Chapter 4

A self-evaluation procedure for secondary school students to improve self-regulated learning of their speaking skills in foreign languages

Abstract

To become autonomous language learners, students must learn to self-regulate their learning. This chapter first explains what is needed to support this self-regulation learning process for speaking skills. From this explanation follow design principles for the teaching approach on the basis of which a concrete self-evaluation procedure for students is proposed. Subsequently, a quasi-experimental study investigated to what extent changes occurred in the students' self-regulation of their own speaking skills and to what extent the students perceived the self-evaluation procedure as motivating and the received feedback and support as adaptive to their needs. From the results can be concluded that during the self-evaluation procedure students' perception of their learning needs did indeed change. Shifts in diagnoses of their own speaking performances and foci of plans for improvement were found. It seemed that students expanded the focus of their diagnoses and plans. It was also found that the perceived need for teachers' assistance decreased and the preference for independence increased. Furthermore, the study showed that students perceived the self-evaluation procedure as motivating. Students in the experimental group found activities to be as tailored to their needs as the control group but students in the experimental group found feedback in lessons in speaking skills in general less tailored to their needs than the control group. However, the students in the experimental group found activities adaptive when they were asked, not about lessons in speaking skills in general, but about a specific cycle of a self-evaluation procedure.

4.1 Introduction

An important goal in foreign language education is to guide students to become autonomous learners (Holec, 1981; Lee, 1998; Little, Dam & Legenhausen, 2017). It is important that students learn to self-regulate their own foreign language skills, "[…] so that they can continue their language development and take increasing responsibility for their learning"(Lee, 1998: 288). This means that students must learn to independently evaluate their current speaking performance, compare this to a desired situation, set goals and draw up a plan to reduce the gap, learn to execute this plan and follow it up with an evaluation, after which the cycle can be repeated (Little et al., 2017). This self-regulated learning process should be supported adaptively, meaning that students receive the help they need (no more and no less) and that

support is phased out gradually until they are able to fulfil all the different parts of the process independently (e.g. Sadler, 1998).

Providing adaptive support on self-regulated learning of speaking skills appears to be difficult, however, in regular teaching in secondary schools. In the complex context of a class situation with a large number of students who have different learning needs, it is challenging for a teacher to monitor each individual student's learning process simultaneously (Keijzer, Perry, Rose & Verheggen, 2011) and to give each student tailored support in the form of feedback and tasks (Chapter 3). Especially for speaking, because of its transient nature, the opportunity for teachers to give feedback passes quickly. It is not surprising, therefore, that descriptive L2- research has shown that teachers do not usually give much feedback, the feedback is not divided equally among the students (Gass & Mackey, 2012) and is not always effective (Lyster, Saito & Sato 2013). Moreover, the quantity and type of feedback does not match students' preferences (Yoshida, 2008), and providing feedback does not always lead to uptake (= learners' responses or self-repair (Gass et al., 2012; Lyster & Ranta, 1997). On the one hand, this is because teachers do not always use feedback techniques that encourage self-correction (Gass et al., 2012; Lyster & Ranta, 1997; Lyster & Sato, 2010). Ideally, feedback not only addresses correctness of the utterances – which can make learners dependent on external feedback (Poehner, 2012; Sadler, 1989) – but it should also invite them to self-correct or it should provide information about how to correct commensurate with the students' ability (Aljaafreh & Lantolf, 1994; Pryor & Crossouard, 2008; Sadler, 1998). On the other hand, the opportunity to improve speaking performance is often lacking. In secondary education in the Netherlands and elsewhere in the world, speaking tasks are often offered only once in a lesson series without additional instruction or practice and with limited time for reflection after interaction (Goh & Burns, 2012; Goh, 2017; Chapter 2 of Van Batenburg, 2018). As a result, in the regular classroom, learners often do not have the chance to practise a speaking activity again and to improve their initial attempt which is a missed opportunity, because reflection, additional input and task repetition can help learners to advance (Bygate, 2001; Goh et al., 2012; Goh, 2017; Van Batenburg, 2018).

In order to improve secondary school students' self-regulation of their speaking skills, we developed a procedure for this study, the self-evaluation procedure, to facilitate diagnosis of current speaking performance by the students themselves, development of a plan to improve it and adaptive support for the execution of the plan. The self-evaluation procedure

was used in several cycles by students to create an iterative learning process of monitoring and improving their own speaking skills. In this chapter we focus on the question of whether our self-evaluation procedure could be an adaptive resource for secondary school students to learn to improve their speaking skills in foreign languages through self-regulation. First, we examine in further detail what was needed to promote self-regulation in speaking skills. We needed to know this in order to derive design principles for the teaching practice. Then we propose a concrete self-evaluation procedure for speaking skills on the basis of these design principles and investigate the extent to which changes occurred in the process of student self-regulation in improving their speaking skills after four iterations of the self-evaluation procedure. We also examine to what extent secondary school students perceived the self-evaluation procedure as motivating and the support they received as adaptive. With this study we hope to contribute to the goal of guiding students to become autonomous learners in learning to speak foreign languages and to provide concrete design principles to support this learning process adaptively.

4.2 Theoretical framework

4.2.1 Self-regulation as a feedback loop

Improving speaking skills can be seen as a goal-directed process that runs through a feedback loop. The core construct of this feedback loop is the reduction of the discrepancy between the learner's perceived current speaking performance and some desired level of performance or goal. This sets off an iterative process. Carver and Scheier (1998) proposed a general feedback loop as a model of self-regulation which we applied to self-regulation in speaking skills (Figure 8, based on Lord, Diefendorff, Schmidt & Hall, 2010: 546; Carver & Scheier, 1998; and Powers, 1973). This model comprises the components of the process of self-regulation and their interrelationships. An autonomous learner goes through all components independently, "[…] taking responsibility for the objectives of learning, self-monitoring, self-assessing, and taking an active role in learning" (Lee, 1998).

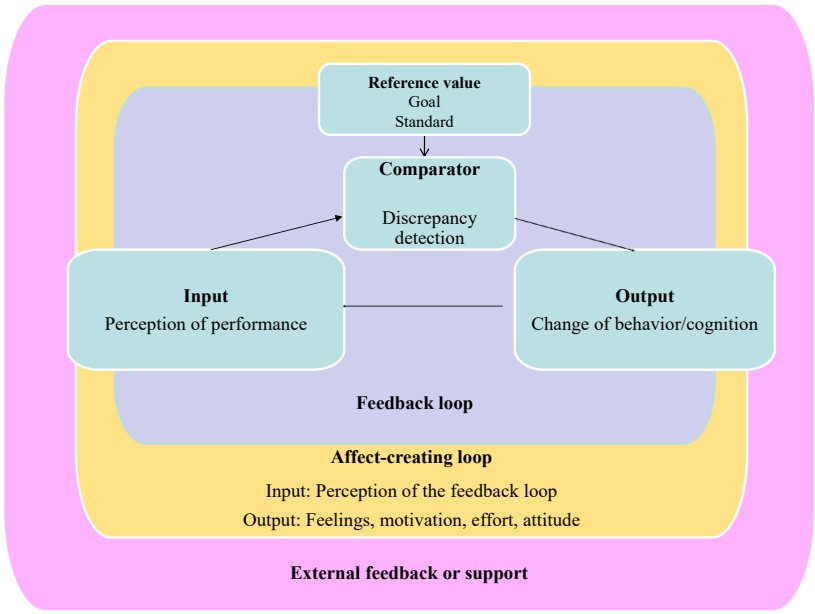

Figure 8: ***The discrepancy reducing feedback loop or model of self-regulation*** **expanded with an affect-creating loop and external feedback or support**

We will illustrate the elements of the feedback loop for speaking skills:

The feedback loop starts with the perception of the learner's own current speaking performance (input). Individual learners compare this perceived speaking performance to a goal or standard that they desire or think is desirable (reference value). Subsequently they try to reduce the discrepancies between the input and the reference value (comparator) by setting new goals and setting plans for improvement in order to improve their speaking performance (output) which is again compared to the desired level of performance (new cycle of input, comparator, output etc.).

Simultaneously with this behaviour-guiding feedback loop, Carver and Scheier (2000: 1717; 2012: 32) suggest that feelings arise via another feedback loop which operates automatically and parallel to the behaviour-guiding loop. This is the affect-creating loop. This second loop "[…] is checking on how well [the first process (the feedback loop, Figure 8 in violet)] is doing at reducing its discrepancies over time" (Carver & Scheier, 2000: 1717). Positive feelings arise when it seems that the goal will be reached in the foreseeable future and negative feelings when the difference between the current and desired situation seems too large to be bridged in time. Carver and Scheier (2000: 1717) explain that "[…] the

perceptual input for the affect-creating loop [(Figure 8 in orange)] is a representation of the rate of discrepancy reduction in the action system over time" and that the output of this affect loop is negative or positive feelings which may in turn influence each element of the behavioural feedback loop.

This process of self-regulation is complex for students. It requires understanding and noticing of different aspects of their own speaking performance (the existing situation), the capability to set goals and plans for improvement and the motivation to carry out these plans. Depending on the learners' capacities, the feedback loop or self-regulation loop can be followed autonomously or with external feedback or support (Figure 8 in pink). External feedback or support can be, for instance, teacher feedback, peer feedback, parents, a course book or another external source (Hattie & Timperley, 2007: 81). The feedback or support may be focused on each element of the feedback loop, thus on the learners' understanding of their own speaking performance (input), on the desired goal or standard (reference value), on the plan for improvement or the improved speaking performance (output) and on the feelings, motivation, effort or attitude which influence the process of self-regulation (affect-creating loop).

The ultimate purpose of this external feedback and support should be to create autonomous learners who are able to self-regulate their own learning process and effectively and independently improve their speaking performance themselves. The process should therefore be an iterative process which leads to increasingly independent and, ultimately, autonomous learners (Little et al., 2017).

4.2.2 Possible needs for external feedback and support to promote self-regulation in speaking skills

Little stated (2007: 26) that "learner autonomy is the product of an interactive process in which the teacher gradually enlarges the scope of her learners' autonomy by gradually allowing them more control of the process and content of their learning." To enhance such an interactive process, insight into what is required for self-regulation is necessary. In this section we will examine what is required for each component of self-regulation in speaking skills (Figure 8) and what students may *need* in the form of support if they cannot yet independently fulfil the requirements of the relevant component of self-regulation in speaking skills.

Input: enhancing noticing and understanding

In the feedback loop as shown in Figure 1, learners perceive their own speaking performance. For this perception to be relevant in self-regulation, the learners must *notice* different aspects of their speaking performance and understand how their prior knowledge, instruction and received feedback relates to different aspects of their speaking performance (cf. Schmidt, 1990, noticing hypothesis).

In common classroom practices, however, because of the volatility of speech, there is often no time for reflection on the oral production. Speaking skills especially, more than other language skills, demand many cognitive processes in a short time: conceptualizing, grammatical and lexical encoding, articulating as well as monitoring (Levelt, 1989). For successful speaking, different types of knowledge (knowledge of the topic, lexicalized items and phrases, morphosyntax, pronunciation, pragmatic knowledge) are needed, including the skill to access the corresponding types of knowledge quickly and efficiently. Another skill that is needed is how to use communication strategies if knowledge is lacking (De Jong, 2020; Goh, 2017). Since attention is limited, learners cannot focus on every aspect of their performance while speaking but introducing immediate feedback or self-reflection would interrupt the flow of communication. Although delaying feedback and reflection until after speaking avoids cognitive overload, due to the volatility of speech it would be too difficult for learners to recall all the details of their performance (e.g. about timing of feedback Lyster, Saito, Sato, 2013; Ellis, 2009). Analysing a recording of one's own speaking performance can be a practical and instructive solution therefore (e.g. Hedge, 2000; Lappin-Fortin & Rye, 2014). It allows students time to reflect on their own speaking skills and to pay attention to more aspects of their speaking performance than is possible while speaking at the same time.

Time for reflection is not always enough. Learners may need support to notice and become aware of the different aspects of their speaking performance (Dlaska & Krekeler, 2008; Goh, 2017; Lappin-Fortin & Rye, 2014; Poehner, 2012). Being aware of the existence of different aspects of speaking is a first step, evaluating those aspects goes further. Research has shown that it is difficult for foreign language learners to assess aspects of their own performance. Low correlations have frequently been found between self-assessments and tests and between self-assessments and other measures shown to be valid and reliable (Blanche, 1988; Poehner, 2012; Ross, 1998). An explanation for such low correlations might be that L2-learners lack the metacognitive and linguistic knowledge to determine the

appropriateness of their utterances. Low self-assessments as well as low correlations with more objective scores were especially reported for self-assessments of grammar and pronunciation (Blanche, 1988). Research has shown that feedback or instruction (about concrete rules for instance) helps learners to assess more accurately (see for example Lappin-Fortin & Rye, 2014; Jones, 1997; Dlaska & Krekeler, 2013). Blanche (1988) stated that self-assessments give information about the extent to which students can appraise their own speaking performance. With this information teachers can tailor their feedback or instruction. The self-assessments give insight into individual learning needs. Learners who can assess accurately enough, do not have to depend entirely on the opinion of teachers (Blanche, 1988) and teachers can gradually reduce their support.

For learners to benefit from instruction and feedback, they need to have the opportunity to gain evaluative experience in an iterative process (Dlaska & Krekeler, 2013; Sadler, 1989). Then learners' ability to assess their speaking performances may improve over time (see for instance Couper, 2003; De Saint Léger & Storch, 2009; Lappin-Fortin & Rye, 2014).

Support can be provided by instruction and feedback but the self-assessment instrument itself can also support learners to assess themselves more accurately (Ross, 1998). Self-assessment instruments can focus the learners' attention on more categories of the speaking performance than they might do without an instrument. Criterion-referenced self-assessment instruments which are tailored to course objectives are helpful for this purpose (Brantmeier, Vanderplank & Strubbe, 2012). Important factors which influence accuracy are being connected with specific curricular content (Brantmeier et al, 2012; Ross, 1998) and doing the self-assessments directly after completing specific tasks (Butler & Lee, 2006).

Furthermore, self-assessment instruments should not only focus on areas for improvement, but also on positive points. Research in positive psychology has shown that reflecting on positive points activates positive emotions that in turn are beneficial to learning (Voerman, Meijer, Korthagen & Simons, 2012).

In short, in order to notice different aspects of their speaking performance, students can be supported by allowing them time to reflect on a recording of their own speaking performance using a self-assessment instrument and by providing input (instruction) and feedback.

The comparator and the reference value

De Bot (1996) argued that L2 learners benefit more from being pushed to retrieve target language forms than from merely hearing the forms in the input, because retrieval and subsequent production can strengthen associations in memory (cf. Swain's pushed output hypothesis, 2005). The activity of reflecting after speaking may strengthen learning even more, because reflecting on both positive points and areas for improvement pushes learners once again to retrieve already internalized target language forms and this time to compare them with their current speaking production.

In order to compare the current with the desired situation, learners have to possess a concept of what they think the desired speaking performance should look like (Sadler, 1989). They compare what they notice and understand of their own speaking performance (the input) with an internal reference or standard. According to Black and William (2009: 15) "the learners' standards will depend in part on their interpretation of the task, on their perception of the criteria and targets for success, on their personal orientation towards the task, and on their view of the time constraints."

In order to support the development of such an internal reference or standard, all kinds of input can be provided through exposure to target exemplars. This positive evidence gives the learner information about what is possible in the language (Lyster, et al., 2013). Support can be provided in the form of models of the desired speaking performance and examples of appropriate linguistic aspects with which the students can compare their own performance (Poehner, 2012; Préfontaine, 2013).

It is of course possible to support students by providing external standards (for instance in the form of a rubric) with criteria for the quality of the speaking performance. These can be descriptive or normative. Descriptive standards can help the learner to see how they can develop (Brantmeier et al., 2012; Little, 2009). In our view, it is important that the external standards are not normative given the purpose of this study. We aimed at stimulating students' reflection on the strengths and weaknesses of their performances, so that aspects associated with success or high quality could be recognized and reinforced, and unsatisfactory aspects modified or improved (Sadler, 1989). The value of comparing resides in the development stimulated through the process of comparing the current with the desired situation (e.g. Bennett, 2011; Orsmond, Merry & Reiling, 2002; Poehner, 2012). The intention is to let learners think about their own performance, their own goals, what is needed and how

to attain new goals. In contrast, normative standards which contain scales or scores serve to rank a performance in comparison to others (Cauley & McMillan, 2010; Clark, 2012; Yin, Shavelson, Ayala. Ruiz-Prima, Brandon, Furtak & Young, 2008) and might direct attention away from the reflection on which aspects can be reinforced or improved towards the question of how good the performance was (Cauley & McMillan, 2010). By using normative standards, performance differences become the most important concern and this may have negative effects on motivation and achievement (Yin et al., 2008; Butler, 1987; 1988).

Output: plan for improvement and improved speaking performance

Students' assessments and self-assessments can be used by teachers to adapt their instruction, activities for improvement and feedback (Black & William, 1998; 2009; Nicol & Macfarlane-Dick, 2006). If the goal is to promote self-regulation of speaking skills, then students should ideally also take the step themselves of taking action to close the gap between their current and desired speaking performance (Cauley & McMillan, 2010; Sadler, 1989). Students must set goals and make plans for improvement themselves on the basis of their own assessment of the different aspects of their speaking performance, (Cauley & McMillan, 2010). Providing opportunities to practise these activities is necessary for learners to gain experience (Sadler, 1989). Self-assessments can encourage student decision-making about what to do and when to do it (Cauley & McMillan, 2010). In addition, support may be needed to stimulate learners to make appropriate plans and to execute these plans.

Firstly, learners can be supported by providing suggestions for activities for improvement, giving choices and asking questions about what they think they need to improve (Cauley & McMillan, 2010). Suggestions for activities for improvement can be deduced from research about developing second language speaking. De Jong, Steinel, Florijn, Schoonen & Hulstijn (2012), for instance, showed that linguistic knowledge about lexical items, chunks, morphosyntax, pronunciation and processing speed are to a large extent important for communicative success. Instruction and activities that encourage the acquisition of this knowledge should be available to the students and be an option for their plans. Processing speed and fluency could be stimulated by practising speaking and automatization (DeKeyser, 1997; Segalowitz & Hulstijn, 2005). Research into task repetition, in which learners do the same or a similar speaking task a few more times, immediately or at a later time, showed positive effects such as more accurate and idiomatic speech and greater

fluency (Bygate, 1996; 2001). This is why Goh (2012; 2017) pleads for repetition of the speaking task as an activity for improvement. She also emphasizes the importance of pre-task planning (2012; 2017). Following Skehan (1998) and Segalowitz (2010), she recommends giving learners time to plan before a task, to think of what to say and how to say it. That helps to free up attentional space during speaking for articulation of ideas, speech monitoring and self-repair (Goh, 2017: 252). Another improvement activity can be learning chunks, fixed phrases, which will lead to more fluent speech, and learning compensating strategies such as asking for repetition, paraphrasing, describing and asking for help, to keep the conversation going (e.g. De Jong, 2020; Goh, 2017).

Secondly, learners can indicate in their plans whether they need help, about what and from whom (teacher or peer) and formulate a request for help (Clark, 2012; Nicol & Macfarlane-Dick, 2006). This activity stimulates awareness of the learning process.

Thirdly, teachers can provide feedback on the improvement plans. The process of self-assessing and making plans generates internal feedback at a variety of levels (i.e. cognitive, motivational and behavioural) (Butler & Winne, 1995; Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Although this internal feedback is invisible, learners' assessment provides information about how they are progressing and how they are regulating this process (Nicol & Macfarlane-Dick, 2006). The output in the form of a learner's plan for improvement with any request for help, and the congruence or lack of congruence of this plan with the assessment of the speaking performance provides information about the degree and kind of support the learner needs in this process. Teachers can use this information to help students self-assess and improve their own performance (Cauley & McMillan, 2010) and also to provide feedback on the self-regulation process itself which can enhance learning (Hattie & Timperley, 2007).

After executing the plan, after activities for improvement and feedback, it is important to offer students the opportunity to do the speaking task again and to have them check whether their speaking performance has indeed improved (Bygate, 2001; Nicol & Macfarlane-Dick, 2006). The improved speaking performance can serve as input for a new feedback-loop, resulting in an iterative learning process which promotes learning. Little (2013: 8) states that "[…] by monitoring our performance we gradually reinforce and/or modify our competences." By replaying and analysing their own production, learners strengthen associations in memory (the 'generation effect', Clark, 1995). Moreover, learners remember information better when they take an active part, rather than having it provided by an external source (deWinstanley

& Bjork, 2004), and this iterative process can provoke positive feelings by giving the students control over the learning process.

Affect creating loop

According to Carver and Scheier (2000: 1717; 2012: 32), feelings arise during the process of self-regulation as learners compare their current performance to the desired situation. Learners evaluate how well they are doing at reducing discrepancies over time, and the negative or positive feelings resulting from this evaluation of progress over time may in turn influence the learning process. Negative feelings such as stress, fear of failure and anxiety may hamper the learning process (Boekaerts, 2010; Bandura, 1997). Especially for speaking skills, research has shown that anxiety often plays an important negative role (Cheng, Horwitz & Schallert, 1999, Horwitz, Horwitz & Cope, 1986; Simons & Decoo, 2009;). The threat to one's self-image is the main cause of speaking anxiety (Horwitz et al., 1986). A safe environment and insight in one's own capacities help to reduce anxiety (Horwitz et al., 1986; Simons & Decoo, 2009).

Several positive effects of self-assessments have been reported. De Saint Léger and Storch (2009), for instance, found that self-assessment of speaking skills has positive effects on self-confidence and on the willingness to communicate orally in class. By giving students insight and the opportunity to control their own learning process, positive feelings can be provoked and anxiety can be reduced. Ownership of students' own learning process can have a motivational effect (Blanche, 1988; Cauley & McMillan, 2010). In assessments, students are given some control over their learning by giving them the opportunity to reflect on the criteria for the task and on the steps needed to meet the learning goal (Cauley & McMillan, 2010). Furthermore, the opportunity to perform the same task several times helps them to get a better grip on their own learning process (Bygate, 2001) and to perceive progress. The iterative process of monitoring and improving may result in self-efficacy which in turn generates positive feelings about self-regulation of speaking skills (e.g. Bandura, 1997).

4.2.3 Design Principles

The aim of this study was to contribute to the goal of guiding students to become autonomous learners in learning to speak foreign languages and to provide concrete design principles to support this leaning process adaptively. In the previous section we described requirements

for each component of the process of self-regulation (see Figure 8 the feedback loop based on Lord et al., 2010; Carver & Scheier, 1998; and Powers, 1973) and the possible needs for feedback and support. If students have to learn to fulfil all the different parts of the process of self-regulation more and more independently, they have to be given the opportunity to gain evaluative experience (Sadler, 1989) and control or ownership of their learning process in a safe environment (Blanche, 1988; Cauley & McMillan, 2010; De Saint Léger & Storch, 2009). Based on the review outlined above, the following design principles can be drawn up for a self-evaluation procedure that aims to support students to learn to improve their speaking skills through self-regulation:

1. Add a self-evaluation by the student to a speaking activity

a. *Start the self-evaluation procedure with the student's diagnosis of a recording of their own speaking performance.*

   In order to get the students to reflect on various linguistic aspects of their speaking, such as grammar, vocabulary, pronunciation and fluency, as well as communicative competence (getting the message across), provide a self-reflection instrument which the students complete directly after completing the speaking task which enables the self-assessment or diagnosis (Butler & Lee, 2006). The self-evaluation instrument has to contain criteria to evaluate both areas for improvement and positive points (Voerman et al., 2012) in a non-normative manner in order to help them to notice different aspects and to enhance learning (Brantmeier et al, 2012; Cauley & McMillan, 2010) as well as for affective reasons (Yin et al., 2008; Butler, 1987; 1988).

b. Let students make a plan for improvement

   Self-regulation should be further enhanced by having students make a plan for improvement (Cauley & McMillan, 2010). This would also give students control over their own learning process which may be motivating (Blanche, 1988; Cauley & McMillan, 2010). Suggestions on what kind of activities could be undertaken should be provided for their plans at this stage (Cauley & McMillan, 2010).

c. *Let students indicate their need for assistance*

Let learners indicate whether they want to work autonomously, with a peer, or whether they need assistance from the teacher. Regulating the degree of autonomy by indicating the kind of support they need gives control to learners.

2. Provide adaptive activities for improvement and 3. Provide adaptive feedback

Teachers make inferences about what their students know and can do and adapt their feedback and instruction based on this knowledge (Bennett, 2011). In regular teaching, those inferences about the learners' speaking skills are based on what teachers hear and know from previous experiences in the classroom. The self-evaluation procedure gives teachers additional information provided by the students' diagnoses, plans for improvement and desired working format or requests for help. Teachers can scan the evaluations for discrepancies with their own inferences, tailor their feedback and propose learning activities aligned with learners' current level and degree of self-regulation (Sadler, 1998). If students are already independent learners, they can select and arrange improvement activities themselves.

1-3. *After executing the plan for improvement, let the students redo the same or a similar speaking activity with self-evaluation*

Give students the chance to repeat the same (or similar) speaking activity to find out whether they have progressed and to put into practice what they have learned. Task repetition can help learners to advance (Bygate, 2001; Goh & Burns, 2012; Goh, 2017). Then a new cycle of monitoring and improving can begin.

4.3 Research aim and research questions

The overarching research aim was to ascertain to what extent a self-evaluation procedure can be an adaptive resource for secondary school students to learn to improve speaking skills in foreign languages through self-regulation. Since the aim of the self-evaluation procedure was to support secondary school students to fulfil all of the different parts of the process of self-regulation more and more independently, to reflect on various aspects of their speaking performance and to make appropriate plans, the following specific research question was formulated:

A. To what extent did the students' perception of their learning needs change during the self-evaluation procedure?

It was intended that students could execute their plans with the activities and feedback they needed to improve their speaking performance during the self-evaluation procedure. Therefore, the following research question was formulated:

B. To what extent did the students consider feedback and activities for improvement provided during the self-evaluation procedure to be adaptive to their needs?

Finally, as mentioned in the theoretical framework, an affect-creating loop operated parallel to the process of self-regulation and influenced it (see Figure 8, Carver & Scheier, 2000). Findings described in the Theoretical Framework section of this chapter suggest that positive feelings could be provoked and anxiety reduced through students' control, ownership and insight into their own learning process during the self-evaluation procedure. The self-evaluation procedure was designed to enable them to learn from both positive and negative points and to gain control over the learning process. As a result, speaking anxiety might be expected to decline and positive feelings to be provoked. The following research question was formulated:

C. To what extent did students experience the self-evaluation procedure as motivating and did their speaking anxiety change during the course of iterations of the self-evaluation procedure?

4.4 Method

4.4.1 Participants

Experimental group

The study was conducted among 329 students learning a foreign language at regular secondary schools in the Netherlands. Their foreign language teachers (two German teachers, four English teachers, three French teachers and two Spanish teachers) were recruited by Leiden University Graduate School of Teaching. They were asked in an e-mail whether, in the context of a study, they and their students were prepared to fill in questionnaires on speaking skills in the classroom and to try out a new approach for adaptive feedback and support which consisted of s*elf-evaluation procedures* in order to improve students' speaking skills. The self-evaluation procedures were carried out at 10 different schools offering three different types of secondary education: two year-2 vwo classes, two year-3 vwo classes, one year-4 vwo class, three year-5 vwo classes, two year-6 vwo classes, one year-2 havo class, three year-3 havo classes, one year-4 havo class, three year-5 havo classes and one year-3 vmbo class. For the purpose of this study, it was not necessary to select students with the same level of language skills, background, motivation or other variables, because we were particularly interested in the extent to which students, who are learning a foreign language at regular secondary schools and may have all kind of different characteristics, perceive the feedback and activities for improvement provided as tailored to them.

Control group

The control group consisted of 369 students learning a foreign language at regular secondary schools in the Netherlands. Their teachers were recruited by Leiden University Graduate School of Teaching and asked in an e-mail whether they and their students were prepared to fill in two digital questionnaires on speaking skills in the classroom in the context of a study into adaptive feedback and support for speaking skills. 329 students of 17 modern foreign language teachers (five German teachers, five English teachers, six French teachers and one Spanish teacher) completed both questionnaires (pre- and post-measurement) about teaching speaking skills in year-3 vwo, year-4 vwo, year-5 vwo, year-6 vwo, year-2 havo, year-3 havo, year-4 havo, year-5 havo, and year-3 mavo/vmbo classes.

4.4.2 Research instruments

Two sets of instruments were used:

- A diagnosis tool and a questionnaire provided in the form of an app for mobile phones (Appendix I), completed by the students of the experimental group. The diagnosis consisted of an audio recording of the speaking performance and questions in which the students evaluated their speaking performances, made plans for improvement and indicated preferences for a working format (in order to answer research question A, see below) (Appendix I, part A, B and C). In addition, the app contained questions designed to evaluate the feedback and activities for improvement provided during the specific cycle of the self-evaluation procedure (in order to answer research question B, see below) (Appendix I, part D).
- A digital questionnaire (Appendix II) which consisted of questions about adaptivity (Appendix II, part A), speaking anxiety (Appendix II, part B), and motivation for the self-evaluation procedure (Appendix II, part C, questions X-Y) (in order to answer research questions B and C, see below).

Below we describe these instruments in more detail.

Instruments to answer research question A about changes in the students' perception of what they need to improve their own speaking skills during the self-evaluation procedure

Students' diagnosis with plan and desired working format

Each diagnosis (see Appendix I) contained the following elements:

- Analysis by student of positive points and areas for improvement: five categories of the speaking performance (message, vocabulary, grammar, pronunciation and fluency) could be evaluated as 'positive point', 'question mark (?)' or 'area for improvement'.
- Plan for improvement: for each category of the diagnosis (getting the message across, vocabulary, grammar, pronunciation, fluency), two activities for improvement could be chosen from a list of 12 activities. As argued in the theoretical framework in this chapter, these activities were focused on acquiring and automatizing linguistic knowledge about vocabulary, chunks (fixed phrases), grammar and pronunciation. These activities were categorized under the headings: vocabulary, grammar and pronunciation. Activities for task repetition and automatization for which the aim was to speed up the speaking process and improve fluency,

were categorized under the heading fluency. Pre-planning by giving learners time to think of what to say and how to say it, was an activity categorized under the heading 'getting the message across'.

- Desired teaching/learning format: individual work, pair work, teacher's assistance.

Instruments to answer research question B about adaptivity

Digital questionnaire to measure the extent to which students considered feedback and activities for improvement in speaking lessons to be adaptive to their learning needs

In a digital questionnaire (Appendix II.A.) students of the experimental and control group scored the extent to which they considered feedback and activities for improvement to be adaptive on a 7-point Likert-scale. Three items concerned the feedback they received and three items concerned activities for improvement.

Digital questionnaire in an app to measure the extent to which students considered feedback and activities for improvement in a specific cycle of the self-evaluation procedure to be adaptive to their learning needs

A questionnaire (presented in an app after each specific self-evaluation-procedure cycle, see Appendix I, part D, questions 7-10) asked whether during a specific self-evaluation procedure cycle student's intended plan for improvement had been implemented, and whether feedback and activities for improvement had been sufficient to improve their speaking performance.

Instruments to answer research question C about motivation and speaking anxiety

Digital questionnaire to evaluate each students' activities making up the self-evaluation procedure

Learners' motivation for the different students' activities making up the self-evaluation procedure consists of three components derived from Fishbein and Ajzen (2010), namely: attitude toward the behaviour; perceived behavioural control (beliefs about the factors that may facilitate or impede performance of the behaviour); and intention to do all or part of the self-evaluation in future. The motivation for each of the following students' activities were scored on a 7-point Likert scale in the experimental group (Appendix II.B):

- recording and re-listening to their own speaking performance (Questions 1-3)
- doing a self-evaluation of their own speaking performance (Questions 4-7)

- producing a plan for improvement (Questions 8 – 11)
- executing the plan for improvement (Questions 12 – 15).

Dutch translation of the FLCAS, the Foreign Language Classroom Anxiety Scale adopted from Horwitz, Horwitz & Cope, 1986.

This questionnaire consisted of 33 items to measure speaking anxiety in class on a 5-point Likert scale (Appendix II.C).

4.4.3 Procedure

As a baseline measurement, all students of the experimental group (n=329) and the control group (n=369) completed the digital questionnaire on their perception of adaptivity of feedback and activities for improvement (Appendix II.A) at the start of the research. In addition, 171 students of the experimental group and 369 of the control group filled in the questionnaire about speaking anxiety (Appendix II.C).

Subsequently, the students of the experimental group were given speaking skills lessons in the foreign language using the self-evaluation procedure over four months while the control group did not use the self-evaluation procedure for training speaking skills in foreign languages. The students of the experimental group did one or several self-evaluation cycles. One cycle of a self-evaluation procedure consisted of performing a speaking activity followed by the student's self-diagnosis with plan for improvement (Appendix I). On the basis of the self-diagnosis and plan for improvement, the student received adaptive feedback and activities for improvement from their teacher in order to improve their speaking performance. The cycle was completed with an end-diagnosis of the performance of the speaking activity. From 281 students of the experimental group 1,024 self-diagnoses were collected. They included two plans for improvement each (Appendix I). Additionally, the adaptivity of the feedback and activities for improvement provided during the implementation of the plans was evaluated by the students for 339 self-evaluation procedures in total (Appendix I).

After four months, a post-test was carried out. Both the experimental (n=225) and control group (n=329) once more completed the digital questionnaires about adaptivity of feedback and activities for improvement (Appendix II.A) and speaking anxiety (Appendix II.C) (179 from the experimental group and 329 from the control group). In addition, 225 students

from the experimental group filled in a questionnaire about learners' motivation for the different parts of the self-evaluation procedure (Appendix II.B).

Dropouts

In the experimental group, the questionnaire about speaking anxiety was administered separately from the questionnaire about adaptivity, in order not to tax the students too much. As a consequence, not everyone from the experimental group filled in both parts of the questionnaires in the post-measurement (46 students of the experimental group did not fill in the second part of the questionnaire). This unfortunately led to more data loss than in the control group who had one questionnaire in the post-measurement containing both parts.

4.4.4 Analysis

Research question A

To answer research question A - *To what extent did the students' perception of their learning needs change during the self-evaluation procedure?* - we examined

1. to what extent were there shifts in students' observation of different aspects in their diagnoses of their speaking performances and in their plans in consecutive cycles of the self-evaluation procedure;
2. to what extent did they go through the self-regulation cycle more independently after several self-evaluations.

and performed the following analyses:

Understanding/noticing in students' self-evaluations

To answer the question to what extent were there shifts in students' observation of different aspects in their diagnoses of their speaking performances and in their plans in consecutive cycles of the self-evaluation procedure (research question A.1), we chose to compare the first cycle with the fourth cycle, because we assumed that learners have to do several cycles to make progress in learning how to monitor their own learning process. Moreover, the learners differed in the number of cycles they did (from 1 (n=281) to 9 cycles (n=1)). In the fourth cycle half of the initial number of learners' data were still complete (n=142). The variety of cycles was caused by choices made by the teacher about the pace of doing speaking activities with

self-evaluations during the period of the intervention. The pace was tailored to the students, the curriculum and other differences in the context.

In order to evaluate changes within learners, we included the same 142 students for the first and fourth cycles in the analyses and used the McNemar-Bowker test for matched pairs. In this way, we examined whether the evaluations in the first cycle were focused on certain categories of errors and positive points in particular by counting the frequencies of areas for improvement and the frequencies of positive points the students had noticed in their first cycle for the categories (message, vocabulary, grammar, pronunciation and fluency) and compared these frequencies with frequencies in the fourth cycle for each category.

In the same way, we used the McNemar-Bowker test to investigate to what extent shifts in focus in the plans for improvement occurred.

Development in independence

To answer the question to what extent the students went through the cycle of self-regulation more independently (research question A.2) we used the McNemar-Bowker test for matched pairs to analyse the extent to which students requested their teacher's assistance in the fourth cycle compared to the first cycle.

Research question B

To answer research question B - *To what extent did the students consider feedback and activities for improvement provided after their diagnosis to be adaptive to their needs?* – we performed two different analyses:

Adaptivity of activities for improvement and feedback from the learners' perspective – pre- and post-test

We first calculated the reliability (Cronbach's alpha) of the items in the questionnaire that should measure the same constructs, namely: the construct adaptive activities for improvement and the construct adaptive feedback. Both proved to be reasonably reliable (adaptive activities for improvement $\alpha = .68$; adaptive feedback $\alpha = .76$ after deleting one item).

With participants in this study taught by different teachers, the data were structured hierarchically. Since teachers could influence differences in the extent to which students

found support and feedback adaptive, linear multilevel analyses were applied, with teacher added as random variable. Linear Mixed models (SPSS version 25, using the Satterwaithe's approximation to calculate denominator degrees of freedom) were carried out in order to investigate whether there were differences in the degree of adaptivity between the experimental and the control group (factor between subjects) and whether there were differences between the pre- and post-test (factor between subjects), in a first analysis for activities for improvement and in a second analysis for feedback. The same students participated in the pre- and post-tests. Because student-ID was not recorded in the questionnaires, the pre- and post-test scores could not be linked to individual students and therefore, this factor was treated in the analyses as a between-subjects factor.

For both analyses, we tested whether adding teacher (as random intercept) as well as adding a random slope for time per teacher contributed significantly to the model. In order to do this we compared the simpler with the more complex models by comparing the difference in Log Likelihood and chi-squares. In this way, we tested whether some teachers elicited higher scores than others in general (teacher as random intercept) and whether the difference between pre-test and post-test would be different for different teachers (by adding the random slope). Analyses revealed that teacher contributed to both models, and that a random slope for time also contributed to the model for adaptivity of activities for improvement. These models are reported on below. Whenever significant interactions were found between time and group, we carried out post-hoc analyses (data split by group), to interpret this interaction.

Adaptivity of activities for improvement and feedback after a cycle of the self-evaluation procedure – intermediate tests

The number of times the students found the *activities for improvement* to be adaptive were compared to what would be expected by chance (50%) using the binomial probability function. This analysis was also carried out for *feedback*.

Research question C

To answer research question C – *To what extent did students experience the self-evaluation procedure as motivating and did their speaking anxiety change during the course of iterations of the self-evaluation procedure?* – we examined:

1. to what extent the students were motivated to carry out the different main activities of the self-evaluation procedure;
2. to what extent students' speaking anxiety changed after carrying out the self-evaluation procedure;

and we performed two different analyses:

Learners' motivation for the different students' main activities of the self-evaluation procedure

We first calculated the reliability (Cronbach's alpha) of the items in the questionnaire that should measure the same constructs, namely the different students' main activities of the self-evaluation procedure: 1) recording and re-listening to their own speaking performance; 2) doing a self-evaluation of their own speaking performance; 3) producing a plan for improvement; 4) executing the plan for improvement. All proved to be reliable (respectively: $\alpha = .81$; $\alpha = .79$; $\alpha = .80$; $\alpha = .79$).

As in the analysis above for adaptivity, the data were structured hierarchically and linear multilevel analyses were applied, with teacher added as random intercept. Mixed repeated measures analyses (SPSS version 25, using the Satterwaithe's approximation to calculate denominator degrees of freedom) were carried out in order to investigate whether there were differences in students' motivation between the separate components of the self-evaluation procedure. Because student-ID was recorded for the four components, this factor was treated as a within-subjects factor in the analyses.

As before, we tested whether adding teacher (as random intercept) contributed significantly to the model by comparing the simpler with more complex models by comparing the difference in Log Likelihood and chi-squares. Analyses revealed that teacher contributed to the model.

Since significant differences were found between students' motivation for the different main activities, we carried out pairwise comparisons with Bonferroni adjustment, to interpret the differences.

Speaking Anxiety

We investigated to what extent students' speaking anxiety changed in a pre-test and after a few cycles (post-test). To do this we first calculated the reliability (Cronbach's alpha) of the 33

items in the questionnaire that was used to measure the same construct, namely: speaking anxiety in the classroom (Cronbach's alpha = ,911).

Again, linear multilevel analyses were applied, with teacher added as random intercept in order to investigate whether there were differences in the degree of speaking anxiety between the experimental and the control group (factor between subjects) and whether there were differences between the pre- and post-test (factor between subjects). As already mentioned, the same students participated in the pre- and post-tests but because student-ID was not recorded, this factor was treated as a between-subjects factor.

For the analysis, we tested whether adding teacher (as random intercept) as well as adding a random slope for time per teacher contributed significantly to the model by comparing the simpler with more complex models by comparing the difference in Log Likelihood and chi-squares. Analyses revealed that only teacher contributed to the model and, therefore, this model is reported on below.

4.5 Results

4.5.1 Results for research question A

To what extent did the students' perception of their learning needs change during the self-evaluation procedure?

Understanding/noticing in students' self-evaluation

Table 4.1 shows the frequencies and percentages of the diagnoses in the first cycle and the fourth cycle for each category (message, vocabulary, grammar, pronunciation and fluency) for students that filled in the self-evaluation in cycle 1 and cycle 4 (n = 142).

Table 4.1

Diagnoses

	Cycle 1				Cycle 4			
Category	**Area for improvement**	**Posi-tive**	**Don't know**	**Total**	**Area for improvement**	**Posi-tive**	**Don't know**	**Total**
Message	5	120	17	142	7	111	24	142
	(3.5%)	(84.5%)	(12.0%)	(100%)	(4.9%)	(78.2%)	(16.9%)	(100%)
Vocabulary	18	95	29	142	24	83	35	142
	(12.7%)	(66.9%)	(20.4%)	(100%)	(16.9%)	(58.5%)	(24.6%)	(100%)
Grammar	36	67	39	142	20	79	43	142
	(25.4%)	(47.2%)	(27.5%)	(100%)	(14.1%)	(55.6%)	(30.3%)	(100%)
Pronunciation	39	49	54	142	38	61	43	142
	(27.5%)	(34.5%)	(38.0%)	(100%)	(26.8%)	(43.0%)	(30.3%)	(100%)
Fluency	55	55	37	142	30	72	40	142
	(35.2%)	(38.7%)	(26.1%)	(100%)	(21.1%)	(50.7%)	(28.2%)	(100%)

Table notes: Frequencies (and percentages) of the diagnoses (area for improvement, positive or don't know) for each aspect of the speaking performance (categories: message, vocabulary, grammar, pronunciation and fluency) in cycle 1 and cycle 4.

These results indicate that in general (both in cycle 1 and cycle 4), learners were mainly positive about the different aspects of their speaking performances, and especially about getting their message across (84.5% and 78.2%) and their vocabulary (66.9% and 58.5%). The students in our sample were least satisfied with fluency and pronunciation in both cycles. 'Don't knows' occurred the least for message and the most for pronunciation and grammar.

The McNemar-Bowker Test shows a significant shift for the category grammar (X^2 (3) = 8.57, p =.036): satisfaction for grammar increased from cycle 1 to cycle 4 (see Table 4.1). A trend was found for fluency (X^2 (3) = 7.72, p = .052): for this category satisfaction also tended to increase from cycle 1 to cycle 4. Other shifts were found to be non-significant (for message X^2 (3) =3.24, p = .355; vocabulary X^2 (3) = 3.70, p = .296; pronunciation X^2 (3) = 3.15, p = .370)

Table 4.2 shows the frequencies and percentages of the focus of plans for improvement in the first cycle and the fourth cycle for each category (message, vocabulary, grammar, pronunciation and fluency).

Table 4.2

Focus of plans for improvement

Category	Cycle 1	Cycle 4
Message	90	76
	(31.7%)	(26.8%)
Vocabulary	43	42
	(15.1%)	(14.8%)
Grammar	55	33
	(19.4%)	(11.6%)
Pronunciation	58	71
	(20.4%)	(25.0%)
Fluency	38	62
	(13.4%)	(21.8%)

Table notes: Frequencies (and percentages) of the focus of the plans for improvement (categories: message, vocabulary, grammar, pronunciation and fluency) in cycle 1 and cycle 4 and shifts form cycle 1 to 4.

Although in general the learners evaluated message as positive (Table 4.1), Table 4.2 shows that most plans for improvement focused on the message. Most students planned to think about what to say in advance, before doing the speaking activity and to note keywords and expressions.

The McNemar Bowker test indicated that the students shifted in their focus from cycle 1 to cycle 4 (X^2 (10) = 28.42, p =.002). As can be seen from Table 4.2, the greatest increase occurred for fluency (increase of 8.4%) and the greatest decline for grammar (decline of 7.8%).

In the first cycle the plans were least focused on fluency, and in the fourth cycle the least on grammar.

Development in independence

Table 4.3 shows the frequencies and percentages for the preferences for assistance from the teacher, collaboration with a peer or independent learning while executing the plan for improvement in the first and fourth cycles.

Table 4.3

Preferences for teacher's assistance, peer-work or independent learning

	Cycle 1	**Cycle 4**
Teacher's assistance	41	18
	(14.4%)	(6.3%)
Peer	115	101
	(40.5%)	(35.6%)
Independent	128	165
	(45.1%	(58.1%)

Table notes: Frequencies (and percentages) of preferences for teacher's assistance vs. peer work vs. independent learning in cycle 1 and cycle 4.

A McNemar test showed a significant shift from cycle 1 to cycle 4 (X^2 (3) = 17.14, p =.001). As can be seen from Table 4.3, this shift was mostly due to a decline in the need for teachers' assistance and a significant increase in preference for independence.

4.5.2 Results for research question B

To what extent did students experience feedback and activities for improvement as adaptive?

Adaptivity of activities for improvement and feedback from the learners' perspective - pre- and post-test

To establish whether there were differences in perceived adaptivity of the activities for improvement between the experimental and the control group and whether there were differences between the pre- and post-tests, the experimental group was compared to the

control group. Table 4.4 shows the means and standard deviations for adaptivity of the activities for improvement, resulting from the pre- and post-tests. Multilevel analyses revealed that neither the effect of time (F (1, 19.090) = 1.106, p = .306), nor the effect of group (F (1, 21.926) = .019, p = .893), nor the interaction (F (1, 19.090) = 2.852, p = .108) were significant. These results indicate that there was no effect of the intervention of the self-evaluation procedure on the students' perception of adaptivity of the activities for improvement.

Table 4.4

Adaptivity of activities for improvement

	Pre-test			**Post-test**		
Group	**Means**	**SD**	**n**	**Means**	**SD**	**n**
Experimental Group	4.19	1.07	329	3.93	1.17	225
Control Group	4.20	1.17	369	4.21	1.20	329

Table notes: Means, SD = standard deviations, n= number, for the adaptivity of activities for improvement on pre- and post-tests for the experimental and control group.

The experimental group was compared to the control group to investigate whether there were differences in the perceived adaptivity of the feedback between the two groups and whether there were differences between the pre- and post-tests. Table 4.5 shows the means and standard deviations for the adaptivity of the support, resulting from the pre- and post-tests. Multilevel analyses revealed that the effect of time (F (1, 1249.167) = 28,379, p <.001), and the interaction (F (1, 1249.167) = 10,005, p = .002) were significant. The effect of group (F (1, 22.841) = 2.416, p = .134) was not significant. In a post-hoc test, we found that only for the experimental group, was there a significant effect for time (p < .001), there was no effect for the control group (p = .096). These results indicate that the students in the experimental group perceived feedback aimed at improving their speaking skills as less adaptive in the period of the intervention (when the self-evaluation procedure was carried out) than before the intervention.

Table 4.5

Adaptivity of feedback

		Pre-test			Post-test	
Group	**Means**	**SD**	**N**	**Means**	**SD**	**n**
Experimental Group	4.38	1.18	329	3.76	1.25	225
Control Group	4.53	1.26	369	4.37	1.28	329

Table notes: Means, SD = standard deviations, n= number, for the adaptivity of feedback on pre- and post-tests for the experimental and control group.

Adaptivity of activities for improvement and feedback after a self-evaluation-procedure cycle – intermediate tests

Table 4.6 shows, from 339 self-evaluation-procedure cycles, whether or not the students considered the activities for improvement and feedback provided during a specific cycle to be adaptive. It appeared that both activities for improvement and feedback were considered adaptive by the students more often than one would expect on the basis of probability calculations (50%), both $p's < 0.001$.

Table 4.6

Adaptivity of activities for improvement and feedback after a self-evaluation-procedure cycle

	Total	**Not adaptive**	**Adaptive**
Activities for improvement	n=339	83 (169.5)	256 (169.5)
Feedback	n=339	102 (169.5)	237 (169.5)

Table notes: Observed frequencies (and expected frequencies) of *self-evaluation-procedure* cycles in which activities for improvement or feedback were or were not found to be adaptive by students.

4.5.3 Results for research question C

To what extent did students experience the self-evaluation procedure as motivating and did their speaking anxiety change during the course of iterations of the self-evaluation procedure?

Learners' motivation for the different students' main activities of the self-evaluation procedure

Table 4.7 shows the differences in students' motivation for the different main activities of the self-evaluation procedure: 1) recording of and re-listening to their speaking performances; 2) doing a self-evaluation of their speaking performances; 3) producing a plan for improvement; 4) executing the plan for improvement. Mixed repeated measures analyses revealed that there were differences between these components (F (3, 225.0) = 19.96, p < .001). Follow-up pairwise comparisons revealed that the differences between all four parts of the self-evaluation procedure were significant (p's <=.027). As can be seen from Table 4.7, these results indicate that producing a plan for improvement and especially executing a plan for improvement were the most appreciated components of the self-evaluation procedure.

Table 4.7

Learners' motivation for the different components of the self-evaluation procedure

	Mean	Standard Deviation	N
1) Recording and re-listening to their own speaking performance	3.28	.17	225
2) Doing a self-evaluation of the own speaking performance	3.60	.16	225
3) Producing a plan for improvement	3.77	.16	225
4) Executing the plan for improvement	3.96	.15	225

Notes: Means, standard deviations, and n = number, for students' motivation for the different parts of the self-evaluation procedure

Speaking Anxiety

The experimental group was compared to the control group to establish whether there were differences in speaking anxiety between the two groups and whether there were differences between the pre- and post-tests. Table 4.8 shows the means and standard deviations for speaking anxiety, resulting from the pre- and post-tests. Multilevel analyses revealed that neither the effect of time (F (1, 1037.877) = .000, p = .997), nor the effect of group (F (1, 22.095) =2.231, p = .149), nor the interaction (F (1, 1037.877) = .184, p = .668) were significant. These results indicate that there was no effect of the intervention of the self-evaluation procedure on the students' speaking anxiety. Overall scores (between 2.58 and 2.75) were slightly lower than neutral on speaking anxiety but the standard deviations suggest quite large differences.

Table 4.8

Speaking anxiety

	Pre-test			Post-test		
Group	**Means**	**SD**	**n**	**Means**	**SD**	**n**
Experimental Group	2.74	.56	171	2.75	.63	179
Control Group	2.59	.60	369	2.58	.63	329

Table notes: Means, SD = standard deviations, n= number, for speaking anxiety on pre- and post-tests for the experimental and control group.

4.6 Conclusions and discussion

Guiding students to become autonomous learners in learning to speak foreign languages is an important goal in foreign language education (Holec, 1981; Lee, 1998; Little, Dam & Legenhausen, 2017), but difficult to realize in regular classroom settings in secondary schools (chapter 3). The aim of this study was to investigate whether a self-evaluation procedure could be an adaptive resource for secondary school students to learn to use self-regulation to improve their foreign language speaking skills. We first outlined what is needed to promote

self-regulation in speaking skills in order to derive design principles for the teaching practice. We then proposed a concrete self-evaluation procedure for speaking skills based on these design principles which we implemented in a number of secondary schools. We investigated the extent to which changes occurred in students' perceptions of learning needs as they tried to improve their speaking skills after four iterations of the self-evaluation procedure and to what extent the students perceived the self-evaluation procedure as motivating and the feedback and activities for improvement as adaptive to their needs. Below, we discuss our results and possible explanations for each research question in more detail.

Research question A *To what extent did the students' perception of their learning needs change during the self-evaluation procedure?*

An important goal of the self-evaluation procedure was to support secondary school students to become more and more independent in fulfilling all the different parts of the self-regulation process. The results showed that the perceived need for teachers' assistance decreased and the preference for independence increased in the fourth cycle compared to the initial round. This suggests an improvement in self-regulation.

Regarding the diagnoses and plans for improvement, some perceptions remained the same, but shifts in diagnoses and focus of plans were also found. In both cycles, the secondary school students generally evaluated many aspects of their speaking performance as positive. The number of 'don't knows' did not change much but areas for improvement decreased in the fourth cycle. That might indicate that the students were more satisfied with their speaking performances in the later cycles because they had improved their speaking skills.

We found from the diagnoses that students were particularly positive about getting the message across. Nevertheless, most plans for improvement still aimed at improving how they got their message across in both cycles (although there was a decline in focus on the message in the fourth cycle). After all, getting the message across is the most important goal of communication and therefore students' focus will be on that goal. Moreover, a closer look showed that most of the plans for getting the message across were very useful pre-plan activities (Goh, 2017; Goh & Burns, 2012; Skehan, 1998).

In both cycles most 'don't knows' were found for grammar and pronunciation. The results did not show a significant decline in 'don't' knows'. However, shifts were found for grammar in diagnoses and plans for improvement: learners were more positive and least

negative about the grammar in their speaking performances in the later cycle and there was a decline in focus on grammar in the plans for improvement from the first to the fourth cycle. Research has shown that learners find it difficult to assess themselves (Blanche, 1988; Poehner, 2012; Ross, 1998), especially on grammar and pronunciation, which might be due to their lack of the metacognitive and linguistic knowledge needed to determine the appropriateness of their utterances (e.g. Blanche, 1988; Lappin-Fortin & Rye, 2014; Dlaska & Krekeler, 2013). A tentative explanation may be that in the first cycle students did indeed have difficulties assessing their grammar and pronunciation, but as they went on, they improved and gained more insight into their grammatical competence stimulated by the self-evaluation procedure.

Besides an increase in satisfaction about grammar in the diagnoses, increasing satisfaction about fluency was also found. While there was least focus on fluency in the plans for improvement in the first cycle, in later rounds, an increase in plans for fluency was found. It seems that students expanded the focus of their plans. The shifts in focus (decline for grammar and for getting the message across, increase for fluency and pronunciation) might indicate that the students had broadened their awareness of different aspects of their speaking performance during the cycles of the self-evaluation procedure.

Research question B *To what extent did students experience feedback and activities for improvement as adaptive?*

Adaptivity of feedback and improvement activities were investigated in two ways: in pre- and post-measurements among an experimental and a control group and in intermediate questionnaires each time directly after the accomplishment of a specific cycle of the self-evaluation procedure carried out by the experimental group.

The results of the pre- and post-measurements showed that, in the period in which the self-evaluation procedure was used, the students found the activities to improve speaking skills equally adaptive and the feedback less adaptive than in regular teaching practice. However, the findings from the intermediate questionnaires indicated that students mainly considered that both improvement activities and feedback were tailored to their needs in the specific self-evaluation procedure cycles.

One way to interpret these data is that the questionnaire in the pre- and post-measurement addressed a whole period of time, whereas the interim questionnaire focused

on a specific cycle of the self-evaluation procedure. The perception of the feedback and improvement activities would be more concrete in such interim questionnaires than when asked about feedback and activities in general over a longer period of time.

Another explanation for the decline in the post-measurement might be that students had become more critical through the self-evaluation procedure, that is, more conscious by thinking about whether they were getting what they needed.

Research question C *To what extent did students experience the self-evaluation procedure as motivating and did their speaking anxiety change during the course of iterations of the self-evaluation procedure?*

The results showed a mean of 3.65 for appreciation of the different main activities of the self-evaluation procedure on a scale of 5. It was expected that recording and re-listening to their own speaking performances would be the least motivating because listening to one's own voice can be a strange experience, but it was still appreciated with an average score of 3.28. Making and executing a plan for improvement was most appreciated.

We did not find any change in speaking anxiety after the self-evaluation procedure. As with the questionnaire on adaptivity, the questions addressed anxiety about speaking in the classroom in general, rather than anxiety during a self-evaluation-procedure cycle specifically. Additionally, it may be that the period of the intervention (four months) was too short to bring about a change in speaking anxiety. Longitudinal research in which the self-evaluation procedure is used structurally for a long time should show whether the procedure has an effect on the level of speaking anxiety.

Limitations of this study

The ultimate goal of the self-evaluation procedure was to help students to improve their speaking skills themselves. A limitation of this study was that we did not measure whether there was any improvement in speaking skills. The scope of this study was constrained by its aim which was to help students *to learn to self-regulate* their speaking skills. Therefore, we did not measure speaking skills and cannot say whether and to what extent they may have improved.

Another limitation is that this study focused on the perceptions of the students. Shifts in evaluations and plans were found. However, we cannot conclude that these shifts mean

that the students had learned to assess themselves better and make better plans, nor that the self-evaluation procedure actually had an impact on the improvement of their speaking skills. In order to draw such conclusions, follow-up research is required, which would include not only the students' assessments, but also the assessment of an expert, for example the teacher. Moreover, longitudinal research would be required to establish changes in the degree of self-regulation or the improvement of speaking skills over time.

Furthermore, we did not investigate which feedback and which improvement activities the students actually received. Follow-up research in which the teachers' choices regarding the kind of feedback and the concrete provision of learning activities based on the students' plans, would provide more insight into how teachers tailor their feedback and activities to learners' needs.

Another limitation of this research was that the study focused on self-regulation by a heterogeneous group of secondary school students. Further research should be carried out in order to identify any differences in terms of year and language. It might be that lower level students differ from higher classes in meta-cognitive skills and therefore would differ in, for instance, independence and need for assistance during the self-evaluation procedure.

Implications

Despite the limitations of this study, we think it contributes to the development of knowledge about guiding students to become autonomous learners in learning to speak foreign languages. Other researchers have also argued for an iterative learning process in which learners gradually become more independent in self-regulating (e.g. Little, 2017) and some have already proposed a cycle of refection and task-repetition in order to improve speaking skills (Goh, 2017; Goh & Burns, 2012). This study adds concrete design principles to realize such an iterative learning process and proposes how students could actually go through a process of self-regulation independently by means of a self-evaluation procedure.

This self-evaluation procedure differed from the more common self-assessments in a number of respects. First of all, the evaluation addressed a specific speaking performance and not students' speaking skills in general. Second, the speaking performance was recorded which enabled the students to listen back to their own speaking performance. Third, instead of normative use of self-evaluation, a self-evaluation instrument was used which contained

non-normative criteria that focused students' attention on both areas for improvement and positive points of different aspects of their speaking performance.

Another break from common practice was that the self-evaluation procedure in this study not only consisted of a diagnosis of the speaking skills, it also contained a plan for improvement produced by the students and where necessary students' requests for teacher's assistance.

Finally, the purpose and use of self-assessment was slightly different from other formative uses. In formative uses, self-assessments are often used by teachers to adapt their teaching (e.g. Black & William, 1998; 2009; Nicol & Macfarlane-Dick, 2006). In this study, however, the diagnoses with plans not only provided information for the teacher, but also aimed to support students to self-regulate their speaking skills. On the basis of their diagnoses, the students themselves had to design and implement their own learning pathways, indicating where they needed help in order to enable the teachers to align their feedback and learning activities.

In conclusion, the results of this study showed that during the self-evaluation procedure students' perceptions of their learning needs did indeed change and that students found the procedure to improve their self-regulation of their speaking skills in foreign languages both adaptive and motivating. We therefore recommend use of the design principles of the self-evaluation procedure for teaching practice in secondary schools. We hope that follow-up research into their effects will be carried out.