Effect of formulaic sequences on fluency of English learners in standardized speaking tests

Jennifer François, Middle Tennessee State University
Mohammed Albakry, Middle Tennessee State University

Abstract

While fluency in second language speaking can be a challenging construct to measure, it is important to identify the discourse features that contribute to it. This small-scale classroom research project examined the effect of formulaic language sequences on fluency as measured by computer-based speaking tasks of young English learners. Thirty-six speech samples were collected as a part of standard instruction in grades 5-8 in a medium-sized public school district in the Southeastern US. The speech samples were analyzed using Praat speech analysis software to identify the mean length of fluent run for fluency and coded formulaic language sequences for discourse function. Findings indicated that the use of formulaic sequences is a significant predictor of fluency in the data set (p = .015) and that the most frequently used formulaic sequences were those used for clarification and to compare and contrast. Finally, the article discusses pedagogical implications for second language instruction, specifically for improving fluency on standardized computer-based speaking assessments.

Keywords: Computer-based Speaking Assessments, Fluency, Formulaic Sequences, Young English Learners

Language(s) Learned in This Study: English


Introduction

English learners (ELs) make up an increasing portion of U.S. public school populations, growing from 3.8 million in 2000 to 4.9 million in 2016 (U.S. Department of Education, 2019). As a requirement of the 2015 federally-mandated act of the U.S. government known as the Every Student Succeeds Act (ESSA), school districts are required to measure the English language proficiency of these students annually to monitor growth. Relying increasingly on computer-based testing platforms, most states are using assessment products for elementary and secondary ELs from World-Class Instructional Design and Assessment (WIDA), a testing and standards consortium whose assessments are implemented in 40 U.S. states and territories. As a result, it has become imperative to investigate how to adapt to the evolving technology of computer-based standardized assessments to meet the needs of the students—particularly in relation to speaking assessments.
While former iterations of WIDA’s speaking assessments were administered and scored locally by teachers who worked directly with the students in a conversational setting, the revamped assessment introduced in 2017 replaced the local test administrator with a virtual testing administrator with no human interaction and instituted remote scoring of the digitally recorded speech samples. As a result, the performance expectations of the speaking subtest of the 2017 WIDA ACCESS 2.0 moved away from measuring communicative competence to measuring the ability to produce concise, impromptu monologues on unprepared academic topics in 50 seconds or less. In school districts across the US implementing the ACCESS 2.0 assessment, these changes caused the scores of students to plummet (see Table 1). The drop in speaking scores in particular was so dramatic that some states had to lower their exit criteria to avoid unnecessarily retaining students in an EL program and preventing them from participating in higher-level classes needed for on-time graduation (Mitchell, 2017).

Table 1
Change in WIDA ACCESS 2.0 Scores by Domain in Tennessee

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCESS 2016</td>
<td>$M = 4.69$</td>
<td>$M = 4.13$</td>
<td>$M = 4.12$</td>
<td>$M = 3.38$</td>
</tr>
<tr>
<td>ACCESS 2017</td>
<td>$M = 4.49$</td>
<td>$M = 2.93$</td>
<td>$M = 3.34$</td>
<td>$M = 3.14$</td>
</tr>
</tbody>
</table>

*Note. $M =$ Mean, Standard Deviations were requested from the Tennessee Department of Education but not received.*

While preparing students for the expectations in participating in the speaking assessment may sound like “teaching to the test,” it is rather about preparing students to develop the new skills needed to participate in the assessment under the exigencies of new technological conditions. Moreover, preparing students for using the different set of discourse skills related to this new technology is necessary since it has not been previously part of instructional standards. This preparation can not only help students be more successful in assessment situations, it can also serve as a building block for preparing students for the rigor of giving longer oral presentations in content area classes.

Research Questions

This study focuses on the occurrence of formulaic sequences (henceforward FS) and their effect on spoken fluency. The functions these sequences might serve in helping young English learners structure spoken discourse could contribute to improvement in student performance on computer-based speaking assessments. The study seeks to answer these research questions:

1. Could the use of FS predict fluency for elementary and middle school EL students?
2. What are the most frequent discourse function types attested in the FS used by EL students in the study?

However, to situate the relationship between FS and fluency, we first present an overview of the relevant literature.

Literature Review

Researchers have been defining and categorizing language sequences since the early 1980s. The COBUILD project (Collins Birmingham University International Language Database) at the University of Birmingham in England, developers of a corpus of contemporary texts from which many dictionaries have been published, was one of the first to demonstrate the nature of formulaic language (Wood, 2002). This corpus made visible on a large scale how words occur together in the English language and paved the way for other researchers (e.g., Biber & Conrad, 1999; Ellis, 1996; Ellis et al., 2008; Gray & Biber, 2013; Nattinger & DeCarrico, 1992; Pawley & Snyder, 1983; Wood, 2006) to further extend the study and role of FS in structuring language output.
While Ellis et al. (2008) define FS as “recurrent multiword patterns or formulas” (p. 375), Wood (2002) includes the metacognitive aspect by defining them as “multiword units of language that are stored in long-term memory as if they were single lexical units” (p. 2). On the other hand, Nattinger & DeCarrico (1992) foreground the pragmatic function that FS perform by defining them as “multi-word lexical phenomena that exist somewhere between the traditional poles of lexicon and syntax . . . [that are] associated with a particular discourse function . . .” (p. 1).

Several important features of FS can be identified and extrapolated from these definitions: formulaic language sequences consist of multiple words which are stored in the long-term memory as a single lexical unit whose function in discourse is organizing speech output. Moreover, Pawley & Syder (1983) contribute that these multiword clusters are “lexicalized sentence stems” (p. 189), which not only perform a function in organizing discourse, but also provide reliable patterns to facilitate meaning-focused output (Gray & Biber, 2013; Pawley & Syder, 1983; Wood, 2006).

Only a few robust studies of FS seem to focus solely on second language learning, and the majority of them concern samples of university-level English as a Foreign Language (EFL) students (Boers et al., 2017; Boers et al., 2006; Ha, 2017; Khodadady, 2012; McGuire & Larson-Hall, 2017; Osman & Jusoff, 2009; Rafieyan, 2018; Tavakoli, 2011; Xia, 2018; Yan, 2020; Yılmaz & Koban Koc, 2020). An even smaller number of these studies focus on the use of FS in the speech of young and adolescent English learners (Eyckmans et al., 2015; Mohammadi & Enayati, 2018; Puimège & Peters, 2019). The majority of the relevant studies, however, provide evidence that learning FS has positive impacts on improving fluency (both perceived or objective). They further point to the relative lack of research on this relationship for young learners’ speech production at the elementary and middle school levels focusing specifically on academic speech production (as opposed to conversational speech).

**Methodology**

**Scope and Definition of Formulaic Sequences**

For this classroom research, we have taken a broad view of FS. This view encompasses lexical bundles such as the author of (Biber & Conrad, 1999), formulaic speech such as you know (Nattinger & DeCarrico, 1992), and lexical frames similar to because X and Y (Wood, 2006), including complex sentence stems like there are Y differences between X and S (Pawley & Syder, 1983) (see Figure 1).

**Figure 1**

*The Scope of Formulaic Sequences*

While the scope may seem relatively wide, all of these varying categories of linguistic units contribute to
how students perform on speaking tasks (Boers et al., 2006; McGuire & Larson-Hall, 2017; Mohammadi & Enayati, 2018; Khodadady, 2012; Rafieyan, 2018; Tavakoli, 2011; Yilmaz & Koban Koc, 2020; and Yan, 2020; to name a few) and are found in the speech samples under investigation. Moreover, they all incorporate multi-word units that are not novel speech produced by the student; and, therefore, they have all been considered in this classroom research project to reflect the broad range of their use by the students, similar to Wood’s (2006) study. This loose definition is also supported by Wray’s (2002) definition of FS as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (p. 9).

**Participants and Tasks**

**Participants**

The data collected for this small-scale classroom research project seeks to investigate FS use in second language speech production in English. Student speech samples were collected from 15 upper elementary and middle school EL students (9 males and 6 females in Grades 5-8). These samples were collected during regular ESL instruction. The students were primarily L1 Spanish speakers in a suburban, medium-sized public school district. The speech collected over an academic school year from August 2018 to May 2019 initially included 41 samples of speech. However, five speech samples were eliminated from the analysis due to poor audio quality, leaving a total of 36 samples for analysis, representing a mean of $M = 2.4$ samples per student. The number of samples collected per student differs due to the nature of classroom instruction, including student absences on the day of the assignment, school-wide events that impacted regular instruction, or other events beyond teacher control. Other impacting reasons also included transient student populations and students enrolling or withdrawing at times during the U.S. school year which do not necessarily correspond to beginning and ending times in their home country’s school calendar. In spite of the variety of English proficiency levels, the data collected shows that the frequency of mean length of fluent run per sample had a normal distribution (see Figure 2), demonstrating that students who may have contributed multiple samples and may have been stronger in FS did not affect the data.
Figure 2

Histogram: Mean Length of Fluent Run Per Sample

Note. $N = 36$ (sample size), $M = 8.05$ (mean), $SD = 3.08$ (standard deviation). This histogram provides the raw data associated with the collected speech samples. It shows that the overall mean length of fluent run per speech sample was normally distributed.

Tasks

The speaking tasks included science, mathematics, social studies, and language arts topics; and they elicited expository and argumentative speech (see Table 2). The WIDA English Language Development Standards, adopted by the school district involved in this study, contain five standards incorporating the language for Social and Instructional purposes, the language of Language Arts, the language of Mathematics, the language of Science, and the language of Social Studies (Board of Regents of the University of Wisconsin System, 2020a). Additionally, the WIDA Can Do Descriptors for Speaking encompass four levels: recount (expository), explain (expository/compare and contrast), argue (argumentative), and discuss (Board of Regents of the University of Wisconsin System, 2020b). These two guiding documents provide a framework for districts, schools, and teachers in providing quality ESL instruction according to best practice and research-tested methodologies.

Because the data set was collected during an academic year of instruction, the tasks represented in the data set comprise all content areas (apart from social and instructional language) and all levels of the Can Do Descriptors (apart from discuss). Social and instructional language tasks were not included in the data set because these skills are practiced outside the scope of academic language use. Furthermore, the Can Do Descriptor level of discuss was not included in the tasks as the speaking assessment is completed without a conversation partner. While a direct connection between the tasks used in the project and the exact content of tasks on the WIDA ACCESS 2.0 Speaking Assessment cannot be made due to confidentiality agreements, the type of academic speech and the subject area content of the tasks are directly aligned with
the English language development standards and key use levels adopted by WIDA (Board of Regents of the University of Wisconsin System on behalf of WIDA, 2016; Board of Regents of the University of Wisconsin System, 2020a).

Table 2
Sample of Academic Speaking Prompts

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Type and Content Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which character do you think changed the most in the book, <em>Origami Yoda</em>? Be sure to explain your reason for thinking this by giving examples from the text.</td>
<td>Argumentative Language Arts</td>
</tr>
<tr>
<td>How have shifts in Earth’s plates changed Earth’s surface? What does the author mean by puzzle pieces of the Earth’s crust?</td>
<td>Expository Science</td>
</tr>
<tr>
<td>What natural resources are necessary for manufacturing plastic bottles and explain how the manufacturing process is harmful for the environment?</td>
<td>Expository Science and Mathematics</td>
</tr>
<tr>
<td>Do you think the Alcatraz escapees survived? Support your opinion with evidence from the resources we used in class.</td>
<td>Argumentative Social Sciences</td>
</tr>
<tr>
<td>Explain the personification in the line <em>the waves beside them danced</em> in the poem “I Wandered Lonely as a Cloud” by Wordsworth.</td>
<td>Expository Language Arts</td>
</tr>
<tr>
<td>What are fossil fuels, and what is one disadvantage of using them?</td>
<td>Expository Science and Mathematics</td>
</tr>
</tbody>
</table>

The speaking tasks were presented individually to the students through Google Classroom using Screencastify, a Google Extension screencasting tool. They included a written representation of the task on the screen, as well as a Screencastify video, which provided the content stimulus necessary to respond to the prompt. Students used district-issued Chromebooks and headsets with integrated microphones to record their responses to the prompts and could complete the tasks at their own pace with the mean response time, \( M = 56.03 \) seconds.

Measures and Procedures

The speech samples were assigned anonymized numbers to remove personally identifiable data and transcribed by hand using Google Voicetyping with three or more repetitions and then analyzed through Praat, a speech analysis software that allows users to analyze very small segments of recorded speech (Boersma & Weenink, 2019; N. H. De Jong & Wempe, 2009) to verify the accuracy of the transcription.

Fluency Measure

Admittedly, the role fluency plays in the scoring of the standardized speaking assessments is unclear. According to WIDA, there is no objective measure of fluency calculated as part of the overall speaking assessment score. Inquiries with WIDA in an attempt to get a clear picture about their evaluation system for the speaking assessment led to an email response that stated “no quantitative measure of fluency is used
to evaluate speaking assessment responses” (A. Traverse, personal communication, July 24, 2019). In a similar vein, WIDA stated that raters are trained and calibrated on a rating rubric, and interrater reliability is between 75-100%. Nevertheless, to investigate the relationship between FS and fluency in this project, an objective measure of fluency was needed.

While there are several aspects necessary to measure fluency accurately, such as the quantity of speech produced, rate of speech production, and disfluencies (N. H. De Jong & Mora, 2019; N. H. De Jong et al., 2015; N. De Jong & Perfetti, 2011; Ginther et al., 2010; Rossiter, 2009; Towel, et al., 1996), a measure was needed for this project that would isolate the impact of the use of FS on fluency to strengthen the robustness of the analysis. As a result of this, the mean length of fluent run was chosen as the measure of fluency as it is a measure used by prior researchers to determine the effect that the use of FS has on fluency (see; N. De Jong & Perfetti, 2011; Ginther et al., 2010; Pawley & Syder, 1983; Towel et al., 1996; and Ushigusa, 2009). Mean length of fluent run takes into consideration the three major components of fluency: quantity, rate, and disfluencies in production.

For this study, mean length of fluent run was calculated using Praat and then manually verified to remove disfluencies from the calculation. To calculate mean length of fluent run, only fluent lexical syllables spoken between silent pauses in each speech sample were counted. Dysfluent lexical utterances such as repetitions, self-corrections, or lexical but non-meaningful fillers (e.g. um, eh, and ah) were excluded.

**FS Measure**

The standard FS were identified using the Michigan Corpus of Academic Spoken English (MICASE), counting all multiword examples that appear in the MICASE with a mean frequency per 10,000 words at or above $M = 0.65$ (see Table 3 for examples). This measure was chosen as the cut-off frequency based on previous research indicating that .40 per 10,000 words should be used (Biber et al., 2004) and going even as low as .10 per 10,000 (Bardovi-Harlig et al., 2015). Wray (2002) even postulates that frequency cut-offs are arbitrary because a phrase that has already been accepted as an FS could have a lower frequency than found in a corpus. After a close reading of the student speech samples, discourse functions of the standard FS (attested in MICASE) as well as sequences taken directly from the prompts were identified and coded according to the following categories: focus, sequence, clarify, compare/contrast, argue, and summarize (Biber & Conrad, 1999; Nattinger, 1980; Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Wood, 2002; and Xu, 2018).

**Table 3**

*Examples of Standard FS*

<table>
<thead>
<tr>
<th>Formulaic Language Sequence</th>
<th>Mean MICASE Tokens per 10,000 Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>according to</td>
<td>2.04</td>
</tr>
<tr>
<td>this is what</td>
<td>1.76</td>
</tr>
<tr>
<td>compared to</td>
<td>1.36</td>
</tr>
<tr>
<td>says in the</td>
<td>0.78</td>
</tr>
<tr>
<td>but then</td>
<td>2.34</td>
</tr>
<tr>
<td>tell you about</td>
<td>1.12</td>
</tr>
<tr>
<td>that is the same</td>
<td>0.79</td>
</tr>
<tr>
<td>show how</td>
<td>0.97</td>
</tr>
<tr>
<td>means by</td>
<td>17.17</td>
</tr>
<tr>
<td>you know</td>
<td>34.65</td>
</tr>
</tbody>
</table>
Standard FS were coded S, prompt FS were coded P, and original student response material sequences were coded O. Coded pauses were noted with an ellipsis ( . . . ) and interrupted thoughts or utterances coded with a dash ( -- ). Student responses were written in italics in one continuous line, and FS were coded using an interspersed format because the codes are short and distinguishable from words (Edwards & Lampert, 1993).

**Praat Measurement**

The speech files had to be converted to .WAV files (a file format used to store audio files on a PC) using a free conversion software available on the internet and then uploaded into the Praat software. After loading the speech files into Praat, the samples were annotated by selecting “To TextGrid (silences)” from the Annotate dropdown list. To determine pauses between runs, the default settings in Praat were used to distinguish between silences and sounds in each sample (N. H. De Jong & Wempe, 2009). However, due to the variances in natural speech speed of the students (N. H. De Jong et al., 2015), all samples were manually checked by listening to the sample a minimum of three times and corrected for pauses incorrectly identified by the software.

Finally, fluent syllables were counted manually to avoid the inadvertent dropping of additional syllables due to irregular variations in pronunciation or the inclusion of non-fluent syllables by Praat, such as repeated word onsets such as the **wi in wi-wiggle** (see **Figure 3** for an example). Due to the small sample size, the manual verification and counting of syllables was practical to implement. A baseline measure of temporal fluency counted in words per minute was calculated (Ginther et al., 2010), and formulaic language sequences, both standard and prompt (Pawley & Syder, 1983), per number of runs in a speech sample were manually coded and calculated.

**Figure 3**

Praat Spectogram Analysis of Speech

*Note.* This is an example of a syllable repetition which is coded as “sounding” speech in Praat: *Then, like, you, you, you try and . . .* with the you repetition circled in red. Two of these syllables were manually removed from the mean length of fluent run syllable count as repetitions.
**Statistical Analysis**

To answer the first research question (can the use of FS predict fluency?), linear regression was used with fluency represented by mean length of fluent run as the dependent variable and the number of FS, including standard and prompt sequences, per run used as the independent variable.

**Results and Discussion**

The results are discussed in relation to the two research questions that concern the effect and type of FS use on fluency for computer-based speaking tasks in the data set under examination. Students in the project who used more FS in their speech produced more fluent speech. Regression analysis indicated that formulaic sequences significantly predicted speech fluency ($\beta = .40$, $t(36) = 2.56$, $p = .015$) and formulaic sequences also explained a significant proportion of variance in fluency scores ($R^2 = .16$, $F(1, 36) = 6.57$, $p = .015$). The use of the FS in this sense frees up more cognitive capacity for the student to focus on the word choice, register, and cohesion of the response, which leads to an overall improvement of the speech and fluency for each task (Boers et al., 2006; Eyckmans et al., 2015; McGuire & Larson-Hall, 2017; Mohammadi & Enayati, 2018; Pawley & Syder, 1983; Rafieyan, 2018; Yan, 2020; Yilmaz & Koban Koc, 2020). This is particularly relevant for students participating in standardized computer-based speaking assessments due to the format of the testing platform, where students are presented with previously unprepared stimulus materials. The cognitive ability to encode these language sequences as one lexical unit and retrieve them from long-term memory could play a role in improving speaking fluency on computer-based speaking assessments that require impromptu speech production on previously unprepared topics. The academic quality of the speech produced also can be improved due to the resulting surplus cognitive capacity (Pawley & Syder, 1983).

**Textual Analysis**

To answer research question two (which discourse functions do FS in the study represent?), this subsection offers qualitative analysis to provide insight into learner speech during this type of oral tasks and how fluency seems to be associated with it. Regarding the most frequent discourse function types attested in the FS used by EL students in the project, the breakdown of the results is shown in Figure 4.
Figure 4

Percentage of Discourse Functions of Formulaic Language Sequences

Comprising nearly 50 percent of the FS utilized in the corpus, clarify was the most frequently used function of the formulaic sequences. The clarifying formulaic sequences were used to elucidate or highlight information in the student’s response or were used to further explain or target information that the student found necessary for the response to precisely address the prompt. One example of clarify from the corpus is:

“The author of (S) ‘I wandered lonely as a cloud’ (P) used personification (P). For example (S), ‘the waves beside them danced . . . ’(O)”

In this utterance, the student used “for example” to provide a clarifying example in response to the prompt. Upon examination of the entire utterance, it is possible to identify other FS, including standard and prompt, as well as original material produced by the student. This exemplifies Pawley & Syder’s (1983) strings of standard formulaic sequences which leave slots for the insertion of prompt or original language sequences to complete a string of utterances and unite them into a connected and fluent response. For example, “the author of X” utilized the standard FS “the author of” in combination with information from the prompt citing the name of the poem being analyzed. Furthermore, the student completed the utterance with “used personification,” which is a prompt sequence taken from the prompt and continued with the clarifying formulaic sequence “for example.” Then the student cited a line from the poem that illustrates the example of personification chosen by the student and included as original content. This example demonstrates not only how the student used the words “for example” to clarify the response, but also how strings of FS can be connected to prompt and original language sequences and inserted into slots to complete responses to speaking prompts.

The second most frequent discourse function represented in the speech samples is compare/contrast. These sequences are used when comparing or contrasting two elements from the speaking tasks. One example of compare can be seen in this example:
“They're both same (S) because they [Earth’s plates] and puzzle pieces... connect... like... they connect (P/S/O)”

In this example, the student used “they’re both the same” to demonstrate how the two items in question from the prompt are similar. The student further used “because X and Y,” a standard sequence that additionally involved a prompt sequence from the prompt “they [Earth’s plates] connect” and original student material “puzzle pieces” to explain how the two items being compared are similar.

An example of contrast is evident in the following excerpt:

“First... first, the whales live in the ocean... But on the other hand (S) elephants live in... on the like... on the ground...”

In this example, the student used the standard FS “on the other hand” to illustrate a difference between whales and elephants.

Another example of contrast is:

“There’s three differences between (S) whales and... whales and Komodo dragon (O)...”

In this example, the student used the standard sequence “there are Y differences between X and S” to indicate that the response would contain a number of examples detailing the differences between the two animals under investigation. The student inserted the number of examples that will be explained in the response, as well as the names of the two animals taken from the speaking prompt as original contributions to the response.

Sequencing is the third most represented discourse function of the speech samples and includes formulaic sequences that organize information in a certain order related to the content of the prompts. Some examples of student language that illustrate the function of sequencing include:

“This is harmful for... the environment because if we keep using oils... then s- uh soon (O/S) there will be no oil on the Earth...”

The standard FS “if X... soon” leaves a slot between “if” and “soon” for the student to insert original language for responding to the prompt. As can be seen through this example, sequencing not only involves steps in a process or procedure but can also be implemented in situations demonstrating cause and effect relationships.

A further example of sequencing can be seen in the following example:

“I think the Alcatraz escapees [unknown word] didn’t make it (P) because (S) the first reason... um... the tides were like so high (O/S)...”

In this sequence, the student used several language sequences to structure the response. “I think... because” is a standard formulaic sequence that utilizes the function of argue. “Alcatraz escapees... didn’t make it” is an adapted prompt sequence taken from the prompt and adapted by the student, followed by “the first reason X” which is an example of a standard FS used for organizing speech in an ordered sequence and followed by the student’s original material to complete the first reason. The student followed up with a second reason:

“The second reason (S) I think... they um didn't make it (P/O) is because (S) they found... their um... important... um personal belongings in the ocean (O)...”

Again, the student utilized similar standard and prompt FS with original input to formulate an impromptu response. “The second reason” fulfills the function of sequence, while “I think... because” performs the function of argue. “They didn’t make it” is the repetition of the student’s adaptation of the prompt language sequence from the prompt. Students find it helpful to use formulaic language sequences for the argue discourse function to structure speech that focuses on expressing an opinion related to the speech prompts.
The last two less frequent discourse functions in the corpus are focus and summarize. They are located in the initial or final segments of the speech samples to either focus the listener’s attention on the topic of the response or to summarize the response. Some examples from the function of focus from the corpus include,

“One kind of (S) pollution . . . is dead plants and animals . . .”

In this sentence, “one kind of” is the standard FS used by the student to focus the listener’s attention on the topic of the response.

Finally, an example of summarize includes,

“This is why (S) I think (S) Dwight changed (P/O) the most (S) . . .”

In this example, the student used “this is why” as a means of providing a summarizing conclusion to the speaking prompt. “This is why” is a standard FS used in this sample as a function of summarizing. “I think” is an additional standard sequence that is utilized in this utterance as a summary to an opinion prompt. “Dwight changed” is a prompt language sequence taken from the speaking prompt with the name of the character added as an original contribution by the student. “The most” is a standard language sequence that is also found in the prompt that is used in the function of clarify. This example further demonstrates how the student was able to connect a combination of standard and prompt language sequences, as well as insert original material, to produce a coherent and fluent sentence serving the discourse function of summarizing for the response to an argumentative speaking task.

Conclusions and Pedagogical Implications

The results of this study indicate that the use of FS in this set of elementary and middle school EL speech samples was a significant predictor of fluency measured as mean length of fluent run. Through the use of multiword formulas, language learners can possibly utilize the extra time saved in processing to think ahead and formulate an answer, thereby providing additional cognitive capacity to focus on improving other aspects essential for academic speech such as tone and cohesion of the speech output.

The analysis of the discourse function types of the student speech samples further yielded insights into how FS are used by English learners to produce responses to computer-based speaking assessments. The standard sequences often serve as sentence frames, leaving slots for students to insert sequences from the prompts combined with original material created by the student to provide fluent responses. In other words, the sequences function as matrixes that enable them to incorporate complex grammatical constructions in their speech before they might even be fully able to understand how these grammatical structures function in the language, contributing to the improvement in fluency and the perceived proficiency level of a speaker (Boers et al., 2006; Eyckmans et al., 2015; McGuire & Larson-Hall, 2017; Mohammadi & Enayati, 2018; Pawley & Syder, 1983; Rafieyan, 2018; Yan, 2020; Yilmaz & Koban Koc, 2020).

The results point to the importance of incorporating explicit instruction of formulaic language sequences starting in elementary school to help prepare students to produce academic speech in the computer-based speaking assessment environment. Instruction should cover standard language sequences that can be memorized for easier retrieval in the testing situation and emphasize the use of prompt language sequences taken from the speaking tasks themselves. The former involves direct instruction in the implementation of standard formulaic language sequences. In contrast, the latter involves training students on how to break apart speaking prompts to facilitate using the language structures and sequences found in them to plan and deliver the spoken response.

The instruction could also take the form of fluency workshop activities that involve teaching students to recognize the type of speaking task being elicited by a prompt, whether expository or argumentative. Students would then be provided a list of sample standard formulaic language sequences to choose from to construct the response based on the discourse function needed. Implementing other fluency workshop activities such as dictogloss and speech shadowing could also provide scaffolded and structured practice of
formulaic language sequences used in authentic contexts (Lindstromberg et al., 2016; Thomson, 2017; Wood, 2009).

Dictogloss activities provide students with targeted language structure practice by integrating listening, writing, and speaking skills. First, the teacher reads a text containing the targeted FS. Students then listen to the text multiple times, once without taking notes and additional times to note keywords and language structures. Finally, students work collaboratively in pairs or small groups to reconstruct the contents of the text to provide a summary.

Speech shadowing activities involve students listening to a short text containing target FS and practice repeating the text as close to the sample as possible, matching pronunciation, speed, and prosody. Furthermore, implementing speaking task performance exercises such as the ones for which the speaking samples from the data set in this study were collected is an example of providing students the opportunity to practice the skills and knowledge gained from the fluency workshop activities. Students can be presented with sample speaking tasks and required to respond to the task in an environment simulating the testing situation and incorporating the FS practiced during the workshops.

**Limitations and Directions for Further Study**

Evidence tells us that ELs use FS and that these are generally accessed and produced faster than non- (or less) formulaic sequences. The question remains, what does this finding tell us about language learning itself (i.e., the mechanisms by which one becomes more proficient, including increased fluency) if we find that more fluent speech contains more sequences that are frequent (and, by inference, formulaic) in a data set? It is not clear what might be driving these associations, as we do not know whether these word sequences were learned as a sequence or became more formulaic/sequence-like as they were used more and more but had been learned by other mechanisms. This warrants further investigation to analyze the use of FS in speech samples and measure improvement in fluency based on the use of FS.

Another limitation of the study that needs to be acknowledged is the small sample size of the data set. While this small-scale data analysis of speaking performance tasks of upper elementary and middle school English learners in the U.S. public school setting demonstrated significant results of formulaic language use on speaking fluency, the next steps would involve the analysis of speech samples on a larger scale to determine if the results are generalizable across a wider population. Finally, it would be necessary for future studies to examine the effect, especially the long-term effect, of direct instruction of formulaic language sequences on student speech production and oral fluency development. Such studies would be beneficial for designing and implementing informed, evidence-based instructional practices to improve speech output on standardized speaking assessments.

**References**


**About the Authors**

Jennifer François is a K-12 ESL Specialist in a middle Tennessee public school system. Her research in curriculum and instruction investigates the intersection of technology, language learning, and teacher professional development.

**E-mail:** jlm2ad@mtmail.mtsu.edu

Mohammed Albakry is Professor of English and Applied Linguistics at MTSU. His research in language studies is interdisciplinary and often takes up the intersection between language and education, language and identity, and discourse and society. He published numerous peer-reviewed articles in a range of academic journals including *World Englishes* and *Educational Theory*, among others.

**E-mail:** mohammed.albakry@mtsu.edu