Conceptualizing and Measuring Second Language Speaking Proficiency

Lee Shzh-chen Nancy

Abstract

Previous research on second language (L2) speaking proficiency has used different operational definitions and measurements of speaking proficiency (Housen & Kuiken, 2009). In the past decades, researchers have often conceptualized second language speaking proficiency in terms of three constructs: complexity, accuracy, and fluency (CAF). However, the interaction between the CAF triad and the extent to which the constructs represent speaking proficiency have been highly controversial (Skehan, 2009). In addition, different studies have produced controversial results concerning speaking proficiency and its development because different measurements and analytical tools were used (Norris & Ortega, 2003). These controversies in definitions and measurements are problematic because they make comparisons across studies difficult. Therefore, the definitions of speaking proficiency and operationalization of its measurement need to be reconsidered. This paper reviews literature on the conceptualization of second language speaking proficiency in terms of complexity, accuracy, and fluency. It also introduces measurements used by major studies of second language speaking proficiency.

1. Introduction

Currently, English as a lingua franca is becoming the common mean of communication for speakers of different first languages, and this is the foundational reason why English speaking ability is becoming a global educational goal. This thinking also holds true in Japan; however, “Why can’t Japanese students speak English?” is probably one of the most frequently asked questions among educational policy makers in Japan in the 21st century. The Japanese Ministry of Education, Culture, Sports, Sciences and Technology (MEXT) has been endeavoring to develop the speaking ability of Japanese students since communicative language teaching was officially implemented in Japanese high schools in 1989. However, despite continuous attempts, English speaking proficiency development continuous to be a difficult task for many Japanese students (Apple, 2011). Speaking is often considered the most valuable language skill, as being able to speak a target language is often equated with being ‘proficient’ in that language (Hughes, 2011). However, speaking is also often considered the most difficult skill to develop compared to reading, listening, and writing (Gan, 2014). One of the reasons that EFL students struggle to development their oral English proficiency is because of their limited exposure to the target language. Another reason is that many EFL teachers do not know exactly what to teach and how to teach speaking (Hughes, 2011). Research on L2 speaking proficiency

is difficult in part because different researchers have used different operational definitions of speaking proficiency (Housen & Kuiken, 2009). In the past three decades, researchers have often conceptualized oral proficiency in terms of three constructs: complexity, accuracy, and fluency (CAF). However, the conceptualization of these three constructs, interaction between them, and the extent to which they represent second language speaking proficiency are highly controversial (Skehan, 2009). In addition, different researchers have used different measurement instruments and analytical tools and as a result, they have produced controversial results concerning second language speaking proficiency and its development (Norris & Ortega, 2003; Ortega, 2003). The controversies in conceptualization and measurements are problematic because they make comparisons across studies difficult. This paper aims to review previous studies concerning the conceptualization and measurement of second language speaking proficiency to dismantle the ongoing controversies. It hopes to provide researchers, teachers, curriculum developers, material designers, school administrators, and policy an opportunity to reconsider the research and teaching of oral communication English for second language learners.

2. Literature Review

2. 1 Conceptualizing Second Language Speaking Proficiency

Distinguishing more proficient second language (L2) speakers from less proficient speakers has created a continuous debate among researchers (Bygate, 2009; Ellis, 2009; Housen et al., 2012; Iwashita, 2010). Generally, it has been assumed that proficient L2 speakers have the ability to use complex grammatical forms and to speak more accurately and fluently (Ellis, 2009). However, past researchers have proposed varying operational definitions of second language speaking proficiency (Ellis, 2009; Hughes, 2011). It can be conceptualized differently depending on the discourse that speaking is being analyzed such as pragmatics, linguistic, functional, interactional, conversational, and sociocultural discourses (Ellis & Barkhuizen, 2005). From a linguistic perspective, researchers in the past three decades have often conceptualized second language speaking proficiency in terms of complexity, accuracy, and fluency (Ellis & Barkhuizen, 2005). Despite the CAF triad has been widely recognized, the conceptualization of L2 speaking proficiency is still controversial because it is not always clear what CAF indicators entail.

Complexity is the extent to which target language production is elaborated and varied (Ellis, 2003). It is also considered to be the most controversial dimension of the CAF triad (Michel, 2017) as it can be influenced by task difficulty (Robinson, 2001).Complexity can be divided into cognitive complexity and linguistic complexity (Housen et al., 2012). Cognitive complexity is learner dependent and concerns elements such as aptitude, memory span, motivation, and first language background, whereas linguistic complexity is language dependent and concerns elements such as morpho-syntactic structures, rules, and patterns (DeKeyser, 1998). Linguistic complexity can be further divided into four dimensions: lexical (words and collocation levels), morphological (inflectional and derivational

levels), syntactic (sentential, clausal, and phrasal levels), and phonological (segmental and suprasegemental levels) complexities (Koizumi & In'nami, 2014). Skehan (2009) argued that lexical performance needs to be defined as an independent fourth construct on top of the existing CAF triad (Malvern & Richards, 2002; Skehan, 2009). However, the addition of the lexical construct is also controversial because it has also been argued that lexical performance does not determine L2 speaking proficiency as much as it does on L1 speaking proficiency (Skehan, 2009). Therefore, most researchers focused on syntactic complexity when they conceptualize complexity of second language speaking proficiency (Norris & Ortega, 2009).

Accuracy is considered to be a straightforward construct of the CAF triad as it is the ability to produce error-free speech (Lennon, 1990). Error is defined as deviations from the native-speaker norm (Housen & Huiken, 2009). Accuracy was added by Skehan (1989) on top of the previously defined complexity and fluency dichotomy to make the existing CAF triad. Speakers who prioritize syntactic accuracy tend to use the forms they have internalized and therefore can become resistant to using more complex and less familiar target language forms (Skehan, 2009). Although the definition of accuracy is straightforward, there are also controversies as to what criteria are used for the choice of norms and how far away the deviations are from the chosen norms (Michel, 2017).

Fluency was traditionally used as the general indicator of language proficiency as fluent speakers are often considered to be successful speakers (Iwashita et al., 2008). Fluency was characterized by easiness, quality, and smoothness of language production, and it included elements of accuracy and complexity (Hilton, 2008; Lennon, 1990; Riggenbach, 1991). The more recent definitions of fluency focused on the speed of the target language produced naturally in real time without unneeded pausing or dysfluency markers, such as hesitations, false-starts, or reformulations (Ellis, 2003; Michel, 2017). Automaticity is a key component of oral fluency because automatized speakers can more speedily retrieve items from memory, encode grammatical forms, and correct their own erroneous output than less automatized speakers (Segalowitz, 2003). While fluency is also a multi-dimensional construct (Lennon, 2000), it is considered to be relatively uncontroversial compared to complexity and accuracy (Michel, 2017). Nevertheless, some aspects of oral fluency are considered to be more closely related to personal traits than language proficiency itself (de Jong et al., 2015).

To date, there has been no consensus as to which of the fluency, accuracy, or complexity construct is a stronger indicator of speaking proficiency as the weighting of these constructs varies depending on how speaking proficiency is conceptualized (Larsen-Freeman, 2009). For example, accuracy and fluency are closely related because fluent speech entails the application of accurate processing mechanisms in learners (Kormos & Dénes, 2004). The inter-relationships among the CAF constructs are controversial because learners can produce fluent but grammatically inaccurate speech, or speak fluently but lack a varied range of sentence structures, or speak accurately but not fluently

(Housen & Kuiken, 2009; Michel, 2017; Skehan, 2009). Therefore, it is impossible to conceptualize one single construct without referring to the other two as all three constructs are interrelated (Hilton, 2008).

2.2 Measuring Second Language Speaking Proficiency

Second language speaking proficiency has been measured using various qualitative and quantitative methods (Iwashita, 2010). Different researchers have used different measurements to determine the degree to which oral production is complex, accurate, and fluent. Unsurprisingly, there is no agreement as to which measures most accurately measure CAF because there are different learning purposes, learners, and contexts (Housen & Kuiken, 2009; Norris & Ortega, 2009). Most researchers, however, agree that speaking proficiency needs to be measured multidimensionally using multiple constructs and each construct needs to be measured using multiple methods (Norris & Ortega, 2009). Studies in speaking proficiency development which used single measurement dimension (e.g., only fluency) have produced more positive results from the effects of intervention than studies that used multi-dimensional measurements of complexity, accuracy, and fluency together (Bygate, 1996). For example, studies that measured speaking proficiency development using only fluency measurements have generally produced positive results from interventions (e.g., Ahmadian & Tavakoli, 2011; Bygate, 2001). On the other hand, studies that used syntactic accuracy measurements have very rarely produced positive results (see Gass et al.,1999 for an exception). This lack of clarity makes the measurement of speaking proficiency difficult and also complicates comparisons across studies (Ellis, 2009; Housen et al., 2012; Iwashita, 2010). Foster, Tonkyn, and Wigglesworth (2000) attempted to propose reliable measurement units by examining definitions and criteria for selecting measurement units in past studies in four leading SLA journals: Applied linguistics, Language Learning, Studies in Second Language Acquisition, and TESOL Quarterly. They outlined the biggest problem was the lack of definitions and explanation of measurement units. Among 87 studies they examined, only half of those studies provided some definitions and explanations of their measurement units (Foster et al., 2000).

2.2.1 Measuring Syntactic Complexity

It has been theoretically and empirically justified that syntactic complexity needs to be measured using multiple measurements (Norris & Ortega, 2009; Yuan & Ellis, 2003). Most studies measured syntactic complexity using multiple measurements because one component of complexity (e.g., subordination) can stabilize while another component (e.g., global complexity) continues to develop (Scott, 1988). The speaking process starts from the expression of ideas by coordinating and sequencing single words, sentences, and clauses to an expansion by which the subordination is added as a resource to express logical connections of ideas, and finally to the emergence of grammatical

metaphor, which leads to the advanced level of language with complex phrases (Halliday & Mathiessen, 1999). Therefore, three sub-dimensions need to be measured to capture the development of complexity across these processes: global complexity, complexity by subordination, and complexity by sub-clausal or phrasal elaboration (Norris & Ortega, 2009).

Global complexity can be measured using mean length of T-units using the number of words divided by the number of T-units (Mochizuki & Ortega, 2008; Norris & Ortega, 2009). A T-unit is defined as an independent clause and all its dependent clauses, which means that a T-unit can be one independent clause, or it can be one independent clause and one or more dependent clauses combined together (Hunt, 1965). Besides T-units, C-units and AS-units are also commonly used to measure speaking complexity because they are often considered to be more appropriate for analyzing oral data containing ungrammatical segments (Foster et al., 2000; Norris & Ortega, 2009). Complexity by subordination can be measured by counting all clauses in the oral data and dividing them over a specified unit (e.g., clauses per T-unit, clauses per C-unit or clauses per AS-unit) (Elder & Iwashita, 2005). Finally, complexity by sub-clausal or phrasal elaboration can be measured using mean length of clauses (Scott, 1988). However, while it is important to measure syntactic complexity multidimensionally using multiple measurements, more measurements is not always better than fewer measurements. It has been argued that there can be overlaps and redundancies in the syntactic complexity measurement metrics as some measurements are measuring the same elements of complexity (Norris & Ortega, 2009).

2.2.2 Measuring Syntactic Accuracy

Accuracy is considered to be the most straightforward CAF construct as it is a measure of error-free usage of the target language (Housen & Kuiken, 2009; Michel, 2017; Mochizuki & Ortega, 2008). It can also be considered as a measure of deviations from native-speaker norms (Housen & Kuiken, 2009). Grammatical accuracy can be measured in terms of global accuracy (Foster & Skehan, 1996) and specific types of errors (Wigglesworth, 1997). The global accuracy measurement is considered to be the most comprehensive approach to measuring syntactic accuracy because all errors are included despite the difficultly in establishing a consistency in the coding of errors (Iwashita et al., 2008). Global accuracy is often measured by calculating the percentage of error-free T-units or percentage of error-free clauses.

Error free T-units are T-units that are free from grammatical errors, including both specific types of errors as well as other syntactic errors, such as word order errors and the omission of words. Finally, the calculation of error-free T-units can also include syntax, morphological, and lexical choice errors (Iwashita et al., 2008). On the other hand, measurements of specific types of errors have analyzed linguistic features such as verb tenses, third person singulars, plural markers, prepositions, and articles (Wigglesworth, 1997). However, while measurement of specific errors can offer detailed

descriptions of erroneous target forms, they cannot represent learners' holistic accuracy performance (Iwashita et al., 2008). In addition, it is more difficult to generalize the research findings of specific error measurements to other contexts (Michel, 2017).

2.2.3 Measuring Oral Fluency

Fluency is the measurement of smoothness, rapidness, and effortless usage of the target language (Michel, 2017). It is also considered to be the most reliable quantitative measure of speaking proficiency (Kormos & Denes, 2004; Mora, 2006). Oral fluency is commonly measured by speed fluency (speech rate), repair fluency (dysfluency markers), and breakdown fluency (pauses) (Chambers, 1997; Freed, 2000; Lennon, 1990; Mora, 2006). Speech rate is a common indicator of speed fluency and it refers to the number of syllables produced per minute while articulation rate also refers to the number of syllables per minute but excluding pausing time (Tavakoli & Skehan, 2005). Speech rate is considered to be a valid measurement of speed fluency because it considers different word lengths (Kormos & Denes, 2004; Mora, 2006). Mean length of runs is another measure of speed fluency where it measures the average number of syllables between pauses (de Jong, 2016). Despite the cut-off point of pause length is controversial, it is accepted that mean length of run is the measurement of the number of syllables between pauses of 0.25 seconds and longer (Kormos & Dénes, 2004; Lennon, 1990). Repair fluency can be indexed by measurements such as reformulations, repetitions, false starts, and replacements (Skehan, 2003). Finally, breakdown fluency is measured by filled and unfilled pauses.

The role of filled and unfilled pauses in measuring breakdown fluency is controversial as previous studies with smaller number of participants indicated that measuring filled and unfilled pauses can help to distinguish fluent speakers from non-fluent speakers (Freed, 2000; Lennon, 1990; Riggenbach, 1991). However, other studies with larger number of participants found that filled and unfilled pauses do not correlate with overall ratings of oral fluency (Kormos & Dénes, 2004). The measurement of filler pauses include sounds such as *mmm*, *eeeh*, *aaah*, *ano*, and *eto*. Some L2 learners naturally use more filler pauses in their speech than others as it is considered to correlate with their L1 proficiency (de Jong et al., 2015). The measurement of unfilled pauses is more ambiguous as different researchers defined it differently ranging from 0.28 to 3.0 seconds (Riggenbach, 1991; Towell, 2002). For measuring unfilled pausing, mean length of pauses is calculated by dividing the total length of pauses above 0.2 seconds by the total number of pauses above 0.2 seconds (Towell et al., 1996).

3. Conclusion

This paper reviewed existing literature on the conceptualization and measurement of second language speaking proficiency. There are different definitions of L2 speaking proficiency because there are different learning purposes, learners, and contexts. While there are different definitions of L2

speaking proficiency, it can still be conceptualized in terms of complexity, accuracy, and fluency (CAF). However, despite L2 speaking proficiency has been conceptualized, there was no agreement as to which measurements most accurately and effectively measure the CAF triad. Past studies reviewed in this paper suggest that each CAF construct needs to be measured multidimensionally using multiple measurements. Complexity can be measured in terms of global complexity, complexity by subordination, and complexity via sub-clausal or phrasal elaboration. Accuracy can be measured in terms of global accuracy and specific error accuracy, using percentage of error free T-units and the number of specific errors such as verb tenses, third person singulars, plural markers, prepositions, and articles. Fluency can be measured in terms of speed fluency, repair fluency, and breakdown fluency using speech rate, the number of reconstructions (e.g., reformulations, repetitions, false starts, and replacements) as well as unfilled and filled pauses. While controversies over the conceptualization and measurement of second language speaking proficiency continue, it is important for teachers and researchers not to be discouraged by these controversies and to try to work toward dismantling them.

References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, *15*(1), 35-59. doi:10.1177/1362168810383329

Apple, M. T. (2011). *The big five personality traits and foreign language speaking confidence among Japanese EFL students* (Doctoral dissertation). Available from ProQuest Dissertations & Theses database (3457819). Retrieved from https://search-proquest- com.libproxy.temple.edu/pqdt/docview/874271493/FC5C1BA3039417EPQ/1?accountid=14270

Bygate, M. (2001). Effects of task repetition on the structure and control of language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks, second language learning, teaching and testing* (pp. 23-48). London, England: Routledge.

Bygate, M. (2009). Teaching and testing speaking. In C. Doughty & M. H. Long (Eds.), *Handbook of second and foreign language teaching* (pp. 412-440). New York, NY: Blackwell.

Chambers, F. (1997). What do we mean by fluency? *System*, *25*(4), 535-544. doi:10.1016/S0346-251X(97)00046-8

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, *11*(4), 367-383. doi:10.1017/S0272263100008391

de Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, *36*(2), 223-243. doi:10.1017/S0142716413000210

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge, England: Cambridge University Press.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219-238). Amsterdam, The Netherlands: Benjamins.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, England: Oxford University Press.

Ellis, R. (2009). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246. doi:10.1111/j.1473-4192.2009.00231.x

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford, England: Oxford University Press.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323. doi:10.1017/S0272263100015047

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354-375. doi:10.1093/applin/21.3.354

Freed, B. F. (2000). Is fluency, like beauty, the eyes, of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor, MI: The University of Michigan Press.

Gan, Z. (2013). Understanding English speaking difficulties: An investing of two Chinese populations. *Journal of Multilingual and Multicultural Development*. *34*(3), 231-248. doi:10.1080/01434632.2013.768622

Gass, S., Mackey, A., Alvarez-Torres, M., & Fernández-García, M. (1999). The effects of task repetition on linguistic output. *Language Learning*, *49*(4), 549-580. doi:10.1111/0023-8333.00102

Halliday, M. A. K., & Matthiessen, C. (1999). *Construing experience through meaning: A language-based approach to cognition*. London, England: Cassell.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal*, *36*(2), 153-166. doi:10.1080/09571730802389983

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. doi:10.1093/applin/amp048.

Housen, A., Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurements and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam, The Netherlands: Benjamins.

Hughes, R. (2011). *Teaching and researching speaking*. New York, NY: Routledge.

Hunt, K. W. (1965). Grammatical structures written at three grade levels. NCTE Research Report No 3. Champaign I11: National Council of Teachers of English.

Iwashita, N. (2010). Features of oral proficiency in task performance by EFL and JFL learners. In M. T. Prior, Y. Watanabe, & S. Lee (Eds.), *Selected proceedings of the 2008 second language research forum: Exploring SLA perspectives, positions, and practices* (pp. 32-47). Honolulu, HI: Second Language Research Forum.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, *29*(1), 24-49. doi:10.1093/applin/amm017

Koizumi, R., & In'nami, Y. (2014). Modeling complexity, accuracy, and fluency of Japanese learners of English: A structural equation modeling approach. *JALT Journal*, *36*, 25-46.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164. doi:10.1016/j.system.2004.01.001

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 579-589. doi:10.1093/applin/amp043

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387-417. doi:10.1111/j.1467-1770.1990.tb00669.x

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*(1), 85-104. doi:10.1191/0265532202lt221oa

Michel, M. (2017). Complexity, accuracy, and fluency (CAF). In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 66-94). London, England: Routledge.

Mochizuki, N., & Ortega, L. (2008). Balancing communication and grammar in beginning level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research*, *12*(1), 11-37. doi:10.1177/1362168807084492

Mora, J. (2006). Age effects on oral fluency development. In C. Munoz (Ed.), A*ge and the rate of foreign language learning* (pp. 65-88). New York, NY: Multilingual Matters.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555-578. doi.org/10.1093/applin/amp044

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492-518. doi:10.1093/applin/24.4.492

Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes*, *14*(4), 423-441. doi:10.1080/01638539109544795

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27-57. doi:10.1093/applin/22.1.27

Scott, C. M. (1988). Spoken and written syntax. In M. Nippold (Ed.), *Later language development: Ages nine through nineteen* (pp. 49-95). Austin, TX: Pro-Ed.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Malden, MA: Blackwell.

Skehan, P. (1989). *Individual differences in second language*. London, England: Edward Arnold.

Skehan, P. (2009). Modeling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, *30*(4), 510-532. doi:10.1093/applin/amp047

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-276). Amsterdam, The Netherlands: Benjamins.

Towell, R. (2002). Relative degrees of fluency: A comparative case study of advanced learners of French. *International Review of Applied Linguistics*, *40*(2), 117-150. doi:10.1515/iral.2002.005

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, *17*(1), 84-119. doi:10.1093/applin/17.1.84

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*(1), 85-106. doi:10.1177/026553229701400105

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1-27. doi:10.1093/applin/24.1.1