is derived in five sections (see Appendix B). The first

section is called the 'Basic Statistics' where the total words, total pruned words and total time is listed down. The rest four sections are Complexity, Accuracy, Lexical and Fluency measure sections. The complexity section based on the coding done on the speech data produces information such as number of AS units, total number of clauses, number of independent and dependent clauses, number of before and after clauses, ratio of clauses per AS unit, number of pruned words per AS unit and number of words per clause. The Accuracy section provides data on the ratio of error- free clauses, errors per 100 words, syntactic, morphological, lexical, phonological discourse and pragmatic errors per 100 words.

The lexical section of the CALF system provides data on the significant F- score and lexical density. The fluency measures that are derived out of the output are raw speech rate, pruned speech rate, mean length of run, phonation time, number of mid-clause pauses and silences per 100 words, number of independent and dependent clause pauses, clause silence and average clause length. It also provides data on number of filled pauses, pseudo filled pauses, false starts, reformulations, repetitions and replacements.

For this study, the researcher has chosen the distinctive indicative measures of Complexity, Accuracy, Lexical and Fluency as defined in the 'constructs' section of the paper. Under the complexity section, the two measures that are taken into consideration for the analysis are number of AS Units (T- units) and the ratio of clauses per AS unit. The accuracy of the speech data is assessed by the ratio of error- free clauses and number of errors per 100 words which is comprehensive of syntactic, morphological, lexical and phonological errors. The F- score stands as a whole representative of the lexical measure and the raw speech rate stands for the fluency in the speech.

Reliability of CALF tool

The CALF tool being a measurement tool of the complexity, accuracy, lexis and fluency of the language data, the data was coded twice at an interval of two weeks and the results were tabulated. Since the nature of the measured data is of continuous type and because the speech data sample was the same in two different administrations, Intraclass Correlation Coefficient (ICC) was used to establish the reliability of the CALF tool in SPSS (see Table-1). The model adopted in the correlation is model 3 as the raters of interest were of only the researcher and it was two-way mixed.

Table-1: Details of Respondents

	Intraclass Correlation[b]	95% Confidence Interval			F Test with True Value 0		
		Lower Bound	Upper Bound	Value	df1	df2	Sig

Single Measures	.914a	.765	.970	22.317	14	14	.000
Average Measures	.955c	.867	.985	22.317	14	14	.000

As seen in Table 1, a high degree of reliability was found between the CALF measurements. The average measure ICC was .955 with a 95% confidence interval from .867 to .985 (F (14, 14) = 22.317, p<0.001). ICC value of 0.75 and upwards shows excellent level of conformity.

Discussion

The data obtained from the CALF system could be of great significance for measuring the oral proficiency of the participants for research purposes. In cases of development or proposal of specific pedagogical methods targeting improving the students' language oral proficiency, CALF tool could indeed be a great way to establish its effectiveness. Though the present study is highly limited in terms of sample size and context, it is safe to state that the CALF tool has the potential of replacing the numerable speaking rating scales which are qualitative, subjective and generic in nature when compared to the distinct indicative measures obtained from the CALF system.

The CALF tool cannot be taken to classroom spaces directly as the process involved in transcribing and coding the speech data is extensive. However, the CALF tool with further research can be made into an essential assessment tool in English language school education by automating the transcribing and coding stages in deriving the oral proficiency measures. In that way, the present lack of oral production assessment in the curriculum design could also be at a later point bridged through optimization of the CALF tool. A tool like CALF is very much in alignment with the National Education Policy 2020 where the organized education system is taking steps forward in this digital era by promoting virtual classes. In such a scenario CALF could prove to be an immense value in the ICT integrated online assessment of the English language production.

Conclusion

Oral speaking assessment of English language in school education is a topic that has been ignored in education policies and has resulted in severe repercussions on the English language proficiency level in the students and inequity in English language education in the country. This paper introduced the four main oral language constructs which are globally approved by the applied linguistics research fraternity. The relatively new automatic analytic tool for measuring the complexity, accuracy, lexis and fluency of oral production is piloted in a context of interest in the study. The CALF tool was piloted in the Government schools in Tamil Nadu among the sixth standard students over a period of three weeks. The students' speech data on a specific

topic was recorded, transcribed, coded and analyzed twice to obtain the indicative measures of complexity, accuracy, lexis and fluency. The Intraclass Correlation Coefficient (ICC= 0.955) was obtained from SPSS on the CALF measures and the results showed excellent level of conformity (0.75 and upwards), thereby establishing the reliability of the tested tool. The research paper has in detail demonstrated the use and reliability of an ICT based resource for testing the efficacy of oral language proficiency target teaching and learning English language education pedagogies for research purposes. It also has established the need to bring in advancement in the CALF tool which could possibly be an effective tool for instantaneous oral language assessment tool for school education and professional language eligibility examinations.

References

Bui, G. & Skehan, P. (2016). CALF: An automatic analytic tool for complexity, accuracy, lexis and fluency. Available at http://www.censpothk.com/calf/

Bui, G. & Skehan, P. (2018). Complexity, accuracy and fluency. In J. I. Liontas, T. International Association & M. DelliCarpini (Eds.) The TESOL encyclopedia of English Language teaching. doi:10.1002/9781118784235.eelt0046.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquistion. Applied Linguistics, 30 (4), 461-473.

Larsen-Freeman, D. (2006). 'The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English,' Applied Linguistics 27/4: 590- 619.

Skehan, P. (1996). A framework for the implementation of task- based instruction. Applied Linguistics, 17(1), 38- 60.

MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk (3rd ed). Mahwah, NJ: Lawrence Erlbaum Associates

Ministry of Human Resource Development (MHRD). (2019). Draft National Education Policy. Retrieved from https://innovate.mygov.in/wpcontent/uploads/2019/06/mygov15596510111.pdf

National Council of Educational Research and Training (NCERT). (2006). Position Paper: National Focus Group on Teaching of English. Retrieved from http://www.ncert.nic.in/new_ncert/ncert/rightside/links/pdf/focus_group/english.pdf

Pallotti, G. (2009). CAF: Defining, Refining and Differentiating Constructs, Applied Linguistics, 30(4), 590- 601.

Ramanathan, H. (2016). English education policy in India. In R. Kirkpatrick (Eds.), English Language Education Policy in India (113- 126). London: Springer.

Appendix A

A Sample of the Coded Speech Data

@Begin
@Languages: eng
@Participants: ARV Participant
@ID: eng|text|ARV||male|||Participant|||
@Time Start: 00:00:00
*ARV: My name is Aravind .
%mor: n:prop|My n|name cop|be&3S n:prop|Aravind .
%gra: 1|2|MOD 2|3|SUBJ 3|0|ROOT 4|3|PRED 5|3|PUNCT
%snd: <00:00:00><00:01:00>
%ARV: My name is Aravind .errfr ::: |
*ARV: VIth standard Kurichi School .
%mor: n:prop|VIth n|standard n:prop|Kurichi n:prop|School .
%gra: 1|2|MOD 2|0|INCROOT 3|4|NAME 4|2|APP 5|2|PUNCT
%snd: <00:01:00><00:03:00>
%ARV: VIth standard Kurichi School .err_a_s ::: |
*ARV: Mother name Jaya .
%mor: n:prop|Mother n|name n:prop|Jaya .
%gra: 1|2|MOD 2|0|INCROOT 3|2|APP 4|2|PUNCT
%snd: <00:03:00><00:06:00>
%ARV: er (1.0) Mother name (0.90)) Jaya .err_s_s ::: |
*ARV: Dad name John .
%mor: n:prop|Dad n|name n:prop|John .
%gra: 1|2|MOD 2|0|INCROOT 3|2|APP 4|2|PUNCT
%snd: <00:06:00><00:11:00>
%ARV: er (1.0) Dad {name} * name er (2.0) John .err_s_s ::: |
*ARV: Age eleven.
%mor: n:prop|Age .
%gra: 1|0|INCROOT 2|1|PUNCT
%snd: <00:11:00><00:27:00>
%ARV: er (10.0) Age # eleven .err_s_s err_m_p ::: |
*ARV: Friends Santhosh, Guruprasad, Hariharan, B Santhosh, Saran, Aravind .
%mor: n:prop|Friends n:prop|Santhosh cm|cm n:prop|Guruprasad cm|cm n:prop|Hariharan cm|cm n:prop|B n:prop|Santhosh cm|cm n:prop|Saran cm|cm n:prop|Aravind .
%gra: 1|2|NAME 2|0|ROOT 3|2|LP 4|2|ENUM 5|2|LP 6|2|ENUM 7|2|LP 8|9|NAME 9|2|APP 10|9|LP 11|2|ENUM 12|11|LP 13|2|ENUM 14|2|PUNCT
%snd: <00:27:00><00:39:00>
%ARV: er (6.0) Friends er Santhosh er Guruprasad er Hariharan B er Santhosh Saran Aravind .err_s_s ::: |

*ARV: Rasagulla .
%mor: n:prop|Rasagulla .
%gra: 1|0|INCROOT 2|1|PUNCT
%snd: <00:39:00><00:43:00>
%ARV: (3.0) Rasagulla . err_a_s ::: |
*ARV: Cartoon .
%mor: n:prop|Cartoon .
%gra: 1|0|INCROOT 2|1|PUNCT
%snd: <00:43:00><00:54:00>
%ARV: (4.0) Cartoon . err_a_s ::: |
*ARV: Police miss .
%mor: n:prop|Police n|miss .
%gra: 1|2|MOD 2|0|INCROOT 3|2|PUNCT
%snd: <00:54:00><00:58:00>
%ARV: er(2.0) Police er (2.0) miss . err_a_s ::: |
@End

Appendix B

A Sample of the Output Derived from CALF Tool

Basic Statistics

1	Total words	36
2	Total pruned words	27
3	Total Time	0.966667

Complexity Section

4	Number of AS Units	9
5	Total number of clauses	9
6	Number of independent clauses	9
7	Number of dependent clauses	0
8	Number of before dependent clauses	0
9	Number of middle embedded dependent clauses	0
10	Number of after clauses	0
11	Ratio of clauses per AS unit	1
12	Number of pruned words per AS unit	3
13	Number of words per clause	3

Accuracy Section

14	Ratio of error-free clauses	0.111111
15	Errors per 100 words	18.51852
16	Syntactic errors per 100 words	14.81481

17	Morphological errors per 100 words	0
18	Lexical errors per 100 words	0
19	Phonological errors per 100 words	3.703704
20	Discourse errors per 100 words	0
21	Pragmatic errors per 100 words	0

Lexical Section

22	F-score	52.5
23	Lexical Density	18.51852

Fluency Measure

27	Raw Speech Rate	37.24138
28	Pruned Speech Rate	27.93103
29	Mean Length of Run	1.588235
30	Phonation Time	0.45
31	Number of Mid-Clause Pauses per 100 words	37.03704
32	Average Mid-Clause Pauses Length	0.8613
33	Mid-Clause Silence Total per 100 words	118.1481
34	Number of Independent Clause Pauses per 100 words	0
35	Average Independent Clause Pause Length	0
36	Independent Clause Silence Total per 100 words	0
37	Number of Dep. Clauses Pauses per 100 words	0
38	Average Dependent Clause Pause Length	0
39	Dependent Clause Slience Total per 100 words	0
40	Number of Filled Pause per 100 words	0
41	Number of Pseudo Filled Pauses	0
42	False Starts per 100 words	3.703704
43	Reformulation per 100 words	0
44	Repetitions per 100 words	3.703704
45	Replacements per 100 words	0