ERASMUS UNIVERSITY ROTTERDAM

Learning not applied

A research into the effects of feedback and education on over-precision

Master Thesis

Machiel van de Kieft
Supervised by Chen Li
Student number: 358400

Rotterdam, July 30th 2016
Erasmus School of Economics
Specialization: Behavioral Economics

Over-precision is a robust phenome of overconfidence. When people display over-precision they show excessive precision in their beliefs. This paper examines the effect of hard and easy questions, the effect of feedback and the effect of education on over-precisions to test if education affects the effect of feedback on over-precision. To test this, an online questionnaire was sent to 109 respondents. They were randomly assigned to either the control or treatment group. The both groups received 10 interval questions and the treatment group received corrective feedback. The results revealed that both groups displayed over-precision. There was no significant difference between the groups with education and feedback. Also, no general effect of feedback or education was found. There was, however, evidence that respondents answered easy questions more accurate than moderate questions, which were, in turn, answered more accurate than hard questions. So the hard-easy effect was found for over-precision. According to these results, it is suggested that with hard tasks, an indication should be given to suggest the difficulty of the task so that people can take this into account when they solve the task. It also suggests that at least on short term, there is no reason to give feedback because it does not improve performance.

Table of Contents

1. Introduction

A lot of research has been done on cognitive biases (Kahneman & Tversky, 1972; Baron, 2007; Hilbert, 2012). This research shows that humans often deviate from expected behavior and that they do not act "rationally". *Rationality* can be defined as using (only) relevant information to asses options and maximize the potential outcome (Simon, 1957). One of the reasons why people violate rationality, are these biases. Because of the biases, a sub-optimal option becomes the most appealing option for a person (Shafir & LeBoeuf, 2002).

One of these biases is the *overconfidence effect*. This effect refers to a systematic judgement error when the correctness of a response to a knowledge based question is assessed (Pallier, et al., 2002). A part of the overconfidence bias is over-precision. *Over-precision* can be seen as an overestimation of one's ability to accurately determine an interval for a question (Healy & Moore, 2008).

A lot of research has already been done in the area of over-precision, and to what effect different factors affect over-precision. There has been research in the effect of feedback (Arkes, et. al., 1987), possible remedies (Haran & Moore, 2010), anchoring & conversational norms (Moore, Carter, & Yang, 2015) and more. Most of this research, however, looks at one aspect, or several aspects, but in isolation of each other.

In a study by Lichtenstein and Fischhoff they asked the question: "*Do those who know more know also more about how much they know?"* In this study they found that in general there was no effect. Only the graduates had a better resolution than the undergraduates (Lichtenstein & Fischhoff, 1977). In this paper a similar question will be researched:

- *Does the level of education affect the effect of feedback on over-precision?*

In other words, this paper will research if the effect of feedback on over-precision, as found by Arkes et al. is different for lowly educated people compared to people with a high level of education. People with a higher level of education (usually) learn easier compared to people with a lower level of education (Dirksen, 2011). Receiving feedback and applying this knowledge afterwards by altering one's behavior can be seen as a learning process. Therefore a difference between different education levels is expected.

In the next section, literature on over-precision and the effect of feedback will be reviewed. Bases on this literature, several hypotheses will be composed to look for an answer for the above mentioned questions. Then this question and the hypotheses will be tested with the results from an online survey. In this survey, the participants will be asked to answer 10 interval questions. The participants will be randomly assigned to one of two groups. One group will get feedback after each question and the other group will not.

2.1 Over-precision

Overconfidence exists and can be measured in three different ways (Healy & Moore, 2008). First of all there is over-placement. *Over-placement* takes place when a person performs a task together with other people. If the person believes he performed better than other people, without a founded reason, he displays over-placement (Croson, Croson, & Ren, 2008). Over-placement is an estimation mistake in performance compared to others. Secondly, people can display overestimation. When a person displays *overestimation*, he overestimates his ability to perform a certain task in a certain area. Overestimation is an estimation mistake in performance in a certain domain. At last, there is over-precision. *Over-precision* takes place if a person is exceedingly certain of his beliefs (Moore & Healy, 2007). Over-precision is the measurement in which this paper takes in regard. Out of these three measures, over-precision is the most robust phenomenon. Other factors, such as age, income, et cetera only have a small effect or no effect at all, on over-precision (McKenzie, Liersch, & Yaniv, 2008). On gender, the opinions are divided. A study by Bhandari and Deaves found a significant effect for gender on over-precision (Bhandari & Deaves, 2006).

According to a study (Alba & Hutchinson, 2000) there is a difference in what people shows that they know, and what they think they know. People think that they know more than what they actually know. This finding is the basis for overconfidence in general, and it is thus also one of the underlying principles for over-precision.

Over-precision is often tested with quantitative questions and confidence intervals (Alpert & Raffai, 1982). In most cases experiments take a confidence interval of 90%, and with this 90% confidence interval, people answer right in less than 50% of the cases. This is in general. With these experiments, the role of the hard-easy effect should be taken into account (Larrick, Burson, & Soll, 2006). No evidence of the hard-easy effect has been found for over-precision (Healy & Moore, 2008), but the hard-easy effect has been found for overconfidence in general (Justin, Winman, & Olsson, 2000). The *hard-easy effect* shows that judgement of the difficulty of tasks is often inaccurate. Hard tasks are perceived as easier than they actually are because of this underestimation of the difficulty of the tasks; people perform inadequately on hard tasks. The opposite is true for easy tasks. People overestimate the difficulty of the tasks and because of that they often over perform. The hard easy effect could possibly also be measured when testing for over-precision because if people have a fixed probability of being right (90%), but if they estimate the chance of being right too high with the hard questions, it means they will think they have a confidence interval of 90%, but it is actually a lot lower. This, in turn, affects the answers they give, and thus has an effect on over-precision. Therefore the following hypothesis will be tested:

Hypothesis 1:

H0: There is no measurable effect of over-precision

H1: There is a measurable effect of over-precision

Hypothesis 2:

H0: The difficulty of the questions does not affect the over-precision effect

H1: The over-precision effect increases with the difficulty of the questions

2.1.1 Underlying factors

There are several possible reasons for people to exhibit over precision (Moore, Carter, & Yang, 2015), namely: anchoring, conversational norms and naïve intuitive statistics and bias balance.

Anchoring:

The *anchoring bias* is the tendency for people to rely on first (or previous) encountered information when making a decision, and adjust their opinion from that particular piece of information. This information is (mostly) irrelevant for the current decision (Tversky & Kahneman, Judgment under Uncertainty: Heuristics and Biases, 1974). For over-precision this would mean that people first estimate a value. This value is the person's best guess for a stated number, which they think is the "correct" answer. After this estimate people deviate from it to create the interval, but not enough, which is why people show over-precision (Block & Harper, 1991). For example, if a questionnaire exists of multiple questions, and all of these questions have a quantitative measure, people may take the answer of the previous question, or a previous encountered number, and deviate from there. If they get a question about the population of china, and the previous question was about the length of the Nile, they will deviate from this number. In general, people do not deviate enough from their anchored first estimate and because of that they end up with an interval which is too narrow. This increases their chance of being over-precise.

Conversational Norms:

When people follow *conversational norms* they provide information with informative judgments (Grice, 1975). This often comes at the cost of the accuracy of the information.

When people do this, they try to be more informative. For example, if a person is asked the amount of inhabitants of the city of Rotterdam, he says 600.000. This information is inaccurate, because Rotterdam has 610386 inhabitants, but it is *informative* because the estimate is close. If he would have answer the amount of inhabitants lies between 100.000 and 1.000.000 inhabitants, this information would be *accurate*, but not informative. Because people try to be informative instead of accurate (Yanif & Foster, 1992), the chance of over-precision increases.

Naïve Intuitive Statistics:

When using statistics, people tend to be naïve in using them. They are considered naïve because people often use small samples as a starting point and they believe that these small samples are representative for the situation they face when making a choice, when often they are not representative. They ignore the error rate and the population dispersion is underestimated (Tversky & Kahneman, Belief in the Law of Small Numbers, 1971). This has a chance of creating over-precision, but there is also a possibility that the small sample size leads to an accurate estimated sample.

Bias Balance:

Another explanation for over-precision is a way for the brain to make future rewards more attractive (Frederick, Lowestein, & O'Donoghue, 2002). In general, our brain is biased towards rewards now over future rewards. By being over-precise people mostly have a smaller amount of required time in mind than is actually needed to complete a task. Therefore, the expected effort of this task is lower than the actual effort required. This makes the payoff for a future reward, compared to the effort required, bigger. The future reward seems more attractive than it actually is, which increases the chance of choosing a future reward over a reward now.

Misunderstanding of confidence interval (Fieller, 1954) can also be a reason why over-precision is measured. This, however, is not an underlying reason for over-precision, but a flaw in experimental design. In this case, it is possible that over-precision is measured, even though it is not there, because people do not understand the value of the percentages given in a confidence interval

The effect for overconfidence, and thus over-precision, was found to be a lot stronger for hypothetical situations compared to normal situations. This suggests that the overconfidence found by people is due to the context, and therefore a framing effect (Armor & Sackett, 2006).

Over-precision is, however, often measured and encountered in forecasting situations, which are hypothetical.

2.1.2 Consequences of over-precision

Because of over-precision, people tend to show stronger over-precision with probabilistic events (Gigerenzer, Hoffrage, & Kleinbolting, 1991). This means that people have a view of the future in which only a small part of the probable outcomes is considered. This can lead to flawed choices. For example, if people need to determine the strength of their walls to protect them against strong winds every 10 years. If they do this with a 98 confidence interval, but they are only correct 7 out of ten times, this means their homes are getting destroyed because of the role of over-precision during the decision stage. With forecasting, people also tend to be too sure of their own predictions. Therefore, they will not seek out help/advice and they will spend no, or too little, time planning for contingencies (Aukutsionek & Belianin, 2001).

People are not only over-precise when experiencing uncertainty. They also show over-precision with perceptual task. In a study (Mamassian, 2008), it was shown that people are over precise when they have to hold an object and estimate the weight with a confidence interval, or when they have to state the temperature.

Over-precision also has an indirect effect on other people. If people are over-precise in a group discussion, these people are seen as competent. The reason they appear as competent is that people believe that others know their own knowledge, as long as there is no reason to assume otherwise (Anderson, Brion, Moore, & Kennedy, 2012).

2.1.3 Summary

Over-precision takes place if a person is exceedingly certain of his beliefs. People think that they know more than what they actually know. Over-precision is often tested with quantitative questions and confidence intervals. Because people misinterpret their confidence level, and the chance of this is higher for hard questions than for easy questions, it is possible that over-precision is measured because of the hard-easy effect. Underlying factors for over-precision are the following: *Anchoring:* people rely on first (or previous) encountered information when making a decision, and adjust their opinion from that particular piece of information. *Conversational Norms:* people prefer to give informative information over giving accurate information. *Naïve Intuitive Statistics:* people often use small samples which are considered representative for the situation, even though they are not. *Bias Balance:* future rewards are more attractive because of lower expected effort. It is also possible to measure over-precision because people misinterpret confidence intervals or because the options are hypothetical. The consequences of over-precision are that only a small part of the probable outcomes is considered for future events. Also for perceptual tasks, people display over-precision, which means they (often) do not accurately estimate/measure quantities. In groups, over-precision makes people seem competent, because their information, even though not accurate, can be seen as more informative.

2.2 Feedback

According to John Hattie (Hattie & Timperley, The power of feedback, 2007), *feedback* can be seen as: *"information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding."* In other words, *feedback* is a process where a person participates in a task, and after the task information is given about his performance, either in the form confirming the person's performance (positive feedback) or in contradicting his performance and telling what the right answer is or where he can improve (negative feedback). *Feedback* is information, with which the receiver can choose (either conscious or unconscious) to continue or alter his behavior according to the information received (Butler & Winne, 1995). In this paper corrective feedback, also known as knowledge of results will be used. *Corrective feedback* is feedback which is directly performance related and states if the results of the performed task where either correct or incorrect.

It is important that the source of the feedback is recognized by the person as authoritative and knowledgeable source, because otherwise the person will not take the received feedback information into consideration and he keeps relying on the correctness of their own performance/answer.

It is argued that feedback has a greater effect over time compared to the instant effect (Leunis, 2013). The reason that feedback has a greater effect over time is that altering behavior as a result of feedback is a form of learning. The learning curve usually shows an increased output over time. A possible reason for this is that a certain aspect becomes more dominant in a person's mind after hearing it more than once (Adler & Clarck, 1991).

According to a study by Erev and Barron in 2005, the maximum payoff does not increase for a person after they receive feedback when they face a decision from experience. Decisions from experience are here interpreted as the process where subjects are confronted with a problem, and from there they have to deduct the underlying characteristics with the feedback of their earlier choices.

2.2.1 Underlying factors

Feedback works because it applies to a person's intrinsic motivation (Vallerand & Reid, 1988). The information that a person receives (feedback), gives him the possibility to re-evaluate his efforts and options in a decision. Feedback reduces the gap between current knowledge of the person and the factual knowledge (the exact answer to a task) (Hattie & Timperley, The power of feedback, 2007).

2.2.2 The feedback effect model

Kluger and Denisi developed a Feedback system where feedback could be given (as can be seen in figure 1) in two forms (Kluger & DeNisi, 1998). When a person completes a task he can either get positive or negative feedback about his performance.

When he gets positive feedback, the people will ask himself the question if there is a possibility to gain other self-goals (to gain self-realization). If this is possible, he will raise the standard for the task and therefore, the person will increase his effort to attain these goals. If this is not the case, he will reduce the effort in a similar next task.

When he gets negative feedback, the person will increase effort in the next task. After this the person will re-evaluate if the extra effort improves performance. If this is the case, he will maintain the increased effort. If this is not directly the case, the person will evaluate if he believes there is a reasonable change to improve the performance he will shift his attention to a learning process instead of expecting direct results. If this is not the case, the person will reduce the effort back to normal, or sometimes even less effort is given.

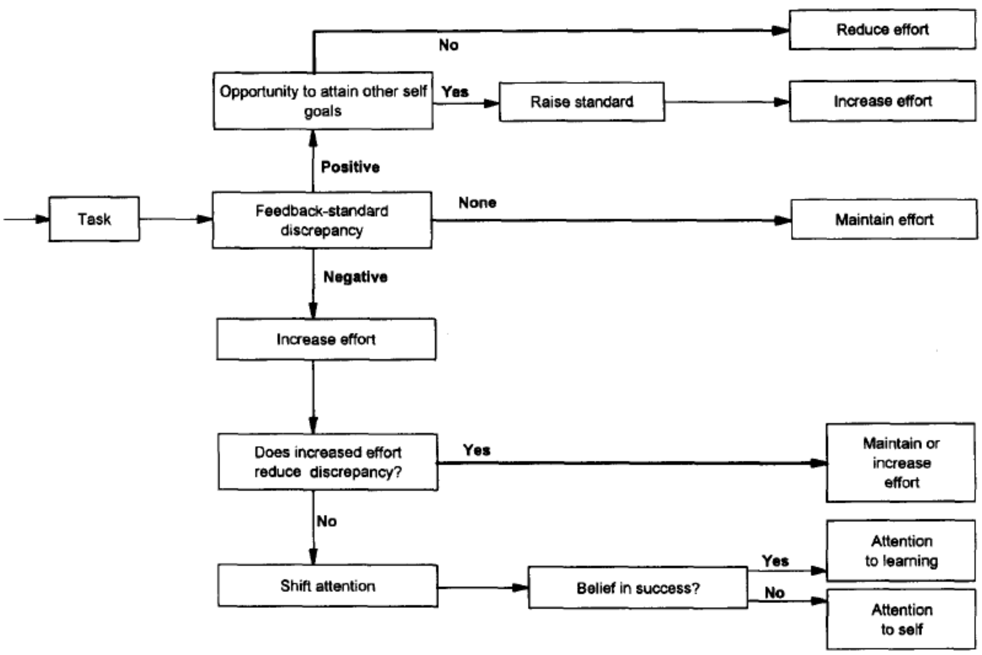


Figure 1 – Kluger & DeNisi 'Feedback effect model'

In the above described model, there is however one critical factor missing. This is the factor that people can reject the feedback and decide not to take the received information into consideration (Kulhavy, 1977). Therefore the following hypothesis will be tested:

Hypothesis 3:

- *H0: There is no measurable effect from feedback on over-precision*
- *H1: Feedback alters the level of over-precision*

2.2.3 Types of feedback

There are several ways to communicate corrective feedback (Tedick & Gortari, 1998). In this paper, explicit correction will be used. *Explicit correction* is a clear indication if the performance/answer was either correct or incorrect and the correct answer is provided.

In a summary of studies done in the area of feedback (Hattie, Influences on Student Learning, 1999), it was shown that feedback has an effect on a person's performance. There are, however, lots of different ways to give feedback, which give different effects. These studies, for example, found that the most effective channel to communicate this feedback is either via audio, via video or via a computer. There are also a lot of studies which indicate that feedback interventions have no influence on a person's performance (Brehmer, 1980). There is still a lot of controversy in the field of feedback and its effects.

When using feedback to improve a person's performance, rewards were negatively correlated with the effect of feedback. The reason for this is that tangible rewards destroy intrinsic motivation (Cameron, Banko, & Pierce, 2001) and intrinsic motivation is one of the underlying reasons why feedback has an effect on performance.

2.2.4 Summary

Feedback is a process where information is given about a person's performance in a task. This can be positive (acknowledging performance), or negative (correcting performance). Because of intrinsic motivation a person re-evaluates his efforts and options in a decision. With positive feedback, people will reduce effort in a similar next task unless it is possible to gain other self-goals. With negative feedback, people will increase effort in the next task. After this people will re-evaluate their performance maintain the increased effort when improved, switch to learning if improvement is possible, or attain original effort no improvement is possible. There are a lot of channel to communicate feedback such as audio, video, computer, etc. each has its own advantages and disadvantages. Rewards were negatively correlated with feedback because they destroy intrinsic motivation. *Explicit correction* is a clear indication if the performance/answer was either correct or incorrect and the correct answer is provided, and this type of feedback will be used in the experiment.

2.3 Feedback, Over-precision and Education

When over-precision is measured, this can be combined with feedback. This is usually done by having a group with interval questions who do not receive any form of response after a question is completed, and a group who get an indication about the accuracy of their answers after each question (Arkes, et. al, 1987). A research done by Mannes and Moore (Mannes & Moore, 2013) showed that when feedback was given, people became less over-precise when they had given an inaccurate interval and the received feedback about their mistake. So feedback can help to eliminate over-precision. This is only the case if the person provides an inaccurate interval. If the person provides an accurate interval, his over-precision will increase instead of decrease (Haran, Tenney, & Moore, 2015).

A study by Colmant and Pulford (Pulford & Colmant, 1997) found that there is a difference in de effect of feedback for the related to the difficulty of the questions. For easy questions, no significant effect was found of feedback on over-precision. For moderately difficult questions, also no significant effect was found of feedback on over-precision. For hard questions, a reduction in over-precision was found after giving feedback. So giving feedback only reduces over-precision in hard task where the person has a hard time estimating the answer.

According to Dirksen, people with a higher education are in general more able to learn than people with a lower education. Also people with a higher education have a broader knowledge. This can lead to an increased effect for feedback on over-precision because receiving feedback and taking the feedback into account in a future decision is a form of learning. This is mostly the case when people receive negative feedback. When they are confronted with a similar task in the future, people can alter their behavior because they remember the negative feedback and decide to alter their decision because of the feedback. The broader knowledge that higher educated people usually have increases the change of them knowing the answer, or having a better base to estimate the answer.

Hypothesis 4:

- *H0: The level of education does not affect the over-precision effect*
- *H1: The level of education affects the over-precision effect*

Hypothesis 5:

- *H0: The level of education does not affect the effect of feedback on over-precision*
- *H1: The level of education affects the effect of feedback on over-precision*

A problem in the real world is that most of the time, there is no direct feedback. Most of the time, decision makers receive feedback or the result of their actions too late and the do not directly link the results to their past decision and therefore do not alter their behavior (Pfeffer, 1992).

3 Experiment

3.1 Subject pool

In the survey 67 men and 42 women participated. The participants were people with a Dutch nationality. The mean of the age of the participants is 24.23 year. The survey was conducted online at the participants' place. Therefore this can be seen as a *framed field experiment* (Lusk, Pruitt, & Norwood, 2006). The subjects are real people (not only students), the experiment was not conducted in a lab or isolated environment and the participants were aware that the participated in the experiment. No (monetary) reward was given for participating in this experiment. Almost all of the participants were approached via Social media. A few of the participants were asked to participate by mail.

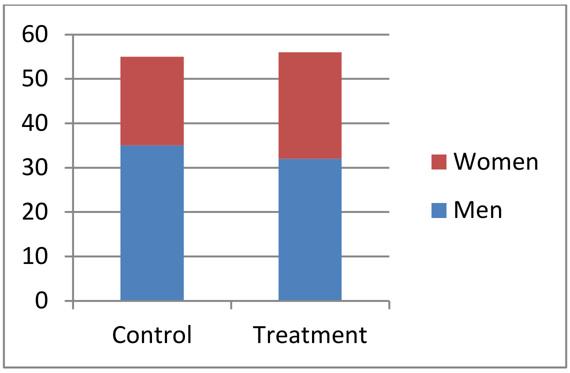


Figure 2 - Subjects

3.2 Design

As soon as the participants started the survey, they were given a short explanation of how to answer interval questions. This was followed by a training question, which had to be answered correctly before the participants could start the actual questions. The questions were questions with a numerical answer, which required the participant to give a lower and an upper bound, such that this represented a 90% confidence interval. The questions were categorized in three difficulty levels: easy, intermediate and hard. The questions with an easy difficult level were questions such as: "*How many legs does a mosquito have", where the* answer of the question was either known or relatively easy to estimate for most participants. The questions with a moderate difficulty were questions such as: "*How many meters long is the Rhine"*, or "*How many keys does a piano have"*, where the answer of the question was not known for (most of) the participants, but where they should be able to form an estimate by using reference points to deduct the answer from there. For example, a piano has seven octaves and an octave has 12 keys, thus coming on 84 keys, which is pretty close to the answer 88. Another method could be estimating the length and the size of the keys to form an estimate. With the length of the Rhine, participants could, for example, think that the Rhine starts in Switzerland, which is about 700 kilometers away from the Netherlands. But it does not go straight, and it also flows in the Netherlands, thus they should be able to generate an estimate. The questions with a hard difficulty, such as: "*How many ants are there on Earth",* where the answer is generally not known by the participants and where it is very difficult to form an estimate on the actual answer. The order of the questions was not randomized, but the difficulties were mixed. The survey can be found in Appendix A.

The participants were randomly assigned to either the treatment group or in the control group. Both groups had to answer the same questions. The treatment group received feedback after each question. This feedback was either: You are RIGHT, the correct answer is…, or You are WRONG, the correct answer is …, depending on if the correct answer was within the interval given by the participants. The control group did not receive any feedback after the questions.

3.3 Results

Table 1 – Correct answers

	Treatment	Control
Observations	55	54
Mean	5,53	4,91
Std. Dev.	2,227	2,157
Max	10	9
Min	1	0

Hypothesis 1: Over-precision

To measure over-precision, the participants were give one point to if the answer to a question was within their interval and otherwise they received zero points. By testing if the sum of the points significantly differed from 9 (10 questions with a 90% confidence interval should result in an average of 9 correct questions with no over-precision), over-precision was measured. A t-test was used to compare the mean of the sum of the points to 9. For the treatment group the mean of the sum of the points was 5.53. There was also a significance level of 0.000, and therefore the H0 of no over-precision can be rejected. Thus there was over-precision for the treatment group. For the control group the mean was 4.91. The significance level was 0.000, thus the H0 of no over-precision can be rejected. Over-precision is also prevalent for the control group. The standard deviation was around 2 for both the control and treatment group. 2 is a relatively large standard deviation, thus the participants scores are spread.

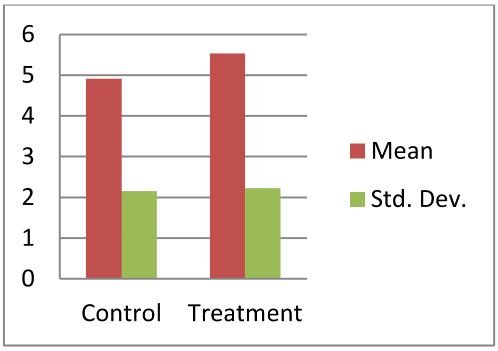


Figure 3 – Mean/ Standard Deviation (t-test)

Hypothesis 2: Hard-easy effect

When looking at the hard-easy effect, the number of correct answers for questions with an easy, moderate and hard difficulty level. A correct answer generated one point for a participant, and an incorrect answer generated zero points. Therefore, the mean of the amount of correct answers lies between one and two, where one is the same as no correct answers and two is the same as all answers correct. For the treatment group, the mean of the easy questions was 0.758, the mean of the moderate questions was 0.594 and the mean for the hard questions was 0.309.

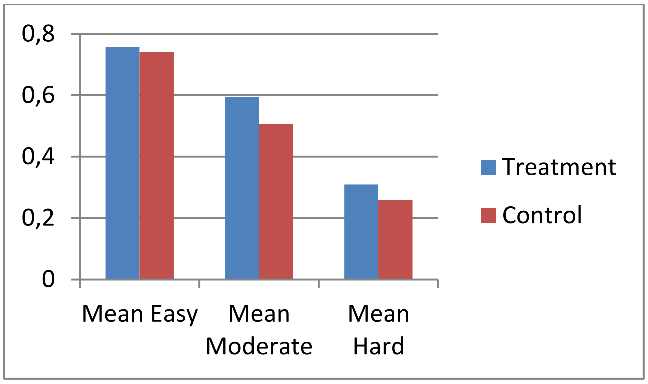


Figure 4 – Difficulty level means

The within subjects effects test shows that there is a significance difference between at least one of the different difficulty levels with all the significance levels at 0.00. The repeated measures ANOVA test showed a significance level of 0.003 when comparing easy to moderate and it shows a significance level of 0.000 when comparing easy to hard and when comparing moderate to hard. All of these significance levels are smaller than 0.05, and thus the H0: no difference in the means can be rejected and a significant difference between the means of different difficulties can be assumed. For the control group, the different means were respectively 0.741, 0.506 and 0.259. The repeated measures ANOVA test showed a significance level of 0.000 for comparison between each of de different difficulties, which are all smaller than 0.05. The H0 can also be rejected for the control group, and the hard-easy effect can be measured for the control group as well as the treatment group, because there is a significant difference, and the mean of the easy questions is larger than the mean of the moderate questions, which in turn is larger than the mean of the hard questions.

Hypothesis 3: Feedback effect

The Mann-Whitney U test showed a significance level of 0.161 when comparing the correct answers of the control group with the feedback group. The results of 0.161> 0.05, and thus no significant difference can be found between the treatment group and the control group. The H0: *There is no measurable effect from feedback on over-precision*, cannot be rejected.

Hypothesis 4: Education

When we look at the effect of the level of education on over-precision, the kruskal Wallis test showed a significance level of 0.165 and de Jonkheere Terpstra test showed a significance level of 0.213. Both are larger than 0.05. Therefore the H0 of different means cannot be rejected. There is no significant reason to assume that education has an effect on over-precision. Also for these test, 3 results were not used, because there was a fourth education level; high school. Al the participants who had high school as their highest education were however, assigned to the treatment group. Therefore no comparison was possible.

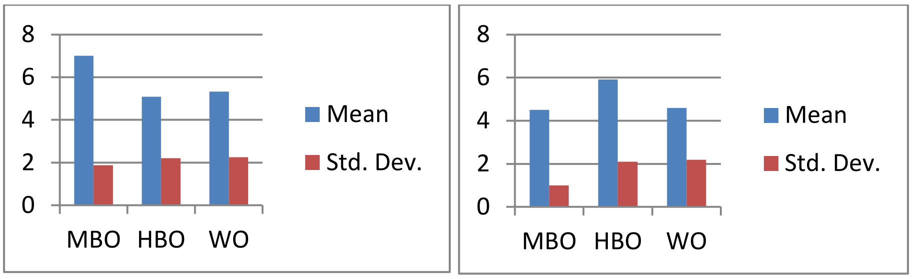


Figure 5 & 6 – Correct answer education (Treatment: left, Control: right)

Hypothesis 5: Education/Feedback effect combined

When we look at the effect of education on the effect of feedback on over-precision with a between-subject ANOVA, there is no significant difference for the variables education, group and group*education (resp. 0.397, 0.179, 0.188 are al larger than 0.05). Here, Group refers to the treatment and control group. Education level is divided into Low (MBO), intermediate (HBO) and high (WO). Therefore the H0 of no significant effect from education on feedback on over-precision cannot be rejected.

Table 2 – Between-Subject ANOVA

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	39,192	5	7,838	1,669	0,149
Intercept	1564,704	1	1564,704	333,092	0,000
Group	8,61	1	8,61	1,833	0,179
Education	8,763	2	4,381	0,933	0,397
Group * Education	20,475	2	10,237	2,179	0,118

4. Discussion

4.1 Implications

The experiment does provide significant evidence on over-precision. Also it shows that participants display a larger effect of over precision for questions with high difficulty, compared to questions with an intermediate difficulty, which in turn generate more over-precision than questions with an easy difficulty. This would indicate that the participants are better able to estimate their answers for easy, compared to hard questions. It could also be the case that the participants are not better in estimating their answers for easy questions, but that they are better able to set the upper and lower boundary given a certain confidence interval. This implicates that when people are participating in harder tasks, their ability to interpret the difficulty of this task is insufficient. Therefore, it is suggested that with hard tasks, an indication should be given to suggest the difficulty of the task. Then it would be easier for people to take the difficulty of a task into account. These findings are contradictory with earlier findings where the hard-easy effect was not found when looking for over-precision.

When looking at the effect of the level of education, the test did not show any evidence to assume education has an effect on over-precision. Assuming that higher educated people have a broader knowledge and or a higher intelligence, this suggests that the over-precision effect is not dependent on the prior knowledge of a person. It also suggests that higher educated people are not more able to correctly set a confidence interval compared to lower educated people. Therefore, knowledge and intelligence have no significant effect on over-precision. It could also be that higher educated people are better able to estimate and set boundaries with confidence intervals, but that, compared to lower educated people, they value informative information even more. Conversational norms could be a reason why no significant effect was indicated for the different education levels.

There is no reason to assume a significant effect from feedback on over-precision. One reason for why no effect could be found is that feedback does not affect over-precision. This is in line with the research from Erev and Barron in 2005, where they say that the maximum payoff does not increase for a person after they receive feedback when they face a decision from experience. There is also another possibility. According to the feedback model by Kluger and Denisi, people increase their effort and try to perform better after they fail at a task, and they decrease their effort if they succeed at a task. It is possible that the effect of positive feedback cancels out the effect of negative feedback. If feedback has no effect on over-precision, this indicates that people do not alter their behavior when they are confronted with their mistakes, and that no short term learning takes place. People do not change their mind about their ability to complete a task, even when there is compelling reason to do so. This finding is contradictory with earlier research on the subject. Arkes et al, for example, found that feedback affected over-precision.

No significant evidence can be found to answer positively to the question: *Does the level of education affect the effect of feedback on over-precision?* The lacking evidence of the combined effect of feedback and education on over-precision indicates that the learning effect for over-precision is the same for all education levels. It was found that feedback (or learning) did not have a general effect on over-precision. Therefore it is safe to assume that it also has no effect for the different education levels.

This research shows that people show more over-precision for hard tasks compared to easy task. It also shows the education level does not affect over-precision and neither does feedback. There is also no effect for a combination of these factors.

4.2 Limitations

The subject pool of the experiment mainly consisted of highly educated people (WO) and higher educated people (HBO). There was a relatively small sample of people with a low education, which could affect the results, especially with testing for the effect of education. This is visible in figure 5 & 6, which show a relatively large difference between (MBO) and (HBO) and between MBO with and without feedback. Also the total subject pool was relatively small. Because of this, people who answered differently had a greater effect, which could have affected the results. A larger or at least a more evenly distributed subject pool could fix this problem.

The subjects did not receive a reward for completing the survey. Therefore, they were driven by extrinsic motivation. On the one hand, this is preferred because it is necessary for the feedback effect to work, but on the other hand, this could be a reason which caused people to be too specific because they preferred to be informative rather than accurate, which could have been avoided by using a reward. A possible way to fix this is by rewarding participants with a monetary reward if the complete the survey and seriously consider the tasks, but their amount of right answers does not affect the rewards. This way they have extrinsic motivation to seriously consider the task, but the intrinsic motivation on the individual question level is not destroyed, which makes the feedback affect possible. The serious answering can be measured by looking at the time spent answering a question and the answers given.

There was also a design flaw in the experiment. The answer box was small, which made it impossible to see the answer people gave with large numbers. This could cause people to give a different number than the number which they actually meant to give as an answer. This was mostly relevant for the hard questions, because the answers here were large numbers. A larger box or questions with a smaller answer can fix this problem.

In this experiment, the hard-easy effect was found and no feedback effect was found. Other research in this area found the opposite. Possible reasons could be that the easy questions were too easy and the hard questions were too hard, because the moderate questions were answered correctly about 50%, which is similar with the effect of over-precision in earlier research. This could be fixed by using questions which are not as easy and hard as the ones used in this survey, but still have a noticeable difference in their difficulty level.

The results showed no significant effect for feedback, even though the effect was found by other research. This could be because, as mentioned earlier, positive and negative feedback cancel each other out. This could be tested separately testing positive and negative feedback.

With an online survey there is also a lack of control for the experimenter. People could have used other sources to help them answer the questions. However, in this research it is unlikely because participants who answered a lot of questions correctly were not specific, and when the completion times of the survey was taken into account, there was not enough time to consult other sources. This risk could be eliminated however, if the experimenter was present during the survey.

4.3 Further research

Further research is suggested in the area of the hard easy effect and the effect of feedback. The hard easy effect can be tested by lowering the differences in difficulty between easy, moderate and hard questions to see if this can account for the difference. For feedback a look at the effect of positive feedback and the effect of negative feedback is also suggested because even though no general feedback effect is found, it is possible that these specific affect can be found, especially since earlier research found an effect from feedback on over-precision. Also the effect of feedback with repetition (long term) can be researched, because literature suggests that the long term effect of feedback, with repetition, is stronger than the short term effect.

5. References:

Adler, P. S., & Clarck, K. B. (1991). Behind the Learning Curve: A Sketch of the Learning Process. *Management Science*, 267-281.

Alba, J. W., & Hutchinson, J. W. (2000). What consumers know and what they think they know. *Journal of Consumer Research*, 123-156.

Alpert, M., & Raffai, H. (1982). A progress report on the training of probability. In D. Kahneman, P. Slovic, & A. Tversky, *Judgement under uncertainty: Heuristics and biases* (pp. 188-207). Cambridge: Cambridge University.

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement account of overconfidence. *Journal of personality*, 718-735.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, 133-144.

Armor, D. A., & Sackett, A. M. (2006). Accuracy, Error, and Bias in Predictions for Real Versus. *Journal of Personality and Social Psychology*, 1369-1383.

Aukutsionek, S. P., & Belianin, A. V. (2001). Quality of forecasts and business performance. *Journal of economic Psychology*, 661-692.

Baron, J. (2007). *Thinking and deciding.* New York City: Cambridge University Press.

Bhandari, G., & Deaves, R. (2006). The Demographics of Overconfidence. *The Journal of Behavioral Finance*, 5-11.

Block, R. A., & Harper, D. R. (1991). Overconfidence in Estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational behavior and human decision processes*, 188-207.

Brehmer, B. (1980). In One Word: Not From Experience. *Acta Psychologica*, 223-241.

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 245-281.

Cameron, J., Banko, K. M., & Pierce, W. D. (2001). Pervasive negative effects of rewards on intrinsic motivation. *Behavioral Analitics*, 1-44.

Croson, D. C., Croson, R., & Ren, Y. (2008). *How to Manage an Overconfident Newsvendor.* Dallas: University of Texas.

Dirksen, J. (2011). *Design for How People Learn.* berkeley: New Riders.

Erev, I., & Barron, G. (2005). On Adaptation, Maximization, and Reinforcement Learning Among Cognitive Strategies. *Psychological Review*, 912–931.

Fieller, E. (1954). Some Problems in Interval Estimation. *Journal of the Royal Statistical Society*, 175-185.

Frederick, S., Lowestein, G., & O'Donoghue, T. (2002). Time Dischounting and Time Preferene: A Critical Review. *Journal of Economic Literature*, 351-401.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic Mental Models: A Brunswikian Theory of Confidence. *Psychological Review*, 506-528.

Grice, H. P. (1975). Logic and Conversation. In C. e. Al, *Syntax and semantics* (pp. 41-58). London: University College London.

Haran, A., Tenney, E. R., & Moore, D. A. (2015). Overprecision in judgement. In W. Blackwell, *The Wiley Blackwell Handbook of Judgment and Decision Making, 2 ..., Volume 1* (pp. 182-211). Chichester: John Wiley & Sons.

Haran, U., & Moore, D. A. (2010). A simple remedy for overprecision in judgement. *Judgment and Decision Making*, 467-476.

Hattie, J. (1999, August 2). Influences on Student Learning. Auckland, New Sealand: University of Auckland.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Eduational Research*, 81-112.

Healy, P. J., & Moore, D. A. (2008). Bayesian Overconfidence. *Psychological Review*, 502-517.

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin*, 211–237.

Justin, P., Winman, A., & Olsson, H. (2000). naive EMpiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect. *Psychological Review*, 384-396.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology 3*, 430–454.

Kluger, A. N., & DeNisi, A. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science*, 67-72.

Kulhavy, R. W. (1977). Feedback in Written Instruction. *Review of Educational Research*, 211-232.

Larrick, R. P., Burson, K. A., & Soll, J. B. (2006). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Elsevier*, 76-94.

Leunis, K. (2013). *Choice Anomalies & Learning effects.* Rotterdam: Erasmus University Rotterdam.

Lichtenstein, S., & Fischhoff, B. (1977). Do Those Who Know More Also Know More about How Much They Know? *ORGANIZATIONAL BEHAVIOR ANn HUMAN PERFORMANCE*, 159-183.

Lusk, J. L., Pruitt, J., & Norwood, B. (2006). External validity of a framed field experiment. *Economic letters*, 285-290.

Mamassian, P. (2008). Overconfidence in an Objective Anticipatory Motor Task. *Psychological Science*, 601-606.

Mannes, A. E., & Moore, D. A. (2013). A Behavioral Demonstration of overconfidence in judgement. *Psychological Science*, 1190-1197.

McKenzie, C. R., Liersch, M. J., & Yaniv, I. (2008). Overconfidence in interval estimates: What does expertise buy you? *Elsevier*, 179-191.

Moore, D. A., Carter, A. B., & Yang, H. H. (2015). Wide of the mark: Evidence on the underlying causes of overprecision. *Elsevier*, 110-120.

Moore, D., & Healy, P. j. (2007). *The Trouble with Overconfidence* . Pittsburgh.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, G., Stankov, L., et al. (2002). The role of individual differences in the accuracy of confidence judgements. *The Journal of General Psychology*, 257-299.

Pfeffer, J. (1992). *Managing with Power: Politics and Influence in Organizations.* Boston: Harvard Business School Press.

Pulford, B. D., & Colmant, A. M. (1997). Overconfidence: Feedback and item difficulty effets. *Elsevier*, 125-133.

Shafir, E., & LeBoeuf, R. A. (2002). RATIONALITY. *Annual Review of Psychology*, 491-517.

Simon, H. A. (1957). A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 99-118.

Tedick, D., & Gortari, B. (1998). Research on Error Correction and Implications for Classroom Teaching. *ACIE*.

Tversky, A., & Kahneman, D. (1971). Belief in the Law of Small Numbers. *Psychological Bulletin*, 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 1124-1131.

Vallerand, R. J., & Reid, G. (1988). On the relative effects of positive and negative verbal feedback on males' and females' intrinsic motivation. *Canadian Journal of Behavioural Science*, 239-250.

Yanif, I., & Foster, D. P. (1992). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of experimental psychology*, 424-432.

6.1 Appendix A (Survey)

An English translation of the survey is provided below:

In this questionnaire, there are ten questions which should be answered in such fashion that you are 90% certain that the correct answer lies between the upper and lower bound. This means that out of the ten questions, you should answer around 9 questions correctly (90%). *Note: none of these questions are supposed to be tricky. Some of the questions are easier and some are harder.*

Please do not use Google or other any other means of help to answer these questions.

We will start with an example:

You need to answer this question with 100% certainty. If you have no clue to the answer, you can fill in between 0 and 9999999999999999999.

Example question: How many flights of stairs does the Domtoren in Utrecht have?

Easy:

1) How many inhabitants does the Netherlands have? (17,011,010)

2) At which temperature does alcohol evaporate? (78)

3) How many legs does a mosquito have? (6)

Intermediate:

4) Out of how many (whole) countries does Europe consist? (44)

5) In which year did Columbus discover America? (1492)

6) How many meters is the Rhine? (1,230,000)

7) How many keys does a piano have? (88)

Hard:

8) How many liters of water pass the Niagara Falls per minute? (170,000,000)

9) How many products does the Coca Cola Company sell in a year? (693.500.000.000)

10) How many ants are there on Earth? (321,035,624,829,901,000)

- The answers are in brackets
- The order of the questions was respectively: 1 – 10 – 6 – 2 – 5 – 7 – 9 – 3 – 4 – 8

6.2 Appendix B (Results)

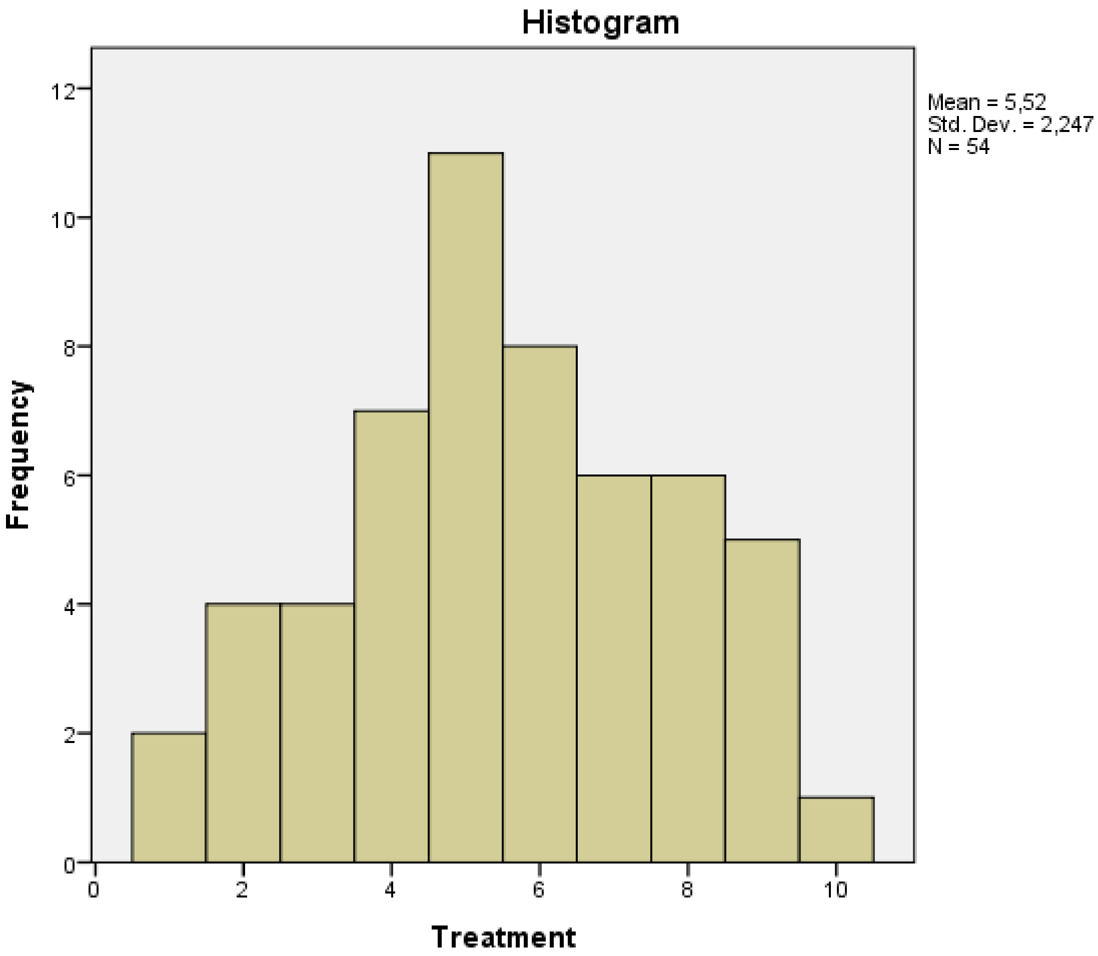
Type of test

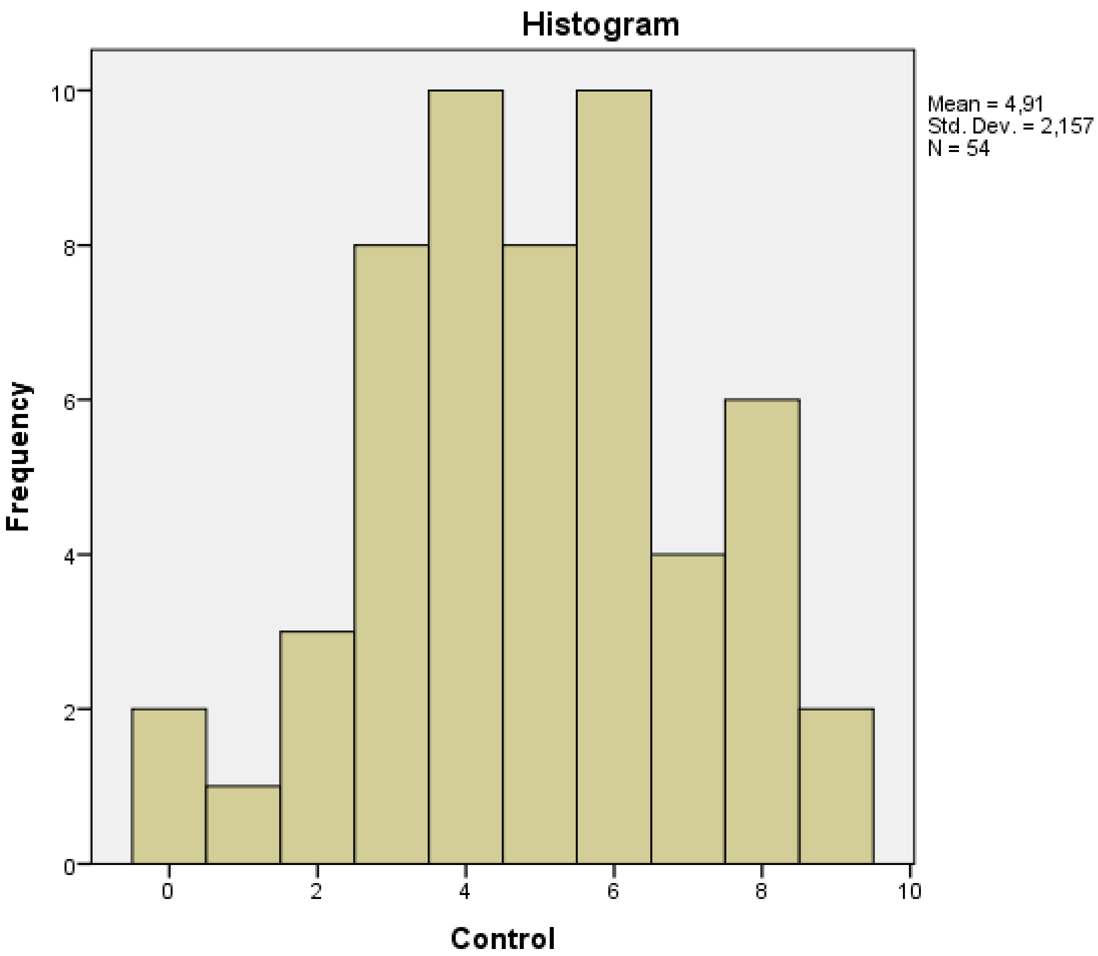
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Treatment	,110	54	,152	,968	54	,165
Control	,107	54	,178	,969	54	,168

a. Lilliefors Significance Correction

The Shapiro Wilkoxon test shows that with a 95% confidence interval the H1 hypothesis of non-normal distribution can be rejected for the treatment group as well as for the control group.





In the histograms we can also see that the control group is a little skewed to the right, with a slightly lower mean

Descriptives

			Statistic	Std. Error
Treatment	Mean		5,52	,306
	95% Confidence Interval for Mean	Lower Bound	4,91	
		Upper Bound	6,13	

	5% Trimmed Mean		5,54	
	Median		5,00	
	Variance		5,047	
	Std. Deviation		2,247	
	Minimum		1	
	Maximum		10	
	Range		9	
	Interquartile Range		3	
	Skewness		-,062	,325
	Kurtosis		-,655	,639
Control	Mean		4,91	,293
	95% Confidence Interval for Mean	Lower Bound	4,32	
		Upper Bound	5,50	
	5% Trimmed Mean		4,95	
	Median		5,00	
	Variance		4,652	
	Std. Deviation		2,157	
	Minimum		0	
	Maximum		9	
	Range		9	
	Interquartile Range		3	
	Skewness		-,134	,325

Kurtosis	-,325	,639

Feedback effect on over-precision

To compare the treatment groups with the control groups for differences in averages/median, the Mann-Whitney U test will be used.

- *H0: There is no measurable effect from feedback on over-precision*
- *H1: Feedback alters the level of over-precision*

Test Statistics[a]

	Treatment
Mann-Whitney U	1256,000
Wilcoxon W	2741,000
Z	-1,401
Asymp. Sig. (2-tailed)	,161

With a confidence interval of 95% there is no reason to reject the H0 (0.161 >0.050), and therefore there is no significant difference in general between feedback and no feedback.

Hard-easy effect

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lo
Difficulty	,996	,572	2	,751	,997	1,000	

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent vari proportional to an identity matrix.

a. Design: Intercept

Within Subjects Design: Difficulty

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are dis the Tests of Within-Subjects Effects table.

Mauchly's Test is satisfied, thus:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Difficulty	Sphericity Assumed	16,998	2	8,499	45,698	,000
	Greenhouse-Geisser	16,998	1,993	8,529	45,698	,000
	Huynh-Feldt	16,998	2,000	8,499	45,698	,000
	Lower-bound	16,998	1,000	16,998	45,698	,000
Error(Difficulty)	Sphericity Assumed	61,002	328	,186		
	Greenhouse-Geisser	61,002	326,856	,187		
	Huynh-Feldt	61,002	328,000	,186		
	Lower-bound	61,002	164,000	,372		

There is a difference in the amount of correctly answered questions between the different difficulties of the questions.

Estimates

Measure: MEASURE_1

Difficulty	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	1,758	,033	1,691	1,824
2	1,594	,038	1,518	1,670
3	1,309	,036	1,238	1,380

As can be seen in the table above, difficulty 1 (easy) has a higher mean, and thus a higher amount of correct answers, than difficulty 2 (moderate), which, in turn, has a higher mean than difficulty 3 (hard).

Pairwise Comparisons

Measure: MEASURE_1

(I) Difficulty	(J) Difficulty	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	,164*	,049	,003	,045	,282
	3	,448*	,047	,000	,336	,561
2	1	-,164*	,049	,003	-,282	-,045
	3	,285*	,047	,000	,171	,398
3	1	-,448*	,047	,000	-,561	-,336
	2	-,285*	,047	,000	-,398	-,171

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Al the difficulties differ significantly from each other. Combined with the previous table, it is reasonable to assume that the 'easy-hard effect' has a significant effect for the treatment group.

Estimates

Measure: MEASURE_1

Difficulty	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1,741	,035	1,673	1,809
2	1,506	,039	1,428	1,584
3	1,259	,035	1,191	1,327

Pairwise Comparisons

Measure: MEASURE_1

(I) Difficulty	(J) Difficulty	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	95% Confidence Interval for Difference[b] Upper Bound
1	2	,235*	,050	,000	,114	,355
	3	,481*	,046	,000	,371	,592
2	1	-,235*	,050	,000	-,355	-,114
	3	,247*	,050	,000	,127	,367
3	1	-,481*	,046	,000	-,592	-,371

2	-,247*	,050	,000	-,367	-,127

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

b. Adjustment for multiple comparisons: Bonferroni.

The same is true for the control group.

Education effect on over-precision

Test Statistics[a,b]

	Goed
Chi-Square	3,606
df	2
Asymp. Sig.	,165

Jonckheere-Terpstra Test[a]

	Goed

Number of Levels in Opleiding	3
N	54
Observed J-T Statistic	273,500
Mean J-T Statistic	340,500
Std. Deviation of J-T Statistic	53,833
Std. J-T Statistic	-1,245
Asymp. Sig. (2-tailed)	,213

The Kruskal-Wallis test and the Jonckheere-Terpstra test show that within the control group, the level of education has no significant effect on over-precision with a 95% confidence interval.

Test Statistics[a,b]

	Goed
Chi-Square	2,900
df	2
Asymp. Sig.	,235

Jonckheere-Terpstra Test[a]

	Goed
Number of Levels in Wat is je hoogst genoten opleidingsniveau?	3

N	52
Observed J-T Statistic	281,000
Mean J-T Statistic	315,500
Std. Deviation of J-T Statistic	50,720
Std. J-T Statistic	-,680
Asymp. Sig. (2-tailed)	,496

The same holds for the treatment group.

Effect of education on the effect of feedback on over-precision

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Goed

F	df1	df2	Sig.
,677	5	100	,642

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Group + Opleiding + Group * Opleiding

There is no significant evidence to assume the error variances are not equal

Tests of Between-Subjects Effects

Dependent Variable: Goed

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	39,192[a]	5	7,838	1,669	,149	,077
Intercept	1564,704	1	1564,704	333,092	,000	,769
Group	8,610	1	8,610	1,833	,179	,018
Opleiding	8,763	2	4,381	,933	,397	,018
Group * Opleiding	20,475	2	10,237	2,179	,118	,042
Error	469,751	100	4,698			
Total	3342,000	106				
Corrected Total	508,943	105				

a. R Squared = ,077 (Adjusted R Squared = ,031)

The variables education, group and education*group are not significant on a 95% confidence interval. Thus the H0 cannot be rejected.

- *H0: There is no measurable effect from level of education on the effect of feedback on over-precision*
- *H1: The level of education alters the effect of feedback on over-precision*

Therefore, according to the data, there is no reason to assume that education has an effect on the effect of feedback on over-precision.

Over-precision

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean

Goed	54	4,91	2,157	,293

One-Sample Test

	Test Value = 9					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Goed	-13,944	53	,000	-4,093	-4,68	-3,50

With a significance level of 0.000, which is smaller than 0.05, the H0 of no over-precision can be rejected