The Effects of Feedback Valence on Challenge and Threat States, and the Moderating Role of Self-Efficacy

Master's Thesis

Maaike Taheij (5908949)

Utrecht University

Master Educational Sciences

Faculty of Social & Behavioural Sciences

Supervisor: Dr. Eva Janssen

Second assessor: Prof. dr. Liesbeth Kester

Version: Final

Date: 13-10-2022

Word count: 7994

Abstract

How students respond to feedback is critical for subsequent learning. Plausibly, students respond differently to (negative) performance feedback because they react to stress differently. The Biopsychosocial Model (BPSM) of challenge and threat offered a potential explanation for the individual differences in stress responses to feedback. This study investigated whether the BPSM can be used to gain insight into the affective responses (self-reported challenge and threat) to positive and negative valanced feedback, and if individuals' self-efficacy can explain individual differences in these responses. This was researched through an online experiment where Dutch University students answered problem-solving tasks on which they received manipulated performance feedback. Results showed that participants who received negative feedback perceived more threat states and fewer challenge states towards the problem-solving tasks than those who received positive feedback. Furthermore, results showed that self-efficacy towards problem-solving tasks positively relates to challenge states, and mixed results were found regarding threat states. No moderating effect of self-efficacy on the relationship between feedback valence and affective responses was found. These results encourage further research to investigate how negative performance feedback could be less threatening to students. In addition, it is relevant to re-examine the moderating effect of self-efficacy, given the study's limitations.

Keywords: Feedback valence, Performance feedback, Self-efficacy, Affective Responses, Biopsychosocial model of challenge and threat

The Effects of Feedback Valence on Challenge and Threat States, and the Moderating Role of Self-Efficacy

In academic environments, the idea is accepted that feedback is an essential component of the learning cycle (Weaver, 2006). Providing students with feedback can influence their ability to perform a task and accurately judge their own performance (Fotheringham, 2011). Feedback is information provided by an agent, for example, a teacher, regarding one's performance or understanding (Hattie & Timperley, 2007). The primary purpose of feedback is to reduce discrepancies between current understandings and performance and the desired goal. The provided feedback has a certain valence: positive or negative. It has a positive valence when it informs students that their performance was correct or higher than expected. Feedback has a negative valence when it informs students that their performance is incorrect or lower than expected (Raaijmakers et al., 2017).

Despite that feedback is employed with the intention to improve performance, feedback does not always have the desired effect on learners (Ilgen & Davis, 2000; Poulos & Mahony, 2008; Winstone et al., 2021). Research shows that how students interpret feedback and act upon those interpretations is critical for subsequent learning and performance (Poulos & Mahony, 2008). Brown and Creaven (2017) argue that it is plausible that individuals respond differently to (negative) performance feedback because they react to stress differently. The Biopsychosocial Model (BPSM) of challenge and threat offers a potential explanation for the individual differences in stress responses to feedback (Blascovich, 2008). Receiving feedback elicits a certain affective response, which can be referred to as challenge and threat states. Challenge states are suggested to facilitate performance, whereas threat states would hinder performance (Gildea et al., 2007; Moore et al., 2012) as these lead to counterproductive learning behaviours and reduced motivation (Belschak & Den Hartog, 2009).

The present study explored whether the BPSM can be used to gain more insight into the affective responses to positive and negative feedback and if individual differences in these responses can be explained by individuals' self-efficacy. For example, negative feedback might be less threatening for students with higher levels of self-efficacy than for those with lower levels of self-efficacy. Theoretically, these questions are relevant as we still know little about how feedback valence affects students' challenge and threat responses and the variables that might explain the individual differences in these responses. Most feedback studies are limited to investigating the recipients' performance improvement rather than affective aspects (Kim & Lee, 2019). The present study should strengthen the theory and explain why negative feedback is more threatening to some students than others. Practically, the issue is relevant as students are often subjected to negative feedback. With a better understanding of why negative performance feedback can evoke feelings of threat in some and feelings of challenge in others, feedback interventions can be evaluated and improved.

Theoretical Framework

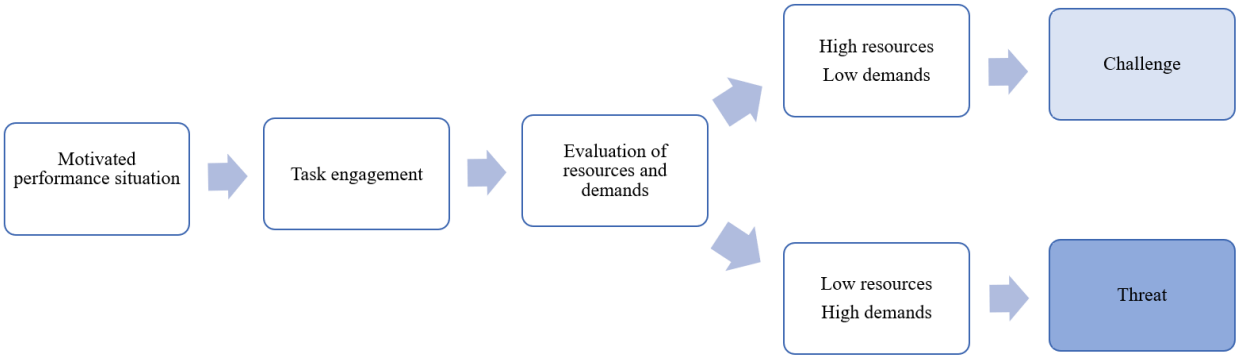
The Biopsychosocial Model of Challenge and Threat

The BPSM provides a possible theoretical approach for understanding individuals' psychological responses to feedback, as it provides a rationale for a connection between specific psychological states and physiological responses, for example, increased heart rate due to test-taking (Blascovich, 2008; Blascovich & Tomaka, 1996; Seery, 2013). These responses can occur when individuals engage in motivated performance situations where they try to reach a self-relevant goal, causing them to be actively engaged in the task (e.g., test-taking; Seery, 2013). Task engagement is a prerequisite component for challenge and threat states. When one is task-engaged, evaluations of individual resources and situational demands determine to what extent experiences of challenge and threat occur (Figure 1; Seery, 2011; 2013). Resources are an individual's tools to execute a task, including knowledge, skills,

abilities and external support. Demands appraisals include the perception of danger, uncertainty and the required effort to execute the task (Tomaka et al., 1997; Blascovich et al., 2003; 2008). A challenge state is experienced when individuals evaluate resources to meet or exceed evaluated demands, whereas a threat state is experienced when evaluated demands exceed evaluated resources (Blascovich & Mendes, 2000; Seery, 2011). These are not two dichotomous states but two ends of a bipolar continuum. Therefore, relative differences (e.g., greater versus lesser challenge) can be discerned, providing a nuanced view of challenge and threat states (Blascovich, 2008; Seery, 2011).

Figure 1

Overview of BPSM of Challenge and Threat (Seery, 2011; 2013)



Challenge and threat states can be measured via physiological cardiovascular measurement tools. For instance, heart rate variability and cardiac pre-ejection period are markers of the autonomic nervous system's response reflecting task engagement (Martin et al., 2021; Moore et al., 2012; Seery: 2011). In addition, challenge and threat states can be psychologically measured through self-reports (Scholl et al., 2018). The measured challenge and threat states in the self-report are relative, based on the idea that the relationship between resources and demands is evaluated as being more (or less) available in a situation.

Challenge and Threat States and Performance

Empirical and predictive studies in psychology, across various situations, have revealed that a challenge state typically facilitates performance, whilst a threat state hinders

performance (Mendes et al., 2007; Seery et al., 2010). A review study by Hase et al. (2019) examined whether a challenge state is associated with superior performance rather than a threat state. Across 38 published articles that conceptualised challenge and threat states in a manner congruent with the BPSM, support emerged for the performance benefits of the challenge state. These results imply that teachers may benefit from promoting a challenge state and preventing a threat state while providing performance feedback to students. However, it is not evident what the effect of performance feedback is on individual affective responses (challenge and threat).

Feedback Valence and Affective Responses

Research shows that how students interpret feedback and act upon those interpretations is critical for subsequent learning and performance (Hase et al., 2019; Poulos & Mahony, 2008). These interpretations are influenced by factors such as emotion (Forgas et al., 1990; Ilgen & Davis, 2000). Feedback in a positive valence will generally lead to positive emotions (e.g., pride, relief), whereas feedback in a negative valence will generally result in negative emotions (e.g., frustration, disappointment) (Lazarus, 1991; Raaijmakers et al., 2017). Moreover, positive emotions are typically associated with a challenge response, and negative emotions are typically associated with a threat response (Jones et al., 2009).

Considering that positive valence informs students that their performance was correct or higher than expected, and negative valence informs students that their performance was lower than expected or incorrect (Raaijmakers et al., 2017), feedback presumably influences the challenge and threat responses accordingly. Providing students with positive or negative formulated feedback is likely to influence these affective responses, as the stress appraisal leads to differences in the responses to stressors (Brown & Craeven, 2017). Thus, it seems plausible that positive feedback stimulates a challenge response and negative feedback a threat response. Although this has not been researched before, it is relevant, considering the

relation between challenge and threat on performance (Mendes et al., 2007; Seery et al., 2010). In addition to studying the effects of feedback valence on perceived challenge and threat, it is relevant to study individual differences, which may explain why some respond better to negative feedback than others (Nease et al., 1999). As previously mentioned, the BPSM lends itself to explaining individual differences in stress responses (Blascovich, 2008), which is important since people react differently to (negative) feedback. Based on the BPSM, it may be argued that individual differences in self-efficacy be an explanatory factor for the differences in affective responses.

Self-efficacy and Affective Responses

Self-efficacy refers to an individual's belief that one's resources meet or exceed the demands of a task (Bandura, 1997; Schunk & DiBenedetto, 2014). It is a self-appraisal of one's ability to master a task and is shaped by self-beliefs about own skills within a specific context (Zimmerman, 2000). This construct can be seen as the resource appraisal that fits the BPSM model (Figure 1) and influences challenge and threat responses (Jones et al., 2009; Uphill et al., 2019). In addition, research shows that more self-efficacious students participate more readily, work harder and persist longer when they encounter difficulties than less self-efficacious students (Bandura, 1997: Schunk & Zimmerman, 2007), thus are likely to evoke a challenge state (Rossato et al., 2018). High self-efficacy beliefs might enable students to view negative feedback as challenges rather than threats, resulting in students taking advantage of the learning opportunities in feedback (Putwain et al., 2013; Adams et al., 2020). However, it is not investigated whether different affective responses to negative performance feedback are related to the individual's differences in relatively high and low self-efficacy beliefs.

Present Study

In this study, participants answered problem-solving tasks and received manipulated feedback on their performance after each task. The aim of this study was to examine the effect

of manipulated performance feedback (positive valence vs negative valence) on students' affective responses (self-reported challenge and threat states towards problem-solving tasks), and whether self-efficacy (towards problem-solving tasks) can explain individual differences in the assumed relation between feedback valence and affective responses. Therefore, the following sub-questions are formulated and tested: (Q1) What is the effect of feedback valence on students' affective responses?; (Q2) What is the relation between self-efficacy feelings and students' affective responses?; (Q3) Do students' self-efficacy feelings toward problem-solving tasks moderate their affective responses to feedback? In addition, it is explored whether negative feedback affects self-efficacy to decrease. This seemed plausible considering the theory of challenge and threat, as the available perceived resources are likely to decrease.

It was hypothesised that (H1) positive feedback would result in a higher relative challenge than negative feedback; (H2) the higher one's self-efficacy, the higher one's relative challenge would be; (H3) the effect of feedback on the affective responses would be moderated by individuals' degree of self-efficacy. Specifically, it was expected that the effect of negative feedback on the relative threat response would be stronger for those with relatively lower self-efficacy than for those with relatively higher self-efficacy.

Method

Design

The emotional impact of feedback was investigated via a between-subjects design. The dependent variable in this design was *the affective response* (relative challenge and threat), and the independent variables were *feedback* (consisting of two conditions; positive and negative) and *self-efficacy*. Participants were randomly assigned to one of the two conditions: predominantly positive feedback or predominantly negative feedback, where the participants received performance feedback after each problem-solving task. Predominantly positive

feedback led participants to believe their performance was correct on all tasks except for the fourth (PPPNP), and predominantly negative feedback vice versa (NNNPN). As the feedback was manipulated, it did not dependent on participants' actual performance. Self-efficacy was measured in a pre-test (and post-test for explorative research). Challenge and threat states were measured after the five tasks whilst participants were under the impression that they had to perform another five tasks, which they did not. This questionnaire asked them about their feelings of challenge and threat towards the expected problem-solving tasks.

Participants

Before data collection, a power analysis was performed using G*Power 3.1.9.7 (Appendix A) for sample size estimation (Faul et al., 2007). The power analysis was based on a conservative scenario to ensure enough participants were acquired. Therefore, an ANOVA (fixed model, main effects and interaction) 2x2 between-subjects design was used, with two factors: feedback condition and self-efficacy, in which self-efficacy was considered a categorical variable (low vs high self-efficacy). Results indicated that the required sample size to achieve a power of .80 for detecting a medium effect ($\eta^2 = .06$) (Cohen, 1988), with a significance criterion of $\alpha = .05$, was $N = 125$.

Dutch university students were recruited with a convenience sample to participate and were approached via the researchers' network (e.g., social media, email). The acquired data were anonymised and stored at YourData to ensure the safety and confidentiality of participants' details. Before participating in the present study, each participant signed the informed consent form. Participants did not get financial compensation and could stop participating in the experiment without giving a reason. The faculty's ethical review board (FERB) approved this study.

Of the 188 participants, 42 were removed prior to analyses because they 1) did not complete the experiment ($n = 30$); 2) took longer than the cut-off time of 30 minutes to

complete the experiment (n = 11); 3) were younger than 17 years old and therefore too young to participate (n = 1). With the remaining sample of N = 146 participants is the achieved power 0.86 (α = .05) for detecting a medium effect (η^2 = .06) (Cohen, 1988; Appendix A). The experimental condition with predominantly positive feedback (PPPNP) consisted of 74 participants, of which 55 were female, 16 were male, and 3 indicated differently with a mean age of 23.64 years (SD = 5.53, Min = 17, Max = 51). The experimental condition with predominantly negative feedback (NNNPN) consisted of 72 participants, of which 51 were female, 19 were male, and 2 indicated differently with a mean age of 23.43 years (SD = 5.35, Min = 18, Max = 50). Six participants mentioned awareness of the feedback manipulation when they were asked, after the experiment, what they thought the goal of the experiment was. Analyses were performed with and without these participants to check whether their inclusion affected the results (Appendix B). As the results did not significantly deviate, it was chosen to report the results of the whole sample (N = 146).

Instruments and Measures

All materials were programmed and presented in the online Qualtrics Survey Software (https://www.qualtrics.com).

Problem-Solving Tasks

The problem-solving tasks used, were adopted from the translated version of the Cognitive Reflection Test (CRT) by Pennycook and Rand (2019), which combined the CRTs by Frederick (2005) and Thomson and Oppenheimer (2016). The problem-solving tasks require deliberative thought to override the intuitive answer (e.g., 'A bat and ball cost €1.10 in total. The bat costs €1.00 more than the ball. How much does the ball costs?' the intuitive answer is that the ball costs 10 cents. The correct answer is that the ball costs 5 cents, and thus the bat costs €1.05). Since these tasks are complex, it was hard for the participants to estimate whether their answers were correct or not. This was important for the credibility of the

feedback since they were manipulated regardless of their actual performance. Two problem-solving tasks were shown to the participants before the test cycle so that they were able to make self-efficacy estimates (which were then measured using a questionnaire). The test cycle followed in which the participants completed five tasks, where a four-option multiple-choice answer format was used (Appendix C). Participants' task performance could not be interpreted meaningfully as the participants received repeated manipulated feedback. Task performance was, therefore, not reported in the result section but can be viewed in Appendix D.

Feedback

Feedback consisted of a written message appearing on the screen after completing the problem-solving task, presented as "Your answer was correct." (positive feedback) or "Your answer was incorrect." (negative feedback). The feedback valence was once reversed so that the feedback would remain credible. The timing of this reversal was based on the hypotheses of Raaijmakers et al. (2017). It was expected that the effect of feedback valence might be pronounced when participants develop performance expectations. Raaijmakers et al. (2017) argued that with a valence reversal early in the experiment (e.g., on the second task), participants do not build a consistent expectancy of their performance as the feedback valence changes from the first to the second task, and the second to the third task. As self-efficacy expectations are developed through repeated success (Bandura, 1977), reversed feedback on the fourth task in a predominantly positive feedback valence (PPPNP) should have less impact than when the reversal feedback is provided earlier in the experiment, for example, on the second task (PNPPP). In other words, the effect of positive feedback valence might be stronger when the task reversal occurs later than earlier. Likewise, it was expected that the effect of a predominantly negative feedback valence might be stronger when the reversal occurs later since the positive feedback after repeated failure is more likely to be ascribed to

chance (NNNPN). Although Raaijmakers et al. (2017) found no significant difference in the effect of reversal timing, the reasoning to put the reversal on the fourth task was adopted.

Self-efficacy

Self-efficacy was measured using the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich et al., 1990). The MSLQ is a self-report questionnaire widely used in educational research and measures the construct self-efficacy using an eight-item scale. A 7-point Likert scale (1 = strongly disagree to 7 = strongly agree) is used with these items. The items were previously translated into Dutch by Agricola et al. (2020) and showed good reliability (α = .89). The questionnaire of Agricola et al. (2020) was adjusted to the context of this study (Appendix E). Hence, it measures the students' self-efficacy toward their ability to solve the problem-solving tasks (e.g., 'I'm confident I will do very well in the upcoming problem-solving tasks'). The adjusted questionnaire was validated via a Confirmatory Factor Analysis (CFA) (Appendix F). The 8-item questionnaire did not fit the data, with the model fit criteria: a non-significant goodness-of-fit test (χ2), CFI and TLI > .90, and RMSEA < .10 (Hooper et al., 2008). Inspection of factor loadings revealed that two items were deviant from the other items and were therefore removed. CFA confirmed the new one-factor model, consisting of 6-items, a good fit (χ2 = 16.03, p = .07, CFI = .99, TLI =.98, RMSEA = .07). Cronbach's Alpha was used to measure reliability, with a α = .93 the internal consistency was found excellent (George & Mallery, 2003). Consistent with Pintrich et al. (1990) were the sum scores of the items used to reflect students' self-efficacy. The range of possible scores was 6 to 42, and a higher sum score indicates a higher level of self-efficacy.

Affective Responses

The affective challenge and threat responses were measured using the self-report questionnaire by Scholl et al. (2018). This questionnaire consists of 12 items measured with a 7-point Likert scale (1 = strongly disagree to 7 = strongly agree). Six items assessed challenge

responses (e.g., 'I feel very stimulated), and six assessed threat responses (e.g., 'I feel a little threatened'). These items were adapted from the Stress Appraisal Measure (Peacock & Wong, 1990). Scholl et al. (2018) validated the questionnaire with factor analyses, which showed that challenge and threat items loaded on one factor, explaining 46.4% of the variance. After reversing the threat item scores, all items were averaged in a relative challenge index with a considered good reliability, α = .89.

The questionnaire from Scholl et al. (2018) was translated and adjusted to the context of this study. For example, the items were introduced as: 'As I approach the next cycle of problem-solving task' (Appendix G). A CFA one-factor model was performed for the adjusted questionnaire, which resulted in a poor fit. Inspection of factor loadings revealed that four items had low factor loadings (< .40) and were therefore removed. The exclusion of these items did not result in a good model fit. An Exploratory Factor Analysis (EFA) was used to identify the underlying structure of the questionnaire (Appendix H). The EFA showed that items loaded on three factors. After oblique rotation, it was checked which items belonged to which factor. During this process, another item was removed because of the lack of distinctiveness and low factor loadings. A final EFA validated the three-factor model consisting of 7 items ($\chi 2$ = .46, p = .93). The newly generated factors are labelled based on the corresponding item's content: f1: Challenge (2 items); f2: Threat, feelings of fear (which refers to the fear of not having enough resources; 2 items); f3: Threat, feelings of intimidation (which refers to an emotional state: 3 items). The reliability of the factors is considered acceptable (Challenge, α = .65) and good (Threat feelings of fear, α = .86 and Threat feelings of intimidation, α = .83).

By identifying the three-factor model, challenge and threat could not be measured on one scale. Three factors were separately used as the dependent variable to measure challenge and threat states. Sum scores of the items were used to reflect students' challenge (range 2 to

14), threat feelings of fear (range 2 to 14), threat feelings of intimidation (range 3 to 21), where a higher score indicates a higher level of the relevant variable.

Procedure

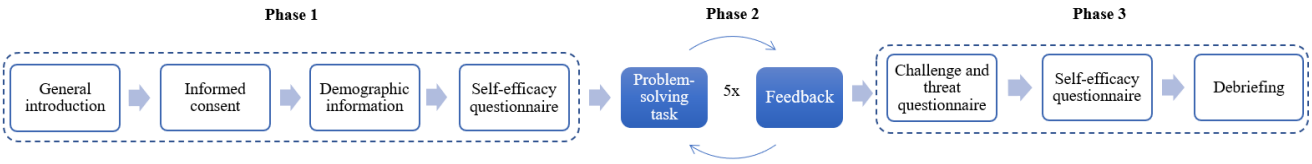
Participants received a login code to one of the two experimental conditions in Qualtrics. The experiment was divided into three phases (Figure 2: Appendix I). In the first phase, the participants received a general introduction to the experiment (Appendix J), where the design and intention of the experiment were explained. In addition, the participants read and signed the informed consent (Appendix K). Hereafter, participants filled in demographic information (gender and age) and read two problem-solving tasks, followed by a self-efficacy questionnaire to indicate their confidence in solving the problem-solving tasks in the test.

The second phase consisted of a test in which participants answered five problem-solving tasks, with each problem immediately followed by either negative or positive feedback (Appendix C). They were informed that they needed to solve each task mentally (e.g., not using paper and a pencil) within one minute. Qualtrics was programmed to warn the participants when they exceeded the allotted time.

In phase three, the participants answered two questionnaires: a self-efficacy questionnaire and a challenge and threat questionnaire. Whilst answering these questionnaires, the participants were under the assumption that they had to answer five more problem-solving tasks, which they did not. Participants had to believe this, as the questionnaires focused on their self-efficacy and affective states toward the anticipated problem-solving tasks. After answering these questionnaires, the participants were asked about the perceived purpose of the experiment, which was followed by a debriefing with the actual purpose. In addition, it was explained that the feedback was manipulated, and the correct answers to the problem-solving tasks were provided (Appendix L).

Figure 2

Visual Presentation of Procedure



Analyses

The hypotheses were tested using three hierarchical regressions for the dependent variables: challenge, threat feelings of fear, and threat feelings of intimidation. Three regression analyses were used as factor analyses revealed that challenge and threat could not be measured using one scale but with three. In the hierarchical regression, variables were added in separate steps to evaluate whether adding variables improves a model's ability to predict the criterion variable and to investigate a variable's moderating effect (Tabachnick & Fidell, 2013). Before each hierarchical regression, it was checked whether the data met the assumptions (Field, 2013). Additionally, due to non-normally distributed data, nonparametric tests were used to compare self-efficacy scores over time and between the feedback conditions. IBM SPSS Statistics v28 was used to analyse the data.

Results

Descriptive Statistics

Table 1 shows the descriptive statistics of both feedback conditions' pre- and post-test self-efficacy scores and affective responses (challenge, threat feelings of fear, and threat feelings of intimidation). Considering the range of possible self-efficacy scores (6 to 42), the pre-test self-efficacy scores were interpreted as relatively high for both feedback conditions (positive: M = 33.43, SD = 5.52; negative: M = 34.07, SD = 5.37), indicating that the participants started the experiment with confidence in their ability to solve the problem-solving tasks. A Mann-Whitney U test showed that these pre-test self-efficacy scores did not

differ significantly between the feedback conditions (U = 2479.00, p = .47). The participants from the conditions started the experiment with comparable self-efficacy feelings.

Table 1

Descriptive statistics

Feedback condition	Variable	*M*	*Mdn*	*SD*	Min	Max
Primarily positive (PPPNP)	SE pre-test	33.43	35	5.52	14	42
	SE post-test	32.49	32	5.45	21	42
	Challenge	9.57	9	2.23	4	14
	Threat fear	4.80	4	2.33	2	11
	Threat intimidation	6.73	6	3.19	3	16
Primarily negative (NNNPN)	SE pre-test	34.07	35	5.37	19	42
	SE post-test	23.21	23	8.50	6	42
	Challenge	7.44	7	2.92	2	14
	Threat fear	7.76	8	3.39	2	14
	Threat intimidation	9.97	10	4.22	3	17

Note. SE = self-efficacy. PPPNP (n = 74), NNNPN (n = 72). The range of possible scores was 6 to 42 for SE, 2 to 14 for challenge and threat fear, and 3 to 21 for threat intimidation.

Investigating the Effect of Feedback Valence and Self-Efficacy

Three hierarchical regressions were performed to investigate to what extent challenge and threat states could be predicted by feedback condition (Q1), pre-test self-efficacy (Q2), and interaction between feedback condition and pre-test self-efficacy (Q3). The regression analyses were reported separately for each dependent variable (challenge, threat feeling of fear, threat feelings of intimidation). The hierarchical regressions were performed in the same way. The independent variable feedback condition was entered as the predicting variable in the first stage utilising a dummy variable, with 0 = primarily positive feedback (PPPNP) and 1

= primarily negative feedback (NNNPN). In the second stage pre-test self-efficacy scores were entered as the prediction variable. The interaction (feedback condition * pre-test self-efficacy) was entered in the third stage to test self-efficacy's moderation effect.

Assumptions

Before conducting each hierarchical regression analysis, it was checked whether the data violated the following assumptions: multicollinearity, outliers, normality, homoscedasticity, linearity, and independence (Field, 2013). To avoid potentially problematic high multicollinearity with the interaction term, the predictor self-efficacy was standardised to Z-scores before the interaction between feedback condition and self-efficacy was added to the regression (Aiken et al., 1991). Analyses of collinearity statistics showed that this assumption was met, as VIF scores were well below 10 and tolerance scores above 0.2 (Field, 2019).

Mahalanobis distance was used to find possible multivariate outliers. For some participants, high values of distances were found. Only one of these showed a *p*-value of < 001, which could be considered an outlier. However, Cook's distance for this case was lower than 1 (*Di* = .007). Cook's distance indicates the overall influence of a respondent on the model, where values greater than 1 indicate influential respondents (Cook & Weisberg, 1982). Therefore, this case was not considered problematic and has not been removed from the dataset.

The assumption of normality was checked using Shapiro-Wilk analyses for both experimental conditions. A significance level of $p < .05$ indicates that the data significantly deviates from the normal distribution and violates the normality assumption. For regression 1 this assumption was met by the positive feedback condition ($W = .97$, $p = .10$) but violated by the negative feedback condition ($W = .95$, $p = .01$). This assumption was violated for both conditions when testing for regression 2 and 3. However, the Q-Q and P-P plots showed a

rather normal distribution of scores (Appendix M). Due to these plots and the sample size (N = 146), it was assumed that the scores were sufficiently normally distributed.

Furthermore, the assumption of homoscedasticity was checked using the scatterplots of standardised Residuals on the Y-axis and Regression Standardized Predicted Value on the X-axis. No obvious signs of funnelling were found, suggesting the assumption of homoscedasticity has been met for each regression (Appendix M). The scatterplot of standardised residuals against the predicted values was used, with the P-P plots, to confirm the assumption of linearity. Lastly, the Durban-Watson statistic showed that the assumption of independence of observations was met, as the obtained value was within the range $1.5 < d < 2.5$, which implies that there is no auto-correlation in the data (Glen, 2016).

Hierarchical Regression: Challenge States

A three-stage hierarchical regression was conducted with challenge as the dependent variable (Table 2). Model 1 provides information about the effect of feedback valence (positive or negative) on perceived challenge. As shown in table 2, feedback condition contributed significantly to the regression model, F (1, 144) = 24.52, $p < .001$, $R^2 = .15$, with a medium effect size (small effect < .02; medium effect < .13; .14 < large effect < .26; Cohen, 1988). In line with H1, participants who received positive feedback showed higher challenge scores (M = 9.57) than those who received negative feedback (M = 7.44; Table 1). This indicates that participants who received positive feedback perceived the upcoming problem-solving tasks more as a positive challenge than the participants in the negative feedback condition. Feedback condition accounted for 15% of the variance in challenge scores, which means that 85% of the variation in challenge scores cannot be explained by feedback conditions alone.

Model 2 shows the relation between pre-test self-efficacy scores and students' perceived challenge states, controlled for feedback condition. Introducing the self-efficacy

predictor explained an additional 5% of the variation in challenge states. The significant change in R^2 is considered a medium effect (Cohen, 1988), F (1, 143) = 17.26, $p < .001$. In line with H2, the positive relation indicated that the higher self-efficacy, the more challenge the participants experienced.

Finally, in Model 3 the interaction (feedback condition * pre-test self-efficacy) was added to the regression model to test the potential moderating effect of self-efficacy on the relationship between feedback condition and challenge states. Model 3 did not significantly contribute to explaining variance in challenge states, F (1, 142) = 11.42, $p < .001$, $\Delta R^2 = .00$. In contrast with H3, no significant moderating effect was found for the interaction between feedback condition and pre-test self-efficacy on challenge scores ($\beta = .01$, $p = .99$). In other words, the effect of feedback condition was not significantly affected by participants' levels of self-efficacy toward the problem-solving tasks. The strongest predictor of challenge was the feedback condition which recorded a higher beta value ($\beta = -.39$, $p < .001$), than self-efficacy ($\beta = .22$, $p = .004$).

Table 2

Hierarchical Regression for Variables predicting Challenge States

	R	***R²***	**Δ*R²***	**Δ*F***	***B***	**SE**	**β**	***t***
Model 1	.38	.15	.15	24.52**				
Feedback condition					-2.13	.43	-.38**	-4.95
Model 2	.44	.19	.05	8.68*				
Feedback condition					-2.20	.42	-.39**	-5.25
Self-efficacy					.62	.21	.22*	2.95
Model 3	.44	.19	.00	.00				
Feedback condition					-2.20	.42	-.39**	-5.23
Self-efficacy					.62	.29	.22*	2.11
Feedback condition * self-efficacy					.01	.42	.00	.01

Note. All values are rounded to two decimals.

*$p < .05$, ** $p < .001$;

Hierarchical Regression: Threat, Feelings of Fear

Table 3 shows the results of the hierarchical regression analysis with threat feelings of fear as the dependent variable. Again, as shown in model 1, feedback condition contributed significantly to the regression model, F (1, 144) = 38.13, $p < .001$, $R^2 = .21$, with a large effect size (Cohen, 1988). In line with H1, participants who received negative feedback showed higher perceived threat scores (M = 7.76, SD = 3.39) than those who received positive feedback (M = 4.80, SD = 2.33). This indicates that participants who received negative feedback experienced more fear toward the upcoming problem-solving tasks than the participants who received positive feedback.

Model 2 shows that self-efficacy explained an additional 3% of the variance scores of feelings of fear, F (1, 143) = 23.04, $p < .001$, which is considered a medium effect (Cohen, 1988). The negative relation indicated that participants with higher levels of self-efficacy in the pre-test perceived fewer feelings of fear towards the problem-solving tasks, which is in accordance with H2.

The interaction (feedback condition * pre-test self-efficacy) in model 3 did not contribute to explaining variance in feelings of fear states, F (1, 142) = 15.47, $p < .001$, $\Delta R^2 = .00$. In contrast to H3, the results showed no moderating effect of self-efficacy on the relationship between feedback condition and feelings of fear. Thus, the effect of feedback condition was not significantly affected by participants' levels of self-efficacy toward the problem-solving tasks. The strongest predictor of threat (feelings of fear) was the feedback condition which recorded a higher beta value ($\beta = .47$, $p < .001$), than self-efficacy ($\beta = -.19$, $p = .01$).

Table 3

Hierarchical Regression for Variables predicting Threat (Feelings of Fear)

	R	***R²***	**Δ*R²***	**Δ*F***	***B***	**SE**	**β**	***t***
Model 1	.46	.21	.21	38.13**				
Feedback condition					2.97	.48	.46**	6.18
Model 2	.49	.24	.03	6.50*				
Feedback condition					3.03	.47	.47**	6.43
Self-efficacy					-.60	.24	-.19*	-2.55
Model 3	.50	.25	.00	.50				
Feedback condition					3.04	.47	.47**	6.42
Self-efficacy					-.74	.33	-.24*	-2.32
Feedback condition * self-efficacy					.33	.48	.07	.70

Note. All values were rounded to two decimals.

*$p < .05$, ** $p < .001$

Hierarchical Regression: Threat, Feelings of Intimidation

Table 4 shows the results of the hierarchical regression with threat feelings of intimidation as the dependent variable. Model 1 shows, as with the previous regressions, that feedback condition contributed significantly to the regression model, F (1, 144) = 27.51, $p < .001$, $R^2 = .16$, with a large effect size (Cohen, 1988). In line with H1, the results show that participants who received negative feedback showed higher threat scores (M = 9.97, SD = 4.22) than those who received positive feedback (M = 6.73, SD =.19*;* Table 1). This means that participants in the negative feedback condition reported feeling more intimidated by the upcoming problem-solving tasks than participants who received positive feedback.

Adding the predictor self-efficacy (model 2) and the interaction (model 3) to the regression showed no significant improvement to model 1. The presumed negative relation between self-efficacy (controlled for feedback condition) and the perceived threat was not confirmed. Likewise, no significant moderation effect of self-efficacy on perceived threat was

found. This insinuates that H2 and H3 should be rejected, as the results suggest that self-efficacy does not directly affect feelings of intimidation, or moderate the effect between feedback condition and feelings of intimidation.

Table 4

Hierarchical Regression for Variables predicting Threat (Feelings of Intimidation)

	R	***R^2***	**ΔR^2**	**ΔF**	***B***	**SE**	**β**	***t***
Model 1	.40	.16	.16	27.51**				
Feedback condition					3.24	.62	.40**	5.25
Model 2	.42	.18	.01	2.49				
Feedback condition					3.3	.62	.41**	5.36
Self-efficacy					-.49	.31	-.12	-1.58
Model 3	.42	.18	.00	.06				
Feedback condition					3.30	.62	.41**	5.34
Self-efficacy					-.56	.43	-.14	-1.31
Feedback condition * Self-efficacy					.16	.62	.03	.25

Note. All values were rounded to two decimals.

*$p < .05$, ** $p < .001$.

Exploring the Change in Self-efficacy

In addition to the research questions, it was explored if feedback condition affected participants' self-efficacy beliefs towards the problem-solving tasks. Due to the non-normally distributed self-efficacy variables, nonparametric tests were used (Field, 2009). Change in self-efficacy over time (pre-test versus post-test) was analysed using Wilcoxon Signed-Rank tests, and the effect of feedback condition was analysed using Mann-Whitney U tests.

First, the effect of time was investigated independently of the feedback condition. A Wilcoxon Signed Rank test revealed a statistically significant negative change in self-efficacy scores over time, $z = -6.77$, $p < .001$, with large effect size $r = .56$ (Cohen, 1988; Rosenthal, 1991). Subsequent analyses showed that this main effect of time was purely driven by the

participants in the negative feedback condition. A statistically significant negative change in self-efficacy scores over time was found in the negative feedback condition, $z = -6.645$, $p < .001$, with a large effect size ($r = .79$), and no significant change was found in the positive feedback condition, $z = -1.626$, $p = .11$.

As mentioned in the descriptive statistics, no statistically significant difference in pre-test self-efficacy scores between the positive and negative feedback condition was found. A Mann-Whitney U test showed that the self-efficacy post-test scores were significantly lower for the negative feedback condition (*Mdn* = 23) than for the positive feedback condition (*Mdn* = 32), $U = 977.50$, $p < .001$, with a large effect $r = .55$. Participants who received negative feedback had relatively less self-efficacy in the post-test than those who received positive feedback.

Discussion

Providing students with feedback does not always improve their performance. Research suggests that how students interpret feedback and act upon those interpretations is critical for subsequent learning and performance (Poulos & Mahony, 2008). It seems plausible that students respond differently to (negative) feedback because they react differently to stress (Brown & Creaven, 2017), leading to differences in the effectiveness of the feedback. However, not much is known about the effects of feedback on students' affective responses. The BPSM offered a potential explanation for individual differences in stress responses to feedback. This study aimed to examine the effect of manipulated performance feedback (positive and negative valence) on students' affective responses and whether self-efficacy can explain the individual differences in the assumed relation between feedback valence and affective responses. Based on the BPSM (Blascovich, 2008), it was hypothesised that (H1) positive feedback would result in a higher relative challenge than negative feedback; (H2) the higher one's self-efficacy, the higher one's relative challenge

would be; (H3) the effect of negative feedback on the threat response would be stronger for those with relatively lower self-efficacy than for those with relatively higher self-efficacy. In addition, this study aimed to explore whether negative feedback decreases self-efficacy.

An online experiment was set up to research the hypotheses. Dutch university students participated and were assigned to one of the conditions: predominantly positive feedback (PPPNP) and predominantly negative feedback (NNNPN). The participants answered problem-solving tasks, with each task followed by either negative or positive feedback. Their challenge and threat states towards the tasks were measured after answering five tasks while they were under the impression that they had to perform more tasks. Participants' self-efficacy toward the problem-solving tasks was measured in a pre-test and post-test. Results showed an effect of feedback valence, mixed results regarding the direct effect of self-efficacy, and no moderating effect of self-efficacy.

Effect of Feedback Valence on Affective Responses

In line with H1, results showed that participants in the negative feedback condition perceived more threat states (both feelings of fear and feelings of intimidation) and fewer challenge states towards the problem-solving tasks than the participants in the positive feedback condition. These findings indicate that feedback valence can indeed affect learners' feelings of challenge and threat towards a learning task.

Literature on challenge and threat suggests that threat states hinder performance (Gildea et al., 2007; Moore et al., 2012), as these lead to counterproductive learning behaviours and reduced motivation (Belschak & Den Hartog, 2009). The present study showed that negative feedback has the potential to induce threat states, which could have substantial implications for the effectiveness of feedback. Even though providing negative performance feedback to students cannot be avoided, it might be relevant for future research

to investigate under what conditions negative performance feedback is less threatening to learners.

Relation Between Self-Efficacy and Affective Responses

Mixed results were found regarding the relationship between self-efficacy and participants' affective responses. In line with H2, the results revealed that pre-test self-efficacy towards problem-solving tasks positively related to challenge states, indicating that the higher one's self-efficacy, the higher the perceived challenge. Moreover, also in line with H2, a significant negative relationship was found between pre-test self-efficacy and feelings of fear, meaning that participants with lower self-efficacy perceived more feelings of fear towards the problem-solving tasks. Furthermore, no significant relation was found between pre-test self-efficacy levels and feelings of intimidation. These findings indicate that self-efficacy can affect learners' feelings of challenge and partially threat toward a learning task.

Substantively, the results indicate that self-efficacy is a negative predictor of an individual's fear of having insufficient resources. However, self-efficacy appears not to be a predictor of individuals feeling intimidated. The definition of self-efficacy (Bandura, 1997) coincides meaningfully with the construct of feelings of fear, which possibly explains the found negative relationship. High self-efficacy beliefs might have enabled participants to view negative feedback as a more positive challenge and less of a threat than participants with low self-efficacy. The link between self-efficacy and feeling of intimidation (as an emotional state) is less evident in the literature, which plausibly means that these variables are unrelated.

Absence of Moderating Effect of Self-Efficacy

As opposed to H3, the results did not provide evidence that the effect of feedback on the affective responses was influenced by the individual's self-efficacy. It was expected that the effect of feedback would be weaker for people with a higher degree of self-efficacy (Nease et al., 1999; Blascovisch et al., 2003). Based on the results, it should be concluded that

self-efficacy does not explain individual differences in responses to (negative) performance feedback. These results suggest that self-efficacy does not play a role in how students react emotionally to feedback. Therefore, assuming the relationship between challenge and threat and performance (Mendes et al., 2007; Seery et al., 2010), it also implies that self-efficacy possibly does not affect how much students learn after receiving negative feedback. However, based on this study, it is too early to conclude that the moderating effect does not exist. There are methodological objections (e.g. the challenge and threat instrument was not content valid), which indicates that the results of this study should be interpreted with caution. This will be elaborated on in the limitation section. Future research could further explicate the interrelation between these factors.

Change of Self-Efficacy

Explorative analyses revealed that self-efficacy significantly decreased for the participants who received repeatedly negative feedback. This finding is in line with existing literature implying that feedback influences future self-efficacy (Daniels & Larson, 2001; Kim & Lee, 2019). The decrease in self-efficacy can be linked to the BPSM theory of challenge and threat, as the perceived available resources may be decreased over time. In contrast, no significant change in self-efficacy was observed in the participants who received repetitive positive feedback. It should be noted that self-efficacy scores were relatively high in the pre-test; therefore, it is not surprising that no significant positive change was found for the participants who received positive feedback.

As self-efficacy facilitates learning, it should be encouraged. However, this study's findings suggested that a decrease in self-efficacy may have been induced by providing negative feedback. It might be relevant for future research to explore how feedback can be provided without it inducing a decrease in self-efficacy.

Limitations and Recommendations

Several limitations were identified in this study. The first limitation was the validity of the instruments measuring challenge and threat and self-efficacy. Factor analyses showed that the validity of the translated and adjusted questionnaire could not be generalised to the sample. This may be a consequence of the translation, leading to an ambiguity that led participants to interpret the questions differently. The challenge and threat questionnaire was constructed to measure relative challenge and threat on one scale. However, factor analyses revealed that challenge and threat needed to be measured via three different constructs (challenge, threat feelings of fear, and threat feelings of intimidation). Using three different constructs implies that no relative differences, for example, greater versus lesser challenge, could be discerned, providing a more nuanced view.

Furthermore, 5 of the 12 items were removed during factor analyses. Removing items may have diminished the instrument's content validity because not all aspects of challenge and threat were enclosed in the items. The three constructs are now measured, each with either two or three items, which is limited. Future studies should use more items to measure challenge and threat to increase content validity. Besides, it is recommended first to conduct a pilot study of the translated questionnaire so it can be improved before starting the data collection. Nevertheless, it might be that this questionnaire is unsuitable in Dutch, implying that it should only be used in English since it was proven reliable (Scholl et al., 2018). Regarding the self-efficacy questionnaire, two items were removed, which could have affected the instrument's content validity.

A second limitation could be that participants were not in a motivated performance situation while participating in the experiment. Seery (2013) denotes that challenge and threat responses occur when individuals engage in motivated performance situations and try to reach a self-relevant goal, causing them to be actively engaged in the task (Figure 1). Even though there was a significant difference in challenge and threat states between feedback conditions,

participants were possibly not invested in a self-relevant goal as achieving a good score was not paramount. A possible sign of lack of task engagement was the large spread in time of completing the experiment. This could signal that participants who took relatively long to complete the experiment were not task-engaged (e.g., took a break between answering the questionnaires or were simultaneously busy with other activities). Therefore, it was decided to remove the participants that took longer than 30 minutes to complete the experiment. Despite the attempt to remove the participants that allegedly were not task-engaged, it could still be a limitation of this study. If this research were to be conducted again, it could be beneficial to stimulate participants' task engagement. Task engagement could, for example, be enhanced by making the participants believe that time is an important factor in the experiment or by awarding a prize based on participants' performance.

A third limitation could be that participants realised the feedback manipulation. However, there was only an indication found for six participants (Appendix N). These participants did not believe the feedback, which could be due to the type of tasks in the experiment. The tasks in the experiment were taken from Pennycook and Rand's CRT (2019) and are known to provoke intuitive answers which are not correct. In the negative feedback condition, participants did not always accept the feedback as they might hold too firmly to the intuitive answer, while they occasionally gave the wrong answer. Ilgen et al. (1979) define feedback acceptance as the recipients' belief that the received feedback accurately represents their performance. Individuals may question the feedback, particularly negative feedback, as it does not match their efficacy beliefs (Ilgen et al., 1979).

Although analyses showed that excluding these six participants from the dataset had no deviant effect on the results, future research may benefit from using different types of tasks to reduce the chance of manipulated feedback awareness, for example, the so-called Weekday Problems (e.g. 'Suppose five days after the day before yesterday is Tuesday. What day of the

week is yesterday?'; Van Gog et al., 2012). These Weekday Problems are complicated tasks that do not evoke an intuitive answer and are made under time pressure. With these tasks, it would be hard for the participants to estimate whether a given answer is correct or not. It is expected that manipulated feedback on these tasks is more likely to be accepted by the participants.

In addition, it is recommended to add a control condition to the design of future research, where participants would not receive any feedback during the experiment. Adding a control condition to the design allows testing for causal effects. When testing with three conditions, a larger sample is needed to achieve a good power of .80. Hence, it was decided to create only two experimental conditions to make achieving a good power manageable during the limited time period of the thesis. It was important to achieve high power as it represents the chance that null hypotheses are rightly rejected (Faul et al., 2007).

Conclusion

To summarise, the present study provided more insight into the affective responses to feedback. This study revealed that feedback valence significantly affects participants' challenge and threat states and that self-efficacy relates to challenge and feelings of fear. The current study provided various practical implications and input for future research. Practitioners should be aware of how feedback valence influences challenge and threat states of their students. As challenge states are positively linked to performance (Gildea et al., 2007; Moore et al., 2012), it is useful to enhance these feelings. Future research could investigate how feedback could be less threatening to learners and what individual characteristics influence this effect. Additionally, future research could focus on how negative feedback promotes students' self-efficacy. Promoting self-efficacy in students seems beneficial as more self-efficacious students participate readily and persist longer when they encounter difficulties than less self-efficacious students (Schunk & Zimmerman, 2007). Considering the limitations

of this study, it might be relevant to examine whether self-efficacy indeed cannot explain the effect of some threat states, and if self-efficacy is indeed not a moderator between feedback valence and affective responses.

References

Adams, A. M., Wilson, H., Money, J., Palmer-Conn, S., & Fearn, J. (2020). Student engagement with feedback and attainment: the role of academic self-efficacy. *Assessment & Evaluation in Higher Education*, *45*(2), 317-329.

Agricola, B. T., Prins, F. J., & Sluijsmans, D. M. (2020). Impact of feedback request forms and verbal feedback on higher education students' feedback perception, self-efficacy, and motivation. *Assessment in Education: Principles, Policy & Practice*, *27*( ), 6-25. https://doi.org./10.1080/0969594X.2019.1688764

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. sage.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191.

Belschak, F. D., & Den Hartog, D. N. (2009). Consequences of positive and negative feedback: The impact on emotions and extra-role behaviors. *Applied Psychology*, *58*(2), 274-303.

Blascovich, J. (2008). Challenge and threat. In A. J. Elliot (Ed.), Handbook of approach and avoidance motivation (pp. 431–445). New York: Psychology Press.

Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. P. Forgas (Ed.), Feeling and thinking: The role of affect in social cognition (pp. 59–82). Paris: Cambridge University Press.

Blascovich, J., Mendes, W. B., Tomaka, J., Salomon, K., & Seery, M. (2003). The robust nature of the biopsychosocial model challenge and threat: A reply to Wright and

Kirby. Personality and Social Psychology Review, 7, 234-243.

Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In *Advances in experimental social psychology* (Vol. 28, pp. 1-51). Academic Press.

Brown, E. G., & Creaven, A. M. (2017). Performance feedback, self-esteem, and cardiovascular adaptation to recurring stressors. *Anxiety, Stress, & Coping*, *30*(3), 290 303.

Cohen, J. (1988). Set correlation and contingency tables. *Applied psychological measurement*, *12*(4), 425-434.

Cohen, J. (1988). Statistical Power Analysis for the Behavior Sciences, 2nd Ed New York. *NY: Academic Press*

Cook, R. D., & Weisberg, S. (1982). Criticism and influence analysis in regression. *Sociological methodology*, *13*, 313-361.

Daniels, J.A., & Larson, L. M. The impact of performance feedback on counselling self-efficacy and counselor anxiety. Couns Educ Superv. 2001;41(2):120–30.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage, 4th Ed

Forgas, J. P., Bower, G. H., & Moylan, S. J. (1990). Praise or blame? Affective influences on attributions for achievement. *Journal of Personality and Social Psychology*, *59*(4), 809.

Fotheringham, D. (2011). The role of expert judgement and feedback in sustainable assessment: A discussion paper. *Nurse Education Today*, *31*(8), e47-e50.

Frederick, S. (2005). Cognitive reflection and decision making. Journal of Economic Perspectives, 19, 25–42. https://doi.org/10.1257/089533005775196732

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon

Glen, S. (2016). Durbin Watson Test & Test Statistic. Retrieved at 01-06-2021, from: https://www.statisticshowto.com/durbin-watson-test-coefficient/

Gildea, K. M., Schneider, T. R., & Shebilske, W. L. (2007). Stress appraisals and training performance on a complex laboratory task. *Human factors*, *49*(4), 745-758. https://doi.org/10.1518/001872007X215818

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Hase, A., O'Brien, J., Moore, L. J., & Freeman, P. (2019). The relationship between challenge and threat states and performance: a systematic review. *Sport, Exercise, and Performance Psychology*, *8*(2), 123.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods*, *6*(1), pp53-60.

Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology*, *49*(3), 550-565.

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organisations. *Journal of applied psychology*, *64*(4), 349.

Jones, Marc V. and Meijen, Carla and McCarthy, Paul J. and Sheffield, David (2009) A theory of challenge and threat states in athletes. International Review of Sport and Exercise Psychology, 2 (2). pp. 161-180. ISSN 1750-9858

Kim, Eun Jung, and Kyeong Ryong Lee. "Effects of an examiner's positive and negative feedback on self-assessment of skill performance, emotional response, and self-efficacy in Korea: a quasi-experimental study." *BMC medical education* 19, no. 1 (2019): 1-7.

Martin, A. J., Kennett, R., Pearson, J., Mansour, M., Papworth, B., & Malmberg, L. E. (2021). Challenge and threat appraisals in high school science: Investigating the roles of psychological and physiological factors. *Educational Psychology, 41*(5), 618-639. https://doi.org/10.1080/01443410.2021.1887456

Mendes, W. B., Blascovich, J., Hunter, S. B., Lickel, B., & Jost, J. T. (2007). Threatened by the unexpected: physiological responses during social interactions with expectancy-violating partners. *Journal of personality and social psychology*, *92*(4), 698-716. https://doi.org/10.1037/0022-3514.92.4.698

Moore, L. J., Vine, S. J., Wilson, M. R., & Freeman, P. (2012). The effect of challenge and threat states on performance: An examination of potential mechanisms. *Psychophysiology*, *49*(10), 1417-1425. https://doi.org/10.1111/j.1469-8986.2012.01449.x

Nease, A. A., Mudgett, B. O., & Quiñones, M. A. (1999). Relationships among feedback sign, self-efficacy, and acceptance of performance feedback. *Journal of applied psychology*, *84*(5), 806.

Lazarus, R. S. (1991). *Emotion and adaptation*. Oxford University Press.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition, 188, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Peacock, E. J., & Wong, P. T. (1990). The stress appraisal measure (SAM): A multidimensional approach to cognitive appraisal. *Stress medicine*, *6*(3), 227-236.

Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. https://doi.org /10.1037/0022-0663.82.1.33

Poulos, A. and Mahony, M. J. (2008) Effectiveness of feedback: The students' perspective. *assessment and Evaluation in Higher Education*, 33(2), 143-54.

Putwain, D., Sander, P. & Larkin, D. (2013). Academic self-efficacy in study-related skills and behaviours: relations with leaning-related emotions and academic success. British Journal of Educational Psychology, 83, 633-650. doi:10.1111/l.2044-8279.2012.02084.x

Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., & Van Gog, T. (2017). Effects of performance feedback valence on perceptions of invested mental effort. *Learning and Instruction, 51*, 36-46.

Rosenthal, R. (1991). Applied Social Research Methods: Meta-analytic procedures for social research Thousand Oaks, CA: SAGE Publications Ltd doi: 10.4135/9781412984997

Rossato, C. J., Uphill, M. A., Swain, J., & Coleman, D. A. (2018). The development and preliminary validation of the Challenge and Threat in Sport (CAT-Sport) Scale. *International Journal of Sport and Exercise Psychology, 16*(2), 164–177. https://doi.org/10.1080/1612197X.2016.1182571

Scholl, A., de Wit, F., Ellemers, N., Fetterman, A. K., Sassenberg, K., & Scheepers, D.

(2018). The burden of power: Construing power as responsibility (rather than as opportunity) alters threat-challenge responses. *Personality and Social Psychology Bulletin, 44*(7), 1024-1038. https://doi.org/10.1177/0146167218757452

Schunk, D. H., & DiBenedetto, M. K. (2014). Academic self-efficacy. In M. J. Furlong, R. Gilman, & E. S. Huebner (Eds.), Handbook of positive psychology in schools (pp. 115–521). Elsevier.

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. Reading & Writing Quarterly, 23(1), 7–25.

Seery, M. D. (2011). Challenge or threat? Cardiovascular indexes of resilience and vulnerability to potential stress in humans. *Neuroscience & Biobehavioral Reviews, 35*(7), 1603-1610. https://doi.org/10.1016/j.neubiorev.2011.03.003

Seery, M. D. (2013). The biopsychosocial model of challenge and threat: Using the heart to measure the mind. *Social and Personality Psychology Compass, 7* (9), 637–653. https://doi.org.10.1111/spc3.12052

Seery, M. D., Weisbuch, M., Hetenyi, M. A., & Blascovich, J. (2010). Cardiovascular measures independently predict performance in a university course. *Psychophysiology*, *47*(3), 535-539.

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics, 6th (ed.) Boston, MA.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, *31*(3), 379-394.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the Cognitive Reflection Test. Judgment and Decision Making, 11, 99–113.

Tomaka, J., Blascovich, J., Kibler, J., & Ernst, J. M. (1997). Cognitive and physiological antecedents of threat of challenge appraisal. *Journal of Personality and Social Psychology*, *73*(1), 63–72. https://doi.org/10.1037/0022-3514.73.1.63

Uphill, M. A., Rossato, C., Swain, J., & O'Driscoll, J. M. (2019). Challenge and threat: A critical review of the literature and an alternative conceptualisation. Frontiers in Psychology, 10, 1255. https://doi.org/10.3389/fpsyg.2019.01255

Van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied cognitive psychology*, *26*(6), 833-839.

Winstone, N. E., Hepper, E. G., & Nash, R. A. (2021). Individual differences in self-reported use of assessment feedback: The mediating role of feedback beliefs. *Educational Psychology*, *41*(7), 844-862.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, *25*(1), 82-91.

Appendix

Appendix A. Power- analyses

Prior Data collection
F tests - ANOVA: Fixed effects, special, main effects and interactions
Analysis: A priori: Compute required sample size

Input:	Effect size f	= 0.2526456
	α err prob	= 0.05
	Power (1-β err prob)	= 0.80
	Numerator df	= 1
	Number of groups	= 4
Output:	Noncentrality parameter λ	= 7.9787249
	Critical F	= 3.9194646
	Denominator df	= 121
	Total sample size	= 125
	Actual power =	0.8001667

Post hoc
F tests - ANCOVA: Fixed effects, main effects and interactions
Analysis: Post hoc: Compute achieved power

Input:	Effect size f	= 0.2526456
	α err prob	= 0.05
	Total sample size	= 146
	Numerator df	= 1
	Number of groups	= 4
	Number of covariates	= 1
Output:	Noncentrality parameter λ	= 9.3191507
	Critical F	= 3.9082581
	Denominator df	= 141
	Power (1-β err prob) =	0.8581195

Power = .86

Appendix B. Hierarchical Multiple Regressions, excluding 6 participants (*N* = 140)

Table 5

Summary of Hierarchical Regression Analysis for Variables predicting Challenge states

	R	***R²***	**Δ*R²***	**Δ*F***	***B***	**SE**	**β**	***t***
Model 1	.40	.16	.16**	25.71**				
Feedback condition					-2.23	.44	-.40**	-5.01
Model 2	.46	.21	.05*	9.30*				
Feedback condition					-2.33	.43	-.41**	-5.42
Self-efficacy pre-test					.65	.21	.23*	3.05
Model 3	.46	.21	.00	.04				
Feedback condition					-2.33	.43	-.41**	-5.41
Self-efficacy pre-test					.62	.29	.22*	2.11
(Feedback condition * self-efficacy pre-test)					.09	.43	.02	.19

Note. Statistical significance *$p < .05$, ** $p < .001$ (1) Feedback condition; 1 = primarily negative feedback valence condition (NNNPN), 0 = primarily positive feedback condition (PPPNP); (2) Self-efficacy is measured prior to the experiment; (3) Self-efficacy was standardized to Z-scores before adding the interaction term to the regression.

Table 6

Summary of Hierarchical Regression Analysis for Variables predicting Threat (fear) states

	R	***R²***	**Δ*R²***	**Δ*F***	***B***	**SE**	**β**	***t***
Model 1	.51	.26	.26**	48.49**				
Feedback condition					3.32	.48	.51**	6.96
Model 2	.55	.30	.04*	7.37*				
Feedback condition					3.41	.47	.52**	7.30
Self-efficacy pre-test					-.63	.24	-.20*	-2.72
Model 3	.55	.30	.00	.36				
Feedback condition					3.41	.47	.52**	7.27
Self-efficacy pre-test					-.76	.32	-.24*	-2.40
Feedback condition * self-efficacy pre-test					.28	.47	.06	.60

Note. Statistical significance *$p < .05$, ** $p < .001$ (1) Feedback condition; 1 = primarily negative feedback valence condition (NNNPN), 0 = primarily positive feedback condition (PPPNP); (2) Self-efficacy is measured prior to the experiment, (3) was standardized to Z-scores before adding the interaction term to the regression.

Table 7

Summary of Hierarchical Regression Analysis for Variables predicting Threat (intimidation) states

	R	***R²***	**Δ*R²***	**Δ*F***	***B***	**SE**	**β**	***t***
Model 1	.44	.20	.20**	33.54**				
Feedback condition					3.57	.62	.44**	5.79
Model 2	.46	.21	.02	2.62				
Feedback condition					3.64	.62	.45**	5.93
Self-efficacy pre-test					-.50	.31	-.12	-1.62
Model 3	.46	.21	.00	.05				
Feedback condition					3.64	.62	.45**	5.90
Self-efficacy pre-test					-.56	.42	-.14	-1.34
Feedback condition * Self-efficacy pre-test					.14	.62	.03	.23

Note. Statistical significance *$p < .05$, ** $p < .001$ (1)Feedback condition; 1 = primarily negative feedback valence condition (NNNPN), 0 = primarily positive feedback condition (PPPNP); (2) Self-efficacy is measured prior to the experiment; (3) Self-efficacy was standardized to Z-scores before adding the interaction term to the regression.

Appendix C. Problem-Solving Tasks

Introductie voorbeeldvragen:

Vraag 1: Een boer had 15 schapen. Op 8 na gingen ze allemaal dood. Hoeveel zijn er nog over?

a. 7
b. 8
c. 9
d. 6

Vraag 2: Emily's vader heeft drie dochters. De eerste twee heten April en Mei. Hoe heet de derde dochter?

a. Juni
b. Emily
c. Dat kan je niet weten op basis van de gegeven informatie.
d. Anne

Testvragen

Vraag 1: Als 5 broodmachines er 5 minuten over doen om 5 broden te snijden, hoelang zou het dan voor 100 broodmachines duren om 100 broden te snijden?

a. 5 minuten
b. 100 minuten
c. 20 minuten
d. 500 minuten

Vraag 2: In een meer groeien waterlelies. De waterlelies verdubbelen zich elke dag. Als het 50 dagen duurt om het hele meer te bedekken, hoelang duurt het dan om het halve meer te bedekken?

a. 25 dagen
b. 12.5 dagen
c. 49 dagen
d. 37.5 dagen

Vraag 3: Je doet mee aan een hardloopwedstrijd. Als je de persoon op de tweede plek inhaalt, op welke plek sta je dan?

a. 1e
b. 2e
c. 3e
d. Dat kan je niet weten op basis van de gegeven informatie.

Vraag 4: Een knuppel en een bal kosten samen €1.10. De honkbalknuppel kost €1,- meer dan de bal. Hoeveel kost de bal?

a. 5 cent
b. 10 cent
c. 9 cent
d. 7 cent

Vraag 5: Hoeveel kubieke meter aarde zit er in een gat van 3 meter diep x 3 meter breed x 3 meter lang?

a. 27 m^3
b. 0 m^3
c. 9 m^3
d. 3 m^3

Appendix D. Participants performance on problem-solving tasks

Table 8

Overview of participants' (problem-solving) task-performance in the two conditions

	PFC				NFC			
	Correct		*Incorrect*		Correct		Incorrect	
	n	%	*n*	%	*n*	%	*N*	%
Task 1	41	28.1	105	71.9	48	32.9	98	67.1
Task 2	53	36.3	93	63.7	42	28.8	104	71.2
Task 3	59	40.4	87	59.6	62	42.5	84	57.5
Task 4	36	24.7	110	75.3	37	25.3	109	74.7
Task 5	14	9.6	132	90.4	15	10.3	131	89.7

Note. PFC= positive feedback conditions, NFC = Negative feedback condition

Appendix E. Self-Efficacy Questionnaire

Table 9

Self-efficacy questionnaire

Self-Efficacy Vragenlijst	*1 = helemaal niet mee eens, 7 = helemaal mee eens*						
In denk dat ik een goede score ga halen voor deze test.	1	2	3	4	5	6	7
Ik heb er wel vertrouwen in dat ik de problem-solving taken van deze test kan begrijpen.	1	2	3	4	5	6	7
Ik bezit de basis kwaliteiten die ik nodig heb om de problem-solving taken te beantwoorden.	1	2	3	4	5	6	7
Ik vertrouw erop dat ik ook de mogelijk ingewikkeldere problem-solving taken die in deze test aan bod komen kan begrijpen.	1	2	3	4	5	6	7
Ik ben vol vertrouwen dat ik de komende problem-solving taken heel goed ga maken.	1	2	3	4	5	6	7
Voor deze test haal ik gemakkelijk een voldoende, verwacht ik.	1	2	3	4	5	6	7
Ik weet zeker dat ik de vaardigheden die bij deze test horen kan beheersen.	1	2	3	4	5	6	7
Na het lezen van de test beschrijving en het maken van de voorbeeld problem-solving taken, weet ik zeker dat ik de test goed ga maken.	1	2	3	4	5	6	7

*De Items zijn overgenomen uit Agricola et al. (2020), en aangepast aan de context van de huidige studie. Antwoorden op de vragen werden gegeven op een 7-punts Likertschaal.

Appendix F. CFA Self-efficacy Questionnaire

Step 1: CFA one-factor model

Table 10

CFA one-factor model, standardised factor loadings

Item	Standardised factor loading
SE_1	0.768
SE_2	0.708
SE_3	**0.683**
SE_4	0.741
SE_5	0.867
SE_6	0.908
SE_7	0.818
SE_8	0.910

This model has not a good fit. Looking at standardised factor loadings: item 3 is relatively low therefore it was removed.

Step 2. CFA one-factor model excluding SE_3

Table 11

CFA one-factor model, standardised factor loadings

Item	Standardised factor loading
SE_1	0.756
SE_2	**0.686**
SE_4	0.729
SE_5	0.873
SE_6	0.909
SE_7	0.514
SE_8	0.923

This model has not a good fit. Looking at standardised factor loadings: items 2 is relatively low therefore it was removed.

Step 3: CFA one-factor model excluding SE_2

Table 12

CFA one-factor model, standardised factor loadings

Item	Standardised factor loading
SE_1	0.744
SE_4	0.717
SE_5	0.874
SE_6	0.911
SE_7	0.816
SE_8	0.929

Note. Good model fit ($\chi 2$ = 16.03, p = .07, CFI = .99, TLI =.98, RMSEA = .07) with α = .93

CFA one-factor model (without item 3 & item 2) → **Good Fit**

- Goodness of fit = 16.025, p = .066
- CFI = .990
- TLI = .983
- RMSEA = .073
- Explained variance = 0.7017355

Appendix G. Challenge and Threat Questionnaire

Table 13

Challenge and Threat questionnaire

Uitdaging en bedreiging vragenlijst	Focus	Likertschaal
Nu ik de volgende cyclus van problem-solving taken benader..		*1 = helemaal niet mee eens, 7 = helemaal mee eens*
… voel ik me een beetje bedreigd	Bedreiging	1 2 3 4 5 6 7
… ben ik bang dat ik het niet onder de knie heb om de taken goed te kunnen maken	Bedreiging	1 2 3 4 5 6 7
… ben ik bang dat ik niet aan mijn eigen of de verwachtingen of die van een ander kan voldoen	Bedreiging	1 2 3 4 5 6 7
… heb ik veel zin om aan de slag te gaan met de taken	Uitdaging	1 2 3 4 5 6 7
… voel ik me erg uitgedaagd	Uitdaging	1 2 3 4 5 6 7
… weet ik zeker dat ik het onder de knie heb om de taken goed te kunnen maken	Uitdaging	1 2 3 4 5 6 7
Hoe voel je je ten opzichte van het maken van de volgende cyclus problem-solving taken?		*1 = totaal niet, 7 = heel erg*
Overbelast	Bedreiging	1 2 3 4 5 6 7
Bedreigd	Bedreiging	1 2 3 4 5 6 7
Gestimuleerd	Uitdaging	1 2 3 4 5 6 7
Uitgedaagd (als in; een uitdaging die ik aankan)	Uitdaging	1 2 3 4 5 6 7
Geïntimideerd	Bedreiging	1 2 3 4 5 6 7
Aangemoedigd	Uitdaging	1 2 3 4 5 6 7

*De items zijn overgenomen uit Scholl et al. (2018), en aangepast aan de context van de huidige studie. Antwoorden op de vragen werden gegeven op een 7-punts Likertschaal.

Appendix H. Factor Analyses Challenge and Threat Questionnaire

Step 1. CFA, standardised factor loadings

Table 14

CFA one-factor model, standardised factor loading

Item	Standardised factor loading
CT_A1	0.701
CT_A2	0.782
CT_A3	0.722
CT_A4	0.563
CT_A5	**0.206**
CT_A6	0.717
CT_B1	0.630
CT_B2	0.718
CT_B3	**0.375**
CT_B4	0.437
CT_B5	0.765
CT_B6	**0.368**

Step 2. CFA, excluding items CT_A5, CT_B3 and CT_B6

Table 15

CFA one-factor model, standardised factor loadings

Item	Standardised factorloading
CT_A1	0.713
CT_A2	0.805
CT_A3	0.754
CT_A4	0.500
CT_A6	0.698
CT_B1	0.642
CT_B2	0.731
CT_B4	**0.354**
CT_B5	0.755

Step 3. CFA, excluding CT_B4

Table 16

One factor CFA, standardised factor loadings

Item	Standardised factor loading
CT_A1	0.718
CT_A2	0.809
CT_A3	0.763
CT_A4	0.478
CT_A6	0.684
CT_B1	0.643
CT_B2	0.734
CT_B5	0.774

No need for removing another item. However this one-factor model shows not to fit the data. Therefore EFA was conducted.

Stap 4 EFA

Table 17

Three-factor model with Oblique rotation after deleting a5, b3, b6, b4

	Factor1	Factor2	Factor3
CT_A1	0.376	**0.518**	-0.068
CT_A2	**0.877**	-0.053	0.113
CT_A3	**0.819**	0.087	-0.067
CT_A4	-0.064	0.045	**0.669**
CT_A6	0.296	0.022	**0.530**
CT_B1	0.480	0.286	0.312
CT_B2	-0.036	**0.976**	0.029
CT_B5	0.182	**0.389**	0.357

Note. $p = .277$

CT_B1 shows no distinctiveness, it has low factor loadings on all three factors. Therefore it was removed.

Stap 5 final EFA

Table 18

Three-factor model with Oblique rotation after deleting a5, b3, b6, b4, b1

	Factor1	Factor2	Factor3
CT_A1	0.350	**0.525**	-0.032
CT_A2	**0.845**	-0.047	0.143
CT_A3	**0.830**	0.088	-0.073
CT_A4	-0.094	0.070	**0.652**
CT_A6	0.212	0.021	**0.631**
CT_B2	-0.033	**0.962**	0.024
CT_B5	0.166	**0.417**	0.327

Note. Chi square statistics = .46, p = .927

Interpretatie challenge / threat

Item	**Nieuwe factor -Label**
Nu de volgende cyclus van taken nadert.. - … ben ik bang dat ik de taken niet voldoende onder de knie heb om ze goed te kunnen maken (A2)	Threat (gevoel van angst / bang zijn)
Nu de volgende cyclus van taken nadert.. - … ben ik bang dat ik niet aan mijn eigen verwachtingen of die van een ander kan voldoen (A3)	
Nu de volgende cyclus van taken nadert.. - … voel ik me een beetje bedreigd (A1)	Threat (gevoel van bereiding / geïntimideerd)
Hoe voel je je ten opzichte van het maken van de volgende cyclus taken? – Bedreigd (B2)	
Hoe voel je je ten opzichte van het maken van de volgende cyclus taken? – Geïntimideerd (B5)	
Nu de volgende cyclus van taken nadert.. - … heb ik veel zin om aan de slag te gaan met de taken (A4)	Challenge (positieve gevoelens van uitdaging)
Nu de volgende cyclus van taken nadert… - … weet ik zeker dat ik het onder de knie heb om de taken goed te kunnen maken (A6)	

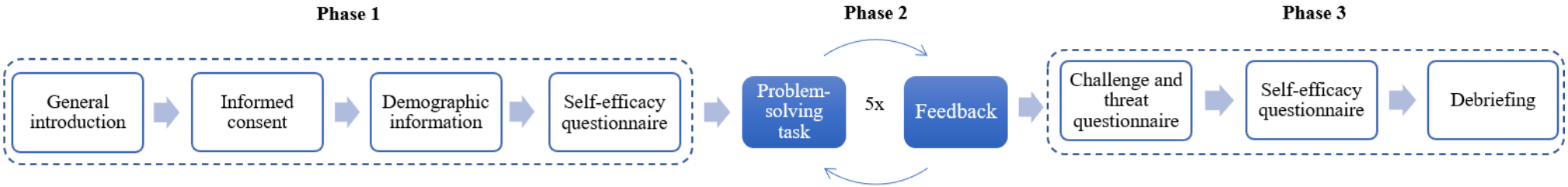
Table 19

Reliability per Factor

Factor	Cronbach's Alpha
Challenge	0.643
Threat Feelings of Fear	0.858
Threat Feelings of intimidation	0.824

Note. Cronbach's Alpha > .9 Excellent; > .8 Good; > .7 Acceptable, > .6

Appendix I. Visual presentation of the experiment process



Appendix J. Information letter

Beste Universitaire bachelor of master student,

Je staat op het punt om deel te nemen aan het onderzoek dat ik, Maaike Taheij, uitvoer in het kader van mijn afstudeerscriptie voor de Master Educational Sciences aan de Universiteit Utrecht. Dit onderzoek vindt geheel online plaats en zal ongeveer 20 minuten van je tijd vragen. Voor deelname ontvang je geen financiële compensatie. Daarbij benadruk ik dat deelname aan dit onderzoek vrijwillig is en dat je op ieder moment kunt stoppen zonder hiervoor een reden te geven. De Facultaire Ethische Toetsingscommissies (FETC) van de Faculteit Sociale Wetenschappen van de Universiteit Utrecht heeft het onderzoek getoetst en goedgekeurd en voldoet daarmee aan de ethische richtlijnen.

Voordat je beslist om mee te doen aan dit onderzoek wil ik je graag middels deze brief een beeld geven wat deelname inhoudt. Mocht je na het lezen van onderstaande informatie vragen hebben, neem dan gerust contact op via het emailadres dat onderaan de pagina vermeld staat.

Onderzoek

Door deel te nemen aan dit onderzoek draag je bij aan het inzichtelijk maken van de ervaringen die studenten hebben tijdens het maken van *problem-solving tasks.*

Je krijgt zo, voordat het onderzoek begint, een oefenfase waarin je twee *problem solving tasks* maakt die vergelijkbaar zijn met de taken die je gaat maken in het onderzoek. Je hebt voor het beantwoorden van iedere taak 1 minuut de tijd. Hierbij wil ik je vragen om de taken te maken **zonder** het gebruik van hulpmiddelen, zoals: internet, rekenmachine of pen en papier. Na de oefenvragen volgt een korte vragenlijst over hoe zelfverzekerd je bent ten aanzien van de taken die je gaat maken in het onderzoek.

Vervolgens start het onderzoek waarin je aan de slag gaat met het maken van 5 korte *problem solving tasks* waarbij je feedback ontvangt op je taakprestatie. Daarna krijg je twee korte vragenlijsten over je ervaringen waarna je deze nog eens 5 korte *problem solving tasks* gaat maken en een vragenlijst. Na het onderzoek krijg je meer gedetailleerde informatie over de doeleinden van het onderzoek.

Vertrouwelijkheid

Na het lezen van deze informatiebrief, zal je gevraagd worden om de geïnformeerde toestemmingsverklaring te tekenen. Daarna zal u gevraagd worden naar demografische informatie (leeftijd, geslacht). Deelname aan dit onderzoek is dus anoniem. Jouw identiteit kan niet worden achterhaald op basis van de informatie die door jou gegeven wordt tijdens participatie aan dit onderzoek. Daarnaast worden er geen IP-adressen opgeslagen. De ruwe data (onderzoeksgegevens) worden gedurende minstens tien jaar bewaard op de UU-server. Dit is in overeenstemming met de richtlijnen van de Vereniging van Nederlandse Universiteiten (VSNU).

Contactgegevens

Voor vragen over het onderzoek kun je terecht bij Valérie Kremer (v.n.a.kremer@students.uu.nl) en mocht je een klacht hebben, dan kan je een email sturen naar klachtenfuntionaris-fetcsocwet@uu.nl.

Alvast bedankt voor uw deelname.
vriendelijke groet, Maaike

Appendix K. Informed Consent

Hierbij verklaar ik de informatiebrief met betrekking tot het onderzoek van Maaike Taheij heb gelezen en dat ik vrijwillig deelneem aan het onderzoek.

Daarbij verklaar ik dat:

Ik begrijp dat dit onderzoek beoordeeld en goedgekeurd door de Facultaire ethische Beoordelingscommissie (FERB) van de Universiteit Utrecht.

Ik begrijp dat het onderzoek bedoeld is om een bijdrage te leveren aan het inzichtelijk maken van de studenten hun ervaringen tijdens het maken van problem-solving tasks.

Ik begrijp dat deelname vrijwillig is en dat ik op ieder moment kan beslissen om te stoppen zonder opgaaf van reden.

Ik ben voldoende geïnformeerd over het doel van het onderzoek en de manier waarop er met de verworven gegeven zal worden omgegaan.

Ik begrijp dat ik bij vragen contact op kan nemen met Valerie Kremer (v.n.a.kremer@students.uu.nl), en een klacht kan indienen via klachtenfuntionaris-fetcsocwet@uu.nl.

Ik onderteken de toestemmingsverklaring

- Ja
- Nee

Appendix L. Debriefing

Debriefing

Beste universitaire bachelor of master student,

Bedankt voor je deelname aan dit onderzoek. Er is **geen tweede ronde problem-solving tasks**. Nu je dit onderzoek hebt **afgerond** wil ik je graag meer vertellen over het doel en de opzet van dit experiment.

Eerder las je dat het doel van dit onderzoek was om inzicht te krijgen in de ervaringen van studenten tijdens het maken van problem-solving tasks. Meer specifiek ben ik geïnteresseerd in de effecten van de feedback valentie (positieve feedback vs. negatieve feedback) op de emotionele ervaring van studenten.

In dit onderzoek heb je **gemanipuleerde prestatiefeedback** ontvangen. Je bent op basis van toeval door de computer ingedeeld in een van de twee conditiegroepen. Groep 1 kreeg voornamelijk positieve feedback (‘je antwoord was juist’) en groep 2 kreeg voornamelijk negatieve feedback (‘je wantwoord was onjuist’). Dit betekent dat **de feedback die je kreeg niet overeenkomt met je werkelijke prestatie, de feedback was dus niet echt.**

Dit onderzoek is zo opgezet omdat ik wil onderzoeken wat de valentie van feedback doet met de emotionele ervaring van studenten ten opzichte van de taken. Ga je de taken als meer uitdagend of juist als bedreigend zien? Daarbij ben ik ook geïnteresseerd of de mate van zelfverzekerdheid (self-efficacy) die je had voor het maken van de taken een invloed heeft op de relatie tussen feedback en de emotionele ervaring (een uitdagend of bedreigd gevoel). Ook ga ik onderzoeken of je zelfverzekerdheid (self-efficacy) ten opzichte van de taken is veranderd (voor vs. na het onderzoek).

Nu dat je de werkelijke doeleinden van het onderzoek weet wil ik je vragen om toestemming te geven om de verworven data te gebruiken.

Ik geef hiervoor toestemming

- Ja
- Nee

Nogmaals bedankt voor je deelname,
Maaike Taheij

Heb je nog vragen of wil je extra informatie over dit onderzoek dan mag je natuurlijk contact opnemen met mij via email: m.taheij@students.uu.nl.

Bij klachten over dit onderzoek kan je contact opnemen via klachtenfuntionaris-fetcsocwet@uu.nl.

Ben je benieuwd naar de werkelijke antwoorden op de problem-solving tasks?

Oefentaken

Vraag 1: Een knuppel en een bal kosten samen €1.10. De honkbalknuppel kost €1,- meer dan de bal. Hoeveel kost de bal?

5 cent

Vraag 2: Lisa's vader heeft drie dochters. The eerste twee heten April en Mei. Hoe heet de derde dochter?

Zij heet Lisa

Onderzoekstaken

Vraag 1: Als 5 broodmachines er 5 minuten over doen om 5 broden te snijden, hoelang zou het dan voor 100 broodmachines duren om 100 broden te snijden?

5 minuten

Vraag 2: In een meer groeien waterlelies. De waterlelies verdubbelen zich elke dag. Als het 48 dagen duurt om het hele meer te bedekken, hoelang duurt het dan om het halve meer te bedekken?

47 dagen

Vraag 3: Je doet mee aan een hardloopwedstrijd. Als je de persoon op de tweede plek inhaalt, op welke plek sta je dan?

2e plek

Vraag 4: Een boer had 15 schapen. Op 8 na gingen ze allemaal dood. Hoeveel zijn er nog over?

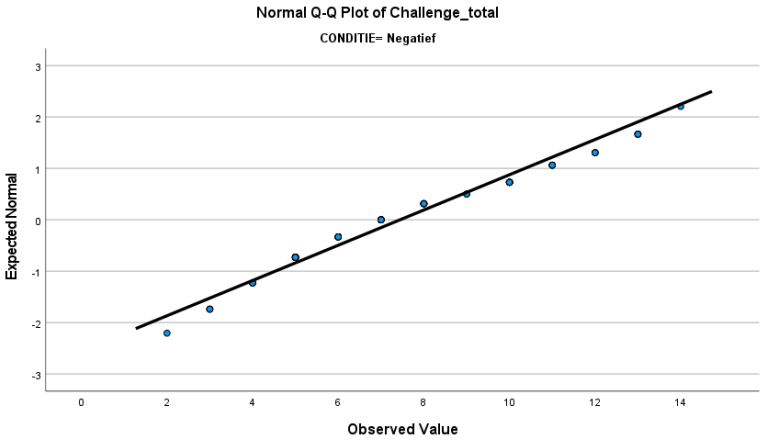
8 schapen

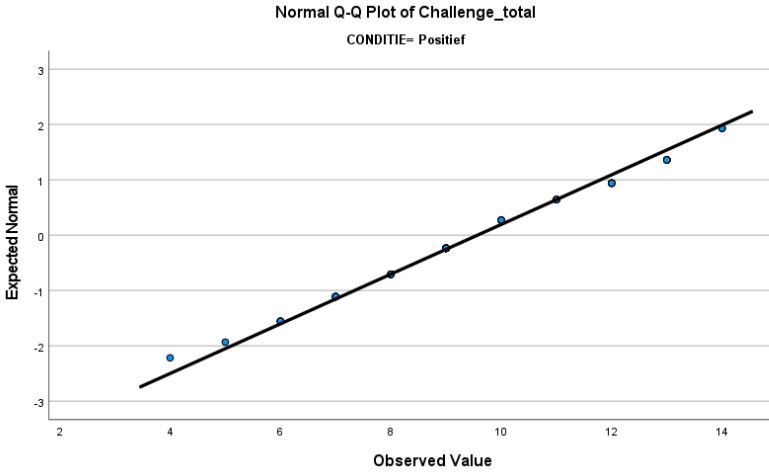
Vraag 5: Hoeveel kubieke meter aarde zit er in een gat van 1 meter diep x 1 meter breed x 1 meter lang?

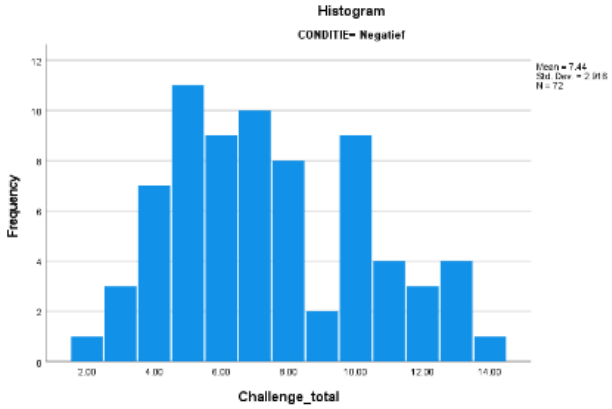
0 m3

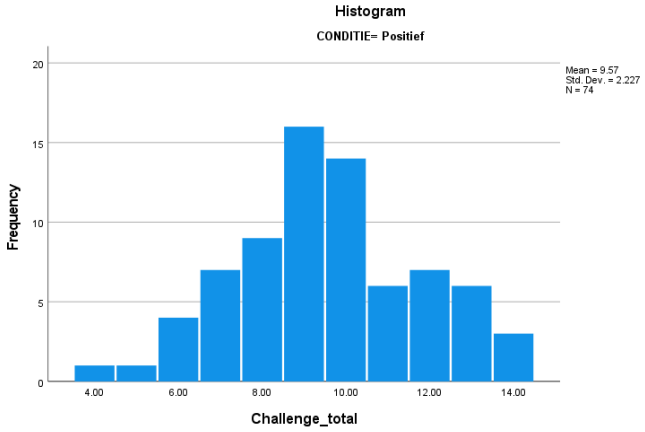
Appendix M. Assumption Plots

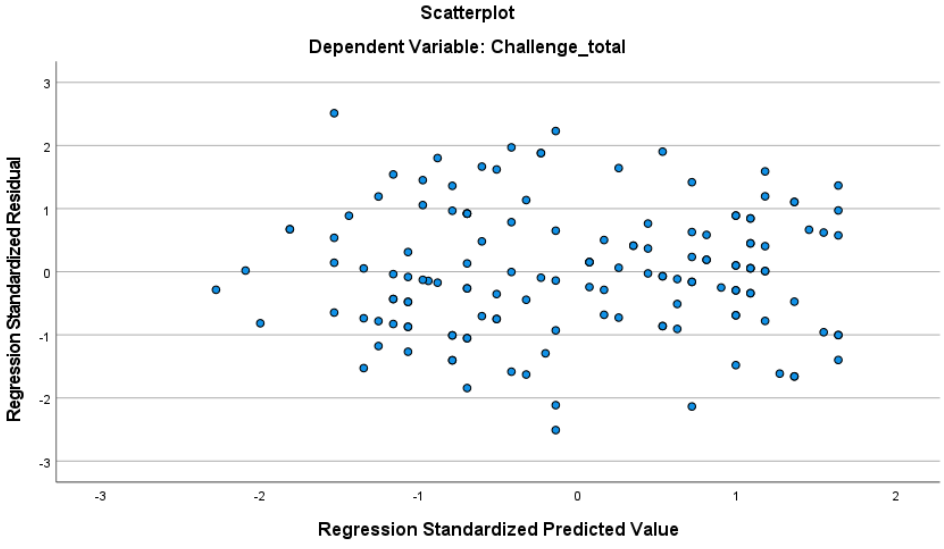
Regression 1. Challenge

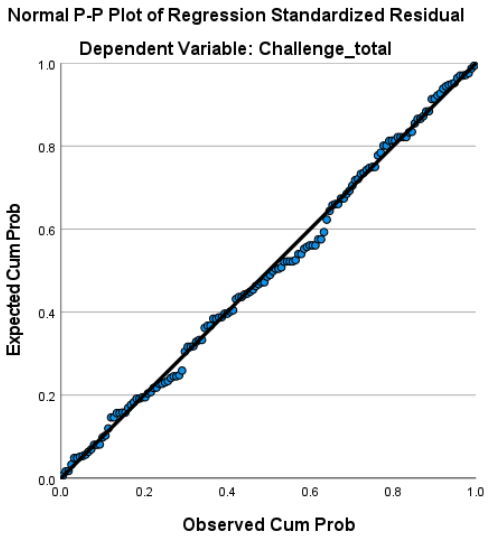




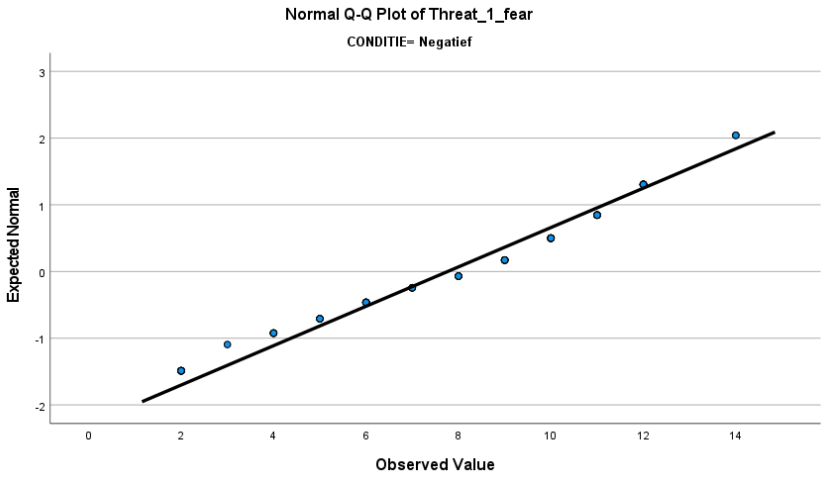


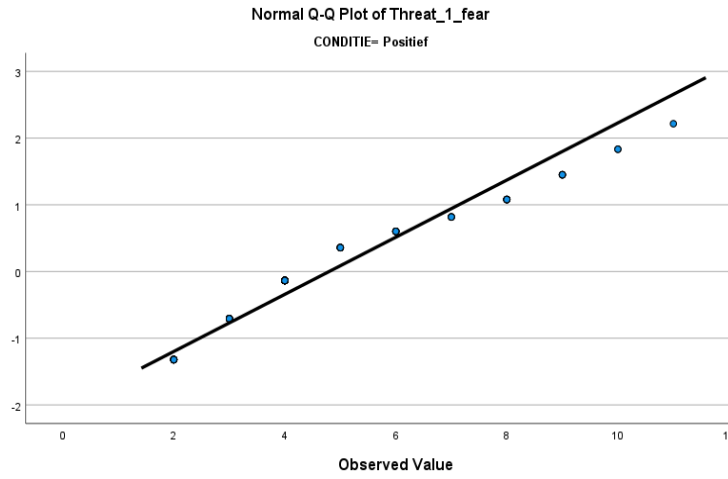


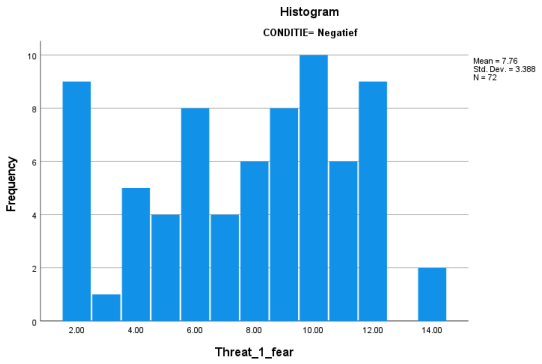


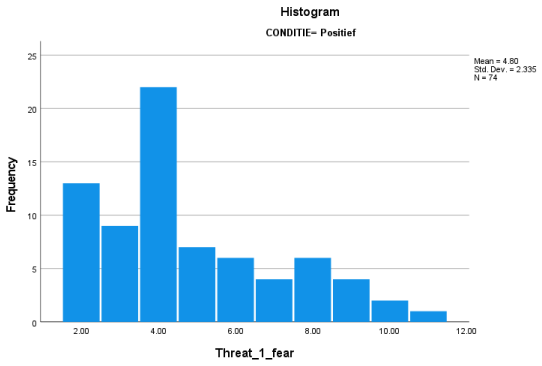


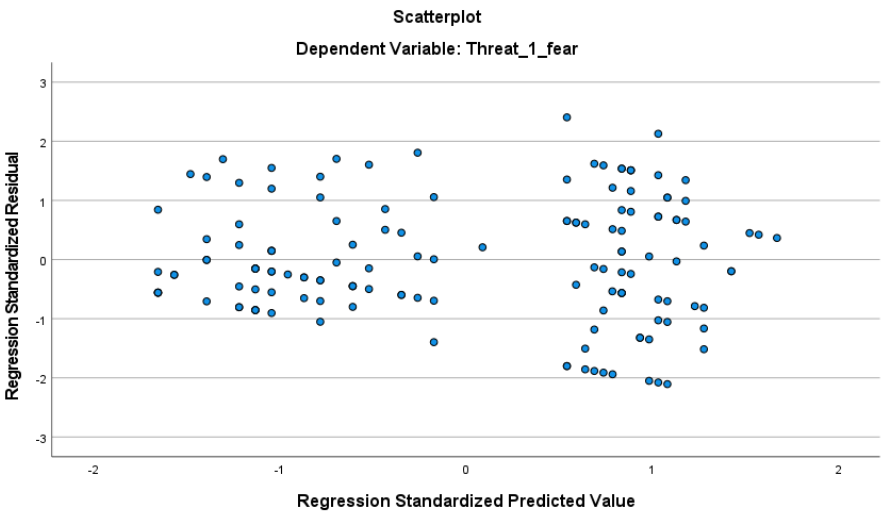
Regression 2. Threat feelings of Fear

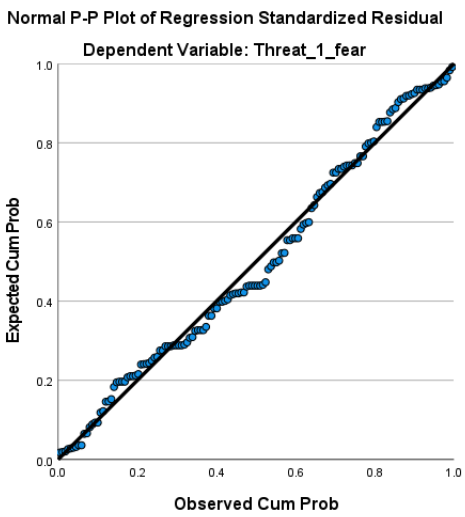




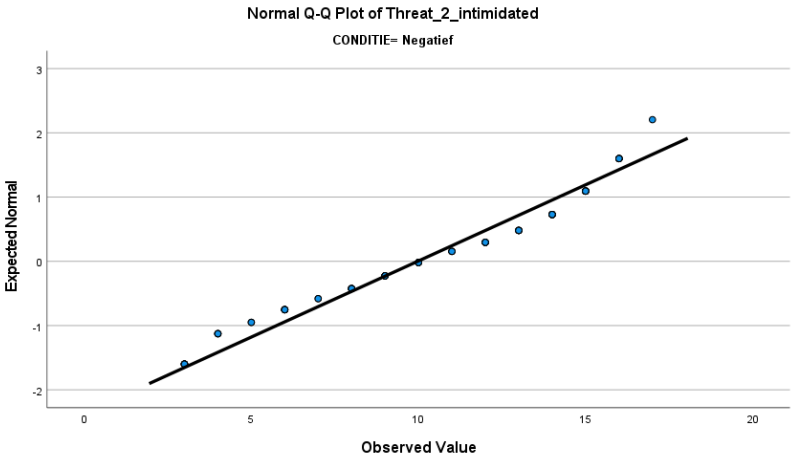


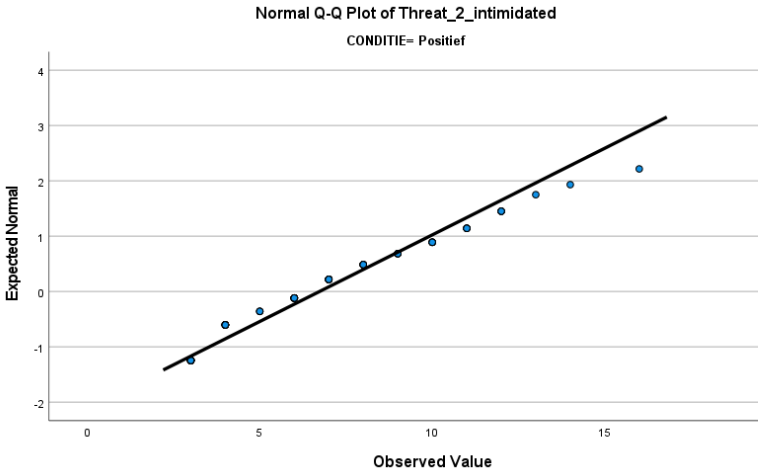


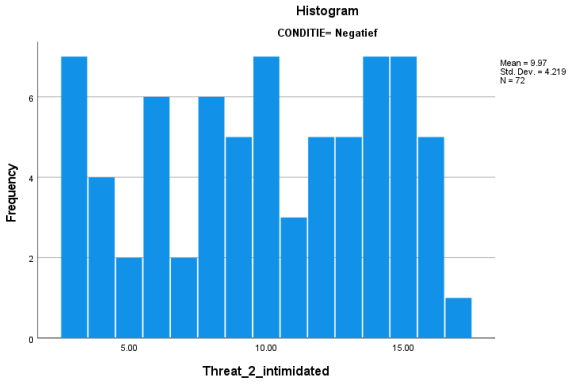


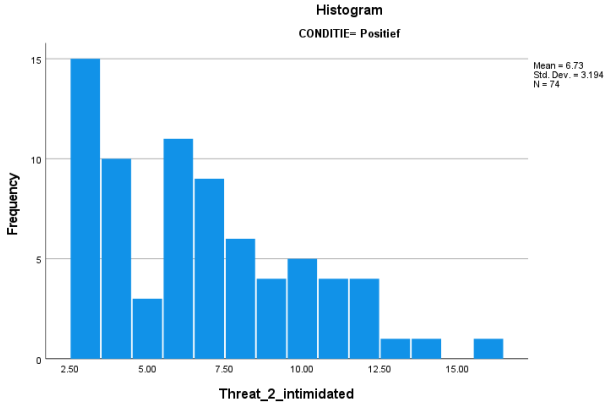


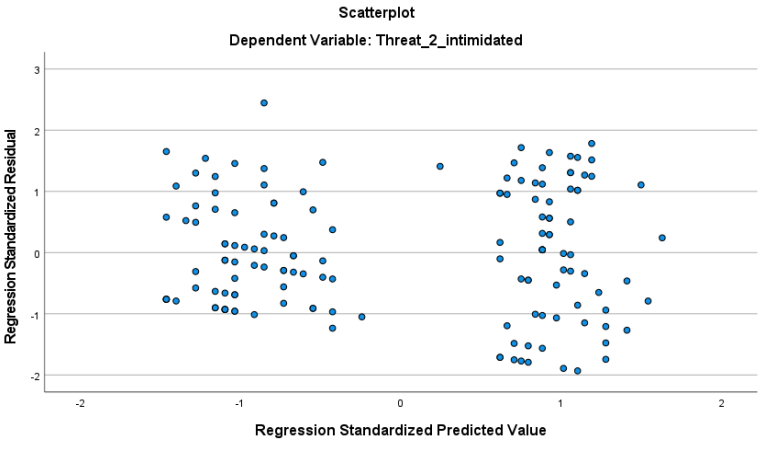
Regression 3. Threat feelings of Intimidation

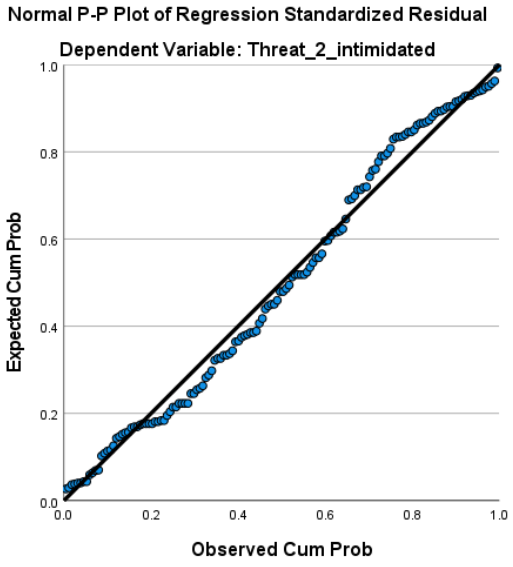












Appendix N. Participants that showed awareness of feedback manipulation

	Taak 1	**Taak 2**	**Taak 3**	**Taak 4 (RF)**	**Taak 5**	
Correcte waarde spss	**1**	**3**	**2**	**1**	**2**	
Participanten NFC						**Comment**
R_WiX2HYhG7ez4HUR	**1**	**1**	**2**	**1**	**2**	Het zelfvertrouwen van de proefpersoon werd naar beneden geprobeerd te halen door te laten zien dat het antwoord op de vraag fout was, ook al was dit soms niet het geval. Ik denk dat de onderzoeker wil weten wat het met mensen doet als ze negatief gestimuleerd worden.
R_RgICVR81rfLgnKN	**3**	**2**	**2**	**2**	**1**	Kijken of je zelfvetrouwen toe of afneemt als je feedback krijgt dat de vraag fout is (terwijl het antwoord misschien goed was)
R_3nx4Y7MS9lemsQt	**1**	**3**	**2**	**1**	**1**	Om te kijken wat de invloed is van als je een goed antwoord invult maar zegt dat het fout is
R_1NxoVVvnM7GGrk3	**1**	**3**	**2**	**1**	**4**	Ik denk dat het doel was om te onderzoeken wat de rol van foutieve feedback is op het zelfvertrouwen van de proefpersoon. Misschien ook gerelateerd aan theorieën rondom gaslighting
R_24Bn0TdOb40UBN3	**1**	**3**	**2**	**1**	**1**	Frustratie oproepen; hoewel de antwoorden die ik gaf correct waren, kreeg ik een melding dat het antwoord incorrect is. Dit is bedoeld om te doen twijfelen aan je zelfverzekerdheid, en te kijken of mensen na afloop van het onderzoek hier anders in staan.

R_1irrnKLghZjkmA0	**1**	**3**	**2**	**2**	**1**	Ik denk dat het doel was om te zien of het laten zien dat een antwoord fout is (al dan niet terwijl het antwoord wel goed was), invloed heeft op je zelfvertrouwen voor vragen die daarna gesteld worden. En dus in het breed, of het direct weergeven van resultaat een (negatieve) impact heeft op de motivatie/het zelfvertrouwen/zelfbeeld van een student (of ander persoon dat beoordeeld wordt).