GÖTEBORGS
UNIVERSITET

The Effects of Different Assessment Methods on EFL-Students' Speaking Performance

A Literature Review

Alicia N. Rådeström & Wilma Åkesson

Ämneslärarprogrammet

Abstract

Nowadays, it can be argued that it is becoming increasingly more important to be proficient in speaking the English language due to globalization. Examining teaching and assessment of oral proficiency is therefore important for establishing how EFL-learners' speaking skills are most effectively taught and assessed. All assessment methods can be categorized as either holistic or analytic, where the teacher either focuses on the overall impression of a student's performance, or evaluates the student in depth through analyzing their performance in relation to different linguistic dimensions. Due to the dominance of the communicative approach in recent years, interactional assessment methods have emerged. Dynamic assessment and peer-assessment involve an active engagement from teachers and students. This literature review explores the benefits and drawbacks of holistic assessment and analytic assessment, and reviews dynamic and peer-assessment as assessment methods for speaking, and what impact they can have on EFL-learners' speaking performances. A total of 15 experimental studies are reviewed and the results show that the holistic approach offers a less time-consuming assessment, while the analytic demands more time. However, the analytic assessment illustrates what areas of development the students have more clearly than the holistic assessment. Furthermore, it is demonstrated that utilizing dynamic assessment and peer-assessment enabled the students to engage in peer-interaction, self-reflection and become more independent in their learning process. It is therefore implied that dynamic assessment and peer-assessment improved the participants' speaking performances.

Table of Contents

1 Introduction

Assessment is a multifaceted area that encompasses many different approaches depending on countries, proficiency levels, age, etc. Assessment can be summative or formative, and the grading in assessment can vary to be more holistic or analytic. In relation to this, assessing and developing the speaking ability have been argued to be one of the most challenging parts of English language teaching and learning (Qayoom et al., 2021; Arroba & Acosta, 2021). The speaking skill is also one of the least tested in a European setting (Borger, 2018). Thus, assessment of oral proficiency can be seen as highly complex, and we therefore consider it of interest to examine how assessment can be employed to improve students' speaking performance.

The investigation on how different Assessment Methods (AM) can be used to enhance EFL-learners' oral performances is not widely researched and to our knowledge there are no previous literature reviews in this field, which is also why we decided to cast a wide net that includes Holistic (HA) and Analytic (AA), as well as Dynamic Assessment (DA) and Peer-Assessment (PA). Considering that oral assessment is a wide-ranging area and it can be viewed from many different perspectives, HA and AA is represented in our literature review since any assessment can be placed on a cline from more or less holistic and analytic. Because of the dominance of the communicative approach in recent decades, the AMs in this paper have been limited to DA and PA, which are facilitated by interaction. Although the studies will be presented under the headings mentioned above it should be acknowledged that there is overlap in that HA and AA applies to all studies.

The findings of this study firstly reviews the concepts of HA and AA. Secondly, we look into DA of speaking and how this can serve to benefit EFL-students. Thirdly, PA and its effect on speaking skills is investigated. Finally, the findings are discussed to determine the current state of knowledge and what impact this might have on teaching and assessing the speaking ability in EFL classrooms.

1.1 Aim and Research Questions

The aim of this literature review is to investigate how different assessment methods affect EFL-learners' performance in speaking exams. To do this, we have formulated three research question that will aid us in answering this:

- What is the difference between using a holistic or analytic assessment of EFL-learners' speaking skills?
- What impact can dynamic assessment have on EFL-learners' speaking skills?
- What impact can peer-assessment have on EFL-learners' speaking skills?

1.2 Method

The literature in this paper was gathered through conducting searches on the ERIC database and Google Scholar with the peer-review filter on. The search words we used were: assessment speaking EFL/ESL; assessing speaking; assessment, EFL OR ESL OR English as a second language OR English language learners, method*, speaking skills or speaking ability; (feedback speaking), EFL; holistic rubrics EFL speaking. A few of the studies were collected by inspecting the reference lists of the studies we found.

The articles that corresponded to our topic, based on their titles and abstracts, were carefully read. The articles that were read and chosen to be represented in our literature review, are the ones that we considered most relevant to our research questions. Further, in our selection of studies, we decided on an EFL-context rather than ESL since it is more applicable to a Swedish context. Although some argue that English in Scandinavia is moving in the direction of ESL, English is not the language of government as in typical ESL locations, e.g. Canada or Jamaica (Graddol, 1997).

All of the studies in this paper are conducted through experimental procedures and based on university students, except for Khabbazbashi and Galaczi's (2020) study, although university level students was not a search criteria. Additionally, the studies originate from a wide range of countries in the world. The selection of using studies with university students can be defended by how oral proficiency in English is not equal around the world. We can expect Swedish upper secondary school students to be more proficient than many university students in other parts of the world, based on the EF English Proficiency Index where Sweden is ranked at seventh place out of 112 countries (EF, 2021).

2 Background

In the description of the purpose of the English subject in Swedish schools it is stated that one goal is to develop students' communicative ability. This ability is multifaceted since it encompasses receptive, productive and interactive skills. The students are expected to express themselves with variation and complexity and in order for the students to develop these abilities, the teacher needs to employ a variety of speaking tasks for assessment (Skolverket, 2020). Furthermore, it requires practice to improve one's EFL speaking skills which strengthens the notion about using more than one AM in EFL-classrooms (Karpovich et al., 2021).

The speaking ability in English can be seen as a significant skill in our globalized world, where the ability to communicate with people of different languages has become increasingly more vital compared to 20 years ago, not least due to the dominance of the communicative approach since the 1970s and 80s and the realization of this action-oriented view of language in the CEFR. The ability to speak English proficiently can not only serve to improve one's performance in the school setting, but it can also contribute to a bigger social network, more job opportunities, facilitate communication with people when traveling in other countries, among other things. Considering that the speaking skill is vital to learn, it is of great importance that we examine how oral proficiency is most effectively taught and assessed in classrooms.

In order to understand why the topic of this literature review has been chosen it is important to be familiar with the terms validity and reliability since they are the pillars of assessment, and both terms signify a type of measurement. Validity measures accuracy, e.g. if an exam tests what it is said to examine, and reliability measures consistency, i.e. if an exam produces the same results when conducted several times. An exam must be both valid and reliable for it to provide students with a fair testing situation; if the exam does not measure what it is supposed to, they do not have the possibility to pass the exam, and if an exam is unreliable, it leads to the students receiving an inconsistent exam (Hughes, 2003).

Another important factor regarding assessment of exams, is inter-rater reliability which means that several raters come to a similar conclusion when assessing the same exam. Low inter-rater reliability does not necessarily have to mean that an exam should not be used since raters' preferences differ, however, it is an indicator that an exam might not be as valid and reliable as it should be. The process of ensuring validity and reliability is what makes it difficult to assess, excluding other existing factors in the classroom (Luoma, 2004).

3 Findings

This chapter contains an investigation of experimental studies that examine assessment of speaking and the methods related to this. To start with, holistic and analytic assessment are reviewed because all AMs, regardless of their approach, are considered to be either holistic or analytic. Second, we have divided the other AMs, DA and PA, into two sections where research is presented accordingly in order to define each method and its effects, and to provide a comprehensive overview of the patterns in holistic, analytic, dynamic and peer-assessment, and how they are connected.

DA and PA are the remaining methods covered in this chapter and they can be analytic and/or holistic depending on which feedback-approach that is utilized and all three sections can be connected by looking at how they complement one another. When PA is used within DA, it functions as a component in the analytic area of DA which makes it possible for the teacher to apply DA in a more holistic manner, provided that the PA is analytic in order to encompass both HA and AA, and vice versa. Therefore, this chapter is of importance to investigate for establishing if and how they can be integrated in EFL speaking assessment.

3.1 Holistic and Analytic Assessment

In this section, studies on HA and AA methods will be presented. According to Khabbazbashi & Galaczi (2020), HA is used when giving an overall score for a student's performance without going into detail for different speaking components. Further, it is indicated that different speaking components are expected to develop at the same time, otherwise the entire score is affected if one component is not as proficient as another. In contrast, Khabbazbashi & Galaczi (2020) state that AA focuses on all speaking components individually which means that if one speaking component is more proficient than another, it does not have as high of an impact on the entire grade because all criterions are aggregated which leads to a more comprehensive grade since the zone of proximal development (ZPD) is not considered in HA.

Metruk's (2018) study included 60 Slovakian EFL-teaching students that were assigned different CEFR-levels based on an unprepared C1-level 15-minute interview with an interlocutor (HA) and an assessor (AA). The grading criteria were: content and organization, grammar, pronunciation, and vocabulary. The aim was to note any statistical difference by aggregating the test-scores from the HA and AA.

The results show that HA leads to a higher grade than other AA-methods, however, Metruk (2018) states that this result does not have to mean that AA is better than HA since the two methods complement each other and can thus be used simultaneously. Nevertheless, HA does not provide students with areas of improvement since it is not as detailed as AA. Metruk (2018) mentions that the results of this study is limited by the assessors and their assessment process as well as their perception of oral proficiency. This is, however, an issue that can be applied to all AMs.

Other than the holistic approach, Khabbazbashi & Galaczi's (2020) introduce another AM called part-marking-models (PMM). PMM is a method that assigns a grade to separate parts of a skill and is similar to AA, however, PMM can be applied to both HA and AA. Thus, PMM is different from AA in the sense that PMM focuses on different parts of an exam while AA focuses on grade criteria, which is more generalizable than PMM but more detail-oriented than HA.

The purpose of the study was to assign participants a CEFR-level by grading performances using a six-level holistic scale covering "coherence and discourse management, language resource, pronunciation, and hesitation and extent" (Khabbazbashi & Galaczi, 2020:341). Each testing part was graded in a holistic manner and the results show that HA is slightly more generous than PMM when grading performances, i.e. HA-grades led to a higher CEFR-level being assigned than the PMM-grade. According to the raters, assigning a holistic grade was quite challenging since individuals rarely fit a single level. Thus it was concluded that HA is not as reliable as PMM due to its generalizability and generosity, which is also supported by Metruk (2018).

Tuan's (2012) study was conducted in Vietnam during four months with 104 university students divided into an Experimental Group (EG), the AA-group, and a Control Group (CG), the HA-group, where both groups had the same, prior the exam, unfamiliar teacher to ensure neutrality. The instruments consisted of a pre- and post-test, a six-scale questionnaire on students' attitudes towards AA, and six speaking tests, the results of which were added up as the final score for the post-test. The analytical scale used for the scoring of the tests contained the following criterias: coherence, content, grammar and structure, language used (e.g. vocabulary), and organization (Tuan, 2012, p. 675).

The results of this study show that the CG did not improve their speaking skills to the same extent as the EG, nor did the CG improve significantly between the speaking tests. This can be explained by the fact that the EG received feedback after every speaking test and, thus,

they were aware of what needed to be improved which made it possible for the students to adjust their strategies in accordance with their weaknesses i.e. ZPD. This means that the EG improved the most from pre- to post-test. The improvement regarding the five criteria was significant for the EG, results for the CG were not disclosed. In conclusion, Tuan (2012) presents that AA has a strong connection to the improvement of EFL-students' speaking skills. Further, he argues that the assessment of students' speaking skills should be viewed as a process rather than only focusing on the end result.

3.2 Dynamic Assessment

DA is an AM originating from Vygotsky's theory about the ZPD within the sociocultural theory which serves as the pillars of DA in the classroom (Çetin Köroğlu, 2019; Ebadi & Asakereh, 2017; Fahmi et al., 2020; Safdari & Fathi, 2020). DA is a method that has to be continuous in order for it to be effective since its purpose is to have constant communication between teacher and student where "teachers act as promoters and provide immediate feedback to whole procedures of a task" (Xiaoxiao & Yan, 2010:25) in order for the students to be able to reach their full potential (Çetin Köroğlu, 2019; Ebadi & Asakereh, 2017; Fahmi et al., 2020; Safdari & Fathi, 2020). The employment of DA can be adapted to fit various tasks but is always dependent on each students' proficiency level and feedback is given accordingly. According to Ebadi and Asakereh (2017), DA is considered a relatively new AM and is distinguished by the interventions of teachers while the learning is occurring.

Chou (2021) carried out the first study in this section, and it is based on the dynamic Teaching Speaking Cycle by Goh and Burns (2012). The participants were 60 Taiwanese university students (B1), divided into an EG and a CG. This study aimed to conclude how the seven steps of the Goh and Burns' cycle (see Appendix) affect different speaking components in the EG and CG. The purpose of this cycle is to improve students' overall speaking ability and in order to do so, students need to be aware of their ZPD, i.e. what they need to improve in order to achieve a higher level of proficiency. This may include accuracy, fluency, communication strategies, and self-correction (p. 21).

The methodology of this study included a quasi-experimental pre-, mid-, and post-test design based on the topics in the course book conducted throughout 18 weeks where 15 weeks were dedicated to dynamic pedagogical employment. The EG received the DA method and the CG received a communicative language teaching (CLT) approach, although both groups used the same textbook. In order to track the students' speaking progress, a

6

questionnaire was used where the scoring was based on an analytic rating scale consisting of the following components: content, fluency, grammar, pronunciation and vocabulary (Chou, 2021). Chou claims that these components are important since they have a strong connection to the three main characteristics of speaking: accuracy, complexity, and fluency. The students' performances during all tests were assessed by two teachers using the same rating scale and the performances were recorded to guarantee inter-reliability (p. 23).

Before the teacher intervention began, a test was conducted to examine how the EG and CG used compensatory, interactional, and rehearsal strategies. The results indicated that there was no significant difference between the EG and CG regarding the previously mentioned strategies, speaking proficiency, and results on the pre- and post-tests.

When inspecting to what extent the students' speaking proficiency had improved the pre-, mid-, and post-tests were compared. It was revealed that the employment of DA led to a significant increase in proficiency regarding all five components, mainly fluency and vocabulary, while the CLT-approach led to a significant proficiency improvement mainly regarding vocabulary. However, the overall results show that both groups improved significantly during the first nine weeks of the semester. After the ninth week the progress for the CG stayed at the same level while the EG continued improving substantially.

The study found that using DA was more effective than non-DA. The EG was provided with mediation and scaffolding to a higher extent than the CG and the results show that teaching without explicit instruction or assessment leads to limited progress, both regarding skills and long-term improvement. However, the Goh and Burns' cycle, and DA in general, is lacking in the fields of improving grammatical accuracy and pronunciation which means that additional learning may be needed in order to reach an even higher proficiency level (Chou, 2021:30-32).

Arfaei Zarandi and Rahbar (2016) conducted the second study of this section on how scaffolding strategies affect the speaking ability of 60 Iranian EFL learners divided into an EG and a CG. The participants were selected through the Oxford Placement Test (OPT) and to establish the students' speaking proficiency level, a speaking section of a Preliminary English Test (PET) was conducted as a pre- and post-test.

The duration of the study was 10 sessions where accuracy and fluency were assessed through different methods that were applied to the pre-test by two raters, which resulted in high inter-reliability. The EG was exposed to several scaffolding practices, e.g. creating and

working towards a goal and receiving appropriate feedback and support, while the CG was exposed to a standard speaking session with routine strategies and teacher interaction.

The results of this study were calculated by comparing the results from the pre- and post-tests, within and between the groups, and it is concluded that the implementation of scaffolding is highly beneficial. The authors conclude that using scaffolding might require a smaller amount of time for the students to significantly improve various areas of English-speaking learning. The authors suggest that intervening scaffolding can be used in several areas of teaching speaking since it is part of communicative English teaching, but doing so increases the workload for teachers additionally.

Ebadi and Asakereh (2017) conducted the third study which researched how DA affects the development of speaking skills and how the effects differ between an advanced learner and a beginner from Iraq and Iran. Ali was proven to be an advanced learner (AL) through the rating on the IELTS scale based on an interview. Ali had been exposed to English from an early age whereas Reza, the beginner (B), had only learnt English in school, therefore Reza viewed himself as a beginner since his speaking skills were quite restricted. The most frequent error for the AL was prepositions, and past tense for the B (p. 4).

The methodology of this study consisted of four stages of DA teaching as well as three interviews during a two-month period. First, the students were interviewed in order to establish their ZPD through a non-DA and a DA pre-test. Second, they took six DA-speaking tasks which were followed by a non-DA and a DA post-test. Stages two to five consisted of a non-DA and a DA diagnostic session, concluding in an unstructured interview and a transcendence session. Everything was recorded, transcribed, and analyzed with the intent of allowing the researchers to closely observe the DA's effect (Ebadi & Asakereh, 2017:6).

Reza's results indicate that DA improved his accuracy significantly when using past tense over time. Additionally, mediation tools (e.g. repetition, explanations, clarification requests) provided by the teacher were needed less over time while the responsiveness to the mediation improved. Furthermore, Reza's self-regulation significantly increased.

For Ali, the results show that there was marginal improvement regarding correct use of prepositions, however; it was also shown that Ali employed overgeneralization which led to fewer correct expressions. The utilization of and responsiveness to mediation, as well as his self-regulation increased throughout the course.

In conclusion, it can be observed that DA had a positive impact on both participants' learning process and, most importantly, their level of self-regulation. DA had a bigger impact

on the B, however, a disclaimer was that past tense may be easier than prepositions due to the rules being more defined regarding past tense. By any means, Ebadi and Asakereh (2017) emphasize that teachers should not use a generalized teaching method for their students since it would not benefit all proficiency levels (p. 16). A disclaimer that this study cannot be generalized is put forth but they do advocate that its results can be used in order to obtain a better understanding for implementing DA on individual students since only two participants are included.

The fourth study of this section was conducted by Safdari and Fathi (2020) where the purpose was to investigate how DA affects fluency and accuracy among 63 pre-intermediate EFL-learners in Iran. The study was conducted during three weeks and a total of eight sessions.

Like Arfaei Zarandi and Rahbar (2016), Safdari and Fathi (2020) used the PET-test with the purpose of including participants whose test scores fell within the set limit. Additionally, they used pre- and post-tests in order to further establish their speaking proficiency and they assessed the students based on two scales that measured fluency and accuracy. Furthermore, the authors employed a self-study assignment called "English Vocabulary in Use" with focus on grammar and vocabulary. Lastly, they conducted interviews where the students were examined in correlation with the effectiveness of DA (pp. 2 & 5-6).

The 63 students were divided into an EG and CG where the EG was exposed to DA while the CG received non-DA by the same teacher during the treatment. The sessions for the EG was conducted through "step-wise mediation from the most implicit type of feedback to the most explicit type" (Safdari & Fathi, 2020:7) whereas the CG received no mediation at all. If the EG made a mistake it was mentioned and assessed but if the CG made a mistake it was simply written on the board and a correct explanation was provided. In the EG, their individual ZPD was met, in contrast to the CG where it was not considered. After the sessions had been completed both groups took the post-test.

The results of this study show that the EG had significantly improved their accuracy while the improvement for the CG was marginal which means that there is a strong correlation between DA and positive effects. However, there was no significant difference between the EG and CG's results regarding fluency. In conclusion, DA affected speaking accuracy to a higher extent than it did fluency and it is evident that DA aided the improvement of accuracy for the EG, this due to mediation being provided when necessary (Safdari & Fathi, 2020:12-14). The authors put forth that further research could include, e.g.,

another age-range, and focus on fluency and accuracy in order to make the conclusions more generalizable together with their own study.

The two remaining studies in this section address student attitudes rather than a measurement of the effects of DA on students' performance, in contrast to the other studies in this section. The combination of the two approaches is relevant because students' attitudes and opinions act as a basis to the successfulness of an AM. If students are positive towards a method it leads to a better teaching and classroom environment which in turn elicits better performance, i.e. a positive attitude equals a positive enhancement in oral proficiency. Thus, a study presenting the students' attitudes in combination with studies that measure improvement offers a comprehensive coverage of the benefits of employing DA.

The fifth study was conducted by Fahmi et al. (2020) who researched how DA affected the eagerness to improve the English-speaking ability of four Indonesian university students, selected through purposive sampling. The methodology included "stimulating information, pretest, post-test, feedback and knowledge expansion, and semi-structured interviews" (p. 781). Furthermore, the DA sessions were recorded in order for the students to increase their self-regulation by watching their own performances and analyzing their thinking process. By doing so, the students became aware of their own errors and were forced to find ways to correct them when speaking. This was the researchers' method to define the ZPD which functioned as a base for the DA-teaching.

The results show that confidence among the students increased which made them more eager to speak even though there was only a slight improvement in terms of correctness. However, the students' fluency shows a significant improvement when comparing pre- and post-tests. Feedback from the teachers was given after every self-assessment occasion which made this improvement possible by: first, having the students assess their own speaking; second, their ZPD being declared; and, third having the opportunity to improve their speaking skills. Thus, the implementation of DA aids students in their self-awareness, self-correction, and self-regulation which emerged from comfortable interactions between student and teacher (Fahmi et al., 2020:782-785). Additionally, the researchers comment on how teachers should avoid underestimating students with low-proficiency speaking skills since students' knowledge cannot be based solely on numbers; the classroom environment and teaching-methods have a big impact on students' development.

Finally, the sixth study to be covered here, was conducted by Çetin Köroğlu (2019) on how interventionist DA affected the testing of speaking skills with 29 Turkish EFL-teacher

students as test subjects that were chosen through a checklist based on the Foreign Language Classroom Anxiety Scale about their attitudes regarding speaking English. Three of the students' teachers applied DA in their teaching and documented their findings.

In this study, 11 different speaking tasks were given to the students, who then used 11 different speaking techniques in groups of four to five people. In order to use the DA method, related vocabulary and a video and dictionary for anxiety regulation were provided to the groups prior to the exam. The tasks included in the exam were the following: argue for an opinion, asking and answering questions, conduct interviews, create stories, explain and give examples, give oral reports, look and describe, role play, and summarize seven different topics. After the exam, written interviews were carried out in order to collect the students' thoughts on DA and its effects on their performance (Çetin Köroğlu, 2019).

The results on the anxiety questionnaire show that students feel more anxious when taking high stakes speaking exams individually than as part of a group. Additionally, it can be concluded that the interventionist DA-tools that were employed positively affected the students' anxiousness and perceived performance. Further, the students' opinions regarding DA were mainly positive due to being given more time to think before speaking, resulting in fewer errors. Moreover, the questionnaire demonstrates that the students preferred DA exams over non-DA exams (Çetin Köroğlu, 2019:26-28).

In conclusion, this study reveals that DA exams are more beneficial to students than traditional exams considering the preparations that were given as part of the exam, it being a group exam, and more time for thinking being provided. This aids in decreasing anxiousness among students since it allows them to adjust the information to their individual ZPD which, in turn, enables them to "reveal their full potential during their speaking performance" (Çetin Köroğlu, 2019:29). In other words, it was found that interventionist DA provided a positive and more comfortable test environment and this suggests that DA can be more effective than other techniques when teaching and assessing speaking. However, DA demands more time and planning for teachers which, in that sense, makes it impractical.

As can be seen in the studies mentioned, DA is proven to have a positive effect on students' general speaking ability in different cultural contexts, age groups and employment methods. In the following section, PA will be addressed which can be connected to DA and scaffolding. When implementing PA, DA is also implemented since it aids the students in recognizing their areas of improvement on their own, but this is supported in PA by peer-feedback rather than teacher-feedback. Thus, the students figure out their zones of

proximal development in a meta-cognitive manner which is also supported in a DA setting, but is not usually the case in non-DA teaching methods.

3.3 Peer-Assessment

PA is often mentioned as an effective tool in language learning which involves an activity where the students provide their peers with feedback on their work. This section reviews six studies that investigate AMs of speaking that make use of PA. Four of the studies use digital devices to facilitate the PA activities.

The first study in this section was conducted at a Russian university and involved 274 first-year students with the intent to look at what impact monologue speaking tasks (MST), paired with peer-interaction and PA, have on oral skills in the English language. The participants were divided into two groups, a CG and an EG where only the EG received treatment and the participants in both groups had approximately the same level of English at the beginning of the period (Karpovich et al., 2021).

The experimental period consisted of four stages, where all participants performed four monologues orally and were graded according to six categories: content & organization, vocabulary, coherence, grammar, fluency & pronunciation, and presentation of MST. The type of MST used in this study focused on descriptions and narratives and had a wide range of subjects (Karpovich et al., 2021).

At the initial stage, the CG was given the task to prepare and present their individual speaking monologue. After two weeks, the CG presented their monologues and were evaluated by the teacher with criteria that were derived from CEFR. The EG had an introductory lecture on the MST and was then given the same task as the CG. Before presenting their monologues, the EG exchanged their written monologues and gave feedback to each other. The evaluation of the presented monologues in the EG was not only conducted by the teacher, but also by the other students in the EG. However, the final score of the speaking task was only based on the teacher's assessment (Karpovich et al., 2021:7-8).

The results of the study show that the EG had substantially better results than the CG. The authors point to the active engagement of peers and teachers as being one factor in this result and they state that this affects the development of speaking skills in a beneficial way. However, the results do not convey which of the teacher- or peer-feedback was most effective. The study shows that peer-interaction can serve to motivate speaking for higher-level learners as well as lower-level learners. Additionally, the authors refer to several studies that show that

speaking anxiety can negatively impact students' oral performances and therefore their overall proficiency in the English language. It is concluded that peer-interaction and assessment is central in terms of improving first-year students' speaking skills (Karpovich et al., 2021).

Another study on speaking assessment looks at the effectiveness of mobile-based peer-assessment (M-PA) on learners' oral proficiency. This study, conducted by Chang and Lin (2020) at a university in Taiwan, involved 60 EFL-students that were divided into one CG and one EG.

The CG used the traditional teacher-only AM, while the EG employed the M-PA method. The students in the EG uploaded their work to a database and rated and gave comments on their peers' work during classes by using mobile devices. The teacher had access to a mode where they could supervise the progress of the PA process. The rating scale that was used included six criteria: fluency, accuracy, grammar, vocabulary, content factors and presentation. The EG participated in an informal interview at the end of the study to collect the participants' views on M-PA. After 12 weeks, a post-test was administered to both groups (Chang & Lin, 2020:672-673).

In terms of improvement of oral proficiency, the post-tests show that the EG outperformed the CG which indicates that the M-PA method enabled the students to learn substantially more than the students who engaged in the teacher-only AM. The authors point to two possible explanations for the improved performances in the EG. Firstly, the students actively engage in an interactive PA process that allows them to recognize their own areas of concern and reflect on how they can continue their development in those areas. Secondly, the teacher was able to view the students' feedback and pinpoint areas that needed extra attention and could therefore guide the students in overcoming these issues (Chang & Lin, 2020).

The informal interviews conducted with the students in the EG show that they in general hold positive attitudes towards the M-PA method and they appreciate the real-time feedback and self-reflection. However, almost all of the participants in the EG view the in-class M-PA as demanding and time-consuming, and some of the participants expressed discomfort regarding providing their peers with feedback on their weaknesses. Thus, it is of significance for the teacher to explain the purpose and value of peer-feedback to the students. Lastly, the authors consider it significant to acknowledge that the teacher has to organize and plan these assessments carefully and the possibility of technical difficulties when there are digital tools involved (Chang & Lin, 2020:676-677).

The next study in this section was conducted by Safari and Koosha (2016) and investigated the effects of speaking video-portfolios as an alternative AM on Iranian EFL-learners. 64 participants were selected for the study based on their OPT-test scores and were divided into four groups: an intermediate learner CG, an advanced learner CG, an intermediate learner EG, and an advanced learner EG.

At the beginning of the study, the participants took a 15-minute speaking test to confirm that the intermediate and advanced students had similar oral proficiency. After three months, the students retook the same test to compare the differences between the results of the groups. The participants in the EGs were also asked to answer questionnaires on using speaking video-portfolios. The only difference in teaching between the groups was that the EGs used speaking video-portfolios paired with peer- and self-assessment. In the EGs, the teachers recorded the speeches and the students provided feedback to themselves and their peers based on the recordings (Safari & Koosha, 2016:104).

The results of the study show that both the intermediate and advanced EGs displayed better oral proficiency than the respective CGs at the end of the treatment. Safari and Koosha (2016) conclude that a relationship between video-portfolios and improved speaking skills in intermediate and advanced learners can be established. Video-portfolios allow the students to monitor their progress to see what they need to develop. The student questionnaires revealed that the EGs conceived the PA activity as useful since it helped them distinguish their problem areas (Safari & Koosha, 2016).

Furthermore, the students view their performances "as a part of the developmental process rather than an end point" (Safari & Koosha, 2016:109). The students consider video-portfolios as a way to practice speaking, rather than an exam, which seemed to decrease their speaking anxiety in class. Thus, speaking portfolios can be utilized instead of traditional speaking exams to decrease anxiety in EFL speaking classes. The findings indicate that video-portfolios can positively influence the students' speaking skills and nearly all students believe that the benefits of using video-portfolios outweigh the drawbacks.

Additionally, the authors emphasize the fact that it is necessary to be well-organized and prepare the teaching sequence far in advance, as well as to properly introduce the students to the PA for them to provide their peers with appropriate feedback (Safari & Koosha, 2016:110).

The fourth study examines the effects of receiving teacher- and peer-feedback on oral performances. The study included 33 EFL-participants in a Turkish university and involved a

pre- and post-test on speaking performances, and a 24-week treatment. The students' speaking self-efficacy levels were also measured using a self-efficacy scale for foreign language speaking (Au & Bardakçi, 2020).

The participants were divided into two groups, group A which utilized peer-feedback, and group B which made use of immediate teacher-feedback. The first speaking test confirmed that the participants were equally proficient in terms of speaking skills. Group A was provided with training sessions on peer-feedback for a week before the treatments began (Au & Bardakçi, 2020).

The pre- and post-test were conducted in the same way where the participants in both groups performed a two to three minute presentation. The teacher provided the students with a topic and evaluated them based on three aspects: language use, delivery, and topic development. All presentations were recorded in case of different scoring of two raters to ensure inter-rater reliability. The peer-feedback group was divided into smaller groups and gave their group members oral feedback after each presentation. This group did not receive any teacher feedback. Group B was assessed and given immediate teacher feedback after their presentations (Au & Bardakçi, 2020, 1458-1459).

Comparing the results of the pre- and post-test, it is demonstrated that both groups improved significantly in their oral performances during the experimental period, where group B had improved their performances more considerably than group A. The authors find the fact that group B performed better than group A to be interesting, since a similar study mentioned in their paper noted a higher oral performance achievement with a PA-only group (Au & Bardakçi, 2020).

The study noticed that feedback is a factor in a higher level of speaking self-efficacy level, which in turn contributed to a better speaking performance. It is thus suggested that increased self-efficacy can promote learners' speaking development and performances in oral exams (Au & Bardakçi, 2020:1464).

Lastly, the authors suggest that students should be provided with more opportunities to practice speaking to develop both their speaking ability and their speaking self-efficacy. Furthermore, the students should be given the opportunity to receive and give feedback to their peers since the results indicated that this can influence the learning experience positively, even though teacher-feedback was proven to be more effective. Additionally, Au & Bardakçi (2020:1466) suggest that teachers should guide students through the process of PA and administer proper feedback-training sessions to ensure the effectiveness of PA.

The next study investigates the impact of digital formative tests on EFL-learners' speaking skills development and compares it to traditional speaking tests. The 52 participants from a Turkish university took two traditional speaking tests and two digital formative tests and were graded based on speaking skills assessment rubrics that included: fluency, accuracy, grammatical structure, pronunciation, and vocabulary (Çetin Köroğlu, 2021).

In the first two test-phases, the participants took two traditional speaking assessment tests. The traditional speaking test consisted of a five- to ten-minute presentation on a given topic, after a five-minute preparation. The last two test-phases consisted of two digital formative assessment tests. At the end of the last phase, the students were handed a written form to gather their opinions on digital formative assessment (Çetin Köroğlu, 2021:110-111).

In advance of the digital formative test, the participants were introduced to various tasks such as role-play, discussion, etc.. The digital test was initiated when four videos with related vocabulary charts were uploaded to the database. The students had two hours to select and prepare a speech about one of the videos before they recorded their speeches. The difference between the summative and formative tests in this study was that "formative assessment emphasizes the learning process rather than the learning product by providing feedback" (Çetin Köroğlu, 2021:107). Peer- and self-assessment are mentioned as useful parts of formative assessment which engages the students in their own learning process.

The findings of this study indicate that digital formative tests improve students' speaking skills and, more specifically, their fluency and accuracy skills, as well as enhancing their vocabulary. Through the frequent testing of this method, the students were exposed to more testing than usual which can be challenging. However, the students practice more which is essential for improving productive skills. The students were also able to focus on the learning goals rather than performance goals (Çetin Köroğlu, 2021).

Unlike the traditional speaking tests, which are of the summative assessment nature, the digital formative test serves the purpose of both learning and assessment, and how to further aid students' learning development. Çetin Köroğlu (2021:119-120) states that the AM requires careful planning and preparation before each assessment cycle, which might result in an extra burden for the teacher. Similarly to the M-PA study, the digital formative tests might entail encountering some technical issues.

The final study investigates student perceptions of using mobile-videos in an EFL presentation class, in contrast to aforementioned studies that measured the improvement of participants' oral performances. It is reviewed in this section since it includes studies on how

mobile-videos can be used to enhance EFL-learners' presentation skills through self- and peer-reflection (Toland et al., 2016).

The study was conducted in a Japanese university where the students tried mobile-video presentations combined with peer- and self-reflective activities. The students were first introduced to model presentations and evaluated these in groups. The students were then able to practice their presentations before a peer recorded their final presentation and uploaded it to the university's learning management system (LMS). Poster sessions are described as "a type of guided interpersonal communicative activity where there are a number of short, interactive student presentations" (Toland et al., 2016:183). After uploading the video, the students engaged in a self- and peer-assessment activity in the LMS. Finally, a questionnaire investigated the participants' perceptions of mobile-videos, self- and peer-reflective feedback, and the benefits and drawbacks with the AM.

Various studies are mentioned by Toland et al. (2016:181-182) that support the fact that mobile assisted language learning can improve EFL-learners speaking skills. It also provides the students with the possibility to be given individualized instruction and feedback.

The 129 respondents of the questionnaire consider the ability to watch their video performances multiple times as a significant advantage and that it helped them to find areas of improvement. Most of the participants view the feedback from PA as helpful and feel that mobile-videos helped them improve their graded presentations. On the other hand, the questionnaire revealed that some of the students experienced anxiety during the video recording and expressed concerns regarding others using their devices. The study states that technical difficulties should also be considered when using this method. Despite this, the authors conclude that the students recognize multiple advantages with using mobile-devices for peer- and self-assessment, and consider this to be an effective tool for teaching and assessing English communication skills (Toland et al., 2016).

The studies in this section employ various tasks and research procedures, but all of the studies make use of peer- and/or self-assessment and they all state that PA has a beneficial impact on students' speaking skills improvement.

4 Discussion and Pedagogical Implications

In this chapter, our findings are first examined in relation to our research questions. Secondly, our findings are analyzed and discussed to illustrate the pedagogical implications that can be derived from our literature review.

4.1 Discussion of Research Questions

I. *What is the difference between using a holistic or analytic assessment of EFL-learners' speaking skills?*

The studies presented show that HA provides a more general perspective on students' performances since it is not as detail-oriented as AA, thus HA tends to be more generous in relation to grades and scores (Khabbazbashi & Galaczi, 2020; Metruk, 2018). Therefore, HA tends to result in poor personalization regarding areas of improvement, i.e. the students' ZPD is not met to the same extent as in AA. Nevertheless, HA can be successfully utilized in combination with AA since its focus is the students' ZPD which elicits a greater number of individualized opportunities to significantly improve one's proficiency, particularly in an EFL-context (Tuan, 2012).

II. *What impact can dynamic assessment have on EFL-learners' speaking skills?*

According to the results of the studies, it can be concluded that DA has a significantly positive effect on EFL-students' speaking proficiency due to the assessment being ZPD-oriented, which also means that DA can impact the teachers' workload negatively. However, a ZPD-oriented AM provides more support and feedback which might require less time for students to improve (Zarandi & Rahbar, 2016).

DA is a method that can be employed in several areas of teaching speaking since it can be seen as part of the communicative English teaching method due to the implementation of scaffolding. However, it has been shown to positively affect fluency to a higher extent than e.g., accuracy and pronunciation (Chou, 2021; Fahmi et al., 2020). Although Safdari and Fathi (2020) agree with fluency being improved, their study shows that accuracy was improved more than fluency which contradicts the other studies' results.

Further, self-regulation has been shown to significantly improve since DA allows students to be autonomous in their learning and adjust their strategies based on their areas of development, by enlightening the students on their ZPD (Fahmi et al., 2020; Ebadi &

Asakereh, 2017). Even though the sample of these two studies were limited, their results aggregated with the results from the remaining studies on DA, showcases the positive impact DA has on EFL-learners' speaking skills.

It has also been noted that confidence and anxiety affect the students' perception of speaking. In short, it was proven that DA had a tremendously positive effect on students' confidence and anxiety which we will address further later on (Çetin Köroğlu, 2019; Fahmi et al., 2018).

In conclusion, every study included in the DA-section has proven that DA is significantly more effective than non-DA. Although the studies' results and methodology differed in some aspects, it is evident that further research with a broader scope is needed which we will explore later on.

III. *What impact can peer-assessment have on EFL-learners' speaking skills?*

The findings of this section found that PA had a central part in improving the students' speaking skills. Several studies note that PA gives the student insight into their own learning process through recognizing their areas of concern (Çetin Köroğlu, 2021; Chang & Lin, 2020; Safari & Koosha, 2016; Toland et al., 2016). It can therefore be argued that PA aids the learner in realizing what they need to improve to reach a higher level of speaking proficiency.

Two of the studies emphasize the need for students to be properly acquainted with PA to ensure its effectiveness (Au & Barkakçi, 2020; Safari & Koosha, 2016). This indicates that the teacher needs time to plan and manage the PA, resulting in an increased teacher workload. Nevertheless, the versatility of PA indicates that once the students have an understanding of PA, it will be beneficial for both teacher and students.

A multitude of AMs in this section were not perceived as speaking assessment exams, rather as an opportunity to practice speaking (Çetin Köroğlu, 2021; Safari & Koosha, 2016; Toland et al., 2016). This can be considered highly beneficial since speaking exams can become an advantageous possibility for not only assessment, but also for the development of speaking skills.

Four of the studies in this section use digital tools to facilitate the PA activity (Çetin Köroğlu, 2021; Chang & Lin, 2020; Safari & Koosha, 2016; Toland et al., 2016). Using digital tools in class requires extra considerations since various issues can arise, such as internet connection, accessibility of devices, etc. It can be argued that we should look to

technology to facilitate teaching and assessment of the English-speaking ability as technological advancements have become a large influence in education.

The studies in the PA section show that these AMs serve to improve EFL-learners' oral proficiency. The research procedures in the studies vary, but all of the studies imply that PA is one factor in the development of the participants' speaking skills.

4.2 Pedagogical Implications

Additionally, the pedagogical implications of our findings reveal that almost all of the AMs that are utilized in the studies require a considerable amount of planning and time management for the teacher which is an aspect to keep in mind (Çetin Köroğlu, 2019, 2021; Chang & Lin, 2021; Safari & Koosha, 2016; Arfaei Zarandi & Rahbar, 2016). On the other hand, considering the fact that assessment and grading contribute to future education, we need to address speaking assessment properly for it to be valid and reliable.

In relation to this, teachers must consider which approach to use when assessing. HA is less time-consuming which is appealing to teachers considering the insufficient time and the amount of students that require feedback, but precision and clear areas of improvement are lost as a result. With AA, each student's ZPD becomes evident to teachers and learners, however, this demands more time. In other words, the predominant utilization of HA eases the workload for teachers but it may restrict the students' improvement. It is possible to incorporate both AA and HA by employing PA in either a holistic manner, e.g two stars and a wish, or analytically through feedback evaluating fluency, accuracy, etc.

As mentioned in the introduction of this paper, oral proficiency is a broad term which includes different abilities. Using different AMs when assessing oral proficiency can strengthen the validity since only one AM cannot measure all of the different abilities that make up oral proficiency. This is exemplified in the DA-study that used the Goh & Burns' cycle that found that grammatical accuracy and pronunciation needed special attention outside of the DA-approach (Chou, 2021:31). Consequently, the teacher must widen the scope and seek alternative AMs.

Anxiety among students' nowadays is very tangible and cannot be avoided in the classroom, and as future teachers we have decided to include this topic since it is highly relevant to take into consideration. Additionally, the topic has been addressed by a majority of the reviewed studies.

DA has been presented to positively affect students' anxiety in speaking exams since the tools used in one study gave the students more time to think, and thus made less errors which in turn lets students focus on more important things than their anxiety. Anxiety is a legitimate issue and should not hinder students from showing their full potential in exam situations. Learning is a process and errors are part of this process which is important to establish. Accommodating anxious students' examination should be done within a reasonable extent in order to avoid affecting the validity of examinations (Çetin Köroglu, 2019; Fahmi et al., 2019; Karpovich et al. 2021; Safari & Koosha, 2016; Toland et al., 2016).

As observed in several studies in this paper, student attitudes on AMs are of utmost importance when deciding on which AM to employ. One approach is to let the students know that their opinions matter, e.g., by giving them the opportunity to choose an AM they feel is the best for the situation. This also encourages students to take control of their own learning process. Students will not have a positive attitude towards the AM if they do not feel comfortable participating in, e.g., PA. Successfully employing PA relies upon a good classroom environment, first when that has been established the challenge of instructing the students can begin. Student attitudes and the classroom environment go hand in hand, one can not be good without the other. Ensuring a good classroom environment is therefore vital (Çetin Köroğlu, 2019; Chang & Lin, 2021).

5 Conclusions

In this literature review we have examined how different AMs affect EFL-students' speaking performance development and have come to the following conclusions. The choice of AM has a clear impact on students' attitudes which, in turn, affects their performance and this cannot be ignored. Even though some AMs increase the workload for teachers it is important to choose an AM that corresponds with the learning goals at hand.

We have noticed several gaps and further research areas. For example, DA and PA has been proven to have a significantly positive effect on student attitudes and performance but further research is needed regarding how to increase the efficiency of these methods on e.g. pronunciation and accuracy. Additionally, in order to make AA more effective, it has to be further researched to decrease the workload for teachers.

Moreover, the scope of future research must be expanded in order to make the conclusions more generalizable and reliable. This may include studies on other levels than university-level, and another focus on proficiency levels since it has been indicated by Ebadi and Asakereh (2017) that the effect of DA differs depending on students' proficiency level. To make this area more applicable for our future teaching, a Swedish perspective would be beneficial since the general proficiency of English in Sweden is higher than the participants' levels of the studies included.

Continuing research on this subject will hopefully lead to speaking being studied further and a consensus on which AMs are most effective for testing different speaking components in the EFL-classroom might be created. This would lead to higher validity and reliability which is crucial for teaching, and a wider range of possibilities in developing students' speaking proficiency. Considering that the speaking skill is essential in today's globalized society, it is vital that future research is dedicated to further investigate the effects of assessment on student speaking development.

Reference List

Arfaei Zarandi, S. Z., & Rahbar, B. (2016). Enhancing Speaking Ability through Intervening Scaffolding Strategies. *Theory and Practice in Language Studies*, *6*(11), 2191. https://doi.org/10.17507/tpls.0611.17

Arroba, J., & Acosta, H. (2021). Authentic Digital Storytelling as Alternative Teaching Strategy to Develop Speaking Skills in EFL Classes. LEARN Journal: Language Education and Acquisition Research Network, *14*(1), 317–343.

Au, H. Y. C., & Bardakçi, M. (2020). An Analysis of the Effect of Peer and Teacher Feedback on EFL Learners' Oral Performances and Speaking Self-Efficacy Levels. *International Online Journal of Education and Teaching*, *7*(4), 1453–1468. http://iojet.org/index.php/IOJET/article/view/895

Borger, L. (2018). *Investigating and validating spoken interactional competence: Rater perspectives on a Swedish national test of English.* University of Gothenburg.

Çetin Köroğlu, Z. (2019). Interventionist Dynamic Assessment's Effects on Speaking Skills Testing: Case of Elt Teacher Candidates. *Advances in Language and Literary Studies*, *10*(3), 23. https://doi.org/10.7575/aiac.alls.v.10n.3p.23

Çetin Köroğlu, Z. (2021). Using Digital Formative Assessment to Evaluate EFL Learners' English Speaking Skills. *GIST Education and Learning Research Journal*. https://doi.org/10.26817/16925777.1001

Chang, C., & Lin, H.-C. K. (2020). Effects of a mobile-based peer-assessment approach on enhancing language-learners' oral proficiency. *Innovations in Education and Teaching International*, *57*(6), 668–679. https://doi.org/10.1080/14703297.2019.1612264

Chou, M.-H. (2021). An integrated approach to developing and assessing EFL students' speaking ability and strategy use. *Language Education and Assessment, 4*(1), 19–37. https://doi.org/10.29140/lea.v4n1.428

Council of Europe (2018). Common European Framework of Reference for Languages [CEFR]. www.coe.int/lang-cefr

Ebadi, S., & Asakereh, A. (2017). Developing EFL learners' speaking skills through dynamic assessment: A case of a beginner and an advanced learner. *Cogent Education*, *4*(1), 1419796. https://doi.org/10.1080/2331186X.2017.1419796

EF (2021). *EF EPI 2021*. https://www.ef.se/epi/

Fahmi, F., Pratolo, B. W., & Zahruni, N. A. (2020). Dynamic assessment effect on speaking performance of Indonesian EFL learners. *International Journal of Evaluation and*

Research in Education (IJERE), *9*(3), 778. https://doi.org/10.11591/ijere.v9i3.20466

Goh, C. C. M., & Burns, A. (2012). *Teaching speaking: A holistic approach*. Cambridge University Press.

Graddol, D. (1997). *The Future of English? A guide to forecasting the popularity of the English language in the 21st century.* The British Council.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University press.

Karpovich, I., Sheredekina, O., Krepkaia, T., & Voronova, L. (2021). The Use of Monologue Speaking Tasks to Improve First-Year Students' English-Speaking Skills. *Education Sciences*, *11*. https://doi.org/10.3390/educsci11060298

Khabbazbashi, N., & Galaczi, E. D. (2020). A comparison of holistic, analytic, and part marking models in speaking assessment. *Language Testing*, *37*(3), 333–360. https://doi.org/10.1177/0265532219898635

Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.

Metruk, R. (2018). COMPARING HOLISTIC AND ANALYTIC WAYS OF SCORING IN THE ASSESSMENT OF SPEAKING SKILLS. *Journal of Teaching English for Specific and Academic Purposes*, *6*(1), 179. https://doi.org/10.22190/JTESAP1801179M

Qayoom, N., Saleem, M., & Mansoor, M. (2021). Objectives of Teaching English Skills and Their Assessment in Indian Schools: A Study on the Divergence. *Journal of Language and Linguistic Studies*, *17*(2), 755–766. https://doi.org/10.52462/jlls.53

Safari, M., & Koosha, M. (2016). Instructional Efficacy of Portfolio for Assessing Iranian EFL Learners' Speaking Ability. *English Language Teaching*, *9*(3), Article 3. https://doi.org/10.5539/elt.v9n3p102

Safdari, M., & Fathi, J. (2020). Investigating the role of dynamic assessment on speaking accuracy and fluency of pre-intermediate EFL learners. *Cogent Education*, *7*(1), 1818924. https://doi.org/10.1080/2331186X.2020.1818924

Skolverket. (2020). Engelska (Ämnesplan). [English – Syllabus]. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/gymnasieprogrammen/amne?url=-996270488%2Fsyllabuscw%2Fjsp%2Fsubject.htm%3FsubjectCode%3DENG%26version%3D6%26tos%3Dgy&sv.url=12.5dfee44715d35a5cdfa92a3

Toland, S. H., Mills, D. J., & Kohyama, M. (2016). Enhancing Japanese University Students' English-Language Presentation Skills with Mobile-Video Recordings. *JALT CALL*

Journal, *12*(3), 179–201.

Tuan, L. T. (2012). Teaching and Assessing Speaking Performance through Analytic Scoring Approach. *Theory and Practice in Language Studies*, *2*(4), 673–679. https://doi.org/10.4304/tpls.2.4.673-679

Xiaoxiao, L., & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics (Foreign Language Teaching & Research Press), 33*(1), 24–40.

Appendix
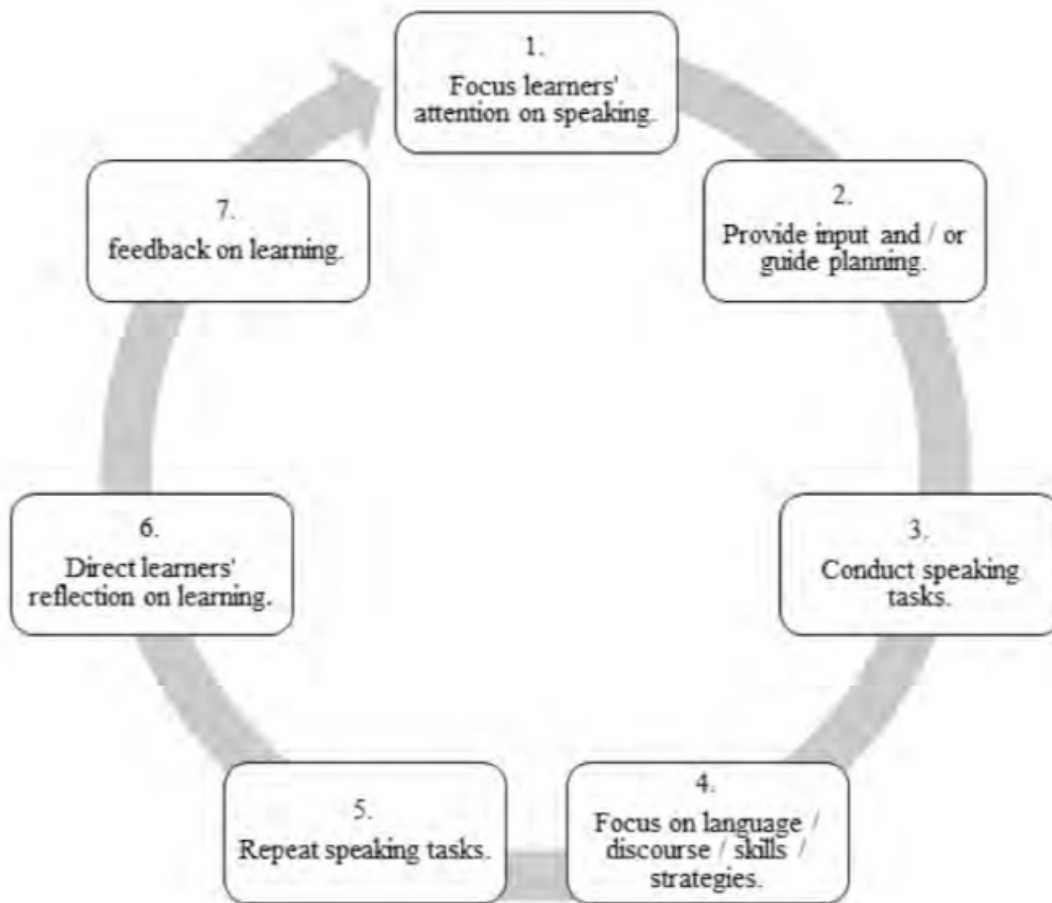
Goh and Burns' *Teaching Speaking Cycle* (2012, p. 153).