Bachelor thesis
Computing Science

Radboud University

Using automated feedback as a base for performing self-regulated learning in programming exercises

Author:
Robin Immel
S4372891

First supervisor/assessor:
Prof. Dr. E. Barendsen (Erik)
e.barendsen@cs.ru.nl

Second assessor:
Dr. J.E.W. Smetsers (Sjaak)
s.smetsers@cs.ru.nl

July 3, 2022

Abstract

The aim of this thesis is to gain insight in the effectiveness of automated feedback as a base for self-regulated learning in an introductory Python course. To do so, we applied a model for self-regulated learning and investigated three qualities of feedback. The course is set in a high school environment and taught to students aged 15 to 17. The automated feedback was provided to students by using Repl.it in combination with regular expressions.

Three types of data, about the correctness and quality of the feedback, were collected and analyzed by use of multiple surveys. The results indicate that students had a neutral to positive attitude towards the qualities of the automatic feedback, suggesting that the automated feedback can be used effectively in this environment.

Contents

Chapter 1

Introduction

The use of formative assessment has been proven to provide a better learning environment for students as well as teachers. Due to the nature of formative assessment however, having many in-depth feedback moments (instead of just one), makes it a time-consuming effort for both students and teachers. To assist teachers with their endeavour to maximize the frequency and quality of the feedback moments, one can find assistance in automated testing and feedback. By supplying students with a framework that supports automated feedback, students can test and evaluate their own work, after which they improve their code, removing repetitive labour from the teacher.

This process can be described as a form of self-regulated learning. A way in which students take control of and evaluate their own learning, without (a lot of) direct interference of their teacher. At universities, this has proven to be a successful way to promote the level of learning.

In this thesis, a (conceptual) model about self-regulated learning made by Nicol and Macfarlane will have the central role.(Nicol & Macfarlane-Dick, 2006) This model describes how students can take control of their own learning by making use of formative assessment and feedback. The model itself is designed for students in higher education. It is argued that the key argument why the model works for these students is that they are already generating their own feedback, and that higher education should build on this ability.(Nicol & Macfarlane-Dick, 2006)

In this thesis however, the model is used to analyze the learning process of students in high school regarding the (automated) feedback. At the high school described in this thesis, the Montessori College, a high level of self-dependence is expected by the students. A big part of this is generating their own goals and internal feedback on how to persuade these goals. Thus fulfilling the key argument which is needed for making use of the model according to Nicol and Macfarlane.

Next to this level of self-dependence, Nicol and Macfarlane describe seven principles of quality regarding feedback which have to be present. In this thesis, the presence of three of these seven principles will be investigated, thus making the model (partly) applicable for high-school students as well. To investigate the presence of these qualities, the following research question is formulated:

To what extent is automated feedback suitable for self-regulated learning in an introductory programming course?

To answer this question, the opinions and workflow of the students mentioned above were examined in a few different ways. These examinations will lead to the conclusion to what extent the qualities of feedback as described by Nicol and Macfarlane are present. If the qualities are deemed present, it can be concluded that automated feedback can indeed serve as a base for self-regulated learning according to the model of Nicol and Macfarlane.

Chapter 2

Background

As mentioned in the introduction, in this thesis, a theoretical model about self-regulated learning made by Nicol and Macfarlane will be applied. The conceptual model describes a circular process in which the learner, a student, gains knowledge through several internal and external processes. The main goal of the model is to describe how the learner uses internal and external feedback to improve their work and work-methods. To be able to use this model to its full extent, preliminary information is required. This knowledge will be given in the following chapter.

First, some definitions and uses of formative assessment will be given. These will be expanded upon as well by giving some background information. In the research by Nicol and Macfarlane, a definition by Black and Williams (2009) was used to provide a base for their theory. We will adopt this definition as well. Afterwards, the use of formative assessment will be linked to the process of self-regulated learning; a process whereby the learner actively takes control of their own learning. Conclusively, both formative assessment and self-regulated learning will be combined to describe and use the conceptual model by Nicol and Macfarlane. In this model, seven principles of quality which feedback must adhere to are described. Due to the focus on the process of the learner in this thesis, three of these points will be focused upon.

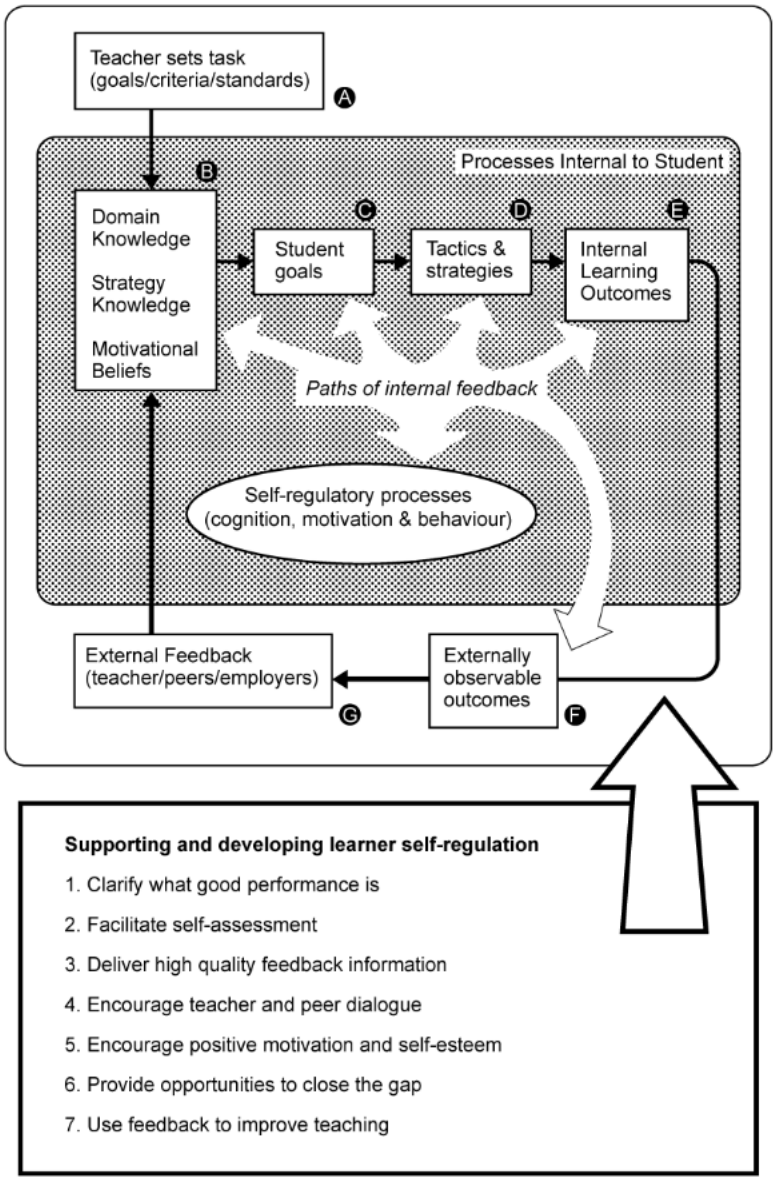

Figure 2.1: The conceptual model for self-regulated learning and feedback principles by Nicol and Macfarlane.

2.1 Formative Assessment

In many educational systems, summative assessment has long been the standard way of grading assignments. It is considered to be the most classical way of grading. The main property of summative assessment is the way in which assignments are given and how these are graded at the end. The assignments often entice knowledge which is given over a longer period of time. At the end of this period, a test or exam is given to test the knowledge level of the student. Generally, this will result in a grade to rate the students' work. This grade will be (part of) the end-result for the student, without getting the chance to improve and learn from their mistakes.
Research shows that when feedback is given at the end of the assignment or test like this, students are less likely to learn from their mistakes. (Black & Wiliam, 2009) It also leads to a less positive way of thinking, especially when said grades are not satisfactory. This will in turn lead to a negative self image and less motivation to improve on their work as well. (Black & Wiliam, 2009) Over the past few decades, the way of teaching and grading has set a different course. The classic way of learning as described above has adapted to a more interactive form. Students are expected to have a more active role in their learning process. One of the methods to enhance this way of learning is the use of formative assessment.

The first mention of the term 'formative' in the literature was made in 1963 by Cronbach. The goal of this research was to enhance curriculum development by adding evaluation moments about it during the course. (Cronbach, 1963) By continuously planning and executing these moments of evaluation, teachers were able to gain empirical knowledge about the effectiveness of their curriculum. Moreover, improvements could be made on the subject, making for a better learning experience. (Dick, 1987) This was however still aimed only at the curriculum, instead of on the teacher and learner as well. In later research , the focus shifted from the assignments and way of teaching to how students perceive them. Considered by many to be a pivotal moment in this shift is the work of Sadler in 1989. He argued that students should be able to assess their own work as well as how to improve on it. (Sadler, 1989) (Nicol & Macfarlane-Dick, 2006) Sadler argued that for students to be able to assess their own work during the act, three conditions should be met.

1. Students possess an appreciation of what high quality work is
2. Students have the evaluative skill to the quality of what they are producing in relation to the higher standard
3. Students develop a store of tactics or moves which can be drawn upon to modify their own work

Concretely, this means that students should be able to fill the gap between their own work and the result a teacher had in mind when making the assignment. For this, students should have a number of (predetermined) steps or moves to iteratively get closer to the required answer. In the research, it was made clear that these skills are hard to develop for students on their own and should therefore be presented by a teacher. (Sadler, 1989)

According to Black and William, both summative and formative assessment are used in practice to complement each other. (Black & Wiliam, 2009) Throughout the learning process multiple moments of feedback are to be used to further enhance the students answer and their concept of learning. After several of these moments, for example at the end of the course, a grade needs to be given accordingly. This is an example of summative assessment.

2.2 Seven principles of good feedback

As stated earlier, to provide an environment in which self-regulation is possible, there still needs to be a form of external feedback. According to Nicol and Macfarlane, this can be by a teacher, fellow students or even an electronic environment. In this thesis, the latter will be used as the main form of external feedback. Through research literature on formative assessment in combination with the self-regulation model, Nicol and Macfarlane identified seven principles of good feedback practice. (Also visible in figure 2.1).

1. Clarify what good performance is
2. Facilitate self-assessment
3. Deliver high quality information
4. Encourage teacher and peer dialogue
5. Encourage positive motivation and self-esteem
6. Provide opportunities to close the gap
7. Use feedback to improve teaching

Due to limited time and resources for this bachelor thesis, the focus will be on the first, third and fifth of these principles. To avoid confusion in the remaining part of this thesis, these principles of feedback which are focused upon will be numbered as follows:

1. Helps clarify what good performance is
2. Deliver high quality information

3. Encourage positive motivation and self-esteem

In the following subsections an in-depth explanation will be given about these three principles of good feedback.

2.2.1 Helps clarify what good performance is

The first principle of feedback which Nicol & Macfarlane mention revolves around the knowledge prior to every self-regulated cycle of learning.

At the start of such a cycle certain goals have to be set by an external factor, e.g. a teachers' explanation. The set goals have to resemble an application of prior knowledge to a certain environment. Whenever these goal are not clear enough, or are perceived differently by the students, the assignment will likely not "connect" and will thus not be interpreted as intended. (Hounsell, 1997) As the model proposed by Nicol & Macfarlane is cyclical, the above also needs to hold for other moments of external input, being the external feedback at the end of the cycle. This means that next to the assignment needing to be clear, the provided feedback should also resemble the goals set above and the prior knowledge of the student. In their paper Nicol & Macfarlane mention several ways to make sure the teacher and students are on the same level.

One of the ways to be sure the students and teacher are on the same line is to use exemplars. These make explicit what is required to develop a program which stands up to the teachers standards. These exemplars can be used to compare the students' solution to the standard.(Nicol & Macfarlane-Dick, 2006)(Orsmond, Merry, & Reiling, 2002) These exemplars can in turn be used to uncover/explain parts of the assignment to guide the student through the intended thinking process.

2.2.2 Deliver high quality information to students about their learning

According to Nicol and Macfarlane, the quality of the feedback when taking self-regulation in mind is as follows: "Good quality external feedback is information that helps students trouble-shoot their own performance and self-correct: that is, it helps students take action to reduce the discrepancy between their intentions and the resulting effects."(Nicol & Macfarlane-Dick, 2006) So to be qualified as good feedback, it needs to give the students relevant information on how to improve their work to be more like the aforementioned standards set.

2.2.3 Encourages positive motivational beliefs and self-esteem

To keep the students working in the cycle of self-regulation as mentioned by Nicol and Macfarlane, it's important to keep them interested in the topic.

This is done by regular and on time feedback. An even more important factor is to have many low-stakes assessments where feedback is given (homework assignments) instead of having a few high stakes where grades are the only thing that matters. (Nicol & Macfarlane-Dick, 2006)

2.3 Self-regulated Learning

A way to implement formative assessment, which has been proven successful, is by making use of self-regulated learning. This entices that the student takes a pro-active role in their learning process by taking control of the process itself.

Nicol and Macfarlane consider the definition of self-regulated learning by Pintrich and Zusho to be leading in their paper: "Self-regulated learning is an active constructive process whereby learners set goals for their learning and monitor, regulate, and control their cognition, motivation, and behaviour, guided and constrained by their goals and the contextual features of the environment." (Pintrich & Zusho, 2002) This definition is built upon the phases described by Zimmerman in figure 2.2 and thus strongly resembles the model as well.

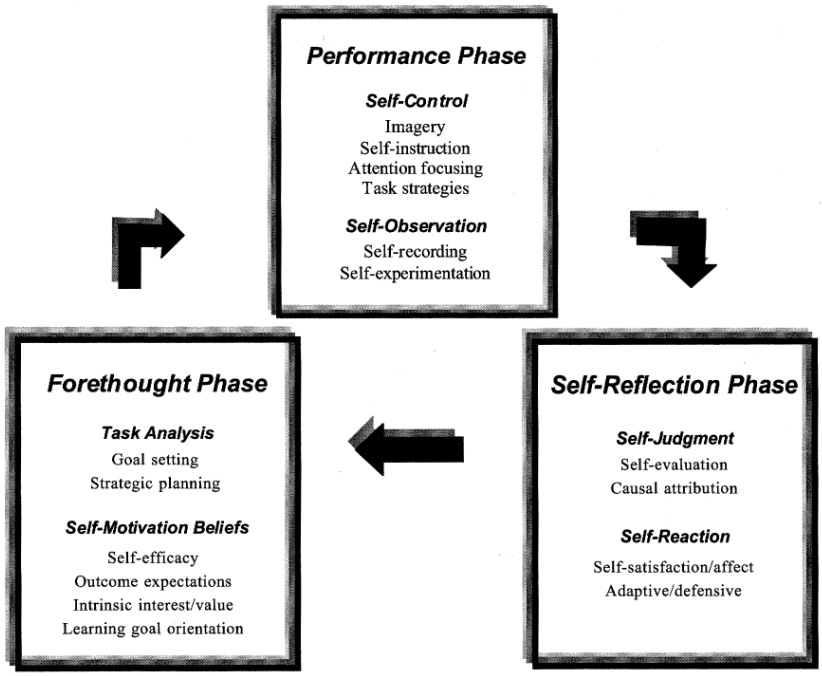

Figure 2.2: The phases of self-regulated learning described by Zimmerman.

Self-regulated learning is a process of a cyclical form. In this, three stages of forethought, performance and self-reflection follow each other. The rest of this section will explain these phases and their properties.

Forethought phase
The first phase in Zimmermans model entices two categories. One is the analysis of what to do and another is the act of self-motivation. Before starting to work on an assignment, a student should first know what is asked of them. (Zimmerman, 2002) According to these ideas, a goal should be set by the student on how to complete the assignment. In this completion, some intermediate steps could be required as well. These intermediate steps should be aspired and thought of by the student before the assignment, creating several sub-goals. Self-motivation entices the student believing the assignment can be completed, having faith in a sufficient end-goal and an interest in the assignment itself.

Performance phase
Secondly, the student needs to actually do the assignment. This process in described in the performance phase and can be divided in two main categories. The act of self-control and self-observation. Self-control is described by Zimmerman as the process of actively pursuing the goals set in the forethought phase. With this, it is important to apply different strategies to solve the assignment with the intended goals in mind.

Self-observation means that the student will have to actively observe their thinking and motivations during the assignment. This will in turn lead to the following phase.

Self-reflection phase
The third and last phase of the cycle described by Zimmerman is the self-reflection phase. The student should reflect on the process of making the assignment. Doing this, the thoughts described above could lead to a reaction in the learning process. These reactions could have two consequences according to Zimmerman. A good outcome of the assignment will lead to the student being satisfied, while a worse result will lead to the student changing their learning process.

Chapter 3

Educational context

Before taking a look at the methods used in this research, some extra educational context will be given as well as some information about the used tools. Firstly the participants will be shortly described. Secondly, the used tool, Repl.it, will be inspected. Thirdly, the given assignments will be described with an example.

3.1 Participants

This research focuses on high-school students of the Montessori College in Nijmegen. As mentioned earlier, the school is special in the fact that it promotes a high level of self dependence. This means that the students are already used to work without a lot of interference of their teacher.

The students have an age ranging from 15 to 17 and are in the second to last class of high-school. The course being given is of an introductory level. Consequently, the students have little to no previous programming experience. The group consists of 12 students.

3.2 Repl.it as Tool

The used tool to teach the students programming is the online tool Repl.it. This tool allows a teacher to create classrooms, assign students and give assignments accordingly. For a **teacher**, the Repl.it assignments are build up out of three major parts.

First, there is a view of the assignment itself. The screen is divided in three compartments. On the right, an explanation about the current assignment can be given. On the left there are an IO-terminal and a (mostly) blank text field for students to code their program. The teacher can provide some code as a framework as well.

After the teacher is satisfied with the assignment, the next step is to provide the automatic unit tests and feedback. For each programming con-

struction which needs to be present in the students' code, the teacher needs to provide a single unit test. This is to be as concise as possible when providing the automatic feedback to the student.

In Repl.it, this automated feedback can accomplished by a combination of unit tests and regular expressions. This way, a more precise form of feedback can be given per case. Eventually, the tests are all of this (simplified) shape:

```
with open("/home/runner/main.py", 'r') as file:
    content = file.read()

    searchObj_1 = re.search(r"(.|\n)*(\+\s?1)(.|\n)*",content,
        re.I)
    searchObj_2 = re.search(r"(.|\n)*(\+=\s?1)(.|\n)*",content,
        re.I)

    if(searchObj_1 or searchObj_2):
      self.assertTrue(True)
    else:
      self.assertFalse(True)
```

On the first two lines, the file containing the students' code is opened and converted to a string. Afterwards two searchObjects are created to apply regular expressions on the content of the students' answers.

These searchObjects will return true if the expression is found in the answer. If this is not the case, the function will return false and thus require feedback to be given to the student.

For **students**, the views of Repl.it are simplified as opposed to the teachers view. Students start on the same assignment view as teachers do. They try to make the assignment and try to submit it. Whenever

3.3 Assignments

As mentioned above, the assignments studied are part of an introductory programming course. They have to be made in Python and include beginner topics with the concluding (and hardest) topic being nested for-loops.

Every week a new topic is explained, where after new assignments need to be made. This will be done over a span of four weeks. The topics per week are sequentially about: assigning and printing variables, strings and how to print them, conditionals, and finally nested for-loops over arrays.

All of these assignments share the same form. They are build up out of three components.

- An **explanation** is given. The theory students learned before will be connected to the theory on the assignment.
- Some **pre-made code** is given as a framework for the students code.

- After submitting, **automated feedback** will be shown to the student.

A students' view of an assignment can be found in the appendix in figure A.1.

Chapter 4

Aim of the study

As mentioned in the introduction, the research question for this thesis is as follows:

To what extent is automated feedback suitable for self-regulated learning in an introductory programming course?

In the preliminaries we have seen that self-regulated learning still requires a form of external feedback for students to succeed in self-regulated learning. According to Nicol and Macfarlane, for external feedback to be considered effective for self-regulated learning, it has to meet seven qualities. For three of these seven qualities their presence will be examined in this research.

Taking the aforementioned in consideration, the research question:

To what extent is automated feedback suitable for self-regulated learning in an introductory programming course?

Is divided into three sub-questions, directly corresponding with the three qualities of feedback:

1. *Does the automated feedback clarify what good performance is?*
2. *Does the feedback deliver high quality information?*
3. *Does the feedback encourage positive motivation and self-esteem?*

Chapter 5

Methods

To investigate the presence of the qualities of feedback mentioned in the sub-questions in Chapter 4 on page 15, four different methods of research were conducted in this research. The decision to make use of four different ways was due to the different nature of the principles as well as to keep the students interested.

The first method was a test for correctness was done on the feedback construction. This was done to rule out mistakes in the automated tests, the assignment or the Repl.it platform itself.

Additionally, examinations were done in the form of questionnaires conducted on the students. These questionnaires were each of a different form for the reason mentioned above. The first questionnaire was of a Likert-scale form. The second one was a form with open questions for the students to answer. The concluding questionnaire consists of two in-depth semi-structured interviews with students. These interviews consisted of the same open questions as the second questionnaire.

Each of these methods will be described in the next section. Each method is split up into two parts; a data collection and a data analysis.

5.1 Data collection

5.1.1 Correctness feedback

Ideally, feedback is always correct. However, it is possible that mistakes were made. Considering this research and which tools were used, the mistakes were either in the Repl.it platform, the automated feedback or the question.

To determine whether or not the given feedback was correct, both the students' answers and expected answer were collected from Repl.it. The expected answer was written by the teacher and thus was considered to be the right answer. The students' answer was submitted after a student deemed their answer correct. Most of the time this was done when no additional feedback was needed to be given. However, students could also hand in their assignment when the tests still gave feedback if they deemed their answer correct already. Either way, this submitted answer was considered to be the students' final answer and was compared to the expected answer.

5.1.2 Likert-scale survey

In the first week, students were asked to answer three questions about a few assignments in the form of a 3-point Likert-scale. As mentioned earlier, some students were faster than others and answered more questions, thus leading to a difference in admissions per assignment. For each assignment made, the students were asked to answer the following questions.

- The check on my answer was correct.
- I know what is meant by the feedback.
- I know how to proceed now because of the feedback.

Originally, these questions were in Dutch. These are added in the Appendix.

5.1.3 Open questions

Survey

In the second week a survey with open questions was conducted. These questions were about one assignment. The survey required students to answer several questions before, during and after the assignment.
The questions before the assignment:

- Do you know what is expected of you?
- What do you think are potential pitfalls?

The questions during the assignment:

- Do you know what is meant by the feedback?

- How does your answer differ with the correct one?
- How are you going to get your answer more like the correct answer?

The questions after the assignment:

- Did you find the provided feedback helpful?
- Did the feedback help you with the assignment?

Originally, the questions were in Dutch. These are added in the appendix in Table A.1.

Interviews

In the final week of the experiment, two groups of two students were interviewed and (audio-)recorded. The questions and structure of the interviews were the same as the open question survey for sake of consistency. This was done to give a more in-depth look at the opinions of the students. As the questions are the same as the above, the structure of the interview will follow the survey as well; first, the students read the assignment. Afterwards, the questions mentioned above were asked before the students started working on the assignment at hand. After these questions, the students started working on the assignment and provided an answer by submitting it. Whenever the students' answer differed from the desired answer, automated feedback was shown by the system. The questions shown above were asked about the feedback. After answering these questions, the students proceeded to finish the assignment.

After the students completed the assignment and the final submission was done, the last questions were asked accordingly.

5.2 Data analysis

5.2.1 Correctness feedback

As explained above, each student had one (final) submitted answer per assignment. Every one of these answers was compared to the relative exemplary answer as provided by the teacher. The comparison was done to provide a sense of correctness regarding the students' answers and the feedback on those answers. The outcome of the comparison can result in three options. The first option is a correct combination of a students' answer and expected answer. In this case, no additional feedback was given to the student as it was not needed. Naturally, this is the desired outcome.

The second option of the comparison, a wrong answer of the student being graded as a correct answer, result in false positives. In this case more feedback was needed to be given, but was not. This led to the student not being given enough feedback and therefore wrongly passing the assignment.

The third option of the comparison is the the opposite of the latter. A correct answer being graded as a faulty answer. This resulted in the student being given feedback which was not needed and thus wrongly failing the assignment.

Whenever a false positive or false negative occurred, the assignment was noted as well as the programming construction on which the test failed. This provided an insight in which common mistakes/pitfalls can be made when making use of automated feedback.

5.2.2 Likert-scale survey

The focus in the first week was on the first two qualities which feedback must meet to be considered good feedback for self-regulated learning.

By asking whether the student knows what was meant by the feedback, it becomes clear whether this feedback was deemed in line with the assignment according to the students. This means that the asked question corresponds with the previously taught material. When it was considered clear, it could be assumed that the feedback fits well with the material and describes the purpose of the assignment. Consequently, this means that the feedback clarifies what good performance is as well. Thus contributing positively to the first sub-question: *"Does the automated feedback clarify what good performance is?"*

By asking whether or not the students know how to proceed by making use of the given feedback, we determine whether the gap between the answer of the student and the expected answer is clear to the student. Effectively, this means that the student can make and execute a plan to reach the desired answer from their own (incomplete) answer. Considering the above, this means that the quality of the feedback was sufficient for the students and

thus contributing positively to the second sub-question: *"Does the feedback deliver high quality information?"*

5.2.3 Open questions

Survey

As stated in the data collection, the second week consisted of an open question survey. These questions were designed to focus on one of the three of the qualities of feedback. Several questions were asked, these are categorized below per quality of feedback in Table 5.1. For each question, a student either gave a negative or a positive answer. To be able to label the data, the students' answers on the survey were inserted into the data analysis program Atlas.ti. The analysis in Atlas.ti was started by giving (parts of) the students' answers an open coding. Due to the surveys (and interviews) being in Dutch, these open codings were done in Dutch as well, to keep the codes as close to the text as possible.

Subsequently, by using axial coding, the open codes were grouped into code groups. Each of these code groups corresponds with one of the qualities of feedback as described in the sub research questions. To be able to distinguish positive and negative responses in the students' opinions, the code groups were split into positive and negative categories as well. This way, a clear division was distinguished whether or not the points of feedback were deemed present by the students.

Table 5.1: Qualities of feedback and corresponding questions

Quality 1	*Does the automated feedback clarify what good performance is?*
	Do you know what is expected of you?
	What do you think are potential pitfalls?
	Did the feedback resemble the taught material?
Quality 2	*Does the feedback deliver high quality information?*
	Do you know what is meant by the feedback?
	How does your answer differ with the correct one?
	How are you going to get your answer more like the correct answer?
Quality 3	*Does the feedback encourage positive motivation and self-esteem?*
	Did you find the provided feedback helpful?
	Did the feedback help you with the assignment?

Interviews

As stated earlier, the interviews in the final week followed the structure as the written open questioned survey. In addition to this, the recorded interviews were transcribed and inserted into Atlas.ti as well. Furthermore, the corresponding structure of questions as mentioned in Table 5.1 were used.

Chapter 6

Results

6.1 Correctness feedback

As stated in the methods, the correctness of the feedback consists of a variety of outcomes regarding the students' answers and the evaluation of these answers. In figure 6.1, the variety of outcomes can be seen. Below, a more in-depth explanation of occurrences per category will be given.

Table 6.1: Student submissions correctness

Situation	Amount
Total submissions	55
Valid passes	32
Invalid passes	1
Valid fails	12
Invalid fails	10

As can be seen in Table 6.1, a total of 55 submissions were done by the students. These submissions were divided into four different categories. Of these categories, valid passes and valid fails are desired categories. False passes and false fails on the other hand are not desirable. The outcome of the correctness is split up into these categories below.

Valid passes
The majority of the students' submissions, 32 of the 55, were correctly submitted and evaluated as such by the tests. Due to the simple programming constructions, students often followed the same method of coding, which coincided with the explained method as well. The only minor differences in the valid passes lied in coding syntax and/or spacings. As the used language, being Python, requires a consistent use of indentation to work properly, the students either used tabs or spaces. These were interpreted differently as

well. Due to the reasons given above, the simpler coding constructions did not give any considerable problems for both the students and the automated feedback.

Valid fails

As seen in Table 6.1, out of the 13 failed submissions, 12 were evaluated as such by the program as well. These fails can be divided into two categories.

Most mistakes turned out to be because of a difference in understanding (between the students and teacher) of what was needed to fulfill the assignment. Within the explanation of the assignment, the teacher stated the desired outcomes. However, the students did not reach the outcome as desired by the teacher.

The other category of fails was due to the students not reading, understanding or following the extra hints provided by the automated tests and feedback. Upon inspection of the students' work, we concluded that their answer would have been correct if they would directly use the given feedback to improve upon it. However, the students were unable to improve their answer as such in this cases.

Invalid passes

There was only 1 invalid pass. In the assignment, the student had to use a print()-function within a for-loop when a particular variable was present. However, the student used a print()-outside the loop construction, instead of inside the loop. As a result, all elements which were evaluated by the tests, including the print(), were present and thus the submission was approved. However, the submission did not correspond with the intended result and should therefore have been evaluated as incorrect.

Invalid fails

10 out of the 55 submissions were invalid fails. The programs were correct and had the desired behavior as described by the teacher, but were evaluated as incorrect by the tests. The main reason for this was due to minor differences in syntax and spacing. As mentioned earlier, Python is very strict when it comes to indentation. At various times, the regular expressions in the test took into account the use of spaces, but not the use of tabs. In addition to this, some tests did not take into account the fact that Python can create Strings with both single quotes and double quotes. Both of these are constructions are correct Python and should be evaluated as such. However, if one of the two was not present and the student used exactly that form, the submission was wrongly evaluated as incorrect. This resulted in the invalid fails.

It is important to note that, partly because this was a basic programming course, major differences in programming methods did not occur. The students often used the same constructions as explained in the assignment.

When more difficult programming assignments would be given, the amount of invalid fails will most likely raise as well.

6.2 Likert-scale survey

As mentioned in the methods, the students were given several three point Likert-scale surveys. Each assignment was graded by students in terms of three questions. This resulted in a negative, neutral or positive result for each question. Herein, questions two and three of the survey correspond directly with the points of feedback indicated below.

Question 1: ***The check on my answer was correct.***

As can be seen in Table 6.2, there was a division as to whether the students considered the automatic feedback to be correct. Considering the two assignments, this had two separate reasons. In assignment 7.2.3 the division is due to an error in the automatic feedback. Part of the assignment was not correctly included in the automated tests. As a result, the assignment continued to provide feedback, while it no longer had to.
In assignment 7.7, the feedback was wrong due to a problem with the underlying program. Due to a misinterpretation of the regular expression by the Repl.it compiler, an exception was thrown to the students. This prevented the students from seeing the correct feedback, so the feedback could not be seen as correct either. For the rest of the assignments, the feedback on their answers was understood and therefore considered correct.

Although this question was asked as a control question to check the correctness of the feedback, the divisiveness of answers above immediately reflects one of the great dangers of automated feedback giving. Due to the nature of automated feedback, i.e. giving the same type of feedback to a larger group of people, technical problems or human errors become a bigger problem than when individual feedback is given. A small mistake or discrepancy immediately reaches the entire group of students, which negatively influences the (self-regulated) learning process of those students. With individual feedback however, an error in the given feedback would often only negatively affect one student.

A second element to note is that despite the errors in assignments, there is still disagreement per assignment about whether the student considers the feedback to be correct. This could indicate a difference in level in understanding of the material, interpretation of the material, or the difference in technical ability when using the tool for automated feedback.

Table 6.2: Likert question 1 results

Assignment	Negative	Neutral	Positive
7.2.3	3	0	3
7.7	4	1	2
7.8	0	0	2
7.9	0	1	2
Total	7	2	9

Question 2: ***I know what is meant by the feedback.***

Table 6.3 shows that the students are generally neutral to positive about their knowledge about the feedback. Consequently, this means that the feedback is understood by the students. This can only be true when the given feedback is in line with the assignment. Subsequently, this indicates that, in the eyes of the students, the feedback is related to the assignment and therefore also fits the study materials.

The negative opinion about the feedback seen in Table 6.3 on assignment 7.7, is in line with the results given on the correctness of the feedback for assignment 7.7 in Table 6.2. As described above, the negative opinion was due to a problem in the Repl.it platform and the manner in which these technical problems are outputted to the student. The feedback gave an extensive error message about the underlying program, which the students were unable to understand due to the complexity.

Considering the overall opinion about the knowledge and understanding of the feedback was positive, we can report that the students deem themselves fit to understand the feedback. According to Nicol and Macfarlane, this means that the feedback corresponds with the previously taught material. Consequently, this means that the given feedback contained high quality information on what the intended answer should look like, with the given assignment and underlying material in mind. Therefore, it can be concluded that the first quality of feedback: *'Does the automated feedback clarify what good performance is?'* was positively met according to the opinion of the students.

Table 6.3: Likert question 2 results

Assignment	Negative	Neutral	Positive
7.2.3	0	2	4
7.7	2	4	1
7.8	0	0	2
7.9	0	0	3
Total	2	6	10

Question 3: ***I know how to proceed, because of the feedback.***

The results, as shown in Table 6.4, were largely in line with the results of question 2. There was a generally neutral to positive consensus. Most of the time, the students deemed themselves able to proceed because of the information given by the feedback. Consequently this means that, in most cases, the students were able to make and execute a plan to enhance their answers. Subsequently, this means that the given feedback is sufficient for students to fill the gap between their answer and the desired answers.

Table 6.4: Likert question 3 results

Assignment	Negative	Neutral	Positive
7.2.3	1	1	4
7.7	2	2	2
7.8	0	1	1
7.9	0	0	4
Total	3	4	11

Overall, the results in Table 6.4 indicate that the students presume they know how to proceed to enhance their answer. Subsequently, this could indicate that the given feedback provided the students with high quality information and thus contributing positively to the feedback quality: *Does the feedback deliver high quality information?*.

6.3 Open questions

Due to a limited amount of answers, the decision was made to combine the results of the written and spoken surveys. This could be done because both the written as the spoken surveys followed the same method and questions.

In these open questions surveys, a total of three written and two spoken interviews were conducted. After taking the interviews and performing the analysis in Atlas, a total of 14 different codes were found. These codes were grouped on the qualities of feedback as described in Table 5.1 and the most notable will be displayed below in **bold** text. Subsequently, an insight into

Quality 1

In seven cases, the students showed a general knowledge about **the overall intention** of the assignment. Being that the course was of a beginners level, this often enticed a construction along the lines of making a for-loop and doing some basic operation inside. On one of the assignments, a student explained the assignment with *"that you can make a bar chart with asterisks if you have a list with several integer values"*, giving an example of a loop.

In addition to having overall knowledge about the assignment, students also showed they were able to **analyze and circumvent pitfalls** in the assignment. For example, the correct use of a function, using a for loop or using assignment-specific commands. This happened a total of nine times. The most common occurrence being that the student had to

"call the function in the right place".

In general, the students found the **feedback to be a good fit with the material associated** with the assignments. The feedback often guided the students into doing minor adjustments in their answers. For example, a student reported that

"their code worked as well ... but the feedback provided a slightly more general answer".

Meaning that in the end their answer was more in line with the answer that the teacher expected of them because of using the feedback.

In contrast with the positive points above, **the feedback was assessed as insufficient** by the students in four cases. According to the students, the feedback provided them with useful information, but due to insufficient knowledge on the material, they *"didn't know where to put it"* or *"how to do it"*. This means that the given feedback was unable to provide the students with information to enhance their answer. Thus not being helpful towards clarifying what good performance is.

Quality 2

Students often mentioned **the feedback provided more information** about their answer. During the interviews and the assignments, they indicated this in nine cases after using the feedback mechanism. After each time the students indicated this, it was often seen that they continued to improve their answer.

Among these improvements, it was common for students to **directly mention the next step to improve their answer**. This happened in twelve cases. The students mentioned, among other things;

"still having to pass a parameter", *"still have to do the printing of numbers"* and *"we still have to call the function"*.

In addition to mentioning these points for improvement, students indicated twice that they there was **an error in their answer**. The next step to improving their answer was not mentioned, but the recognition of the mistake was present. After some thought and consultation between the students, the improvement was made. Students also indicated that they understood the feedback. A student said that

"it is nice that it (the feedback) tells you what to use".

In contrast with the positive points, a more negative image was observed as well. Namely, in five cases, students indicated that they **did not actively use the feedback mechanism**. In three cases, the students reacted surprised to the advice for early review of feedback. In addition, a student indicated that he *"only submits when he sees that it is running"*. This consequently means that the feedback mechanism was not used to improve the answer at all, but was only used as validation for the students' answer afterwards.

Quality 3

Students indicated six times that **the feedback had helped with the assignment**. The main positive point they mentioned here was the repetition of the aforementioned material. For example, one student said that

"somewhere they knew the answer .. but the repetition was nice".

Students also indicated that they **found the feedback useful**. They were often positive about how they *"were not sure how to proceed"* but were helped by the feedback.

On the other hand, there were also some cases where the students felt **unsure as what to do with the feedback**. This made them doubt their answer or even themselves. The most notable confusion was observed whenever a bug in the system occurred. It resulted in irrelevant data being

shown to the student and consequently confusing the students. This led to a negative impact on the learning process of the students. In addition to this, one student also indicated that he

"benefited more from the explanation than from the feedback".

Which indicates that the information in the feedback was deemed less valuable than the information in the explanation of the assignment.

Chapter 7

Discussion

Although the results which emerged from this research lead to some insightful comments, there are limitations to this study which need to be mentioned.

7.1 Structured methods

Due to technical problems with the tool, there was a lack of time in preparing the research. The weeks in which this research could be carried out were fixed because of the structure of the school year of the students. As a result, the first weeks of the study were prepared sub-optimally. Some details about this will be given below.

Three-point Likert

In this research, the choice was made to use a three-points Likert scale. This was done to prevent overwhelming the participants in the first week and keep the focus on learning the process of making the assignment and effectively using the tool for automated feedback. However, it can be argued that a four-point scale would have provided more meaningful results. The aim of this study was to determine whether or not qualities of feedback were present. According to previous research, the clear distinction between these categories would provide enough reason to eliminate the neutral position. (Garland, 1991)

Open questions

On their own initiative, students gave considerably less extensive answers to the open-ended questions than initially expected. The answers were often limited to a few words or even a yes/no. As a result, a lot of relevant information was lost. This way of requesting information from the students

is therefore not ideal and in hindsight should have been done differently in this study.

Participants group

The final limitation of this research lies in its' group size of the test group. The group had a maximum of twelve students. These students all attended the same school and class. This allows us to conclude that the test group consisted of a homogeneous group when regarding age and school level. These two factors combined lead to a low probability of reaching the (theoretical) saturation of this particular topic. Consequently, additional interviews could still lead to new information and thus not be sufficient to reach a sound and valid conclusion with the current results.

Considering all of the limitations above, one could agree to a more structured method to reach more significant results. Our advice for any further research on this topic would be to have a distinct week for the students (and researchers) to get used to the feedback tools at hand. Subsequently, we would advice to strictly use pair programming and stick to in-depth (audio-recorded) interviews instead of using several data collection methods.

Another point of interest might be to conduct the research at multiple classes or even schools to have a higher chance to reach saturation.

7.2 Further research

Additional feedback qualities

To maintain a reasonable scope within this thesis, the focus was put on three of the seven qualities of feedback as developed in the model by Nicol and Macfarlane. These qualities were focused around the student and his/her opinions and ideas about the feedback with regard to the model of self-regulated learning. The ideal situation, of self-regulated learning using automated feedback giving, is one in which all seven qualities as described by Nicol and Macfarlane are present. To ensure such an ideal situation, research will also have to be done into the opinion and ideas of the teacher regarding their material and assignments.

Technical issues

At the time of conducting this research, the programming tool Repl.it was still in development. This resulted in frequent technical problems. For example, parts of regular expressions were misinterpreted by the underlying program. As a result, the input of students was not properly recognized and thus lead to mistakes. Whenever this occurred, a full log about the error was outputted to the students. As a result, we saw that the students closed

down and simply decided to not finish the assignment. This response makes sense according to the research of Nicol and Macfarlane, as the feedback did not describe how the student could improve his work in small steps, and thus does not comply with the ideas of good external feedback (Nicol & Macfarlane-Dick, 2006). However, this is not a desirable behavior, so in the future it might be better to use a different tool all together.

Chapter 8

Conclusions

In this thesis, we tried to answer the question: *"To what extent is automated feedback suitable to support self-regulated learning in an introductory programming course?"*

This study found that the feedback was often thought to be in line with the teachers material and thus *"clarified what good performance was"*. Students gave a neutral to positive opinion about understanding the feedback. We found that students were able to recognize and analyze difficulties in the assignment and understood why the feedback was shown. However, students sometimes indicated that they did not know what to do with the feedback or how to use it.

In addition, a neutral to positive opinion was found about the feedback *"delivering high quality information"*. Students often indicated they could make improvements because of the feedback. Some students indicated that they only use the feedback mechanism as validation, meaning the feedback did not appear in a timely manner.

Furthermore, we found that *"the feedback often helped to encourage the students"*. On the other hand, it was also found on a number of occasions that students became uncertain about their answers. This had a negative influence on the completion of the assignments and thus the learning process of the students.

Finally, *the correctness of the feedback* was examined. 80 percent of student submissions were assessed correctly by the automated tests. The errors in this were mainly due to bugs in the software or errors in the regular expressions. The margin of error is expected to grow when making more complex assignments.

To sum up, the students expressed sufficient knowledge about the assignments and often knew how to efficiently use the automated feedback. However, some negative points were also found. Students sometimes did not know how to use the feedback or did not use it at all. This suggests that the qualities of the feedback were at least high enough to support the use of automated feedback for self-regulated learning.

References

Black, P., & Wiliam, D. (2009, 2). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. doi: 10.1007/s11092-008-9068-5

Cronbach, L. (1963). *Course improvement through evaluation. en pa taylor & dm cowley (eds.) reading in curriculum evaluation (pp. 11-19). dubuque, io: Wm. c.* Brown Company Publishers.

Dick, W. (1987). Formative evaluation: Prospects for the future. *Educational Technology*, *27*(10), 55–57.

Garland, R. (1991). The mid-point on a rating scale: Is it desirable. *Marketing bulletin*, *2*(1), 66–70.

Hounsell, D. (1997). Understanding teaching and teaching for understanding. *F. Marton, D. Hounsell, N. Entwistle (Eds.), The Experience of Learning: Implications for Teaching and Studying in Higher Education (pp.*, 238-257.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*.

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, *27*(4), 309-323. Retrieved from `https://doi.org/10.1080/0260293022000001337` doi: 10.1080/0260293022000001337

Pintrich, P. R., & Zusho, A. (2002). Student motivation and self-regulated learning in the college classroom. *J. C. Smart and W.G. Tierney (Eds) Higher Education: Handbook ofTheory and Research, Volume XVII (New York, Agathon Press).*

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, 119-144.

Zimmerman, B. (2002, 06). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*, 64-70. doi: 10.1207/s15430421tip4102_2

Appendix A

Appendix

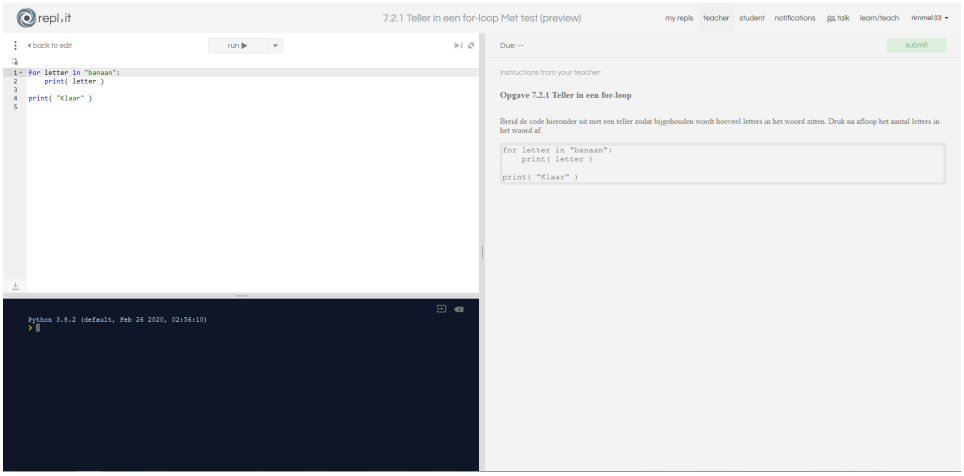

Figure A.1: The students' view in Repl.it

Opgave 7.2.3.2

			
Ik weet wat er bedoeld wordt met de feedback.			
Ik weet door de feedback hoe ik nu verder moet.			
Ik vond de feedback nuttig bij het maken van de opdracht.			

Opgave 7.7

			
Ik weet wat er bedoeld wordt met de feedback.			
Ik weet door de feedback hoe ik nu verder moet.			
Ik vond de feedback nuttig bij het maken van de opdracht.			

Opgave 7.

			
Ik weet wat er bedoeld wordt met de feedback.			
Ik weet door de feedback hoe ik nu verder moet.			
Ik vond de feedback nuttig bij het maken van de opdracht.			

Figure A.2: The likert-scale questionnaire in Dutch

Table A.1: Qualities of feedback and corresponding questions in Dutch

Quality 1	*Does the automated feedback clarify what good performance is?*
	Weet je wat er van je verwacht wordt?
	Wat denk je dat mogelijke valkuilen zijn bij deze opdracht?
	Vind je dat de feedback goed aansluit op de stof?
Quality 2	*Does the feedback deliver high quality information?*
	Weet je wat er bedoeld wordt met de gegeven feedback?
	Weet je hoe jouw code verschilt met het goede antwoord?
	Hoe ga je ervoor zorgen dat je antwoord wel zo wordt als de bedoeling is?
Quality 3	*Does the feedback encourage positive motivation and self-esteem?*
	Vond je de gegeven feedback nuttig?
	Heeft de gegeven feedback je geholpen bij de opdracht? Licht je antwoord toe.