



**The effect of feedback complexity on university students' achievement and intrinsic motivation**

Lotte van den Heuvel (5660823)

Master's Thesis Educational Sciences

First assessor: Dr. Barbara Flunger

Second assessor: Dr. Frans Prins

Date: 07-06-2020

Word count: 7995

**Abstract**

Feedback is an important mechanism in education to improve the achievement and intrinsic motivation of students. Unfortunately, the feedback provided in university education is often perceived by students as inadequate. To uncover whether certain levels of feedback complexity can aid achievement and motivation, and thus be used to make feedback more effective, this study examined the effects of a simple versus complex feedback type. Additionally, the influence of background knowledge and one's academic self-concept were investigated. An online experiment was carried out with 42 university students from the Netherlands. Participants read a text about the academic self-concept and received feedback varying in complexity on practice questions about the text. A week later, they completed a follow-up knowledge test. The results did not show a statistically significant impact of feedback complexity nor a mediating influence of background knowledge on either achievement, motivation or academic self-concept. The effect of reading a text about the academic self-concept could also not be confirmed. In conclusion, it is recommended to replicate the study with better validated measures and a larger sample.

### **Introduction**

Feedback is considered to be a crucial means in facilitating students' development both within and beyond their studies (Ferguson, 2011; Evans, 2013). As Hattie and Timperley (2007) put it: "feedback is one of the most powerful influences on student learning" (p.81). Feedback in an achievement context can be understood as the information provided by an external agent that compares the present achievement of a student with a desired achievement in order to improve the student's cognition, behaviour, and/or motivation (Shute, 2008). In the last forty years, university education has changed to encompass more frequent assessment of and feedback to students (Kreber, Anderson, Entwistle & McArthur, 2014). At the same time, university students often perceive the feedback they receive as inadequate by stating that feedback fails to relate to students' overall performance in a curriculum (Krause et al., 2009; HEFCE, 2011). Also the timing, frequency and consistency of feedback are perceived as suboptimal by university students across the world (Hounsell, 2007). As Archer (2010) noted, students need to regard received feedback as effective for it to have a positive effect on learning.

As a result, improving feedback forms a major educational challenge. Because teachers spend a lot of time providing students with feedback (Stern & Solomon, 2006), and as well-designed feedback can aid the achievement and motivation of these students (Deci, Koestner, & Ryan, 1999; Shute, 2008), it is crucial to find out which specific aspects of feedback make it effective. Therefore, this research will examine the influence of the feedback aspect 'complexity' on academic achievement and intrinsic motivation. Intrinsic motivation is the motivation for an activity in its own sake, meaning that one undertakes the activity out of personal interest and enjoyment (Deci & Ryan, 2010). The complexity of feedback, in turn, is determined by the amount of elements included in the feedback, and how difficult it is to understand (Dempsey, Driscoll & Swindell, 1993). Because the findings

regarding the influence of feedback complexity are inconclusive, more research on this aspect of feedback is needed.

The present study examines the effect of feedback types of varying levels of complexity on academic achievement and intrinsic motivation. Based on extensive literature research, Evans (2013) concluded that feedback in higher education is usually provided in written form (either on paper or digitally) on some type of assessment by an external agent. For that reason, the feedback in this study will be administered 1) based on a quiz about a text, 2) in a written digital format, and 3) by a researcher to ensure ecological validity.

In addition, it is interesting to investigate to what extent the changes in achievement and motivation can indeed be attributed to feedback complexity, and how much is due to simply reading about the topic of the achievement test. For this reason, the text that participants read in this study deals with the academic self-concept (ASC), which is defined as the perception of one's own academic abilities (Bong & Skaalvik, 2003). By measuring students' ASC before and after they read the text, it is possible to uncover the influence of the text itself on achievement and motivation, regardless of the received feedback.

### **Feedback in Achievement Settings**

Feedback can serve to improve students' achievement and motivation (Shute, 2008). When discussing feedback, it is important to distinguish between corrective and explanatory forms of feedback. Corrective feedback is described by Mouratidis, Lens, and Vansteenkiste (2010) as "pointing out faults and weaknesses to improve the learner's skills" (p. 621). Explanatory feedback, in turn, is aimed at changing students' thinking or behaviour patterns in order to facilitate learning (Shute, 2008). This means that feedback including both corrective and explanatory elements is higher in complexity compared to feedback including either one of these elements.

Shute (2008) referred to the simplest form of feedback as 'verification', as it verifies the correctness of students' answers. Thus, verification feedback only entails a corrective element. Complexity increases when feedback becomes lengthier and when it includes more components (Kulhavy, White, Topp, Chan & Adams, 1985). Examples of such components can be: an indication of the correctness of the given answer, a referral to the passage in the learning material where the correct answer can be found, or an explanation of why alternative answers are incorrect. A more complex type of feedback was coined 'elaborated feedback', where an explanation is provided as to why an answer is (in)correct. The elaborated feedback type (including corrective and explanatory elements) has several subtypes, of which the 'response contingent'-type focuses on explaining why incorrect answers are wrong and correct answers are right. Because verification and elaborated feedback differ in their level of complexity and have both been shown to be effective (Kulhavy et al., 1985; Mandernach, 2005). For that reasons, these two feedback types will be used in this study.

### **The Role of Feedback Complexity for Student Outcomes**

Feedback and its distinct aspects have been shown to be an essential predictor of student outcomes, including achievement (Jonsson, 2013) and motivation (Mouratidis, Lens, and Vansteenkiste, 2010). Yet, the effect of some feedback aspects, including complexity, remains unclear (Shute, 2008; Pereira, Flores, Simão & Barros, 2016). This section will highlight some of the conflicting findings on feedback complexity, which form the basis for the hypotheses of this research.

Concerning the influence on achievement, Schimmel (1983) found in his meta-analysis that feedback complexity did not have a significant effect. In their more recent meta-review, Van de Ridder, McGaghie, Stokking, and Ten Cate (2015) did not report significant effects of feedback complexity on academic achievement either. It must be noted that in the latter study, feedback complexity was one of many feedback aspects considered, and the

conclusion on feedback complexity was based on merely one source. Still, another study by Bivens (1964) also found no significant differences in achievement following the provision of simple versus complex feedback, although this may have been due to low motivation of the students involved. Moreover, this study was conducted with 8<sup>th</sup>-grade students of around 13 years of age, instead of university students. A study by Phye & Bender (1989) presented evidence for the notion that there is an inverse relationship between feedback complexity and consequent achievement. The same was found by Kulhavy and colleagues (1985). As an explanation, these researchers argued that an elaborate explanation of an incorrect answer may be the only thing a student remembers, thus leading to an incorrect response. Because more contemporary experimental research is needed, this study will investigate whether the findings by Phye & Bender (1989) and Kulhavy et al. (1985) can be replicated. Based on these earlier results, it is hypothesised that less complex feedback (i.e. verification feedback) will lead to most improvement in achievement. However, as the main goal of feedback is to improve student performance, elaborated feedback is expected to be more beneficial to achievement than receiving no feedback, even though it is more complex.

Regarding intrinsic motivation, various levels of feedback complexity have been found to have differing effects, although the findings are inconclusive. To illustrate this, lower complexity was found by one research team to be more intrinsically motivating, as it reduced the time needed to understand and process the feedback (Goltz, Citera, Jensen, Favero & Komaki, 1989), whereas another researcher argued that increasingly complex feedback leads to higher intrinsic motivation, albeit moderated by self-efficacy, which refers to whether a student believes that he or she is capable of successfully completing a certain task (Narciss, 1999). Regarding this topic, there again is a lack of experimental studies that report actual effect sizes of the extent to which intrinsic motivation is influenced by feedback complexity. Based on these contradicting findings, it is possible that there are distinct processes at play for

verification feedback and elaborate feedback. For verification feedback, the short length and easiness of understanding may be motivating, whereas the belief whether one can successfully understand and use the feedback influences intrinsic motivation in the case of elaborated feedback. This means that regardless of how a student views their own capabilities, verification feedback will always be motivating due to its low complexity. Therefore, it is predicted that verification feedback generally leads to a higher motivation than elaborated feedback. So, the hypothesis for intrinsic motivation is that it will be highest in the case of less complex feedback.

### **Feedback and Student Characteristics**

Student characteristics can also influence achievement and/or motivation (Doubé, Carding, Flanagan, Kaufman and Armitage, 2018). Of particular interest here are background knowledge and the academic self-concept (ASC). The latter is especially intriguing as the experiment text is about the ASC. This section will discuss some of the literature on the two student characteristics in relation to feedback complexity, academic achievement and intrinsic motivation, and the hypotheses derived from that.

**Background knowledge.** Background knowledge has been found to mediate the relationship between feedback complexity and achievement. There appears to be a difference in usefulness of elaborated, response-contingent feedback for students with a high versus low background knowledge base, meaning that this type of feedback may be especially suitable for students with low levels of background knowledge (Doubé et al., 2018). Thus, evaluating the effectiveness of verification and elaborated feedback types in students with varying levels of background knowledge can make the relationship between feedback and achievement clearer. Based on the study by Doubé et al. (2018), it is hypothesised that the effect of feedback complexity is positively mediated by a student's level of background knowledge. The expectation is that verification feedback is even more useful for students with more

background knowledge, as opposed to elaboration feedback, which suits students with less background knowledge better.

Next, background knowledge is also of influence when considering intrinsic motivation. Several researches have shown that motivation to improve one's performance is a prerequisite for the effectiveness of both corrective and explanatory feedback (Guénette, 2007; Evans, Hartshorn, McCollum & Wolfersberger, 2010). So, when students do not intend to use the learned material in the future to improve their performance on the topic, the effect of feedback will be less strong. Hence, background knowledge on the topic is expected to be correlated with motivation, as was also found in an experimental study by Alexander, Jetton and Kulikowich (1995). These authors looked at the relationships between the background knowledge, interest/motivation and performance of 78 university students. The investigated variables were all strongly related, so the authors suggested that further research should look into how and under which circumstances the correlations between background knowledge, motivation and performance occur. Therefore, this paper will examine if a relationship between feedback complexity, background knowledge, achievement and motivation exists. Because Doubé and colleagues (2018) found that verification feedback is more suitable for students with a high level background knowledge, less complex feedback is predicted to lead to the highest outcomes in intrinsic motivation when provided to students with high background knowledge. This means background knowledge will also be investigated as a mediator of the relationship between feedback complexity and motivation.

**Academic self-concept.** The experiment used in this study deals with the academic self-concept. Huang (2011) described the self-concept as the perception of oneself, and it having different sub-domains. One of these domains is the academic self-concept, which is further subdivided into math and verbal self-concept, and does not correlate with the non-academic self-concept (Möller, Pohlmann, Köller & Marsh, 2009). In their meta-analysis of



69 datasets, Möller and colleagues (2009) found that the self-concept for math explained 61% of the variance in math achievement; the verbal self-concept explained 49% of variance in verbal achievements. Hence, the academic self-concept (ASC) is an important factor to take into account regarding academic achievement. The present study regards the global ASC (i.e. how students' view themselves in relation to their overall academics skills).

In light of earlier findings, the ASC is a relevant phenomenon to study, especially as it is the topic of the experiment text. By examining whether changes in students' own ASC occur after reading a text about the ASC, it can be determined whether the complexity of feedback indeed has an effect, so changes in achievement, motivation and ASC cannot simply be attributed to studying a text about the ASC. As it was found that knowledge about motivation can help improve achievement and intrinsic motivation (Boekaerts, 1996; Ginsberg, 2005), it is interesting to investigate whether knowledge about the ASC elicits changes in a student's own ASC, as well. This leads to the hypothesis that reading a text about the ASC will cause changes in one's own ASC across all conditions. On top of that, as it is expected that verification feedback leads to the highest scores in achievement and motivation, it will likely also cause the largest differences in ASC.

### **Present study**

Based on the abovementioned gaps in the literature surrounding the importance of feedback complexity (Shute, 2008), the main research question of this study is as follows:

*What is the effect of feedback complexity on achievement, intrinsic motivation and academic self-concept?* Multiple hypotheses are examined that help to answer the research question.

Hypothesis 1 states the expectation that: *University students' achievement and intrinsic motivation are affected by feedback complexity.* More specifically, hypothesis 1a is voiced as: *Less complex feedback aids achievement most*, which would be in line with the findings of Kulhavy and colleagues (1985). This means that achievement is expected to be

highest for verification feedback, followed by elaborated feedback. So, the worst achievement outcomes are expected for the control condition; hypothesis 1b is: *Less complex feedback aids intrinsic motivation most*, as was found by Goltz et al. (1989). Again, intrinsic motivation is predicted to be highest in the case of verification feedback and lowest for the control group.

Hypothesis 2 describes the predicted effect of background knowledge: *Achievement and intrinsic motivation of university students are affected by feedback complexity, positively mediated by background knowledge*. Moreover, based on findings by Doubé et al. (2018), hypothesis 2a is: *Achievement is highest when background knowledge is high, and feedback is less complex*; hypothesis 2b is: *Intrinsic motivation is highest when background knowledge is high, and feedback is less complex*.

Another question concerns whether the content under study (namely, the ASC) has an effect independent of feedback complexity. Relying on earlier findings by Boekaerts (1996) and Ginsberg (2005), who both found a similar effect of knowledge about motivation on subsequent motivational scores, hypothesis 3 articulates that: *Reading a text about the academic self-concept will elicit changes in one's own academic self-concept*. To further specify the effect, hypothesis 3a states that: *Changes in academic self-concept will be highest when less complex feedback is provided*. Here, the effect is also expected to be strongest for verification feedback, and least strong when no feedback is provided (control condition). In addition, the ASC is included as a covariate in the analysis of the main effect between feedback complexity and achievement/motivation.

By testing the hypotheses and answering the research question, this study aims to contribute to the scientific knowledge base around feedback effectivity. Moreover, the goal of the study is to formulate implications for practice concerning the use of more or less complex feedback.

## Method

### Design

This quantitative research made use of a three-group design including two experimental groups ('verification feedback' and 'elaborated feedback'; see: Table 1), and a control group. The research was conducted in the form of an experiment in which the independent variable of feedback complexity was manipulated. The dependent variables were achievement, intrinsic motivation, and the academic self-concept.

The experiment was distributed to participants via two online surveys. This design was chosen for because a survey increased accessibility and enabled the collection of more quantitative data than would have been possible in a laboratory setting. The experiment consisted of a pre-test with a practice task (survey 1), and a post-test (survey 2). Random assignment was built into the first survey, so participants were randomly sent to the control condition or either of the feedback conditions.

*Table 1. Feedback complexity (Shute, 2008)*

<b>Group</b>	<b>Feedback type</b>	<b>Description</b>
<b>ordered by complexity</b>		
Control	No feedback	Refers to the condition where the learner is presented with a question and is required to respond, but there is no indication as to the correctness of the learner's response.
1	Verification	Also called "knowledge of results" or "knowledge of outcome." It informs the learners about the correctness of their responses (e.g., right-wrong, or overall percentage correct).

2	Elaborated feedback (response contingent)	Elaborated feedback that focuses on the learner's specific response. It may describe why an answer is correct and why incorrect answers are wrong.
---	---	--

---

## Participants

A power analysis was conducted to determine the number of participants needed to avoid a Type II error, in which case the research fails to reject a false null hypothesis due to a lack of statistical power (Neyman, 1950). The “pwr” package (Champely, 2018) in R (R Core Team, 2017) was used to uncover that  $n = 34$  per experimental condition is appropriate for a one-way ANOVA with 3 groups to achieve a statistical significance of  $p = .05$  (Cohen, 2013). Hence, a total  $n$  of at least 102 would be needed. However, merely 42 participants completed both the pre- and post-test. As a result, non-parametric tests were used (see section below: Data analysis). The sample consisted of 30 women and 12 men, with an average age of 24 years. The youngest participant was 19 years old and the oldest was an outlier in terms of age with being 53 years old.

The responses of two participants were not used in the analyses, as they indicated a very low effort on the knowledge post-test. As a criterium for inclusion, respondents had to be at least neutral in their average effort score (meaning 3.5 or higher). In addition, one of these respondents showed a very atypical response pattern by using only the extremes on all Likert scale questions.

## Instruments

### *Feedback Complexity*

The independent variable feedback complexity was manipulated by having participants receive feedback on four practice questions about a 900-word text about the academic self-concept (see: Appendix I). The text was published in a recent paper by Roelle and Nückles (2019) and relevant for university students.

The text was chosen for because its content about the academic self-concept is linked to educational sciences, but not included in the Educational Sciences programme at Utrecht University. Hence, it was expected to be an appealing area of research to the participants without being so familiar that the students would know the correct answers regardless of the provided text. Other topics considered (such as an introduction on educational psychology) were disregarded because several master's students of the Educational Sciences programme indicated these to be too easy and too familiar.

The text was somewhat altered for clarity and brevity. Some examples provided in the text referred to the German schooling system, so these were adjusted to the Dutch system. In addition, the text was divided into four parts. After each part of the text, respondents answered one practice question (e.g. "Describe the structure of the academic self-concept"). The practice questions were open questions, followed by no feedback (control group), the correct answer (verification feedback), or an explanation of the correct and a possible wrong answer (elaborated feedback). The practice questions, answers and feedback were devised by the author of this paper (see: Appendix II).

Furthermore, a manipulation check was used for the experimental conditions (see: Appendix III). This manipulation check served to uncover whether respondents, who were randomly assigned to receive feedback, understood and valued the feedback. It consisted of 9 questions and was answered on a 7-item Likert scale.

### ***Achievement***

During the follow-up survey, participants completed a knowledge test on the academic self-concept that was also used in the study by Roelle & Nückles (2019). The knowledge post-test consisted of eight open questions, of which four measured reproduction of the text that was read earlier, and four measured understanding (see: Appendix IV). The reliability of the overall knowledge test was shown to be satisfactory ( $\alpha = .74$ ), but the reliability of the

subset of understanding question was very low ( $\alpha = .46$ ). Scoring of the answers was done by the author of this study, in accordance with the provided scoring scheme from Roelle's & Nückles' (2019) paper.

### ***Intrinsic Motivation***

The intrinsic motivation of participants was measured before reading the text about the academic self-concept, and after the completion of the knowledge test in the follow-up survey. To measure intrinsic motivation, the "Interest/enjoyment" questions from the Intrinsic Motivation Inventory (IMI, Deci & Ryan, 2003) were used (see: Appendix V). This subtest from the IMI consists of 7 questions, which asked participants to self-report their intrinsic motivation on a 7-item Likert scale ranging from "1 = not at all true" to "7 = very true". Examples of included items are: "Learning about the academic self-concept does not hold my attention at all" and "I would describe the academic self-concept as very interesting".

The IMI is considered to be a highly valid instrument based on confirmatory factor analyses (Monteiro, Mata, & Peixoto, 2015). Cronbach's alpha was calculated to evaluate the reliability of the IMI subtest used in this study. Based on the standardised alpha score, which was used because the subtest is unifactorial, the IMI subtest has excellent reliability ( $\alpha = .90$ ; Gliem & Gliem, 2003).

### ***Background Knowledge***

Background knowledge on the subject of the academic self-concept was measured with a pre-test question asking participants to write down everything they knew about the concept and the processes involved in it. To score the answers on this question, a scoring scheme was devised by the author of this study. As the pre-test for background knowledge only consisted of one question, its reliability could not be calculated. Roelle and Nückles (2019) also did not report anything on the reliability and/or validity of the knowledge pre-test. Other questions relating to background knowledge included one's field of study, having

followed courses that discussed the academic self-concept, and having used theories or articles on the academic self-concept for assignments.

### *Academic self-concept*

To measure participants' academic self-concept and whether it is influenced by learning about the topic, self-reports based on the Academic Self-Concept Questionnaire (ASCQ) were used. From this instrument, devised by Marsh (1990), the sub-test about mathematics was taken and adjusted to studying in general (see Appendix VI). Items included "I hate studying" and "Studying is one of my best skills". It can be assumed that the reliability of the ASCQ was good ( $\alpha = .83$ ), and multiple studies have confirmed the validity of the instrument (Dusek & Flaherty, 1981; Marsh, Byrne & Shavelson, 1988).

### *General Effort*

Effort was measured as a covariate, to control for respondents just clicking through the survey without attempting to answer the questions as well as they could. Four questions were posed to measure general effort. During the pre-test, the questions concerned effort for studying ("I am prepared to put effort into my academics", "I work hard for my academics", "I am serious when working on study assignments" and "I finish all study assignments as best as I can"); in the post-test the questions targeted effort put into the knowledge test about the academic self-concept (Flunger et al., 2015). Based on Cronbach's standardised alpha, the reliability of the effort measure seemed good ( $\alpha = .83$ ). Concerning its validity, the instrument was originally used as an instrument to measure homework compliance (Flunger et al., 2015), so it may be somewhat less valid when used to measure general effort.

### **Procedure**

Students enrolled in Educational Sciences at Utrecht University were approached through stratified purposeful sampling. This sampling technique is used to ensure that participants are included, who vary on certain parameters (Sandelowski, 2000). For this study,

Educational Sciences students were expected to have, on average, a higher level of background knowledge about the ASC. Hence, purposefully approaching this group to participate in the research would ensure a variation in background knowledge compared to people studying different academic subjects. Therefore, the link to the survey was published in the online learning environment of bachelor's as well as master's students majoring in Educational Sciences. The survey link was also published on various social media (such as LinkedIn, Facebook and Instagram) to reach people who studied other subjects.

Participants, who clicked on the link to the pre-test survey, first received information about the purpose of the experiment and an indication of the time it would take to complete it. Participants were asked to provide active informed consent by ticking "agree" and continuing the survey (see: Appendix VIII). No risks were associated with participation. As participation in the study was completely voluntary, there were also no negative consequences of non-participation.

Next, respondents disclosed what they already knew about the academic self-concept as a test of their pre-knowledge and filled out the items on intrinsic motivation, as well as items on their own academic self-concept and effort. After these general questions, participants were randomly assigned to either the control group or an experimental condition. This randomisation was built into the survey and took place automatically. The four parts of the text about the academic self-concept and the accompanying practice questions were shown one by one. In between the texts, feedback was provided to those participants who were assigned to the verification feedback and elaborated feedback groups.

The last part of the pre-test survey consisted of questions asking participants to provide some personal information, such as age, gender, field of study and having used the academic self-concept in courses and assignments or not. Participants also inserted their email address, on which they were contacted a week later to fill out the post-test.



The post-test started with the eight questions comprising the knowledge test on the academic self-concept, followed by a duplication of the general questions from the pre-test (on intrinsic motivation, one's academic self-concept, and effort). The survey ended with a debriefing (see: Appendix VIII) and the option to leave one's email address in order to receive the outcomes of the study. These email addresses were stored separately from the answers on both surveys.

In order to link the anonymous answers on the pre- and post-test, every participant received an entry code for the post-test survey by email (for which they noted down their email address at the end of the pre-test; see: Appendix IX). This code was coupled to their answers on the pre-test, allowing for the answers on both surveys to be matched. When participants did not finish the post-test within one week after receiving the link to it, they were sent a reminder email. As soon as a respondent had finished the post-test survey, their email address was deleted from the database where the survey answers were stored.

## **Data Analysis**

### ***Data Preparation***

In preparing the data for analysis, several items measuring intrinsic motivation and academic self-concept had to be reverse coded, as those were negatively worded. For these measures, as well as for the concept effort, averages were then calculated. In that way, the scores kept the same definitions as the original 7-point Likert scale. For example, every participant's answers on the items making up the intrinsic motivation pre-test were averaged, resulting in a score between 1 and 7. These averages were not rounded off to whole numbers for the analyses.

### ***Statistical Analysis***

To analyse the collected data, statistical analyses were performed using R (R Core Team, 2017). To examine the differences between the three conditions (control group,

verification feedback and elaborated feedback) in terms of achievement and motivation, Kruskal-Wallis tests were used. Non-parametric tests had to be used, as the assumption for data independence was not met, although this is likely due to the small sample size.

Achievement, intrinsic motivation and the ASC made up the dependent variables with feedback complexity as the independent variable. Covariates added to the between-groups tests included the pre-test measures for motivation, academic self-concept, effort, and background knowledge. As it is not yet possible to conduct a non-parametric ANCOVA, linear regression was used instead.

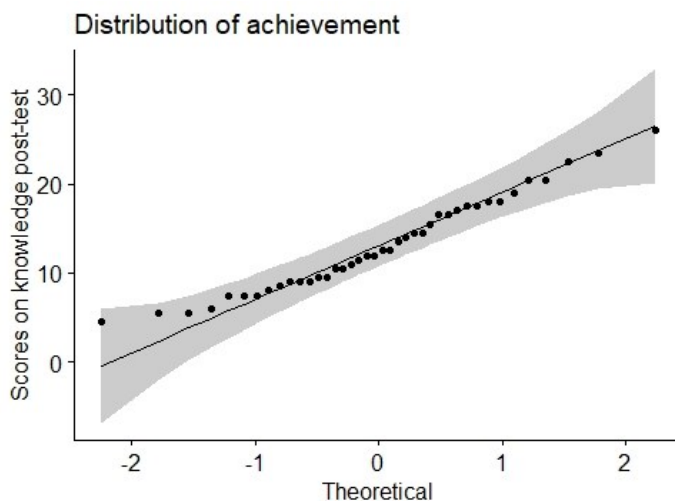
For hypothesis 2, a mediation analysis was also run using linear regression.

Hypothesis 3 was tested for by employing a Wilcoxon signed rank test as the non-parametric alternative for a paired t-test, investigating the difference between pre- and post-test ASC.

## Results

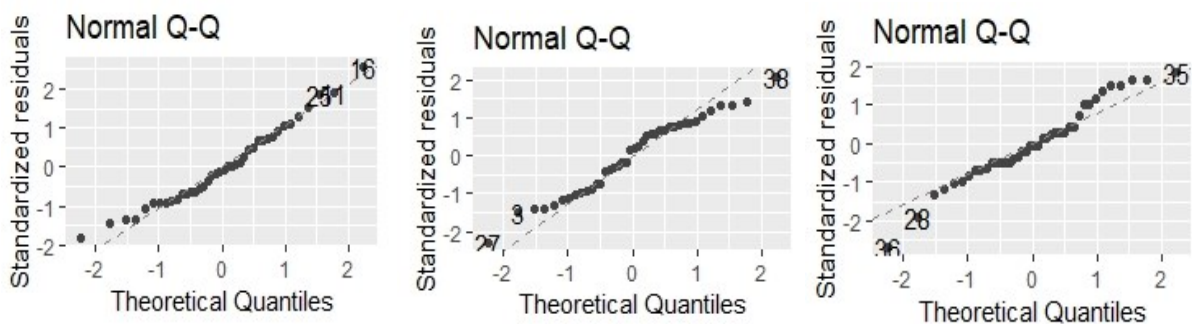
Firstly, the assumptions of normality, linearity, homogeneity of variances and independence of the data were tested for. The Shapiro-Wilk normality test showed that the dependent variables were normally distributed (for achievement,  $W = .97, p = .27$ ; for intrinsic motivation,  $W = .97, p = .37$ ; for academic self-concept,  $W = .98, p = .57$ ). Figure 1 shows the normal distribution of achievement graphically, using a QQ-plot.

*Figure 1 – QQ-plot showing the normal distribution of independent variable achievement*



Based on Bartlett's test, homogeneity of variances can also be assumed for achievement ( $K^2 = 1.83, df = 2, p = .40$ ), intrinsic motivation ( $K^2 = 2.04, df = 2, p = .36$ ) and the ASC ( $K^2 = 5.51, df = 2, p = .06$ ). Similarly, all dependent variables complied with the assumption of linearity based on the normal distribution of their residuals. This can be seen in the corresponding Normal QQ-plots (Figure 2).

Figure 2 – Normal QQ-distributions showing the linearity of the achievement, motivation and ASC variables.



Pearson's Chi-squared test for data independence proved to be significant ( $\chi^2 = 1270.30, df = 585, p < .01$ ). Hence, the assumption of data independence was violated. This means that some of the variables depend on one another and cannot be seen as separate measures. As parametric analyses assume data independence, non-parametric tests had to be used to analyse the data in this study.

Lastly, two tests were run to detect multivariate outliers. Several outliers were identified based on both Crooks distance ( $D_i > 1$ ) and the Mahalanobis distance (see: Figure 3). However, deletion of these outliers would mean that the control group would have an  $n$  smaller than 7, whereas  $n = 7$  is the absolute minimum group size for running a regression analysis with a good prediction level (Knofczynski & Mundfrom, 2008). So, to enable all analyses, the outliers remained included in the dataset. On top of that, it is possible that these datapoints would not have been outliers in a larger dataset.

The descriptive statistics of the dependent variables are presented in Table 2 to 4.

When looking at the means for each group in the experiment, one can already see that the differences in scores for all dependent variables are small. On top of that, as can be seen from Table 5, feedback complexity did not significantly correlate with any variable.

Figure 3 – Mahalanobis distance to detect multivariate outliers

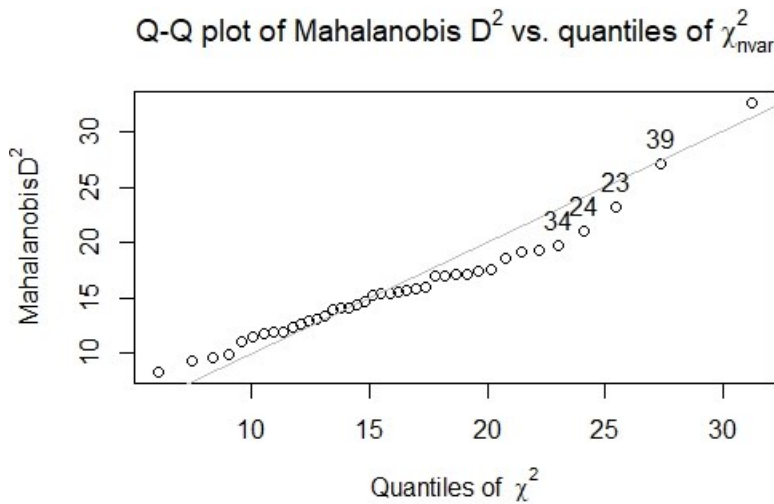


Table 2 – Descriptive statistics for post-test knowledge

	<i>N</i>	Mean	Std. deviation	Minimum	Maximum
Control group	10	14.85	5.58	4.5	23.5
'Verification feedback'	15	11.93	4.28	5.5	19.0
'Elaborated feedback'	15	13.10	6.19	5.5	26.0
Overall	40	13.10	5.37	4.5	26.0

Table 3 – Descriptive statistics for post-test motivation

	<i>N</i>	Mean	Std. deviation	Minimum	Maximum
Control group	10	4.61	.91	3.29	5.86
'Verification feedback'	15	4.32	1.39	1.86	6.86
'Elaborated feedback'	15	4.48	1.07	2.71	6.00
Overall	40	4.47	1.10	1.86	6.86

Table 4 – Descriptive statistics for post-test ASC

	<i>N</i>	Mean	Std. deviation	Minimum	Maximum
Control group	10	5.52	.69	4.7	6.6
'Verification feedback'	15	5.67	.82	3.9	6.6
'Elaborated feedback'	15	5.70	.42	6.6	6.9
Overall	40	5.64	.65	3.9	6.9

Running head: THE EFFECT OF FEEDBACK COMPLEXITY ON UNIVERSITY STUDENTS' ACHIEVEMENT AND INTRINSIC MOTIVATION

Table 5 – Correlation matrix

	Pre-test knowledge	Pre-test motivation	Pre-test ASC	Pre-test effort	Age	Level	Study	Use of ASC	Post-test knowledge	Post-test motivation	Post-test ASC	Post-test effort
Pre-test knowledge	-											
Pre-test motivation		-	.46**							.43**	.48**	
Pre-test ASC		.46**	-								.87***	
Pre-test effort				-								
Age					-	.39*	.36*					
Level					.39*	-						
Study					.36*		-	.46**				
Use of ASC							.46**	-				
Post-test knowledge									-			
Post-test motivation		.43**								-		
Post-test ASC		.48**	.87***								-	.44**
Post-test effort											.44**	-

\* significant p-value at a 0.05  $\alpha$ -level

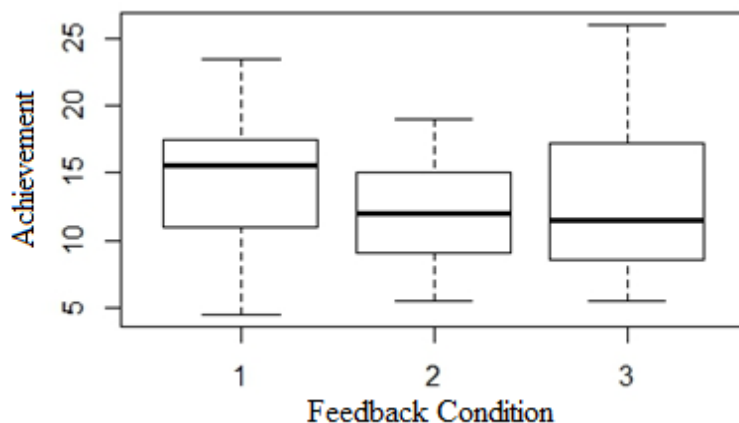
\*\* significant p-value at a 0.01  $\alpha$ -level

\*\*\* significant p-value at a 0.001  $\alpha$ -level

Running head: THE EFFECT OF FEEDBACK COMPLEXITY ON UNIVERSITY STUDENTS' ACHIEVEMENT AND INTRINSIC MOTIVATION

To test the first hypothesis about the influence of feedback complexity on achievement and intrinsic motivation, non-parametric Kruskal-Wallis tests were used. The same test was used for the ASC, in line with hypothesis 3. The Kruskal-Wallis tests examined whether there was a significant difference in scores on the dependent variables due to varying feedback complexity. For achievement, the test did not yield significant results ( $\chi^2 = 1.67, df = 2, p = .43$ ; for a graphical display, see Figure 4). Also concerning motivation ( $\chi^2 = .19, df = 2, p = .91$ ), no significant differences were found. Pairwise comparisons using the Wilcoxon rank sum test showed that there was no significant between-group difference for any of the groups in achievement ( $p = .55$  for control group vs. verification feedback,  $p = .58$  for control group vs. elaborated feedback, and  $p = .88$  for verification feedback vs. elaborated feedback) or motivation ( $p = .87$  for all three pairwise comparisons).

Figure 4 – Boxplot of achievement scores per feedback condition



To conduct covariate analyses, linear regression was used, as no non-parametric equivalent of an ANCOVA is available to date. The regression models showed that achievement was not significantly influenced by any of the covariates when testing the effect of feedback complexity. For motivation and ASC, some significant predictors were found. A model including feedback complexity as a predictor, and pre-test motivation, background knowledge, effort, and sex as covariates was tested for intrinsic motivation ( $F(33,39) = .41, p = .83, R^2 = .13$ ). Only pre-test motivation ( $p = .02$ ) and effort ( $p = .01$ ) significantly explained

some of the variance in post-test intrinsic motivation. The covariate analysis for the ASC included pre-test motivation, background knowledge, pre-test ASC, the use of the ASC, study, sex, and post-test motivation next to the independent variable feedback complexity ( $F(32,39) = .20, p = .68, R^2 = .96$ ). For the ASC, pre-test ASC ( $p < .001$ ) and sex ( $p < .01$ ) were significant covariates. Based on these results, as well as those shown by the Kruskal-Wallis tests, hypothesis 1 had to be rejected, as no evidence for the effect of feedback complexity on either achievement or intrinsic motivation was found.

A mediation analysis, aimed at examining the interaction effect between feedback complexity and background knowledge (hypothesis 2) did not show a significant effect for either achievement ( $F(3, 36) = .99, p = .41$ ), motivation ( $F(3, 36) = .03, p = .99$ ), or ASC ( $F(3, 36) = .54, p = .66$ ). Because no significant influence of background knowledge was found, hypothesis 2 was rejected.

To uncover whether reading and learning about the ASC caused changes in one's own ASC (corresponding with hypothesis 3), a Wilcoxon signed rank test was conducted. On the basis of this test, hypothesis 3 was rejected, meaning that no significant difference in ASC scores before and after reading the text was found ( $V = 282.5, p = .74$ ). A Kruskal-Wallis test ( $\chi^2 = .80, df = 2, p = .67$ ) and pairwise comparison ( $p = .87$  for control group vs. verification feedback,  $p = .87$  for control group vs. elaborated feedback, and  $p = .97$  for verification feedback vs. elaborated feedback) also showed that a potential effect of reading about the ASC was not significantly influenced by feedback complexity.

### Discussion

The main goal of this study was to examine the effect of feedback complexity on university students' academic achievement and intrinsic motivation. To investigate this, participants were divided over a control group and two experimental groups, in which the complexity of feedback was manipulated. As earlier research found background knowledge to

influence on the role of feedback complexity, a mediator analysis was conducted, as well.

Thirdly, the effect of reading a text about the ASC was examined by comparing pre- and post-test self-reports on students' own ASC.

In line with earlier meta-analyses (Schimmel, 1983; Van de Ridder et al., 2015), however, no effect of feedback complexity on subsequent achievement was found, leading to the rejection of hypothesis 1. The absence of an effect may be due to participants not being highly motivated to learn about the ASC. Low participant motivation was also an issue in the study by Bivens (1964), who likewise did not find an effect of feedback complexity on achievement. Of course, adding the lack of findings from the current study to the abovementioned meta-analyses, it is possible that feedback complexity simply does not affect achievement in any circumstance. Alternatively, the text may have been too short and too easy to understand for the feedback to have an effect. In comparison, Kulhavy and colleagues (1985), who found an inverse relationship between feedback complexity and achievement, used a 2,400-word text with 16 multiple choice questions on which feedback was provided.

Regarding intrinsic motivation, some earlier studies have found more complex feedback to be more motivating, whereas others concluded the opposite. As an explanation, feedback preferences of the used samples may have varied. To illustrate that, Narciss (1999) used a sample of undergraduate students studying psychology, education or communication at one university; Goltz and colleagues (1989) based their findings on an even more homogenous sample of workers from one factory. In contrast, the current study included both undergraduate and graduate students from universities throughout the Netherlands. Whereas the participants in the previous samples may have had similar preferences for a certain kind of feedback (possibly because they also received the same type of feedback), the sample in this was more heterogeneous.



Together, the finding that feedback complexity does not affect either academic achievement or intrinsic motivation suggests that these student outcomes are influenced by other factors. For that reason, this study also looked into two student characteristics that could have influenced achievement and motivation.

It was expected that background knowledge would mediate the effect of feedback complexity on student outcomes. This expectation was not met and hypothesis 2 had to be rejected. Interestingly, the measure for background knowledge (in the data analysis called 'pre-test knowledge') did not correlate with previous use of the ASC or field of study, which could be other indicators of background knowledge. This may indicate that the knowledge pre-test was not reliable. In light of earlier findings, it is also possible that the intention to improve one's performance was not present in the participants whilst reading the text and answering practice questions. When students do not intend to use the material to improve their achievement (which means that it has no value for them), feedback has been found to have less effect (Eccles et al., 1983; Evans et al., 2010). But in that scenario, researchers found background knowledge to be correlated with motivation (Alexander et al., 1995), which was not found in the current study. Replication of the current study with more participants and an alternative instrument for measuring background knowledge, could yield more insight into the role background knowledge in relation to feedback complexity.

Furthermore, reading a text about the academic self-concept did not seem to cause changes in student's own ASC, which contradicts hypothesis 3. The high overall score on the ASC in both the pre- and post-test implies that there was perhaps not much room for improvement anyways. Intuitively, one would attribute these findings to the sample consisting of university students, who are likely to know that they are performing well. However, as was also explained in the text used for this study, the Big Fish Little Pond effect states quite the opposite (Marsh, 1987). So, a more likely explanation lies in the relationship between student

involvement and academic self-concept. As House (2000) found in his experimental study, student involvement through for instance volunteering and participating in student clubs was associated with a higher ASC. So, respondents' voluntary participation in this research may be linked to their average high ASC both before and after reading the text. That is to say, there were likely other factors at play that influenced the ASC. A possibility to achieve a change in ASC based on reading a text, would be to include tips on improving one's own ASC. It would be interesting to investigate whether the use of text that is explicitly aimed at bringing about changes in students' ASC would be able to do so.

Furthermore, both pre- and post-test ASC showed a moderate correlation with pre-test motivation. It is not surprising to find the ASC being correlated with intrinsic motivation, as one's self-concept can be regarded as the motivational aspect of self-esteem (Gecas, 1982). However, the ASC scores only correlated with pre-test motivation and not with post-test motivation, a pattern that needs to be studied in more depth. Still, it may be worthwhile for teachers in university education to focus more on promoting a high ASC when they want to increase student motivation.

The finding that students' ASC is significantly predicted by their sex is also not unique. In this study, women on average had a higher global academic self-concept than men. Significant gender differences in ASC were also found by Skaalvik and Skaalvik (2004), although dependent on the academic domain (math versus languages). It would be valuable to study whether preferences in terms of feedback complexity also differ for men and women, which was not the focus of the present research.

All in all, several findings that are well-established in the literature (e.g. the correlation between ASC and motivation) also appeared in this study. So, although no significant influence was found for feedback complexity, the ability to find correlations between phenomena for which there is much evidence, indicates that the design of this study

was adequate. Hence, it may be worthwhile to replicate the study when overcoming some of its limitations, which are outlined below.

### **Limitations**

There are several limitations that need to be considered when interpreting the findings of the present study. First and foremost, the number of participants in this study was too low to reach adequate statistical power, based on the outcomes of the power analysis. A reason for the small sample size is that 56 respondents closed the survey after viewing the first two question groups. These question groups included the measure for background knowledge, and seven Likert scale items about intrinsic motivation for learning about the ASC. A possible explanation for why many respondents closed the survey after viewing the first or second question group, is that they did not find the topic relevant enough to continue. So, twice as many complete responses could have been gathered through ensuring that all participants had an incentive to complete the survey. This could be done by for example rewarding participants with 'test subject hours', a grade for the knowledge post-test or money, as was done in the study by Roelle and Nückles (2019). The low amount of participants in the present study could account for the lack of a statistically significant effect of feedback complexity on achievement, motivation and ASC. Therefore, it is recommended to replicate the study with a larger sample.

Second, the follow-up test formed a barrier to 12 people. Four of them refused to leave their email address behind, so there was no way of contacting them for the post-test. The other eight provided their email address, but simply did not fill out the post-test, even after having received two reminders. Again, this problem could be solved by granting a reward to participants after they completed the full experiment.

Next, there was a wide discrepancy in the time it took participants to conclude the follow-up survey. Some participants finished the post-test on the same day they received the

invitation for it, others only did so after two weeks or, in the most extreme case, a month. The time between the pre- and post-test was not used as a selection criterium to avoid having to exclude more participants, and because it only became clear halfway through data collection. Earlier responses had at that point already been anonymised. Hence, it was impossible to still include the time it took participants to complete the follow-up survey as a covariate. It is, therefore, also impossible to tell whether the variation in time mediated the effect of feedback complexity. If this variable indeed mediated the relationship between feedback complexity and achievement, motivation and the ASC, it would explain the lack of findings in the current analyses. Hence, it is recommended that the time between the experiment and the follow-up is included as a covariate or that it is controlled for in an offline setting. Roelle and Nückles (2019) did this by inviting participants to complete the follow-up test at the university on a set date.

Fourth, the survey software used was not optimal in display and user-friendliness on all devices. Text boxes moved when the surveys were completed on mobile devices and refreshing of the page caused answers to be lost. Because the surveys were quite long and intensive, these issues made that some respondents quit halfway through or did not want to start over again. In addition, the progress bar did not function well: it remained at 30 percent for too long and then jumped to 80 percent. The progress bar was found to be misleading (also causing respondents to quit before finishing the entire survey) and was turned off after a respondent commented about it. However, without a progress bar participants had no idea how many questions there were left to answer, which two people indicated as being frustrating. To make the progress bar function more accurately, it is suggested to show each question separately, instead of clustering multiple questions into a larger question group.

The final limitation concerns the instrument measuring academic achievement, as it only existed out of eight questions. These eight questions were subdivided into four

reproduction questions and four understanding question. So, due to the low number of questions, the understanding questions turned out to have an unacceptable reliability in the study by Roelle and Nückles (2019). The low reliability could also be attributed to the fact that an expository text was used, whilst the post-test included questions that involve relational processing. Because the text focuses readers' attention on the processing of individual phenomena (i.e. aspects of the ASC and processes related to it), the use of a knowledge post-test that asks participants to link these aspects may have led to the low overall achievement scores. As a result, it is recommended that a more reliable achievement post-test is used in future studies on feedback complexity.

### **Conclusion**

The aim of this study was to investigate the relationship between feedback complexity, and achievement, motivation, and the ASC. In addition, the mediating effect of background knowledge on this relationship was examined, as well as the role reading a text about the ASC. Due to a lack of significant findings, however, all hypotheses had to be rejected. The lack of findings may be due to the low amount of participants in the study and the questionable reliability of the measures used for background knowledge and achievement. Despite these limitations, the overall design of the experiment seemed appropriate. It is therefore advised to replicate the study with a larger sample and more reliable instruments to assess pre- and post-test knowledge, so implications for practice can be formulated regarding the use of different levels of feedback complexity.

### References

- Archer, J. C. (2010). State of the science in health professional education: Effective feedback. *Medical Education*, 44(1), 101-108. doi:10.1111/j.1365-2923.2009.03546.x
- Bivens, L. W. (1964). Feedback complexity and self-direction in programmed instruction. *Psychological Reports*, 14(1), 155-160. doi: 10.2466/pr0.1964.14.1.155
- Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. *European Psychologist*, 1(2), 100-112. doi: 10.1027/1016-9040.1.2.100
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. *Educational Psychology Review*, 15(1), 1-40. doi: 10.1023/A:1021302408382
- Champely, S. (2018). pwr: Basic Functions for Power Analysis. R package version 1.2-2. <https://CRAN.R-project.org/package=pwr>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Deci, E.L., Koestner, R., & Ryan, R.M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, 125, 627–668. doi: 10.1037/0033-2909.125.6.627
- Deci, E. L., & Ryan, R. M. (2003). Intrinsic motivation inventory. *Self-Determination Theory*, 267. Accessed online.
- Deci, E.L. and Ryan, R.M. (2010). Intrinsic Motivation. In *The Corsini Encyclopedia of Psychology* (eds. Weiner, I.B and Craighead, W.E.). doi: 10.1002/9780470479216.corpsy0467
- Dempsey, J., Driscoll, M., & Swindell, L. (1993). *Text-based feedback*. In J. Dempsey & G. Sales (Eds.), *Interactive instruction and feedback* (pp. 21–54). Englewood Cliffs, NJ: Educational Technology Publications.
- Doubé, W., Carding, P., Flanagan, K., Kaufman, J., & Armitage, H. (2018). Comparing

- feedback types in multimedia learning of speech by young children with common speech sound disorders: Research protocol for a pretest posttest independent measures control trial. *Frontiers in Psychology*, 9(444), 1-12. doi: 10.3389/fpsyg.2018.00444
- Dusek, J. B., & Flaherty, J. F. (1981). The development of self- concept during adolescent years. *Monographs of the Society for Research in Child Development*, 46(4), 1-67. doi: 10.2307/1166004
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). *Expectancies, Values, and Academic Behaviors*. In J. T. Spence (Ed.), *Achievement and Achievement Motivation* (pp. 75-146). San Francisco, CA: W.H. Freeman.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational research*, 83(1), 70-120. doi: 10.3102/0034654312474350
- Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, 14(4), 445-463. doi: 10.1177/1362168810375367
- Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, 36, 51–62. doi: 10.1080/02602930903197883
- Flunger, B., Trautwein, U., Nagengast, B., Lüdtke, O., Niggli, A., & Schnyder, I. (2015). The Janus-faced nature of time spent on homework: Using latent profile analyses to predict academic achievement over a school year. *Learning and Instruction*, 39, 97-106. doi: 10.1016/j.learninstruc.2015.05.008
- Gecas, V. (1982). The self concept. *Annual Review of Sociology*, 8, 1–33. doi: 10.1146/annurev.so.08.080182.000245
- Ginsberg, M. B. (2005). Cultural diversity, motivation, and differentiation. *Theory into*

*Practice*, 44(3), 218-225. doi: 10.1207/s15430421tip4403\_6

Gliem, J. A., & Gliem, R. R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. *Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education*, Columbus, Ohio, 8-10 October 2003.

Goltz, S. M., Citera, M., Jensen, M., Favero, J., & Komaki, J. L. (1989). Individual feedback: Does it enhance effects of group feedback? *Journal of Organizational Behavior Management*, 10, 77-92. doi: 10.1300/J075v10n02\_06

Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40–53. doi: 10.1016/j.jslw.2007.01.001

Hattie, J. & Timperley, H. (2007) The power of feedback. *Review of Educational Research* 77(1), 81–112. doi: 10.3102/003465430298487

HEFCE. (2011). *National student survey: Findings and trends 2006-2010*. Bristol: HEFCE.

Hounsell, D. (2007). *Towards more sustainable feedback to students*. In: Rethinking assessment in higher education: Learning for the longer term. (eds. D. Boud & N. Falchikov).

House, J. D. (2000). The effect of student involvement on the development of academic self-concept. *The Journal of Social Psychology*, 140(2), 261-263. doi: 10.1080/00224540009600467

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal relations. *Journal of School Psychology*, 49(5), 505-528. doi: 10.1016/j.jsp.2011.07.001

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63-76. doi: 10.1177/1469787412467125



- Knofczynski, G. T., & Mundfrom, D. (2008). Sample sizes when using multiple linear regression for prediction. *Educational and Psychological Measurement, 68*(3), 431-442. doi: 10.1177/0013164407310131
- Krause, U. M., Stark, R., & Mandl, H. (2009). The effects of cooperative learning and feedback on e-learning in statistics. *Learning and Instruction, 19*(2), 158-170. doi: 10.1016/j.learninstruc.2008.03.003
- Kreber, C., Anderson, C., Entwistle, N., & McArthur, J. (2014). *Advances and Innovations in University Assessment and Feedback: A Festschrift in Honour of Professor Dai Hounsell*. Edinburgh University Press.
- Kulhavy, R. W., White, M. T., Topp, B. W., Chan, A. L., & Adams, J. (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology, 10*(3), 285-291. doi: 10.1016/0361-476X(85)90025-6
- Mandernach, B. J. (2005). Relative effectiveness of computer-based and human feedback for enhancing student learning. *The Journal of Educators Online, 2*(1), 1-17. doi: 10.9743/JEO.2005.1.5
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3), 280-295. doi: 10.1037/0022-0663.79.3.280
- Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational psychology, 82*(4), 623-636. doi: 10.1037/0022-0663.82.4.623
- Marsh, H. W., Byrne, B. M., & Shavelson, R. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. *Journal of Educational Psychology, 80*, 366-380. doi: 10.1037/0022-0663.80.3.366
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of

- the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79(3), 1129-1167. doi: 10.3102/0034654309337522
- Monteiro, V., Mata, L., & Peixoto, F. (2015). Intrinsic motivation inventory: Psychometric properties in the context of first language and mathematics learning. *Psicologia: Reflexão e Crítica*, 28(3), 434-443. doi: 10.1590/1678-7153.201528302
- Mouratidis, A., Lens, W., & Vansteenkiste, M. (2010). How you provide corrective feedback makes a difference: The motivating role of communicating in an autonomy-supporting way. *Journal of Sport and Exercise Psychology*, 32(5), 619-637. doi: 10.1123/jsep.32.5.619
- Narciss, S. (1999). *Motivational Effects of the Informativeness of Feedback*. Presented at American Educational Research Association Annual Meeting, Montreal, Quebec, 19-23 April, 1999.
- Neyman, J. (1950). *First Course in Probability and Statistics*. New York: Holt.
- Pereira, D., Flores, M. A., Simão, A. M. V., & Barros, A. (2016). Effectiveness and relevance of feedback in Higher Education: A study of undergraduate students. *Studies in Educational Evaluation*, 49, 7-14. doi 10.1016/j.stueduc.2016.03.004
- Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology*, 14(2), 97-110. doi: 10.1016/0361-476X(89)90028-3
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Roelle, J., & Nückles, M. (2019). Generative learning vs. retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, 111(8), 1341–1361. doi: 10.1037/edu0000345

- Sandelowski, M. (2000). Focus on research methods combining qualitative and quantitative sampling, data collection, and analysis techniques. *Research in Nursing & Health*, 23(3), 246-255. doi: 10.1002/1098-240X(200006)23:3<246::AID-NUR9>3.0.CO;2-H
- Schimmel, B. J. (1983, April). *A meta-analysis of feedback to learners in computerized and programmed instruction*. Paper presented at the annual meeting of the American Educational Research Association, Montréal. (ERIC Document Reproduction Service No. 233708).
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189. doi: 10.3102/0034654307313795
- Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, 50(3-4), 241-252. doi: 10.1023/B:SERS.0000015555.40976.e6
- Stern, L. A., & Solomon, A. (2006). Effective faculty feedback: The road less traveled. *Assessing Writing*, 11(1), 22-41. doi: 10.1016/j.asw.2005.12.001
- Van de Ridder, J. M., McGaghie, W. C., Stokking, K. M., & ten Cate, O. T. (2015). Variables that affect the process and outcome of feedback, relevant for medical training: A meta-review. *Medical Education*, 49(7), 658-673. doi: 10.1111/medu.12744

**Appendix I – Experiment Text**

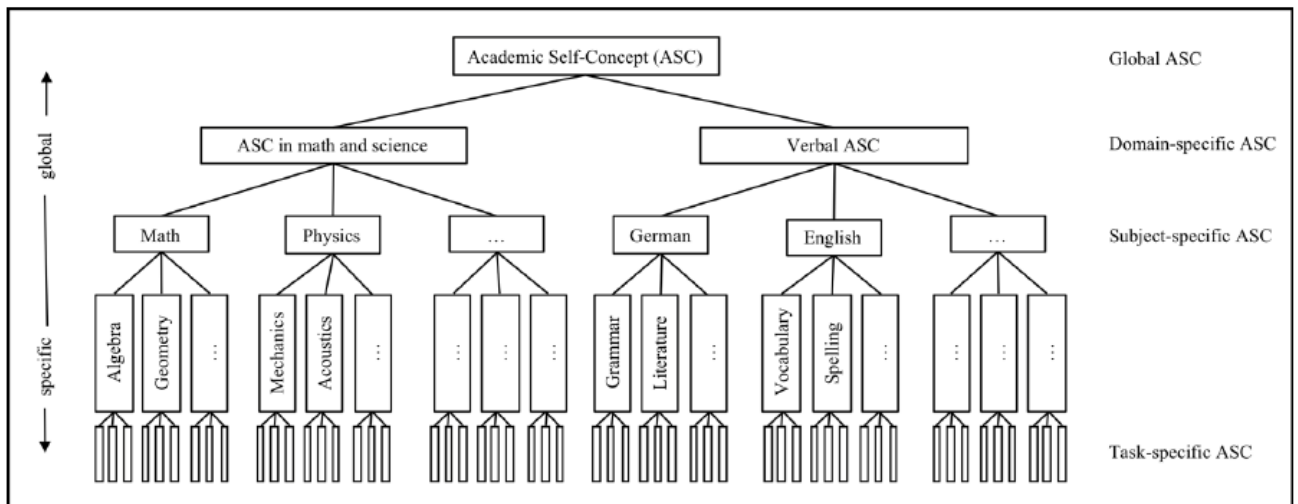
**Part 1**

The Academic Self-Concept

The term **self-concept** denotes assessments and attitudes related to many aspects of one’s own self. Among these assessments and attitudes are not only general emotional evaluations (“What am I good for?”) but also more or less rational assessments of one’s own properties and abilities (e.g. “How pretty/vain/clever/quick am I?”). Traditionally, self-concept research in educational psychology—particularly in relation to students—is very active. This is mostly for two reasons:

1. The attainment of a positive self-concept is regarded as an important educational goal, as the psychological well-being of children, youths, and adults all profit from a positive self-concept.
2. The assumption that a positive self-concept can positively influence the demonstrated performance is empirically sound.

The following text concerns the **academic self-concept (ASC)**. The academic self-concept denotes (not necessarily realistic) knowledge about one’s own academic abilities. According to common theoretical perception, the academic self-concept has a hierarchical structure. Thus, one may divide the academic self-concept into domain-specific self-concepts in mathematical-natural sciences and linguistics, which in turn can be further differentiated into specific self-concepts in the individual school subjects (subject-specific self-concept). These are in turn also subdivided task-specifically (see Figure). With increasing age students gain incremental knowledge about themselves and their abilities. This allows them to recognize their own differentiated strengths and weaknesses.



**Part 2**

Influences on the academic self-concept

It generally holds true that the academic self-concept is further differentiated with increasing age. This allows students to recognize their own differentiated strengths and weaknesses. The fact that the academic self-concept is further differentiated from global to more specific over the course of an academic career leads to the question of how the academic self-concept arises and changes. An important source for the formation of the academic self-concept is the person’s interaction with the environment. Here the primary path is when the judgments of another person are shared directly with the student (e.g. when a fellow student or a teacher

says “you’re good at math”). These opinions of other people can, however, also be indirectly shared with the student. So, for example, if a student is frequently consulted for help with math from their fellow students they may draw the conclusion that they are good at math.

Another significant source for the formation of the academic self-concept are comparisons of one’s own performance with certain **reference norms**. Reference norms are standards with which one compares an existing result (e.g. one’s own result in the last math exam). Three reference norms can be differentiated: (a) the individual reference norm, (b) the social reference norm, and (c) the objective reference norm.

- a) An individual reference norm denotes that one’s current result is measured against one’s previous result.
- b) The social reference norm denotes, by contrast, that one orients oneself to the performance of one or more other persons. In school this often means the class average.
- c) The objective reference norm denotes orientation towards objective considerations, such as pass or fail criterions.

### ***Part 3***

#### High- and low-performing environments

Oftentimes student orient themselves particularly strongly to the social reference norm. The **big-fish-little-pond-effect** phenomenon states that there is a negative correlation between the achievement potential of a class and the academic self-concept of an individual student. Concretely this means that the better a class is, the lower the academic self-concept of an individual student. On the other hand, this also means that the worse a class is, the better the academic self-concept of an individual turns out.

Why is this true? Well, when students find themselves in a high-achieving class, the social comparison often turns out poorly for them. They become a “small fish in a large pond”. In a low-achieving class, a student is instead more likely to become a “big fish in a small pond”. This big-fish-little-pond effect is especially pronounced during the transition from primary to secondary school. Students in the lower-performing group at the end of primary school often experience an increase in their academic self-concept after the change to secondary school. Students in the upper performance group at the end of primary school experience rather a decline in their academic self-concept after the change to a secondary school (VWO).

However, at least for students that change to a high-performing type of school, there is also an effect that counteracts the big-fish-little-pond effect. It has been shown that the self-confidence of being enrolled in a prestigious level of schooling like VWO can raise the ability-self-concept. This mechanism is also known as “**basking-in-reflected-glory**” or a prestige effect. However, this prestige effect is generally significantly less pronounced than the negative effect of the adverse social comparisons in high-performing classes.

### ***Part 4***

#### Effects on performance and motivation

The academic self-concept of learners has diverse effects on their current motivation and the quality of their learning processes. With regard to motivation, the influence of the academic self-concept on the expectations of success is particularly significant. The higher the subject-specific academic self-concept is, the higher the expectation of success concerning the subject. When the expectation of success is higher, a student will have more motivation to invest effort into the subject. The higher the expectation of success, the more motivation a

student has to invest effort into the processing of a task in a given subject. Accordingly, successful learning is likely to increase. In addition, a series of research activities substantiate that students with a higher subject-specific academic self-concept experience—due to their high expectation of success—fewer task-irrelevant thoughts when studying for a subject (e.g. anxiety, self-doubt) and likewise show a greater resilience against arising difficulties than students with a lower subject-specific academic self-concept. In a multitude of studies, it was demonstrated that subject-specific academic self-concepts can positively influence the school performance in the respective subject.

## Appendix II – Practice Questions and Feedback

### *Feedback questions*

Question after part 1:

Beschrijf de structuur van het academische zelfconcept.

*Verification feedback - Het juiste antwoord is: het academische zelfconcept (AZC) heeft een hiërarchische structuur. Het globale AZC kan worden onderverdeeld in domein-specifieke AZC's, die weer verder kunnen worden opgesplitst in vakspecifieke en taakspecifieke AZC's.*

*Elaborated feedback - Het academische zelfconcept (AZC) heeft een hiërarchische structuur. Dit betekent dat het globale AZC kan worden onderverdeeld in specifiekere AZC's. Ten eerste kan je AZC verschillen per domein (e.g. alfa- versus bètavakken). Omdat deze domeinen bestaan uit meerdere vakken, kan je AZC verder worden opgesplitst in vakken (e.g. Nederlands versus Engels) of zelfs in taken (e.g. een opstel schrijven versus stijffiguren herkennen). Dit betekent dus dat je als student niet hetzelfde AZC hebt voor verschillende vakken, ook al heb je een globaal concept van hoe goed je bent in studeren in het algemeen.*

Question after part 2:

Voor haar vorige wiskundetentamen haalde Lisa een 5,5. Nu haalt ze een 6. Op het eerste gezicht lijkt dit goed voor haar academische zelfconcept voor wiskunde. Lisa is echter erg gefocust op de sociale referentienorm (het groepsgemiddelde was een 7) en de objectieve referentienorm (in haar familie wordt een wiskundecijfer van 7,5 als goed gezien). Wat gebeurt er nu met Lisa's academische zelfconcept?

*Verification feedback - Het juiste antwoord is: Lisa's academische zelfconcept zal niet beter worden of zelfs verslechteren.*

*Elaborated feedback - Lisa's academische zelfconcept zal niet verbeteren, omdat zij geen aandacht besteedt aan hoe ze ten opzichte van zichzelf vooruit is gegaan (de individuele referentienorm). In plaats daarvan vergelijkt ze haar cijfer met het groepsgemiddelde, wat aantoont dat ze onder gemiddeld presteert (de sociale referentienorm). Daarnaast vergelijkt Lisa haar cijfer met de standaard die door haar familie gehanteerd wordt en waar ze ook niet aan voldoet (de objectieve referentienorm). Vanwege deze negatieve vergelijkingen, kan haar academische zelfconcept zelfs verslechteren.*

Question after part 3:

Hoewel iedereen zich goed zou moeten voelen door een "big fish in a little pond" te zijn, is het niet altijd het geval dat iedereen zich slecht voelt over het zijn van een "little fish". Leg uit hoe het prestige effect als een buffer kan werken tegen het "Big-Fish-Little-Pond-effect".

*Verification feedback - Het juiste antwoord is: les krijgen op een prestigieus schoolniveau kan het academische zelfconcept doen stijgen, wat de negatieve correlatie tussen prestaties van de groep en individueel AZC tegengaat.*

*Elaborated feedback - Deel uitmaken van een goed presterende groep kan leiden tot een afname van het individuele academische zelfconcept, vanwege de hoge sociale referentienorm ("Big-Fish-Little-Pond-effect"). Aan de andere kant kan les volgen op een prestigieus schoolniveau ook het academische zelfconcept doen stijgen, wat het prestige effect genoemd*

*wordt. Deze effecten balanceren elkaar niet uit, omdat het "Big-Fish-Little-Pond-effect" een groter effect heeft dan het prestige effect. Om die reden kan het prestige effect slechts als een buffer werken tegen de negatieve invloed van het "Big-Fish-Little-Pond-effect".*

Question after part 4:

Beschrijf de relatie tussen het vakspecifieke academische zelfconcept en motivatie.

*Verification feedback - Het juiste antwoord is: hoe hoger het academische zelfconcept voor vak, des te hoger de motivatie voor dat vak.*

*Elaborated feedback - Het academische zelfconcept beïnvloedt je succesverwachtingen, die op hun beurt motivatie beïnvloeden. Dus, de kans is groter dat je verwacht te slagen voor een vak, wanneer je een hoog academische zelfconcept voor dat vak hebt. Doordat je verwacht te slagen, stijgt je motivatie om inderdaad succesvol in het vak te zijn. Het is belangrijk om te noemen dat het academische zelfconcept dus geen directe invloed heeft op motivatie, maar met name succesverwachtingen beïnvloedt.*



**Appendix III – Manipulation Check**

*Feedback structure/value*

Ik begreep de feedback goed.

De ontvangen feedback was te complex.

Het lezen van de feedback kostte te veel tijd.

Ik had graag meer diepgaande feedback ontvangen.

De ontvangen feedback was hulpvol.

Ik heb een voorkeur voor kortere feedback.

Ik denk dat ik de feedback goed begreep, vergeleken met andere studenten.

Ik wil meer van dit soort feedback ontvangen.

De lengte van de feedback maakte dat deze bruikbaar was.

### Appendix IV – Knowledge Test

#### *Reproduction questions (original in German)*

1. Beschreiben Sie die Struktur des schulischen Fähigkeitsselbstkonzepts.
2. Beschreiben Sie die verschiedenen Quellen zur Entstehung/Veränderung des schulischen Fähigkeitsselbstkonzepts.
3. Was sagt der Big-Fish-Little-Pond-Effekt und was sagt „basking-in-reflected-glory“ aus?
4. Beschreiben Sie, inwiefern das schulische Fähigkeitsselbstkonzept relevant für die schulische Leistung ist.

#### *Understanding questions (original in German)*

1. Wenn Sie das schulische Fähigkeitsselbstkonzept eines Schülers einmal zu Beginn der Grundschule (Alter: ca. 6 Jahre), einmal zu Beginn der Sekundarschule (Alter: ca. 10 Jahre) und einmal am Ende der Sekundarschule (Alter: ca. 17 Jahre) messen würden, welche Unterschiede in der Struktur würden Sie erwarten? Begründen Sie.
2. Paula hat in der letzten Mathematikarbeit eine 2 geschrieben. Wenn sie die individuelle Bezugsnorm an ihre Leistung anlegt, verschlechtert sich ihr Fähigkeitsselbstkonzept im Fach Mathematik. Wenn sie die soziale Bezugsnorm anlegt, verbessert es sich. Was können Sie hieraus für den Klassendurchschnitt und Paulas bisherige Leistungen im Fach Mathematik folgern? Begründen Sie.
3. Tina und Tom befanden sich in der 4. Klasse der Grundschule im oberen Leistungsdrittel ihrer Klassen. Tina ist auf ein Gymnasium gewechselt, Tom auf eine Realschule. Welche Unterschiede in Bezug auf die Entwicklung des schulischen Fähigkeitsselbstkonzepts sollten zwischen den beiden nun auftreten? Begründen Sie.
4. Sophie und Lea sind beide in der gleichen Klasse und waren zu Beginn der 6. Klasse exakt gleich gut im Fach Mathematik. Sophie hatte jedoch zu Beginn der 6. Klasse ein höheres Fähigkeitsselbstkonzept im Fach Mathematik als Lea. Wenn Sie nun die Mathematikleistung der beiden Schülerinnen am Ende der 6. Klasse messen würden, welche Unterschiede zwischen Lea und Sophie würden Sie erwarten? Begründen Sie.

#### *Reproduction questions (translated to Dutch)*

1. Beschrijf de structuur van het academische zelfconcept.
2. Beschrijf de verschillende bronnen die het academische zelfconcept vormen/veranderen.
3. Wat is de definitie van het “Big-Fish-Little-Pond-effect” en wat is de definitie van “basking-in-reflected-glory”?
4. Beschrijf in hoeverre het academische zelfconcept relevant is voor studieprestaties.

*Understanding questions (translated to Dutch)*

1. Wanneer je het academische zelfconcept van een student zou meten aan het begin van de basisschool (ongeveer 6 jaar oud), aan het begin van de middelbare school (ongeveer 12 jaar oud) en aan het einde van de middelbare school (ongeveer 18 jaar oud), welke verschillen verwacht je dan in de structuur van het academische zelfconcept? Leg uit.
2. Voor haar laatste tentamen heeft Paula een 8 behaald. Wanneer ze de individuele referentienorm voor haar prestatie gebruikt, verslechtert haar academische zelfconcept. Wanneer ze de sociale referentienorm hanteert, verbetert haar academische zelfconcept. Wat kun je hieruit afleiden over haar vorige prestaties en het klassengemiddelde? Leg uit.
3. In groep 8 van de basisschool bevonden Tina en Tom zich in de hoogst presterende groep kinderen van de klas. Vervolgens is Tina naar het VWO gegaan en Tom naar de HAVO. Welke verschillen in de ontwikkeling van het academische zelfconcept van beide leerlingen verwacht je? Leg uit.
4. Sophie en Lea zitten in dezelfde klas en zijn aan het begin van klas 2 even goed in wiskunde. Toch had Sophie aan het begin van klas 2 een hoger academische zelfconcept voor het vak wiskunde dan Lea. Wanneer je de wiskunde prestaties van beide leerlingen aan het einde van de tweede klas zou meten, welke verschillen verwacht je dan tussen Sophie en Lea? Leg uit.

**Appendix V – IMI**

***Interest/Enjoyment (original in English)***

I enjoy learning about the academic self-concept very much.  
Reading about academic self-concept is fun to do.  
I think the academic self-concept is boring.  
Learning about the academic self-concept does not hold my attention at all.  
I would describe the academic self-concept as very interesting.  
I think that reading about the academic self-concept is quite enjoyable.  
While I was writing down what I know about the academic self-concept, I was thinking about how much I enjoyed it.

***Interest/Enjoyment (translated to Dutch)***

Ik vind het erg leuk om te leren over het academische zelfconcept.  
Lezen over het academische zelfconcept is leuk om te doen.  
Ik vind het academische zelfconcept een saai onderwerp.  
Leren over het academische zelfconcept houdt mijn aandacht niet vast.  
Ik zou het academische zelfconcept omschrijven als erg interessant.  
Ik denk dat lezen over het academische zelfconcept best leuk is.  
Toen ik aan het opschrijven was, wat ik weet over het academische zelfconcept, dacht ik na over hoe leuk ik het vond.

**Appendix VI – ASCQ**

*Academic self-concept for math (original in English)*

Math is one of my best subjects.  
I often need help with math.  
I look forward to math lessons.  
I have trouble understanding math.  
I enjoy math.  
I do badly at math.  
I get good grades for math.  
I never want to do math again.  
I have never been good at math.  
I hate math.

*Academic self-concept (adjusted and translated to Dutch)*

Studeren is één van mijn sterkste vaardigheden.  
Ik heb vaak hulp nodig met studeren.  
Ik kijk uit naar studieactiviteiten.  
Ik heb moeite met het begrijpen van studie-gerelateerde onderwerpen.  
Ik vind studeren leuk.  
Ik presteer slecht op academische tentamens.  
Ik behaal goede resultaten door hoe ik studeer.  
Ik denk er vaak over na om te stoppen met mijn studie.  
Ik ben nooit goed geweest in studeren.  
Ik heb een hekel aan studeren.

**Appendix VII – Effort**

***General effort (original in Dutch)***

Ik ben bereid om energie in mijn studie te steken.

Ik doe erg mijn best voor mijn studie.

Ik werk serieus aan studieopdrachten.

Ik maak alle studieopdrachten zo goed mogelijk af.

## Appendix VIII – Informed Consent and Debriefing

### *Informed consent*

#### INFORMATIE OVER DIT ONDERZOEK EN GEÏNFORMEERDE TOESTEMMING

Dit onderzoek wordt uitgevoerd door Lotte van den Heuvel, master studente Educational Sciences aan de Utrecht Universiteit.

Dit onderzoek bestaat uit een voormeting en een nameting. Eerst krijg je een aantal vragen over motivatie en hoe jij je studie ervaart. Vervolgens lees je 4 korte teksten. Over iedere tekst word je gevraagd om 1 oefenvraag te beantwoorden. Aan het einde van de vragenlijst wordt er gevraagd naar je leeftijd, gender en studierichting. Na een week zul je benaderd worden om de nameting in te vullen. De nameting bevat vragen over de gelezen teksten en meet wederom jouw motivatie. Het is belangrijk dat je ook de nameting doet, anders kunnen je antwoorden niet gebruikt worden.

Er zal niet gevraagd worden naar je naam, dus het onderzoek is anoniem. Bovendien gaat dit onderzoek over groepeffecten, dus hoef je je geen zorgen te maken over jouw individuele prestaties tijdens de nameting. Beide metingen duren niet langer dan 15 minuten.

#### Vrijwillige deelname

Jouw deelname aan dit onderzoek is vrijwillig. Mocht je op een bepaald moment dus niet meer willen deelnemen, dan kun je je terugtrekken uit het onderzoek. Om jouw antwoorden te laten verwijderen, kun je contact opnemen via [l.heuvel@students.uu.nl](mailto:l.heuvel@students.uu.nl). Je hoeft niet uit te leggen waarom je jouw deelname beëindigt. Er zijn geen risico's verbonden aan deelname aan dit onderzoek.

#### Meer informatie

Als je op de hoogte gebracht wilt worden over de uitkomsten van dit onderzoek, kun je aan het einde van het onderzoek je e-mailadres achterlaten. De mailadressen zullen in een apart bestand worden bewaard, zodat je e-mail niet gekoppeld is aan je antwoorden in het onderzoek. Aarzel niet om bij verdere vragen of opmerkingen contact op te nemen via [l.heuvel@students.uu.nl](mailto:l.heuvel@students.uu.nl).

#### Ethische overwegingen

Dit onderzoek is goedgekeurd door de Ethische Toetsingscommissie van de Faculteit Sociale Wetenschappen van de Utrecht Universiteit. Voor klachten over dit onderzoek kun je mailen naar het bovengenoemde e-mailadres. Bij ernstige klachten wordt aangeraden om contact op te nemen met de vertrouwenspersoon van de universiteit: [vertrouwenspersoon-wi@uu.nl](mailto:vertrouwenspersoon-wi@uu.nl).

#### Geïnformeerde toestemming

Door akkoord te gaan met deelname aan dit onderzoek en verder te gaan in deze vragenlijst, bevestig je dat:

- Je de informatie hierboven gelezen hebt en ermee akkoord gaat;
- Je vrijwillig deelneemt;
- Je 18 jaar of ouder bent en studeert aan een universiteit.

Ga je akkoord met de voorwaarden en geef je toestemming tot het gebruik van jouw antwoorden voor dit onderzoek? Indien nee, sluit dan de vragenlijst af.

***Debriefing***

Ontzettend bedankt voor je deelname aan dit onderzoek!

Uitleg over het onderzoek

Het doel van dit onderzoek was om te testen of het ontvangen van verifiërende feedback (het juiste antwoord is...) of uitgebreide feedback (...is het juiste antwoord omdat... + ... is niet het juiste antwoord omdat...) leidt tot betere prestaties en verhoogde motivatie. In de vorige vragenlijst heb je oefenvragen beantwoord over een tekst en daarna een van deze soorten feedback ontvangen (experimentele groepen) of geen feedback ontvangen (controlegroep). Daarnaast wilden we kijken of het lezen van een tekst over het academische zelfconcept veranderingen teweeg brengt in jouw eigen academische zelfconcept.



**Appendix IX – Email to Participants**

Nameting master's thesis onderzoek

Goede morgen,

Een week geleden heb je deelgenomen aan mijn master's thesis onderzoek door de eerste vragenlijst in te vullen. Nu wil ik je vragen om ook de nameting te voltooien.

Daarvoor klik je op de volgende link:

<https://survey1.fss.uu.nl/index.php/997866?lang=nl>

En vul je op de tweede pagina deze toegangscode in: XXXX

Zodra jouw antwoorden voor de nameting ontvangen zijn, wordt je e-mailadres uit onze database verwijderd en zal ik ook deze mail uit mijn verzonden berichten verwijderen. Op die manier kunnen jouw antwoorden niet langer aan jouw e-mailadres gekoppeld worden.

Ik hoop je hiermee voldoende te hebben geïnformeerd en wens je succes bij de nameting.

Alvast hartelijk dank voor het invullen!

Met vriendelijke groeten,  
Lotte van den Heuvel

**Appendix X – FETC Form**

**Section 1: Basic Study Information**

1. Name student:

Lotte van den Heuvel

2. Name(s) of the supervisor(s):

Dr. Barbara Flunger (second assessor: Dr. Frans Prins)

3. Title of the thesis (plan):

The effect of feedback complexity on university students' achievement and intrinsic motivation

4. Does the study concern a multi-center project, e.g. a collaboration with other organizations, universities, a GGZ mental health care institution, or a university medical center?

Yes / No  
If yes: Explain.

5. Where will the study (data collection) be conducted? If this is abroad, please note that you have to be sure of the local ethical codes of conducts and permissions.

University students in the Netherlands

**Section 2: Study Details I**

6. Will you collect data?

Yes / ~~No~~  
Yes → Continue to question 11  
No → Continue to question 7

7. Where is the data stored?

8. Is the data publicly available?

Yes / No  
If yes: Where?

9. Can participants be identified by the student? (e.g., does the data contain (indirectly retrievable) personal information, video, or audio data?)

Yes / No  
If yes: Explain.

10. If the data is pseudonymized, who has the key to permit re-identification?

### Section 3: Participants

11. What age group is included in your study?

University students above the age of 18 (with the expected maximum age lying around 30)

12. Will be participants that are recruited be > 16 years? ~~Yes~~/No

13. Will participants be mentally competent (wilsbekwam in Dutch)? Yes/~~No~~

14. Does the participant population contain vulnerable persons? ~~Yes~~/No  
(e.g., incapacitated, children, mentally challenged, traumatized, pregnant)

15. If you answered 'Yes' to any of the three questions above: Please provide reasons to justify why this particular groups of participant is included in your study.

It is expected that university students are mentally competent, so there will be no ethical issues regarding the choices of caregivers on participation in the study.

16. What possible risk could participating hold for your participants?

None.

17. What measures are implemented to minimize risks (or burden) for the participants?

Data will be anonymized and participants can have their data taken out of the study at all times.

18. What time investment and effort will be requested from participants?

Participants fill out a two surveys (pre-test and post-test) containing an experiment, which will take no longer than 30 minutes to complete in total.

19. Will be participants be reimbursed for their efforts? If yes, how? (financial reimbursement, travelling expenses, otherwise). What is the amount? Will this compensation depend on certain conditions, such as the completion of the study?

No.

20. How does the burden on the participants compare to the study's potential scientific or practical contribution?

Participants can receive the outcomes of the study if they wish. In that way, they can also gain knowledge from participating. Other than that, and contributing to research in general, there are no specific benefits for participants.

21. What is the number of participants? Provide a power analysis and/or motivation for the number of participants. The current convention is a power of 0.80. If the study deviates from this convention, the FERB would like you to justify why this is necessary.

(Note, you want to include enough participants to be able to answer your research questions adequately, but you do not want to include too many participants and unnecessarily burden participants.)

The "pwr" package (Champely, 2018) in R (R Core Team, 2017) was used to uncover that  $n = 34$  per experimental condition is appropriate for a one-way ANOVA with 3 groups to achieve a statistical significance of  $p = .05$ . Hence, a total  $n$  of at least 102 would be needed. Unfortunately, only 42 participants completed both the pre- and post-test survey, so adequate statistical power was not reached. This was taken into account as a limitation when analyzing and interpreting the finding.

22. How will the participants be recruited? Explain and attach the information letter to this document.

Stratified purposeful sampling was used to approach students with varying levels of background knowledge on a topic related to Educational Sciences. Therefore, bachelor and master students enrolled in Educational Sciences at Utrecht University were invited to participate in the research via BlackBoard. Other participants targeted via social media.

23. How much time will prospective participants have to decide as to whether they will indeed participate in the study?

They can decide at any moment that they do not want to participate (anymore).

24. Please explain the consent procedures. Note, active consent of participants (or their parents) is in principle mandatory. Enclose the consent letters as attachments. You can use the consent forms on Blackboard.

Before starting the survey/experiment, participants have to tick a box stating that they consent.

25. Are the participants fully free to participate and terminate their participation whenever they want and without stating their grounds for doing so? Explain.

Yes. They can contact the researcher at any time to have their data eliminated from the sample. No explanation for wanting to have their data deleted is required.

26. Will the participants be in a dependent relationship with the researcher?

Yes / ~~No~~

If yes: Explain. Participants are not aware in which experimental condition they are grouped, and are not fully aware of the goals of the study until the debriefing that follows the experiment. This debriefing will be held on the last page of the survey.

27. Is there an independent contact person or a general email address of a complaint officer whom the participant can contact?

Yes: [vertrouwenspersoon-wi@uu.nl](mailto:vertrouwenspersoon-wi@uu.nl).

28. Is there an independent contact person or a general email address of a complaint officer whom the participant can contact in case of complaints?

Yes: [vertrouwenspersoon-wi@uu.nl](mailto:vertrouwenspersoon-wi@uu.nl).

**Section 4: Data management**

29. Who has access to the data and who will be responsible for managing (access to) the data?

Only the student researcher and the thesis supervisor have access to the collected data. The student will be responsible for managing the data.

30. What type of data will you collect or create? Please provide a description of the instruments.

The collected data consist of general participant information (bachelor/master, field of study, male/female/other, university, age); answers to a short multiple-choice knowledge test based on a text participants have read during the experiment; and information about the participants' level of motivation during the experiment.

31. Will you be exchanging (personal) data with organizations/research partners outside the UU?

Yes / No  
If yes: Explain.

32. If so, will a data processing agreement be made up?

Yes / No  
If yes: Please attach the agreement.  
If no: Please explain.

33. Where will the data be stored and for how long?

Data will be stored in the YODA environment until graduation of the researcher.

34. Will the data potentially be used for other purposes than the master's thesis? (e.g., publication, reporting back to participants, etc.)

No. Information reported back to participants will be of general nature (i.e. overall conclusions as reported in the master's thesis, not individual answers).

35. Will the data potentially be used for other purposes than the master's thesis? (e.g., publication, reporting back to participants, etc.)

Yes / No

If yes: Explain.

Attachment:

Information provided on BlackBoard to invite students to participate (in Dutch)

Beste student,

Wil je inspiratie opdoen voor wetenschappelijk onderzoek in onderwijskunde? Of wil je oefenen met toetsvragen over een onderwijskundig thema? Neem dan deel aan dit onderzoek! Ik ben op zoek naar studenten die willen meedoen aan onderwijswetenschappelijk onderzoek door een vragenlijst in te vullen. Deelname duurt maximaal 2x 15 tot 20 minuten (voor- en nameting). Het onderzoek richt zich op het lezen en verwerken van een verklarende tekst over een onderwijswetenschappelijk relevant onderwerp.

Door deel te nemen kun je kennis opdoen over factoren, die de prestatie van studenten en leerlingen kunnen beïnvloeden.

Mocht je geïnteresseerd zijn, kan ik je na afloop van het onderzoek mijn thesis opsturen. Zo kun je bekend raken met vragenlijsten en onderzoeksmethoden. Klik op de onderstaande link om deel te nemen. Hartelijk dank!

<https://survey1.fss.uu.nl/index.php/251461?newtest=Y&lang=nl>