Discussion Paper Series – CRC TR 224

Discussion Paper No. 150
Project B 05

Student Performance and Loss Aversion

Heiko Karle *
Dirk Engelmann **
Martin Peitz ***

January 2020

*Frankfurt School of Finance & Management, Department of Economics, 60322 Frankfurt am Main, Germany;
Email: h.karle@fs.de
** Humboldt Universitat zu Berlin, School of Business and Economics, 10099 Berlin, Germany.
E-mail: dirk.engelmann@hu-berlin.de
***University of Mannheim, Department of Economics and MaCCI, 68131 Mannheim, Germany.
E-mail: martin.peitz@gmail.com. Also affiliated with CEPR, CESifo, and ZEW.

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Student Performance and Loss Aversion*

Heiko Karle[†] Dirk Engelmann[‡] Martin Peitz[§]

Version: January 22, 2020

Abstract

In this paper, we match data on student performance in a multiple-choice exam with data on student risk preferences that are extracted from a classroom experiment. We find that more-loss-averse students leave more questions unanswered and perform worse in the multiple-choice exam when giving an incorrect answer is penalized compared to not answering. We provide evidence that loss aversion parameters extracted from lottery choices in a controlled experiment have predictive power in a field environment of decision making under uncertainty. Furthermore, the degree of loss aversion appears to be persistent over time, as the experiment was conducted three months prior to the exam. We also find important differences across genders; they are partly explained by differences in loss aversion.

Keywords: Loss Aversion, Decision Making under Uncertainty, Multiple Choice

JEL Classification: C91, D01, D11, D83

---

*We are grateful to Sebastian Ebert, Jana Friedrichsen, Olivier l'Haridon, Katharina Hombach, Ulrich Laitenberger, Vardges Levonyan, Wanda Mimra, Julia Nafziger, Michaela Pagel, Helena Perrone, Eva Ranehill, Gerhard Riener, Nicolas Schutz, Alexander Stremitzer, Stefan Trautmann, Christian Waibel as well as several seminar and conference audiences for their valuable comments and suggestions. Martin Peitz gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TR 224 (Project B05). Dirk Engelmann gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190. *Correspondence to:* H. Karle

[†]Frankfurt School of Finance & Management, Department of Economics, 60322 Frankfurt am Main, Germany; Email: h.karle@fs.de.

[‡]Humboldt Universität zu Berlin, School of Business and Economics, 10099 Berlin, Germany. E-mail: dirk.engelmann@hu-berlin.de

[§]University of Mannheim, Department of Economics and MaCCI, 68131 Mannheim, Germany. E-mail: martin.peitz@gmail.com. Also affiliated with CEPR, CESifo, and ZEW.

1 Introduction

In multiple-choice exams, students have to make risky choices among the possible answer options. With rewards for not answering questions (compared to giving the wrong answer), students have to decide for each question whether or not to answer. To do so, they have to assess how likely they are to pick the correct answer. As we show in this paper, loss aversion enters as an important explanatory factor to make such gambles.

Our field data consist of multiple-choice scores from an introductory economics exam, which is typically taken in the first semester of studies: each student was asked to answer 30 multiple-choice questions and received a score in which correctly answered and unanswered questions entered positively. We match them with data on students' ability and behavioral characteristics including students' loss aversion, which we extracted from an incentivized classroom experiment on lottery choices.[1] We regress our performance measures from the field on students' characteristics. Our main result is that more-loss-averse students are less likely to answer a question.

Since the classroom experiment was conducted around three months prior to the multiple-choice exam, the loss aversion parameter elicited through our experiment appears to be persistent. In addition, loss aversion present in a low-stake environment explains performance in a different, arguably high-stake environment. In our data set, loss aversion hurts students as they take too few gambles. If the goal of the multiple-choice test is to evaluate the knowledge and ability of the student, behavioral parameters such as loss aversion and self-confidence should not affect the score. From this perspective, our results inform the designer of multiple-choice exams of unintended consequences when introducing punishment for wrong answers (i.e. through deductions for wrong answers or, as in the exam we investigate, rewards for not answering a question). The argument in favor of punishment for wrong answers is that it increases the precision, as students who do not assign a high probability to any answer of a particular question decide not to answer this question, such that luck in guessing affects the results less. As we show, the designer faces a tradeoff between precision and bias: Punishment increases precision at the cost of introducing a bias by punishing loss-averse students.

This bias is arguably particularly problematic if it affects different groups of the popula-

[1] Our elicitation method of individual loss aversion builds on Köbberling and Wakker (2005), Fehr and Götte (2007), and Gächter, Johnson, and Herrmann (2007); it is based on Tversky and Kahneman's (1992) cumulative prospect theory and Rabin (2000)'s calibration theorem.

tion to a different extent, because the scheme then not only negatively impacts on loss-averse individuals' performance but also leads to a systematic bias against some groups. To address this issue in the context of gender, and in line with the extant literature, we analyze gender differences in answering questions in multiple-choice exams. Women are less likely to answer a given question conditional on estimated ability and other individual characteristics. This gender gap is partly explained by gender differences in the inferred loss-aversion parameters, suggesting that the punishment scheme introduces a gender bias in the exam results. This is supported by the observation that in the following year, when there was no more reward for not answering questions in the exam, the gender gap was substantially reduced.

To guide our empirical investigation, we provide a simple theory and derive the testable prediction that a higher degree of loss aversion reduces the inclination to gamble. The idea here is that higher expected utility losses due to a larger degree of loss aversion reduce the inclination to accept a gamble. We note that a theory based on risk aversion rather than loss aversion could also provide the prediction of a lower inclination to gamble. However, this is not fully borne out by our data. The payoffs from gambling in the exam have a mixed support (i.e. are positive or negative if we make the plausible assumption that the payoff from not answering a question serves as reference point). This resembles the monetary lottery that we use in the classroom experiment to elicit individual degrees of loss aversion and suggests that loss aversion could play out in a similar way in both environments.

We combine experimental and field data to obtain a unique data set of more than 650 students that includes students' lottery choices, other characteristics of students, and data on their behavior and performance in an exam. At the beginning of the term students of an introductory economics course participated in a classroom experiment consisting of a crude ability test (a cognitive reflection test) and an incentivized problem of lottery choice. Furthermore, we collected information on student characteristics (gender, main field of study, age, self-assessment of confidence). Then, at the end of the term, students took the exam of the introductory economics class. This gives us students' responses to 30 multiple-choice questions.

Our result that more-loss-averse students are less inclined to gamble is robust across a number of empirical specifications. Furthermore, lower response rates affect performance. In our data set, more-loss-averse students perform worse. A causal channel how loss aversion affects performance is that more-loss-averse students are less inclined to gamble when faced with the choice to select between multiple choices with implicit punishment for wrong answers.

In our regressions, we condition results on the level of ability measured by a cognitive reflection test (CRT, Frederick, 2005). This is an imperfect measure of ability and the only direct measure available to us. Therefore, an important concern is that loss aversion may be negatively correlated with unobserved ability (in line with Dohmen et al., 2010), and the effect of loss aversion on gambling and performance may be spurious. If this spurious effect is sufficiently strong, then more-loss-averse students perform worse even conditional on answering a question, while the opposite holds true if the causal effect dominates.

To disentangle the causal from the spurious channel, we consider different subpopulations. One may expect that the spurious channel matters most for students who do not answer all (or almost all) questions and for students of a subpopulation that performs worse. In contrast, we find evidence in support of the causal effect in the subpopulation that is less prone to answer all questions. We interpret this finding as support for our hypothesis that more loss-averse students perform worse because they refrain from making some gambles that would have increased their performance in expectation. In this subpopulation, we observe an above-average fraction of students who do not have business administration or economics as their main field of study. We therefore split the sample into students in business administration or economics and students with other main field of study. We find evidence in support of the causal effect for other main fields of studies and the spurious effect for business administration or economics. The latter can be explained by business and economics students being inclined to answer all questions in any case, which would remove the causal effect.

Our paper relates to several strands of literature. A growing empirical and experimental literature on choice under uncertainty provides evidence that individuals experience loss aversion. Loss aversion was introduced through prospect theory by Kahneman and Tversky (1979) and modified by Tversky and Kahneman (1992). Prospect theory postulates an exogenous (status-quo-based) reference point, while Kőszegi and Rabin (2006, 2007) endogenize the formation of reference points by their concept of expectation-based loss aversion.[2] Our analysis is compatible with both approaches: Either approach gives rise to the same hypotheses that we use to predict students' choices in the exam. Our elicitation of students' loss-aversion parameters follows Tversky and Kahneman (1992); interchangeably, we could elicit them based on

[2]Bell (1985); Loomes and Sugden (1987); and Gul (1991) provide alternative theories that formalize that expectations act as reference points.

Kőszegi and Rabin (2006, 2007).

DellaVigna (2009) provides an overview on empirical and experimental evidence of loss aversion. Work on expectation-based loss aversion includes exchange and valuation experiments (e.g. Ericson and Fuster, 2011), experiments in which participants are compensated for exerting effort in a tedious and repetitive task (e.g. Abeler et al., 2011), and sequential-move tournaments (e.g. Gill and Prowse, 2012). Using field data, there is evidence that expectation-based reference dependence affects golf players' performance (see Pope and Schweitzer, 2011) and cabdrivers' labor-supply decision (see Crawford and Meng, 2011).[3] Regarding evidence from the laboratory, close to our paper is Karle, Kirchsteiger, and Peitz (2015), who show that individual loss aversion parameters elicited through lotteries (as in the present paper), predict consumption choice in an environment (encountered immediately after the lottery choice) in which consumers initially face uncertainty regarding the purchase price. Our paper contributes to this literature by documenting that behavior in a low-stake experimental task has predictive power for behavior in a high-stake non-experimental task several months later.[4]

With a different focus, student behavior in multiple-choice tests has been analyzed in the literature on gender effects. Akyol et al. (2016) analyze student choice in the Turkish University Entrance Exam. They infer from their data that women are more risk-averse. Funk and Perrone (2017) use field-experimental data from an exam in microeconomics to analyze gender effects. They introduce the treatment that each student faces half of the questions with and half without penalty for responding wrongly to a question. Women guess less with punishment than men, which is consistent with our work.[5] However, in Funk and Perrone (2017),

[3] See, in particular, Camerer et al. (1997); Farber (2005, 2008, 2015) for work on cabdrivers' labor-supply decision, which partly challenge the findings of reference dependence. Fehr and Götte (2007) provide evidence on reference-dependence in labor supply from a field experiment with bike messengers. Further evidence on expectation-based reference points includes Loomes and Sugden (1987) and Choi et al. (2007) for choices over lotteries; Post et al. (2008) for gambling behavior in game shows; and Card and Dahl (2011) for disappointment-induced domestic violence. Countervailing evidence is found in Smith (2019), Heffetz and List (2014), and Gneezy et al. (2017). One explanation for negative results of Smith (2019) and Heffetz and List (2014) could be that the way how subjects form expectations varies with details of the experimental design (see Ericson and Fuster, 2014). Experimental results by Engelmann and Hollard (2010) suggest that the endowment effect is caused by aversion to trade rather than by loss aversion.

[4] University examinations arguably constitute a high stake environment in Germany, as students are concerned about their grade and the grade is a key selection criterion e.g. to be admitted to a master program. The grade in "introductory economics" enters the final grade of studies with 4.5 percent of the total and the course and the obtained grade are explicitly listed in the final official transcript.

[5] In a related investigation using a large sample of math tests, Iriberri and Rey-Biel (2019) confirm our finding that female participants leave significantly more questions unanswered than males when answering wrongly is penalized and that this hurts their performance.

women do generally better and benefit from this reluctance to answer questions. This result runs counter to our work, but can be reconciled with the contrasting findings if one allows for the possibility that, in some exams, students systematically underestimate the difficulty of a question. Funk and Perrone (2017) observe the students' university entry grade—this is their measure of ability. They also obtained individual measures of risk aversion from a laboratory experiment performed one year after the exam. In their data set, women have on average higher ability. They find that risk aversion has a zero effect on scores on both parts of the exam, which is in line with our finding that risk aversion does not have a significant effect. Different from Funk and Perrone (2017), we focus on loss aversion.

More closely related, in a laboratory experiment with 406 participants, Baldiga (2014) analyzes the interplay between gender effects and risk attitudes. She collects students' answers to questions in a SAT practise test in history considering treatments with and without penalty. She finds that women answer relatively fewer questions with penalty than men. This gender gap is partly explained by differences in risk attitude, which she extracted in a different part of the experiment. In her laboratory setting in which she observes answers for questions which participants initially did not answer, she obtains a clean estimate of the effect of skipping questions on performance. She finds that skipping questions hurts performance. Our findings are broadly in line with her findings in the sense that with penalty women are less likely to answer questions than men. Baldiga (2014) considers lotteries with mixed domain which are suitable for identifying loss aversion. Her measure of risk attitude is the lowest success probability a subject accepted—a measure which is linked to loss aversion. Different from Baldiga (2014), we extract measures of loss aversions from lottery choices that have been made three months prior to the performance (and not at the same point in time) to explain performance in the field (rather than in the laboratory) when stakes are high. As an important contribution to this literature, we shed light on the different channels how risk preferences affect student behavior, namely whether the causal or the spurious channel dominates; while the former reflects preferences, the latter reflects that risk preferences are correlated with (the unobserved part of) ability.

The paper proceeds as follows. In Section 2, we consider the multiple-choice problem and derive hypotheses on how outcome variables depend on loss aversion. In Section 3, we present the experimental design of the classroom experiment and the collection of the exam data. Section 4 contains the empirical analysis and results. In Section 5, we provide some more context and conclude.

2 Risk Preferences and Behavior in Multiple-Choice Exams

In this section, we provide a theoretical framework to analyze behavior in multiple-choice exams when students are loss averse. We then derive several hypotheses and preview the extent to which these hypotheses are supported by the subsequent empirical analysis.

For each question k, there are several options to answer. We denote by p_{jk} the probability that a student thinks that answer j in question k is correct. Probability $p_k \equiv \max_j\{p_{jk}\}$ is her perceived success probability in case she picks the answer that she believes most likely to be correct, i.e. p_k denotes the probability that a student assigns to correctly answering question k. A utility-maximizing student answers question k if p_k is above a threshold p^* which depends on the student's risk preferences (i.e. risk aversion and loss aversion). If the reverse inequality holds, $p_k < p^*$, a student should not answer this question. In the following, we specify the threshold p^* as a function of a loss-aversion parameter but other parameters capturing, for instance, risk aversion and confidence have qualitatively similar effects on the threshold (we leave them aside for brevity).

In the exam, each student faces 30 questions. We treat them as a sequence of independent decision problems, $k \in \{1, ..., 30\}$, about each of which a student may experience loss aversion.[6] There are four possible answers to each question, $j \in \{1, ..., 4\}$: a correct answer gives 3 points, no answer 1 point, and an incorrect answer gives 0 points, as in the exam in our data set. This defines a student's payoff per question. Thus, a risk- and loss-neutral student should answer the question if her success probability p_k exceeds $1/3$. For instance, $p_k \geq 1/3$ is implied if a student can rule out one of the four possible answers to a question with probability one. If a student, however, is risk-averse or loss-averse, pure randomization is not attractive at $p_k = 1/3$, i.e. the student's threshold for answering a question is larger than $1/3$.

We formalize loss aversion applying the power utility representation of Tversky and Kah-

[6] This means that we postulate that students are narrow bracketers. If students were broad bracketers, i.e. they had a reference point of the total number of points in the exam, more-loss-averse students should answer more question than less-averse student when they are below this reference point. As argued below in Section 4.3, this is inconsistent with our findings.

neman (1992).

$$u_i(z) = \begin{cases} z^\beta & \text{if } z \geq 0; \\ -\lambda(-z)^\beta & \text{if } z < 0; \end{cases} \tag{1}$$

where z denotes the material payoff relative to a reference point; $\lambda > 1$ represents loss aversion; and $\beta \in (0, 1)$ represents diminishing sensitivity—i.e., risk aversion in gains and risk love in losses (and vice versa for $\beta > 1$).[7] In particular, we assume that $u_i(z_k)$ describes student i's utility from question k, where $z_k = x_k - r_k$, and x_k describes the student's score from question k and r_k her status-quo-based reference point.

In our analysis we postulate that students have a status-quo based reference point that equals the score of the safe option (i.e., not answering) which leads to $r_k = 1$ for all questions k. As we prove in Appendix C, the same predictions hold under the loss aversion approach of Kőszegi and Rabin (2006, 2007) according to which students form an expectation-based probabilistic reference point instead of a status-quo-based deterministic one. Since the two approaches give rise to the same hypotheses, they can thus be used interchangeably in our setup.

In the analysis of the main text we assume furthermore that students are risk neutral $\beta = 1$. In Appendix B we redo the analysis accounting for students' loss and risk parameters jointly and show that our main empirical findings are robust.

Under risk neutrality the student with loss aversion parameter λ is indifferent between answering and not answering question k if and only if $p_k \cdot (3 - r_k) + (1 - p_k) \cdot \lambda(0 - r_k) = 1 - r_k$ where the right-hand side results because not answering a question yields 1 for sure. Using that the reference outcome is $r_k = 1$, this translates to the threshold

$$p^*(\lambda) \equiv \frac{\lambda}{\lambda + 2}.$$

Note that for loss-averse students $p^*(\lambda) \in (1/3, 1)$.

Proposition 1. *The threshold p^* above which a student answers a question is strictly increasing in the degree of loss aversion λ.*

The proof of this proposition follows directly from taking the first-order derivative of p^*

[7] For simplicity, we exclude the possibility of different degrees of diminishing sensitivity in the gain and the loss domain.

with respect to λ. Thus, we obtain the prediction that the larger is the degree of loss aversion λ the larger must be the student's success probability p_k in order to answer question k.

Based on Proposition 1, we derive several testable hypotheses. Denote by G_k the cumulative distribution function over success probabilities p_k in the population about question k and g_k its density function. In the following, we will neglect the index k wherever unambiguous. Note that the empirical distribution depends on the particular question.[8] It may also depend on the particular student population. Thus, it may depend on student characteristics including a student's loss aversion parameter λ. To formulate our hypotheses, we assume that G does not depend on λ—we return to this issue after formulating the hypotheses.

Hypothesis 1. *Students are less likely to answer a question the more loss-averse they are.*

Aggregated over all questions we obtain a prediction at the student level about the correlation between the number of unanswered questions m and the loss aversion parameter λ.

Hypothesis 1'. *Students answer fewer questions the more loss-averse they are.*

In our empirical analysis, we will provide strong support for Hypotheses 1 and 1'. Related to Hypothesis 1, we also look at the unconditional probability of giving a correct answer. Take two students who only differ in their degree of loss aversion. This implies that the more-loss-averse student answers fewer questions. He or she does not give an answer to those questions for which he or she thinks that success probability is not sufficiently high. By not answering, his or her probability of giving the correct answer to those questions is obviously zero. Thus, the more-loss-averse student is less likely to give the correct answer taking the average over all questions. This gives rise to Hypothesis 2:

Hypothesis 2. *Students are less likely to give the correct answer the more loss-averse they are.*

Hypothesis 2'. *Students give fewer correct answers the more loss-averse they are.*

Our empirical analysis also strongly supports Hypotheses 2 and 2'. That loss-averse students should only answer if they are more confident about knowing the correct answer leads

[8] Figure 5 in Appendix A.2 reports the answer and correct answer ratios. As the figure shows, questions that are answered less often get on average a lower ratio of correct answers.

to a positive selection effect and implies that conditional on answering a question, students are more likely to give the correct answer the more loss-averse they are. More formally, $E[p \geq p^*(\lambda)] = \int_{p^*(\lambda)}^{1} pg(p)dp/[1 - G(p^*(\lambda))]$ is increasing in λ. This leads us to our Hypothesis 3.

Hypothesis 3. *Conditional on answering, students are more likely to give the correct answer the more loss-averse they are.*

Hypothesis 3'. *Students have a higher ratio of correctly answered to answered questions the more loss-averse they are.*

In our empirical analysis, we do not obtain significant coefficients in the cross section of for Hypothesis 3'. We find support for Hypothesis 3 in one of two subsamples defined by main field of study; in Section 4.3 we will provide an interpretation of these findings.

We derived our hypotheses in a setting in which the degree of loss aversion affects choices through a change of the cutoff above which students answer. We call this the *causal effect*. However, there is a *spurious effect* through which the degree of loss aversion may affect student choices if (the unobserved part of) a student's ability is negatively correlated with the degree of loss aversion.[9] If the spurious effect is present, the cumulative distribution function G differs across student groups with different degrees of loss aversion. In particular, less-loss-averse students are more frequent (relative to more-loss-averse students) among students who have a high probability to choose the correct answer. We will consider families of distribution functions that satisfy the monotone likelihood ratio (MLR) property to show that also the spurious effect implies Hypotheses 1 and 2, but violates Hypothesis 3. Thus, Hypothesis 3 is the key hypothesis to separate between the causal and the spurious effect.

Suppose that the spurious effect is present—i.e., a student's degree of loss aversion is an inverse proxy for her ability—but that the causal channel is closed down—i.e., the threshold p^* is independent of the degree of loss aversion. Denote the distribution of success probabilities of students with loss-aversion parameter λ by $G(p; \lambda)$ and its density by $g(p; \lambda)$. Considering two loss aversion parameters λ_H and λ_L with $\lambda_H > \lambda_L$, we simplify notation and assign the cumulative distribution function G_L with density g_L for students with loss aversion λ_L and G_H with density g_H for students with loss aversion λ_H.

[9]Relatedly, in a lottery-choice experiment with positive domain, Dohmen et al. (2010) find a negative correlation between their measures of risk aversion and ability.

The monotone likelihood ratio property (MLR) is defined as follows in our context: G_L MLR-dominates G_H if and only if g_L/g_H is weakly increasing in p for values of p for which it is defined.[10] MLR says that less-loss-averse students are particularly frequent relative to more-loss-averse students among students who have a very high probability to get it right.[11] We note that MLR implies first-order stochastic dominance (FOSD); i.e., $1 - G_L(p) \geq 1 - G_H(p)$ with $\lambda_H > \lambda_L$ for any p. FOSD says that more-loss-averse students reply less often because they have lower success probabilities.

We first show that FOSD implies Hypotheses 1 and 2. From the definition of FOSD, it is immediate that Hypothesis 1 is satisfied. Furthermore, because of FOSD,[12] students are less likely to give a correct answer the more loss-averse they are; i.e. $\int_{p^*}^1 p dG_L(p) \geq \int_{p^*}^1 p dG_H(p)$ (note that this is the overall probability, not the one conditional on answering). Thus, Hypothesis 2 holds.

We next show that MLR implies a violation of Hypothesis 3. In particular, we show that under MLR the negative correlation between loss aversion and unobserved ability implies that conditional on answering a question, students are less likely to give the correct answer the more loss-averse they are; i.e., for any p^*,

$$\frac{\int_{p^*}^1 p g_L(p) dp}{1 - G_L(p^*)} \geq \frac{\int_{p^*}^1 p g_H(p) dp}{1 - G_H(p^*)} \tag{2}$$

with $\lambda_H > \lambda_L$. For $p \geq p^*$, define the (conditional) cumulative distribution functions $G_L|_{p^*}(p) = [G_L(p) - G_L(p^*)]/[1 - G_L(p^*)]$ and $G_H|_{p_*}(p) = [G_H(p) - G_H(p^*)]/[1 - G_H(p^*)]$ on $[p^*, 1]$. It follows that

$$\frac{G'_L|_{p^*}(p)}{G'_H|_{p^*}(p)} = \frac{1 - G_H(p^*)}{1 - G_L(p^*)} \frac{g_L(p)}{g_H(p)}.$$

Hence, since g_L/g_H is weakly increasing in p for all $p \in [p^*, 1]$, $G'_L|_{p^*}/G'_H|_{p^*}$ is weakly increasing in p for all $p \in [p^*, 1]$ and, thus, $G_L|_{p^*}$ MLR-dominates $G_H|_{p^*}$. This implies that, for every

[10] The ratio takes values in $\mathbb{R}_0^+ \cup \{\infty\}$. For $g_L > 0$ and $g_H = 0$, we assign value ∞. The ratio is not defined if g_L and g_H are equal to zero.

[11] The following examples satisfy MLR: (i) for any λ, G is uniform on a proper subinterval of $[0, 1]$ and loss aversion shifts the support of density g to the left; (ii) alternatively, the upper bound of the support is always 1 and loss aversion shifts the lower bound to the left (in both examples, MLR holds with equality for almost all p where it is defined; g_L/g_H has two upward jumps in the first and one upward jump in the second example); (iii) g has full support on $[0, 1]$ and the family of distributions functions has densities that are linear in p with the slope decreasing in λ.

[12] Recall the equivalent definition of FOSD according to which G_L first-order stochastically dominates G_H if and only if $\int u(p)dG_L \geq \int u(p)dG_H$ for every weakly increasing function u.

p^*, $G_L|_{p^*}$ first-order stochastically dominates $G_H|_{p^*}$ and, therefore,

$$\frac{1}{1-G_L(p^*)}\int_{p^*}^{1} p g_L(p)dp = \int_{p^*}^{1} p dG_L|_{p^*}(p) \geq \int_{p^*}^{1} p dG_H|_{p^*}(p) = \frac{1}{1-G_H(p^*)}\int_{p^*}^{1} p g_H(p)dp.$$

Thus, we have established that Hypothesis 3 is violated if the spurious, but not the causal effect is present.

To summarize, Hypotheses 1 and 2 are compatible with both, the causal and the spurious effect. If only the causal effect is present, Hypothesis 3 must hold. If only the spurious effect is present, Hypothesis 3 is violated. If causal and spurious effects are present, they tend to go in opposite directions regarding conditional performance. Thus, statistically insignificant results when checking for Hypotheses 3 can be explained by the joint presence of causal and spurious effect. We investigate this issue carefully in the empirical analysis below and find support for Hypothesis 3 in some subsamples of the student population.

Behavior may also be driven by risk aversion. A more-risk-averse student should be more inclined to go for the safe bet (no answer) than a gamble. Thus, theory predicts that the threshold probability is also an increasing function of the degree of risk aversion. This would give rise to hypotheses corresponding to Hypotheses 1 and 2 in which loss aversion is replaced by risk aversion. However, as Rabin (2000) argues, risk aversion cannot plausibly explain choice behavior in small-stake lotteries without implying absurd degrees of risk aversion in high-stake gambles. Since we extracted the degree of risk aversion from lottery choice with small stakes and the exam questions arguably involve high stakes, we conjecture that our measure of risk aversion provides little predictive power. A very different reading of our investigation is that also the exam questions constitute low stakes. To address this issue, in Appendix B, we include our measure of risk aversion from low-stake lottery choices. Our results indicate that this measure does not explain the probability to answer a question.

3 Data Collection

In the empirical analysis, we match data from the classroom (September 2013) to data in the field (exam in December 2013).[13] Our aim is to investigate whether student outcomes

[13] We matched students based on student IDs in the experiment and in the exam; we anonymized the data after the matching.

in the introductory economics exam can be explained by student characteristics and inferred preferences with respect to risk and losses. Moreover, we want to assess the degree to which these inferred preferences contribute to a gender bias in the exam.

3.1 Data from the Classroom

Risk Preferences. We elicited a ranking of participants with respect to their choice behavior on both, a mixed domain (including negative and positive payments) and a purely positive domain. The former will be interpreted as loss aversion (see Tversky and Kahneman, 1992, and Rabin, 2000) and the latter as risk aversion.

In particular, subjects have to choose between lotteries and sure payments.[14] There were two series of choices, with six choices each. First, subjects have to make six choices between a lottery that gave a 50-percent chance of winning 4 Euro and a 50-percent chance of losing R, and, on the other hand, a sure payment of zero. R takes values -0.60, -1.20, -1.80, -2.40, -3.00, -4.00 Euro (in series A; see Appendix E). To cover potential losses, each participant received 6 Euro for participating in the survey. Second, subjects have to make six choices between a lottery with a 50-percent chance of winning 4 Euro and a 50-percent chance of winning zero, and, on the other hand, a sure payment of S (in series B; see Appendix E). This payment S takes values $0.40, 0.80, 1.20, 1.60, 2.00$, or 2.40 Euro. These are standard lottery tasks with and without losses.[15] At the end of the experiment, one of the 12 choices was chosen randomly and implemented.

For series A, subject i's choice is characterized by a cutoff value $R_i \leq 0$ such that all lotteries with $|R| > |R_i|$ are rejected, and all lotteries with $|R| \leq |R_i|$ are accepted. Similarly, for series B subject i's choice is characterized by a cutoff value S_i such that for any $S < S_i$, the lottery is chosen, and for any $S \geq S_i$, the sure payment is preferred. These cutoff values characterize our individual measures of loss aversion and risk aversion—the latter is derived in Appendix B.

The power utility representation of Tversky and Kahneman (1992) in equation (1) incor-

[14]Fehr and Götte (2007), Gächter et al. (2007), and Karle, Kirchsteiger, and Peitz (2015) used a similar way of measuring loss aversion.

[15]The fact that both lottery outcomes are equally likely rules out that probability weighting has an effect on our measures of risk preferences (cf. Köbberling and Wakker, 2005). This constitutes an advantage of our elicitation methods of risk preferences over others that use binary lotteries with asymmetric probabilities (e.g. Holt and Laury, 2002).

porates a loss parameter $\lambda > 1$ and a risk parameter $\beta > 0$. We next apply this representation to identify our measures of loss aversion and risk aversion. First, according to Rabin (2000), risk aversion cannot plausibly explain choice behavior in small-stake lotteries without implying absurd degrees of risk aversion in high-stake gambles. In small-stake lotteries, people should therefore be considered as risk-neutral. According to this view and in line with part of the experimental literature (see, e.g., Gächter et al., 2007), in the main part of the paper, we assume that β is equal to one for all students. An individual measure of loss aversion λ_i can then be derived from the cutoff values of series A using the cutoff condition $0 = 1/2 \cdot 4 + 1/2 \cdot (-\lambda_i |R_i|)$, where the reference point equals status quo of zero. The degree of loss aversion of participant i is set equal to[16]

$$\lambda_i = \frac{4}{|R_i|}. \tag{3}$$

We note that λ_i is increasing in R_i , as follows from equation (3).

Other Explanatory Variables. The classroom experiment allowed us to obtain additional variables, which we will use as controls in our empirical analysis. Each student took a cognitive reflection test (CRT) as introduced by Frederick (2005). The outcome of this test constitutes our proxy of a student's general ability.

In addition, we obtain a measure for the students' confidence (cf. Hoppe and Kusterer, 2011). Students are asked about their estimates of the percentage of own correct answers to a set of general interest questions and the average percentage of the others' correct answers. The difference between the former and the latter is our measure of confidence. Furthermore, we obtained the personal characteristics gender, age, and main field of study that we use as further controls.[17]

The experiment was taken early in the first term implying that topics in microeconomics such as risk aversion and expected utility theory have not yet been covered in class.[18] There

[16]Since we observe a finite number of cutoff values, we can assign an interval of loss aversion parameters to each consumer. For convenience, we report the upper bound. Those who did not choose any lottery, are, for convenience, assigned $R_i = 0$ and thus λ_i equal to infinity. Also note that, similar to Proposition 1, we could apply the expected total utility representation of Kőszegi and Rabin (2006, 2007) to derive an alternative but qualitatively similar measure of loss aversion. In that case, an expectation-based reference point instead of a status-quo-based reference point equal to zero had to be used.

[17]The introductory economics course is a mandatory course in economics, business administration, economics education, business law, business informatics and an elective in a variety of other bachelor programs.

[18]An exception are business informatics students who tend to take the course in their third semester. However, they did not take any other economics course prior to introductory economics. There are also a few students in a higher semester retaking the course. We did not obtain access to this information and, thus, could not exclude

was a three-month time span between experiment and the observed behavior in the field. This suggests that any effect between behavioral parameters extracted from the experimental data on actual behavior is rather persistent.

3.2 Field Data

In the field we observed the performance of each student in the final exam of the introductory economics course. This course is taken by more than 1,000 students in economics, business adminstration, business law, economics education, political science, sociology, and business informatics; this class was taught in three sections. At the end of the course, students have to take an exam, which fully determines the grade for the course. The exam took place around three months after obtaining the experimental data. Students who failed or missed the exam could retake it a couple of months later. We decided to use data from the first exam only; we replicated the analysis for the pooled sample confirming qualitatively our results (the significance of some variable drops in a few instances).[19]

As mentioned above, the exam contained 30 multiple-choice questions. For each question, there are four possible answers, one of which is correct and all others are false. Students receive 3 points for each correct answer, 0 points for each wrong answer and 1 point for each question without an answer. Thus, each student can make a total of 90 points;[20] they know that they will pass for sure with at least 50 points, but that the mapping between points and grades will be done ex post (in particular, the threshold to pass may be set below 50 points). Thus, since students do not know whether one additional point or correct answer improves their grade, we believe that students typically do not "strategically" provide answers; i.e., we do not expect them to guess more if they expect to be below the threshold for passing the exam. Hence, we assume they do not answer a question if their subjective success probability

them from the sample, but we know that the number is low since students who fail the first exam after the course take the second exam shortly before the following term. In addition, after a third failure, students are no longer allowed to continue to study. With a failure rate of around 15 percent in an exam this implies that significantly less than five percent retake the course. In addition, since the course material does not change much over time, students who retake the course often ask at the beginning of the term about any changes in the course material and then stop attending (and, thus, will not be in our sample).

[19]We focused on the first exam for a number of reasons: re-sitters might perform differently and would constitute repeated observations; exam questions and possibly the overall exam differed in difficulty. Analyzing the second exam separately is no viable alternative since it provides too few observations.

[20]After the exam was written it turned out that one question did not have a unique correct answer; students were assigned 3 points independent of whether and what they answered. We removed this question from our data set leaving 29 questions with a maximal score of 87.

is below the threshold p^* for their expected probability to answer correctly and they answer otherwise. We observe the individual answers to all questions; in particular, we observe, how many and which questions the student did not answer, as well as how many and which answers are correct.

In the first part of the empirical analysis, *the student* is the unit of observation. Summing over the associated points, we obtain the total number of points a student reached—this is the exam score. From the individual answers we construct a variable that approximates a student's propensity to gamble. Provided that a student was maximizing the expected number of points she should not answer a question if the expected number of points is less than one. Suppose that a question falls into either one of two categories for a student: she either knows the correct answer for sure or does not know the correct answer for sure, assigns different probabilities to the four options to answer, and the option with the highest probability has probability one third. In the former case she would answer for sure and in the latter case the student would be indifferent between choosing the best option and not answering. For example, such a situation arises if she can exclude one of the options and assigns equal probability to the remaining three options. With equal probability assigned to each remaining option, the student should expect to be wrong with probability 2/3. If we observe n wrong answers in a given exam, in expectation, the student should have taken at least 3/2 times n gambles. Given that the maximum probability to be correct will often be higher, this is a lower bound on the number of gambles. The total number of questions where the student has some doubt then is $(3/2)n$ plus the number of unanswered questions m. As our measure of gambling, we define the rate

$$\gamma = \frac{\frac{3}{2}n}{\frac{3}{2}n + m}.$$

Of course, this is a crude measure since we do not observe subjective probabilities of each question. Clearly, apart from introducing noise one may be worried about introducing a bias. According to our hypotheses, loss-averse students are less prone to gamble as they require a higher threshold probability. If this hypothesis is correct, for loss averse students, the number of wrong questions would need to be multiplied by a number larger than 3/2. We indirectly address this issue, as we also regress the total number of answered questions on the degree of loss aversion.

In the second part of the empirical analysis, *each question for every student* is the unit of observation. Here, we view the decision to answer and the choice of the correct answer as probabilistic outcomes.

4 Empirical Analysis and Results

4.1 Descriptives

In our matched data set, we have 646 students of which 367 are male and 279 female. Table 1 reports descriptive statistics from this data set.

Table 1: Descriptive Statistics:

Variable	Obs	Mean	Std. Dev.	Min	Max
Answered Questions	646	23.9954	5.3095	11	29
Correct Answers	646	19.2740	5.7239	5	29
Exam Score	646	62.8266	13.0478	18	87
Propensity to Gamble	646	0.6981	0.2736	0.0811	1
Confidence	645	-0.5189	1.7614	-7	5
Cognitive Reflection	646	1.7665	1.076	0	3
Age	646	19.4593	2.1767	16	37

In the exam, some students answered all remaining 29 questions;[21] the lowest number of answers is 11. This student should know that this may be sufficient to pass the exam.[22] Students answered on average around 24 questions.

As we can see from Figure 1, any number between 11 and 29 questions are answered with a spike at all questions being answered. Students answer on average around 19 questions correctly. As documented in Figure 2, the empirical support of the exam score is the interval $[30, 87]$ plus one outlier at 18. Descriptives on number of answered and correctly answered questions differ by gender with a mean of 25.35 vs. 22.22 (diff = 3.134, p-value < 0.001) and 20.79 vs. 17.28 (diff = 3.517, p-value < 0.001) in favor of male students, respectively (cf. Tables 15 and 16 in Appendix A which provide information on descriptives by gender).

Main field of study is an important control, as student ability correlates with it and the ratio of female students varies by field—Tables 17 and 18 in Appendix A provide information

[21] As mentioned above, one of the 30 questions was not valid and, thus, had to be removed from the analysis.

[22] Even if she did not answer the question which was removed, she could get up to 33 points for 11 correct answers and 19 points for not answering the remaining 19 questions, which gives 52 points and guarantees that she passed.

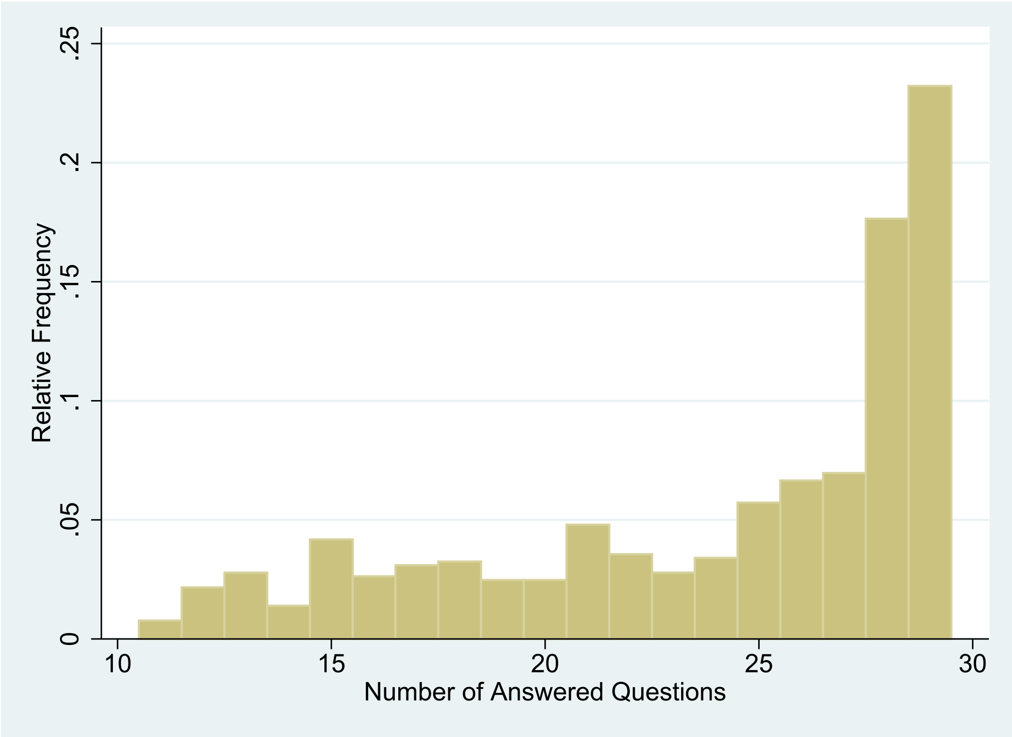

Figure 1: Histogram of answered questions

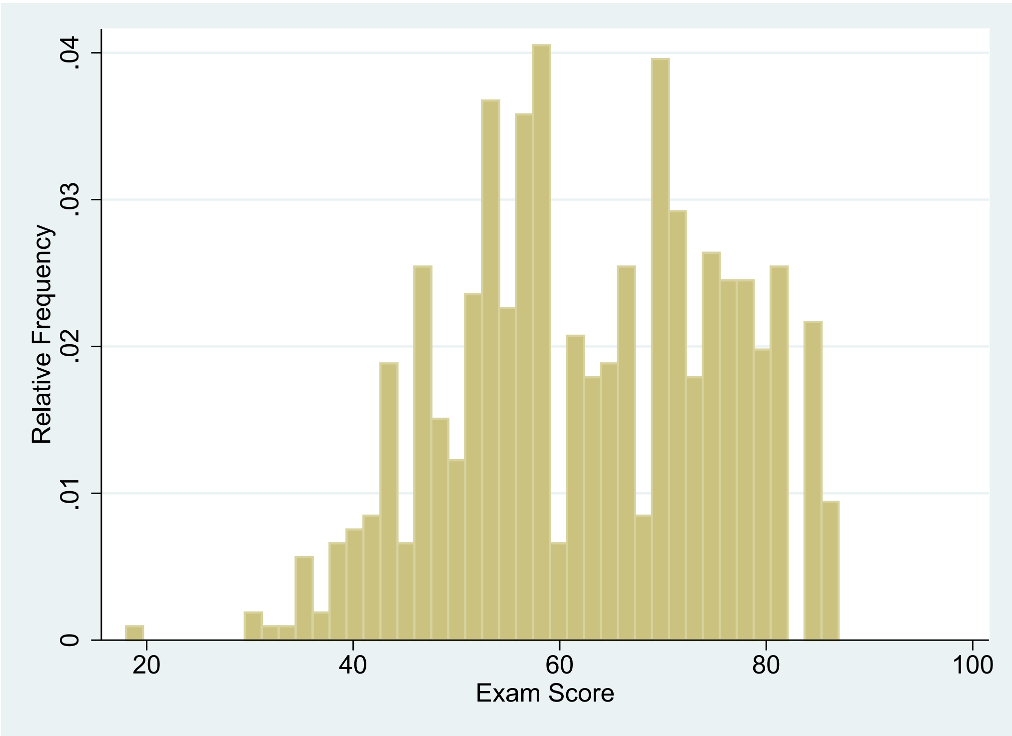

Figure 2: Histogram of exam score

about main field of study and its correlation with our main variables of interest. In particular, we split the sample in two subsamples, students of business administration or economics in one group and all other students including those studying business law or business education in the other group. As Figures 3 and 4 in the Appendix document, exam responses are markedly different in both groups. A large fraction of students in business administration or economics answer all or all but one questions and very few students answer less than 20 questions. By contrast, students from other fields answer between 11 and all questions; the distribution is much less skewed towards answering many questions than in the case of students in business administration or economics. As Figures 5 and 6 in Appendix A document, there is substantial heterogeneity across questions regarding response rates and success rates in the exam.

The data from the classroom experiment allow us to measure individual risk preferences. To avoid that the results depend on outliers, we categorize the measured degree of loss aversion in three categories from "loss-neutral or weakly loss-averse" to "strongly loss-averse".[23] We categorize students as follows

$$\lambda_i^c = \begin{cases} 1 \text{ "loss-neutral or weakly loss-averse"}, & \text{if } \lambda_i \leq 1.67; \\ 2 \text{ "loss-averse"}, & \text{if } \lambda_i \in (1.67, 3.33]; \\ 3 \text{ "strongly loss-averse"}, & \text{if } \lambda_i > 3.33. \end{cases}$$

Table 2 contains the descriptives of the mapping from cutoff values R in lottery series A (defined in Section 3.1) into categories of loss aversion λ^c. According to our categorization, students with cutoff values $R < -2$ are labelled "loss-neutral or weakly loss-averse" and those with $R > -1$ "strongly loss-averse"; students with intermediate cutoff values are labelled "loss-averse".

In Appendix A, we report the distribution of loss aversion parameters and their categorization for men and women separately. As one can see, women are not only categorized more frequently as strongly loss-averse than men and far less frequently as loss-neutral or weakly loss-averse, but women are also more frequently on the more loss-averse side within these categories (see Tables 19 and 20).

In addition, we ask students difficult general interest questions and also ask them to assess their performance relative to the average student. This gives an estimate of students' confidence which we measured in 10% steps (extracted from question 22 and 23 in the ques-

[23]Some students with the highest loss aversion score did not play a single loss lottery. This could be due to a lack of understanding the choice setting or due to excessive scepticism towards lotteries with negative payoff.

Table 2: Descriptive Statistics: Cutoffs in Lottery Series A and Loss Aversion Category

R	λ^c		
	"loss-neutral or weakly loss-averse"	"loss-averse"	"strongly loss-averse"
	1	2	3
-4	60	0	0
-3	53	0	0
-2.4	57	0	0
-1.8	0	199	0
-1.2	0	119	0
-0.6	0	0	77
0	0	0	81
Total	170	318	158

A choice in Lottery Series A is between a lottery with a 50-percent chance of winning 4 Euro and a 50-percent chance of losing $|R|$, and a sure payment of zero.

tionnaire; see Section 3.1).[24] The most-confident students expect to give 50% more correct answers than the average student; the least-confident students expecting to give 70% fewer correct answers than the average student. In the cognitive reflection test, students achieved a score between 0 and 3 with a mean of 1.77 correct answers.

Table 3: Correlation Loss Aversion and Other Explanatory Variables

	Loss Aversion	Confidence	Cognitive Reflection	Gender (F.)
Loss Aversion	1			
Confidence	-0.180***	1		
Cognitive Reflection	-0.207***	0.176***	1	
Gender (F.)	0.343***	-0.364***	-0.307***	1

Table 3: Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

In Table 3, we report the correlation of λ_i^c with the students' gender, cognitive reflection score and confidence. We find a highly significant, negative correlation between loss aversion and confidence. Furthermore, our measure of loss aversion is negatively correlated with students' cognitive reflection score (CRT). This and the fact that CRT is a rather crude measure of ability, gives support to the concern that our measure of loss aversion catches some unobserved low ability of students in line with Dohmen et al. (2010).[25] Female students tend to

[24]The direct measure of overconfidence based on a student's assessment of her performance (see question 22 in the questionnaire) turned out to be less powerful, possibly because the general interest questions were difficult and answers therefore noisy.

[25]Frederick (2005) finds that loss aversion is more prominent among subjects with a low cognitive reflection

be more loss-averse than their male classmates (cf. Tables 19 and 20 in Appendix A). There is no significant correlation of the risk measures with age. Table 18 in the Appendix reports the correlation coefficients of the main variables and main field of study. We note that the correlation coefficients between cognitive reflection score and main field of study (cf. the first column of Table 18) are in line with the average high school grade of students per field in 2013 (recorded in the admission process). In particular, students of business administration or economics have the highest average high school grades (1.34 and 1.51, respectively, on a scale from 1 to 5 with 1 being the highest grade) and studying one of these fields correlates significantly positively with the cognitive reflection score, whereas students of business education have the lowest average high school grade (2.63) and studying business education correlates significantly negatively with the cognitive reflection score.

4.2 Cross-Section Regressions

In this section we take a first shot at loss aversion as an explanatory variable of the students' behavior in the exam and test Hypothesis 1', 2' and 3'.

Table 4: OLS Regression: Number of Answered Questions

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.561***	0.486***	0.504***	0.438***
	(0.171)	(0.166)	(0.170)	(0.167)
Loss Aversion		-1.649***		-1.574***
		(0.382)		(0.383)
Strong Loss Aversion		-1.998***		-1.889***
		(0.475)		(0.474)
Confidence			0.340***	0.314***
			(0.102)	(0.102)
Gender (F.)	-2.026***	-1.544***	-1.600***	-1.177***
	(0.375)	(0.380)	(0.375)	(0.378)
Age	0.057	0.054	0.042	0.039
	(0.094)	(0.094)	(0.090)	(0.090)
Constant	21.916***	23.263***	25.738***	27.119***
	(2.101)	(2.113)	(1.807)	(1.847)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.3855	0.4051	0.3970	0.4145

Table 4: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

score; see his Table 3b. For a survey on the link between risk preferences and cognitive ability, see Dohmen et al. (2018).

Table 4 reports OLS regression results with the number of answered questions as the dependent variable. All independent variables are extracted from the classroom experiment. We report robust standard errors. We find that loss aversion (resp. confidence) have a negative (resp. positive) impact on the number of answered questions.[26] This effect is statistically significant at the 1% level. Cognitive reflection is strongly significant. In all our regressions, we include main field of study as fixed effects. Our reading of the regression results is that we find strong evidence in support of Hypothesis 1'. Our estimates suggest that loss-neutral students answer approximately 2 questions more than otherwise identical students in the highest category of loss aversion (and 5/3 more than those in the middle category). There is a significant gender difference, even after controlling for loss aversion and confidence. We find that the gender effect is partly explained by our measures of loss aversion and confidence (roughly 25% of it by loss aversion (diff $= -0.482$, p-value < 0.001) and 40% by the combination (diff $= -0.849$, p-value < 0.001). Including an interaction term of gender and loss aversion (or gender and confidence) would lead to statistically insignificant coefficients of all gender variables.

Table 5: OLS Regression: Number of Correct Answers Unconditionally

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.715***	0.643***	0.657***	0.593***
	(0.175)	(0.174)	(0.175)	(0.174)
Loss Aversion		-0.893**		-0.811*
		(0.431)		(0.431)
Strong Loss Aversion		-1.673***		-1.563***
		(0.522)		(0.521)
Confidence			0.310***	0.291***
			(0.109)	(0.109)
Gender (F.)	-2.060***	-1.676***	-1.664***	-1.331***
	(0.376)	(0.387)	(0.385)	(0.397)
Age	0.018	0.020	0.006	0.009
	(0.078)	(0.077)	(0.077)	(0.076)
Constant	17.204***	18.035***	21.358***	22.139***
	(1.856)	(1.864)	(1.602)	(1.637)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.4217	0.4311	0.4306	0.4388

Table 5: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

In Table 5, we present the regression results in which the number of correct answers is the

[26]For one student in our sample, the confidence measure was missing. Thus, the number of observations drops from 646 to 645 in columns 3 and 4 in Table 4.

dependent variable. Also in these regressions, loss aversion (resp. confidence) have a negative (resp. positive) impact on the dependent variable. The effect of strong loss aversion (resp. loss aversion) is statistically significant at the 1% (resp. 5% or 10%) level (cf. column (2) and (4)). Cognitive reflection is strongly significant. We interpret these results as strong evidence in support of Hypothesis 2'. Our estimate suggests that ceteris paribus students in the highest category of loss aversion give approximately 1.5 correct answers less than otherwise identical students who are loss-neutral. Also in these regressions, the gender difference is significant and partly explained by our measures of loss aversion and confidence (roughly 20% of it by loss aversion (diff = -0.384, p-value = 0.003) and 35% by the combination (diff = -0.729, p-value < 0.001)).

Table 6: OLS Regression: Number of Correct Answers/ Questions Answered

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.013**	0.013**	0.013**	0.012**
	(0.005)	(0.005)	(0.005)	(0.005)
Loss Aversion		0.020		0.021*
		(0.012)		(0.012)
Strong Loss Aversion		-0.004		-0.003
		(0.015)		(0.015)
Confidence			0.003	0.003
			(0.003)	(0.003)
Gender (F.)	-0.021**	-0.021*	-0.018	-0.018
	(0.010)	(0.011)	(0.011)	(0.012)
Age	-0.001	-0.000	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.766***	0.754***	0.810***	0.795***
	(0.051)	(0.053)	(0.045)	(0.047)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.1329	0.1399	0.1344	0.1416

Table 6: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 6 reports regression results with the ratio of correct answers per questions answered as the dependent variable. We do not find a statistically significant effect (at the 5%) of loss aversion nor confidence on the ratio of correct answers per questions answered. However, the coefficient of loss aversion (but not strong loss aversion) turns positively significant at the 10% level when considering loss aversion and confidence together (cf. column (4)). This implies that we find weak support for Hypothesis 3' in the cross section. Even though loss aversion and confidence have only small impact on the ratio of correctly answered questions, we note that including them turns gender insignificant.

Table 7: OLS Regression: Propensity to Gamble

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.098***	0.084**	0.087**	0.076**
	(0.035)	(0.034)	(0.035)	(0.035)
Loss Aversion		-0.386***		-0.373***
		(0.080)		(0.080)
Strong Loss Aversion		-0.390***		-0.370***
		(0.095)		(0.096)
Confidence			0.064***	0.059***
			(0.021)	(0.021)
Gender (F.)	-0.304***	-0.208***	-0.224***	-0.139*
	(0.075)	(0.077)	(0.077)	(0.078)
Age	0.013	0.012	0.010	0.009
	(0.018)	(0.019)	(0.018)	(0.018)
Constant	-0.407	-0.106	0.231	0.549
	(0.410)	(0.416)	(0.360)	(0.372)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.2939	0.3200	0.3050	0.3291

Table 7: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

In addition to testing Hypotheses 1',2', and 3', we report a number of additional meaningful regressions. In the previous section we introduced another variable related to the inclination to answer a question: the propensity to gamble. Our theory predicts that more-loss-averse students should have a lower propensity to gamble. This is indeed what we find—see Table 7 for the regression results with the propensity to gamble as the dependent variable. The effect is highly significant (at the 1% level). Cognitive reflection is again significant (with "smarter" people guessing more). In the regression in column 4, the gender effect is only statistically significant at the 10% level and loses more than half of its value compared to the specification in which individual loss aversion and confidence are not included as explanatory variables (see column 1, diff = -0.165, p-value < 0.001). Thus, we find that the gender effect, which says that women are more hesitant to answer, is to a large part explained by our measures of loss aversion and confidence. Including an interaction term of gender and loss aversion (or gender and confidence) would lead to statistically insignificant coefficients of all gender variables.

If loss-averse students take too few gambles, a student's performance should be worse if she is more loss-averse.[27] Table 8 reports OLS regressions in which the dependent variable is

[27] Specifically, Hypothesis 3' states that the conditional probability of answering a question correctly increases

Table 8: OLS Regression: Exam Score

	(1)	(2)	(3)	(4)
Cognitive Reflection	1.583***	1.443***	1.466***	1.341***
	(0.414)	(0.415)	(0.417)	(0.418)
Loss Aversion		-1.029		-0.860
		(1.055)		(1.055)
Strong Loss Aversion		-3.022**		-2.800**
		(1.262)		(1.261)
Confidence			0.590**	0.557**
			(0.263)	(0.264)
Gender (F.)	-4.154***	-3.483***	-3.391***	-2.816***
	(0.876)	(0.913)	(0.918)	(0.952)
Age	-0.002	0.007	-0.023	-0.013
	(0.172)	(0.168)	(0.173)	(0.169)
Constant	58.697***	59.843***	67.337***	68.297***
	(4.131)	(4.197)	(3.627)	(3.743)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.3682	0.3744	0.3749	0.3803

Table 8: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 9: OLS Regression: Exam Score including Propensity to Gamble

	(1)	(2)	(3)	(4)
Cognitive Reflection	1.267***	1.173***	1.194***	1.106***
	(0.406)	(0.409)	(0.409)	(0.412)
Propensity to Gamble	3.231***	3.196***	3.118***	3.097***
	(0.506)	(0.516)	(0.516)	(0.526)
Loss Aversion		0.205		0.294
		(1.043)		(1.043)
Strong Loss Aversion		-1.776		-1.654
		(1.233)		(1.233)
Confidence			0.391	0.375
			(0.262)	(0.262)
Gender (F.)	-3.171***	-2.820***	-2.692***	-2.384**
	(0.848)	(0.894)	(0.896)	(0.937)
Age	-0.045	-0.032	-0.055	-0.041
	(0.164)	(0.161)	(0.167)	(0.164)
Constant	60.013***	60.181***	66.616***	66.595***
	(3.937)	(4.008)	(3.472)	(3.601)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.4115	0.4152	0.4146	0.4181

Table 9: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

in the degree of loss aversion. The underlying mechanism is that more loss-averse students are less likely to answer a question than less loss-averse students, given that their expected gain in points is small. Missing out on these relatively small expected gains implies, though, that they should perform worse on average.

students' exam score. We find a significant negative (resp. positive) coefficient of our dummy of strong loss aversion (resp. confidence) on exam score (at the 5% level). This result on the effect of strong loss aversion is in line with Hypothesis 3'. However, this finding could also stem from the effect of unobserved ability being negatively correlated with loss aversion; see our panel estimates in the next subsection in which we try to disentangle those two channels. We also find that the gender effect is partly explained by our measures of loss aversion and confidence, in line with our earlier results (diff = −1.338, p-value = 0.002).

In Table 9 we provide evidence that the significantly negative effect of loss aversion on exam score becomes smaller and statistically insignificant when we include propensity to gamble as an explanatory variable. This suggests that guessing less when it pays is indeed the reason for which loss aversion affects performance. As an alternative to the propensity to gamble one could include the ratio of correctly answered questions to capture the inclination to gamble. However, this would not be very informative because the latter measure also captures competence.

4.3 Panel Data Estimation

In this section we consider a panel with students in the cross section and exam questions in the longitudinal dimension (cf. Table 10 to 14) to test Hypotheses 1, 2, and 3. As an estimation method, we use the random-effect logit model (with field fixed effects and clustered standard errors at the student level).[28] The question-specific regression equation can be written as

$$Pr\left(y_{ik} = 1 | x_{ik}, \nu_i\right) = F(x_{ik}\beta + \nu_i),$$

where ν_i represents the realization of student i's random effect and F is the logistic cdf. Variables y_{ik} and x_{ik} represent the dependent and independent variables per student i and question k, respectively. The vector β contains coefficients of interest.[29] The vector x_{ik} contains student-

[28]We used the same estimation method in the panel data estimations in Appendix B. As an alternative, we also ran Poisson regressions corresponding to those reported in Table 10 to 14, which confirm our qualitative findings.

[29]The joint conditional probability over all 29 questions equals

$$Pr\left(y_{i1}, ..., y_{i29} | x_{i1}, ..., x_{i29}, \nu_i\right) = \Pi_{k=1}^{29} F(x_{ik}\beta + \nu_i)^{\mathbf{1}_{\{y_{ik}=1\}}} \cdot (1 - F(x_{ik}\beta + \nu_i))^{1-\mathbf{1}_{\{y_{ik}=1\}}}.$$

Assuming a normally distributed student-specific random effect ν_i, the log likelihood of the previous equation can be approximated by the Gauss-Hermite quadrature and maximized accordingly.

specific variables such as gender, loss aversion or cognitive reflection score which do not vary by question as well as the time-trend variable in the micro and macro part of the exam which do not vary by student, cf. Table 10 to 14.

Table 10: Random-Effect Logit Regression: Answer a Question

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.286***	0.255***	0.267***	0.240***
	(0.058)	(0.057)	(0.057)	(0.057)
Loss Aversion		-0.590***		-0.565***
		(0.147)		(0.147)
Strong Loss Aversion		-0.684***		-0.645***
		(0.164)		(0.164)
Confidence			0.106***	0.097***
			(0.034)	(0.034)
Gender (F.)	-0.667***	-0.518***	-0.532***	-0.404***
	(0.120)	(0.118)	(0.122)	(0.119)
Time Micro	-0.704***	-0.703***	-0.702***	-0.702***
	(0.098)	(0.098)	(0.098)	(0.098)
Time Macro	-1.350***	-1.350***	-1.347***	-1.348***
	(0.093)	(0.093)	(0.093)	(0.093)
Cognitive Reflection × Micro	-0.132***	-0.132***	-0.132***	-0.132***
	(0.027)	(0.027)	(0.028)	(0.028)
Constant	2.115***	2.610***	2.102***	2.571***
	(0.268)	(0.284)	(0.269)	(0.287)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	18,734	18,734	18,705	18,705

Table 10 : Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

In Table 10, the dependent variable y_{ik} is whether student i answered question k or not. Loss aversion plays a highly significant role for students' answer probability. The coefficient of loss aversion is negatively significant at the 1% level in all specifications, which is in support of Hypothesis 1. Cognitive reflection shows a highly significant coefficient of expected sign in all regressions in this section. Our measure of confidence also shows a highly significant positive coefficient. The coefficient for female students is significantly negative. The size of the effect drops when introducing loss-aversion or confidence as explanatory variable or both. We further introduce a time trend in both, the micro part (the first 15 questions) and the macro part (the second 15 questions) of the exam. The reason is that the lecture is split into a micro and a macro part (taught by different lecturers) and students may start with the micro or the macro part when answering the exam. In all regressions in this section, the corresponding coefficients are significant at a 1% level. They are negative in all columns of

Table 10, indicating an increase in perceived difficulty per question in each part of the exam or an increasing time pressure. An interaction of the cognitive reflection score with a dummy for the micro part shows a highly significant coefficient, whose sign is negative. Since the micro part requires different skills than the macro part, it is not surprising that the effect of the CRT depends on whether a question is from the micro or the macro part. Our main take-away from this regression is that we find strong support for Hypothesis 1 and also confirm the result in the cross-section regression that the gender effect in answering a question is to a large part explained by our measures of loss aversion and confidence (about 40% of it; cf. columns (1) and (4), diff = -0.253, p-value = 0.033).

Table 11: Random-Effect Logit Regression: Unconditionally Correct Answer

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.199***	0.184***	0.189***	0.176***
	(0.033)	(0.033)	(0.033)	(0.033)
Loss Aversion		-0.171**		-0.157*
		(0.084)		(0.083)
Strong Loss Aversion		-0.311***		-0.291***
		(0.096)		(0.096)
Confidence			0.054***	0.050**
			(0.020)	(0.020)
Gender (F.)	-0.369***	-0.299***	-0.299***	-0.239***
	(0.069)	(0.069)	(0.070)	(0.071)
Time Micro	-1.013***	-1.013***	-1.017***	-1.017***
	(0.069)	(0.069)	(0.069)	(0.069)
Time Macro	-1.063***	-1.063***	-1.064***	-1.064***
	(0.076)	(0.076)	(0.076)	(0.076)
Cognitive Reflection × Micro	-0.134***	-0.133***	-0.133***	-0.133***
	(0.020)	(0.020)	(0.020)	(0.020)
Constant	0.911***	1.092***	0.910***	1.077***
	(0.166)	(0.176)	(0.168)	(0.178)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	18,734	18,734	18,705	18,705

Table 11: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 11 reports the estimates of a regression explaining the unconditional probability of providing the correct answer. As columns (2) and (4) show, loss-averse and strongly loss averse students are less likely to give the correct answer. We read this as strong support for Hypothesis 2, which may be due to the causal or the spurious effect. In line with our earlier findings, loss aversion and confidence explain a large part of the gender effect.

As shown in Section 2, a higher coefficient of loss aversion positively affects the students'

Table 12: Random-Effect Logit Regression: Correct Answer, Conditionally On Answering

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.142***	0.137***	0.138***	0.133***
	(0.035)	(0.035)	(0.036)	(0.036)
Loss Aversion		0.109		0.115
		(0.084)		(0.084)
Strong Loss Aversion		-0.049		-0.041
		(0.097)		(0.097)
Confidence			0.014	0.014
			(0.021)	(0.021)
Gender (F.)	-0.143**	-0.138*	-0.124*	-0.120
	(0.069)	(0.073)	(0.074)	(0.076)
Time Micro	-1.081***	-1.081***	-1.089***	-1.089***
	(0.084)	(0.084)	(0.084)	(0.084)
Time Macro	-0.632***	-0.631***	-0.637***	-0.636***
	(0.101)	(0.101)	(0.101)	(0.101)
Cognitive Reflection × Micro	-0.095***	-0.095***	-0.094***	-0.094***
	(0.024)	(0.024)	(0.024)	(0.024)
Constant	1.575***	1.547***	1.579***	1.547***
	(0.170)	(0.182)	(0.171)	(0.182)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	15,501	15,501	15,473	15,473

Table 12: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

response probability due to the causal or the spurious channel, or possibly both. If the causal effect dominates, we should also find empirical support for Hypothesis 3 and thus find that loss aversion is positive and significant for the probability of answering correctly conditional on answering. By contrast, if the spurious effect dominates, we should find that loss aversion is negative and significant for the probability of answering correctly conditional on answering. In Table 12, we report the estimates of students' conditional success probability; i.e., the probability that a student answers a question correctly conditional on answering it. According to these estimates, at a first look, we do not find empirical support for Hypothesis 3, i.e., that a more-loss-averse student has a higher conditional success probability than a less-loss-averse student (cf. the statistically insignificant coefficients for loss aversion and strong loss aversion in columns (2) and (4)). However, we also do not find support for the spurious channel being dominant. Furthermore, not only is loss aversion statistically insignificant, so is confidence.

The key take-away from Table 12 is that we do not find support for Hypothesis 3 for the whole sample. While this may suggest that both the causal and spurious effect essentially cancel each other out, we will take a closer look at the data next and come to a more differentiated conclusion.

An explanation for the insignificance of loss aversion might be related to the observation that a large number of students answered all or almost all questions; cf. Figure 1, in which we see a spike at answering 28 and 29 questions, whereas we do not observe such a spike at high exam scores, cf. Figure 2. The latter spike would have indicated that indeed many students did extremely well at the exam. Yet, this was not the case. The issue may be that, in contrast to what we postulated above, some students feel inclined to answer all questions.[30]

To remove behavior stemming from the temptation to answer all or almost all questions, we will look at two specifications. In our first specification, we look at the sub-sample of answers by students who did not answer at least two questions and find empirical support for Hypothesis 3. However, this sample split is based on choices. To address this concern, in our second specification, we split the sample exogenously according to broad field of study. We do so because we observe that students with main field of study for which introductory economics constitutes a core field course (economics or business administration) often answer all or almost all questions, but students from other fields do so less often (see Figures 3 and 4 in Appendix A). We find support for Hypothesis 3 among students whose major is neither economics nor business administration.

Taking a closer look at the first specification, the sample split is between students who answered 27 and fewer questions and those who answered 28 or 29 questions. This is provided in Table 13 (cf. columns (1) and (2) for the former and columns (3) and (4) for the latter).[31] As Table 13 reveals, the dummy for loss aversion enters significantly positive for students who answer few questions and the dummy for strong loss aversion enters significantly negative for students who respond to many (both at the 5 % level). This suggests that the causal effect is dominant for the former group and that the spurious effect is dominant for the latter group.[32] Overall, we read our findings as providing strong support of Hypothesis 3 for the subsample

[30] One reason may be that students may aim for a particular grade. What they know is that they receive the top grade if they have all or all but one questions correct. However, they do not know the mapping between points collected through correct and unanswered questions and the grades for fewer points as this mapping is not announced in advanced and has been varying over the years. For more discussion, see the end of this section.

[31] Many students apparently deemed questions 18 and slightly less so 14 too difficult and, thus, did not answer.

[32] In the latter subsample, we find that more-loss-averse students perform worse, which is compatible with the spurious effect being dominant. An explanation is that the subsample may consist mainly of observations from students who feel compelled to answer all questions. For these students, the causal effect is suppressed. In our regressions reported in Table 13 and, subsequently, Table 14 we did not include confidence because with split samples it did not show up significantly.

Table 13: Random-Effect Logit Regression: Correct Answer, Conditionally On Answering by Subsamples

	Low Sum of Answers:		High Sum of Answers:	
	(1)	(2)	(3)	(4)
Cognitive Reflection	0.137***	0.142***	0.098*	0.103*
	(0.047)	(0.047)	(0.056)	(0.056)
Loss Aversion		0.302**		-0.125
		(0.118)		(0.108)
Strong Loss Aversion		0.180		-0.342**
		(0.130)		(0.145)
Gender (F.)	-0.130	-0.157	-0.026	0.079
	(0.094)	(0.096)	(0.104)	(0.113)
Time Micro	-1.015***	-1.014***	-1.174***	-1.174***
	(0.111)	(0.111)	(0.127)	(0.127)
Time Macro	-0.739***	-0.736***	-0.484***	-0.484***
	(0.132)	(0.132)	(0.154)	(0.154)
Cognitive Reflection × Micro	-0.166***	-0.166***	-0.015	-0.015
	(0.034)	(0.034)	(0.033)	(0.033)
Constant	1.544***	1.345***	1.827***	1.847***
	(0.200)	(0.223)	(0.320)	(0.334)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	7,959	7,959	7,542	7,542

Table 13: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

of students who do not answer all or almost all questions.

As alluded to above, the splitting of the sample based on the number of answered questions may be criticized, as it is based on endogenous choices. As an alternative approach, we split the sample based on the main field of study. Choice behavior by students with business administration or economics as their major field of study tends to be different from those with a different main field of study—the former have, on average, higher grades in the exam and answer more questions. As Figure 3 and 4 in Appendix A document, being a student in business administration or economics and having a high answer ratio are strongly positively correlated.

To check for evidence for Hypothesis 3, we take a look at regression results in Table 14. In column 2 the coefficient for loss aversion is positive and statistically significant (at the 5% level). Thus students with main fields of study other than business administration and economics perform better conditional on answering a question if they are loss averse. We read this as strong evidence in support of Hypothesis 3 for this group of students. Note that this also rules out that in our sample only the spurious effect is present (which otherwise could have been a sign that our measure of loss aversion only picked up unobserved ability).

Table 14: Random-Effect Logit Regression: Correct Answer, Conditionally On Answering by Subsamples

	Other Fields:		Business Adm'n and Econ:	
	(1)	(2)	(3)	(4)
Cognitive Reflection	0.152***	0.157***	0.137***	0.137***
	(0.054)	(0.055)	(0.047)	(0.047)
Loss Aversion		0.251**		-0.022
		(0.122)		(0.114)
Strong Loss Aversion		0.135		-0.242*
		(0.140)		(0.135)
Gender (F.)	-0.198*	-0.227**	-0.097	-0.029
	(0.103)	(0.106)	(0.092)	(0.097)
Time Micro	-1.037***	-1.036***	-1.113***	-1.111***
	(0.120)	(0.120)	(0.120)	(0.120)
Time Macro	-0.648***	-0.643***	-0.619***	-0.620***
	(0.152)	(0.152)	(0.132)	(0.132)
Cognitive Reflection × Micro	-0.132***	-0.131***	-0.071**	-0.071**
	(0.038)	(0.038)	(0.032)	(0.032)
Constant	1.615***	1.453***	1.856***	1.897***
	(0.226)	(0.256)	(0.178)	(0.195)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	6,212	6,212	9,289	9,289

Table 14: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

By contrast, in the subsample of students in business administration or economics those students with strong loss aversion perform worse conditionally on answering a question (however, it is significant only at the 10% level). This means that, for this group, strongly loss-averse students are more likely to answer a question incorrectly conditional on answering than those who are not loss-averse. Our interpretation of this result is that the spurious effect dominates the causal effect of loss aversion for this group of students. One reason for this result could be that these students feel compelled to aim at answering all or almost all questions, possibly because they aim at having a chance for the top grade. The gender effect is present in this estimation for the sample of students who do not have economics or business administration as their major. The inclusion of confidence is statistically significant but does not alter the above finding.

Overall, we read our findings as evidence in support of the causal effect of loss aversion, at least for students who are rather unlikely to answer all or almost all questions. For other students we do not find such evidence. This may be due to the spurious effect cancelling out or even dominating the causal effect for those students. The causal effect may be absent for business and economics students, e.g. because many of them simply aim to answer all questions. Another reason could be that these students are broad bracketers and, therefore, loss aversion relates to a reference point about the total number of points in the exam. This would imply that more-loss-averse students answer *more* questions than less-loss-averse students, which would imply that Hypothesis 1 is violated for this group of students. This is not what we found, i.e. Hypothesis 1 also holds for business and economics students (see column (4) in Table 32 in Appendix D).

5 Discussion and Conclusion

In this paper, we show that more-loss-averse students are less inclined to answer an exam question than less-risk-averse students if a wrong answer gives a lower score than no response. Thus, if students have the correct probabilistic assessment, more-loss-averse students will perform worse. Loss aversion parameters are extracted from a classroom experiment of lottery choices conducted three months prior to the exam.[33] As we show, loss aversion in such a low-stake environment explains performance in a different, high-stake environment a few months down the road. As we also show, risk aversion does not explain behavior.[34] Differences in the inferred loss-aversion parameters to a large part explain the gender gap that, in line with the literature, we observe in our field data.

According to a university directive, the differential treatment of wrong and no responses was no longer allowed after the academic year 2013/2014, which is the exam year we used in this paper. In the Fall of 2014, we observe that the gender difference in exam score was much lower than in 2013. Controlling for field fixed effects and normalizing the coefficient of the gender dummy by its standard error, we estimate a gender gap in favor of male students of only 4.70 (with a sample of 1008 students) in 2014 instead of 7.16 (with a sample of 936 students) in 2013; the R^2 in the regression was 0.294 in 2014 and 0.380 in 2013, respectively.[35] Assuming that the level of difficulty and the pool of students in both exams was similar, this

[33] Our elicitation method can easily be used in classroom experiments and could even be integrated into surveys because it relies on a small number of lottery choices.

[34] In our questionnaire, we also obtained a non-incentivized measure of risk preferences and a measure of regret (see instructions; questions about behavior I and II). We checked that also these measures do not explain behavior in the exam.

[35] Variables on cognitive reflection and risk preferences are not available for 2014.

finding can be explained by loss aversion: the more-loss-averse gender was less disadvantaged by the new multiple-choice setup according to which incorrect answers were not punished and, thus, answering became the preferred action for all questions irrespective of the degree of loss aversion. This suggests that the exam with punishment for incorrect answers partly measured loss aversion rather than ability, which warrants caution in the use of such punishment. In our sample, this applies particularly to students in fields other than economics and business administration where we found that the causal effect outweighs the spurious effect. Since these are the on average weaker students, loss aversion is likely to make the critical difference between pass and fail for some of these students. This runs counter to the purpose of the exam to assess a student's knowledge.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman, "Reference Points and Effort Provision," *American Economic Review*, 2011, *101* (2), 470–492.

Akyol, Pelin, James Key Key, and Kala Krishna, "Hit or Miss? Test Taking Behavior in Multiple Choice Exams," *NBER Working Paper No. 22401*, 2016.

Baldiga, Katherine, "Gender Differences in Willingness to Guess," *Management Science*, 2014, *60* (2), 434–448.

Bell, David E., "Disappointment in Decision Making under Uncertainty," *Operations Research*, 1985, *33* (1), 1–27.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, "Labor Supply of New York City Cabdrivers: One Day at a Time," *Quarterly Journal of Economics*, 1997, *112* (2), 407–441.

Card, David and Gordon B. Dahl, "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior," *Quarterly Journal of Economics*, 2011, *126* (1), 103–143.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv, "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review*, 2007, *97* (5), 1921–1938.

Crawford, Vincent P. and Juanjuan Meng, "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with RationalExpectations Targets for Hours and Income," *American Economic Review*, 2011, *101* (5), 1912–1932.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, June 2009, *47* (2), 315–372.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "Are Risk Aversion and Impatience Related to Cognitive Ability?," *American Economic Review*, 2010, *100* (3), 1238–60.

_ , _ , _ , and _ , "On the Relationship Between Cognitive Ability and Risk Preference," *Journal of Economic Perspectives*, 2018, *32* (2), 115–134.

Engelmann, Dirk and Guillaume Hollard, "Reconsidering the Effect of Market Experience on the "Endowment Effect"," *Econometrica*, 2010, *78* (6), 2005–2019.

Ericson, Keith M. Marzilli and Andreas Fuster, "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments," *Quarterly Journal of Economics*, 2011, *126* (4), 1879–1907.

__ **and** __ , "The Endowment Effect," *Annual Review of Economics*, 2014, *6* (1), 555–579.

Farber, Henry S., "Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers," *Journal of Political Economy*, 2005, *113* (1), 46–82.

__ , "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers," *American Economic Review*, 2008, *98* (3), 1069–1082.

__ , "Why you Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers," *Quarterly Journal of Economics*, 2015, *130* (4), 1975–2026.

Fehr, Ernst and Lorenz Götte, "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment," *American Economic Review*, 2007, *97* (1), 298–317.

Frederick, Shane, "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, 2005, *19* (4), 25–42.

Funk, Patricia and Helena Perrone, "Gender Differences in Academic Performance: The Role of Negative Marking in Multiple-Choice Exams," 2017.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann, "Individual-Level Loss Aversion in Riskless and Risky Choices," *Discussion Paper, University of Nottingham*, 2007.

Gill, David and Victoria Prowse, "A Structural Analysis of Disappointment Aversion in a Real Effort Competition," *American Economic Review*, February 2012, *102* (1), 469–503.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann, "The Limits of Expectations-Based Reference Dependence," *Journal of the European Economic Association*, 2017, *15* (4), 861–876.

Gul, Faruk, "A Theory of Disappointment Aversion," *Econometrica*, 1991, *95* (3), 667–686.

Heffetz, Ori and John A. List, "Is the Endowment Effect an Expectations Effect?," *Journal of the European Economic Association*, 2014, *12* (5), 1396–1422.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *The American Economic Review*, 2002, *92* (5), 1644–1655.

Hoppe, Eva I. and David J. Kusterer, "Behavioral Biases and Cognitive Reflection," *Economics Letters*, 2011, *110* (2), 97 – 100.

Iriberri, Nagore and Pedro Rey-Biel, "Brave Boys and Play-it-Safe Girls: Gender Differences in Willingness to Guess in a Large Scale Natural Field Experiment," *Working Paper*, 2019.

Kahneman, Daniel and Amos Tversky, "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 1979, *47* (2), 263–291.

Karle, Heiko, Georg Kirchsteiger, and Martin Peitz, "Loss Aversion and Consumption Choice: Theory and Experimental Evidence," *American Economic Journal: Microeconomics*, 2015, *7* (2), 101–120.

Kőszegi, Botond and Matthew Rabin, "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

_ and _ , "Reference-Dependent Risk Attitudes," *American Economic Review*, September 2007, *97* (4), 1047–1073.

Köbberling, Veronika and Peter P. Wakker, "An Index of Loss Aversion," *Journal of Economic Theory*, 2005, *122* (1), 119 – 131.

Loomes, Graham and Robert Sugden, "Testing for Regret and Disappointment in Choice under Uncertainty," *Economic Journal*, 1987, *97*, 118–129.

Pope, Devin G. and Maurice E. Schweitzer, "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes," *American Economic Review*, 2011, *101*, 129–157.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler, "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show," *American Economic Review*, 2008, *98* (1), 38–71.

Rabin, Matthew, "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 2000, *68* (5), 1281–1292.

Smith, Alec, "Lagged beliefs and reference-dependent utility," *Journal of Economic Behavior & Organization*, 2019, *167*, 331 – 340.

Tversky, Amos and Daniel Kahneman, "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty*, 1992, *5*, 297–323.

Appendix

A Further Descriptive Statistics

A.1 Tables

Table 15: Descriptive Statistics: Male Students

Variable	Obs	Mean	Std. Dev.	Min	Max
Answered Questions	367	25.3488	4.6452	11	29
Correct Answers	367	20.7929	5.2641	5	29
Exam Score	367	66.0300	12.4058	18	87
Propensity to Gamble	367	0.7541	0.2623	0.0811	1
Confidence	366	0.03989	1.6360	-4.6	5
Cognitive Reflection	367	2.054	0.9903	0	3
Age	367	19.4092	2.4678	16	37

Table 16: Descriptive Statistics: Female Students

Variable	Obs	Mean	Std. Dev.	Min	Max
Answered Questions	279	22.2151	5.6035	11	29
Correct Answers	279	17.2760	5.6992	6	29
Exam Score	279	58.6129	12.6889	30	87
Propensity to Gamble	279	0.6249	0.2716	0.0968	1
Confidence	279	-1.2518	1.6501	-7	3
Cognitive Reflection	279	1.3901	1.0690	0	3
Age	279	19.5248	1.7252	17	27

Table 17: Descriptive Statistics: Students per field

Field	Obs	Freq.	% Female
Business Administration	249	38.54	37.35
Business Law	136	21.05	42.65
Business Education	107	16.56	68.22
Economics	99	15.33	31.31
Others	55	8.51	43.64
Total	646	100.00	43.19

Table 18: Descriptive Statistics: Correlation coefficients of main variables and field of study

Field	Cognitive Reflection	Loss Aversion	Confidence
Business Administration	0.1549***	-0.0776**	-0.0246
	(0.000)	(0.049)	(0.533)
Business Law	-0.0728*	0.0081	0.0010
	(0.065)	(0.837)	(0.980)
Business Education	-0.3068***	0.0584	-0.0907**
	(0.000)	(0.138)	(0.021)
Economics	0.1830***	0.0111	0.0591
	(0.000)	(0.778)	(0.134)
Others	0.0087	0.0313	0.0856**
	(0.826)	(0.427)	(0.030)

Table 18: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 19: Cutoffs in Lottery Series A and Loss Aversion Category: Male

R	λ^c		
	"loss-neutral or weakly loss-averse"	"loss-averse"	"strongly loss-averse"
	1	2	3
-4	54	0	0
-3	46	0	0
-2.4	37	0	0
-1.8	0	115	0
-1.2	0	63	0
-0.6	0	0	30
0	0	0	22
Total	137	178	52

A choice in Lottery Series A is between a lottery with a 50-percent chance of winning 4 Euro and a 50-percent chance of losing $|R|$, and a sure payment of zero.

Table 20: Cutoffs in Lottery Series A and Loss Aversion Category: Female

R	λ^c		
	"loss-neutral or weakly loss-averse"	"loss-averse"	"strongly loss-averse"
	1	2	3
-4	6	0	0
-3	7	0	0
-2.4	20	0	0
-1.8	0	84	0
-1.2	0	56	0
-0.6	0	0	47
0	0	0	59
Total	33	140	106

A choice in Lottery Series A is between a lottery with a 50-percent chance of winning 4 Euro and a 50-percent chance of losing $|R|$, and a sure payment of zero.

A.2 Figures

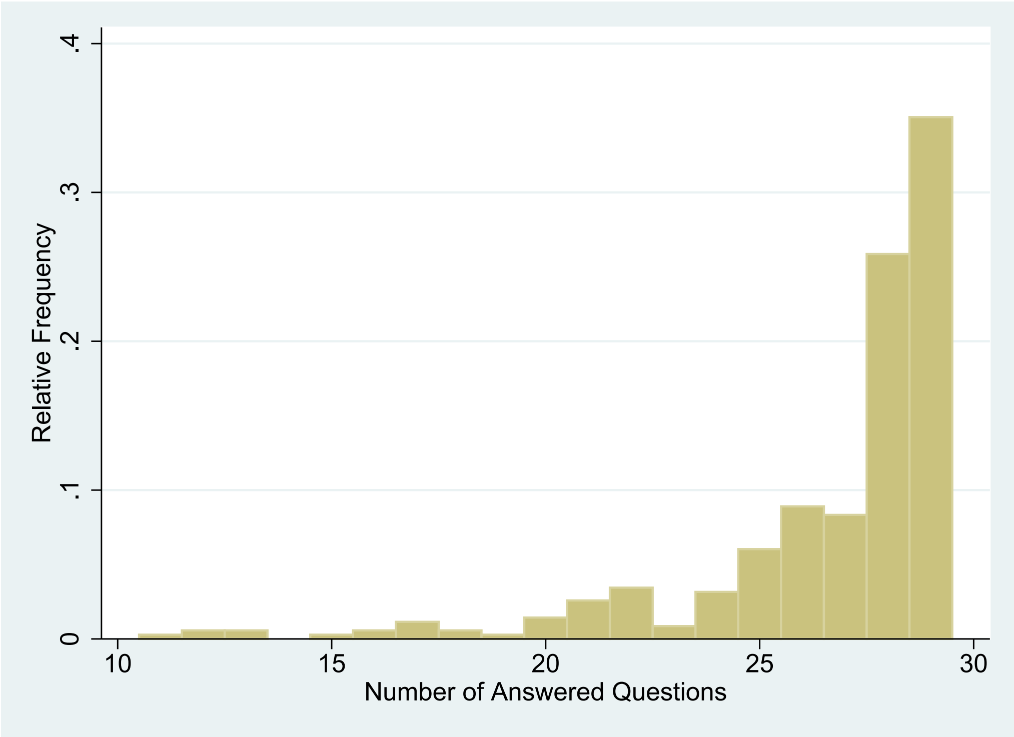

Figure 3: Histogram of answered questions (Business and Econ)

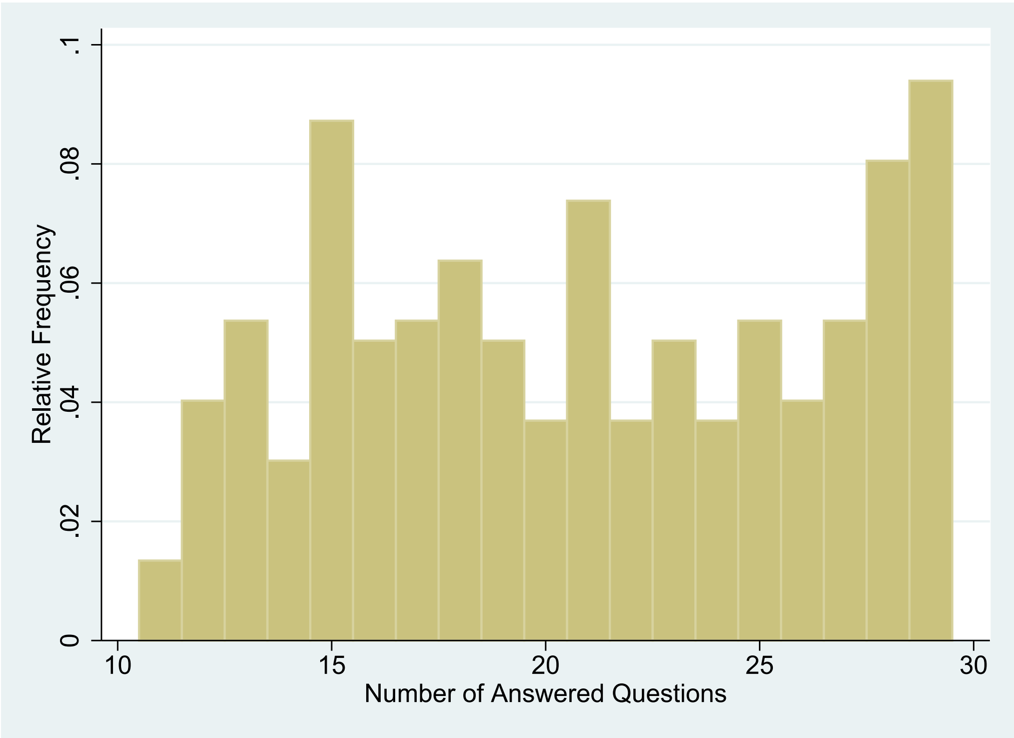

Figure 4: Histogram of of answered questions (no Business or Econ)

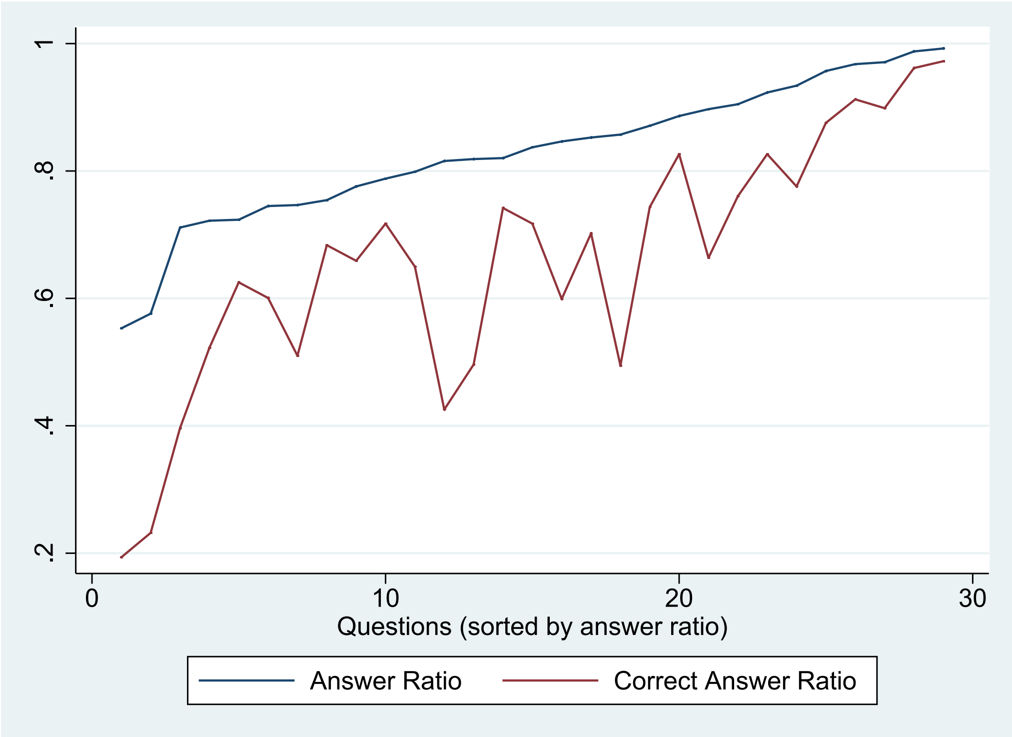

Figure 5: Ratio of answers per question and ratio of correct answers (relative to total number of students) per question; sorted

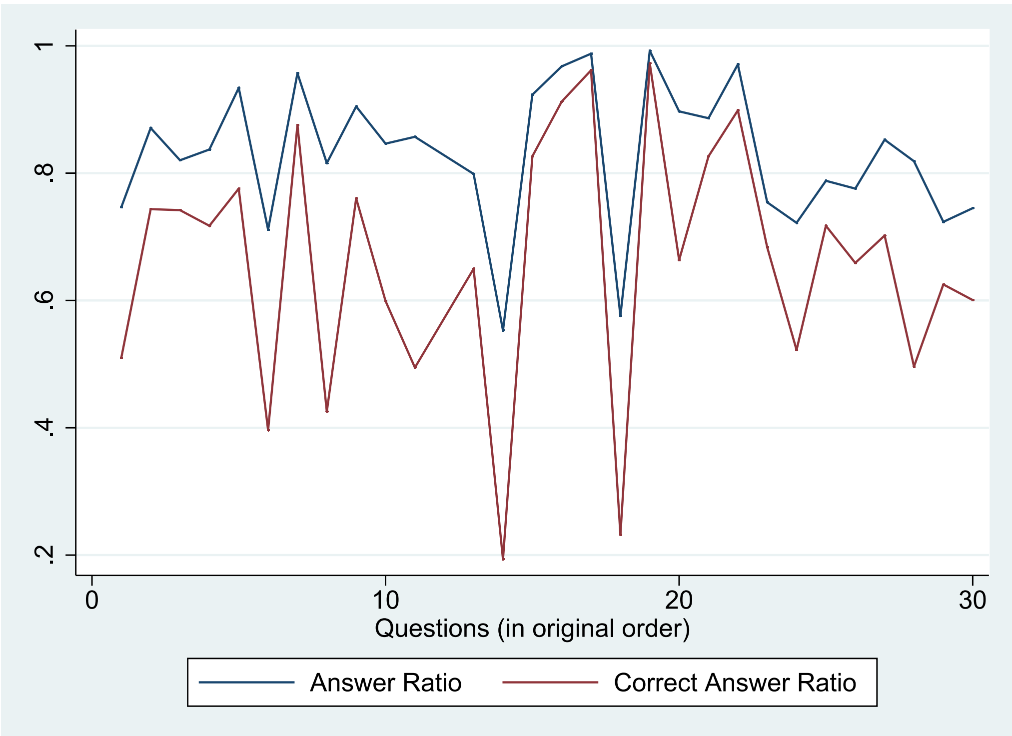

Figure 6: Ratio of answers per question and ratio of correct answers (relative to total number of students) per question; unsorted

Comments on Figures 5 and 6: Figure 5 illustrates that students' perception of the difficulty of a question (measured by ratio of answers per question) and the actual ratio of correct answers are highly positively correlated. The correlation coefficient is 0.8866. The ratio of correct answers in Figure 5 is defined relative to the total number of students instead of the number of students who answered a particular question. This definition implies that the red line can at most touch the blue but never cross it. Questions are sorted by the ratio of answers per question, i.e. from questions which are perceived as difficult to those perceived as easy. Therefore the blue line is increasing by construction. The red line does not always move parallel to the blue line. It does so for the questions which are perceived as the most easy and the most difficult ones but not necessarily for questions perceived as intermediately difficult. Figure 6 is a different representation in which questions are in the original order.

B Alternative Specification of the Degree of Loss Aversion

In this appendix, we provide the results of the empirical analysis that includes individial risk parameters β_i and the curvature-adjusted measure of loss aversion, which incorporates the cutoffs of both, series A and B.

To account for risk parameters different from 1,[36] cutoff values of series B can be used to represent the risk parameter β_i because all those lottery outcomes lie in the positive domain (with reference point $r = 0$).[37] Using the condition that the utility of receiving the safe payment S_i must be equal to the expected utility of getting 4 Euros with a 50-percent chance and zero otherwise, i.e. $1/2 \cdot 4^{\beta_i} = S_i^{\beta_i}$, we obtain as an inverse measure for risk aversion,

$$\beta_i = \frac{\ln(2)}{\ln(4) - \ln(S_i)}. \tag{4}$$

In our empirical analysis, we use $1 - \beta$ as our direct measure of risk aversion, reflecting the

[36] This is in conflict with Rabin (2000)'s critique if the exam constitutes a high-stake environment.

[37] Theoretically, allowing for a reference point larger than zero, lottery series B could be considered as having a quasi-mixed domain. A natural candidate for a reference point larger than zero is the safe payment S_i. In that case, a measure of loss aversion in line with Rabin (2000)'s critique (i.e. $\beta = 1$) could be derived from series B using the power utility representation of Tversky and Kahneman (1992). The cutoff condition $1/2 \cdot (-\lambda_{Bi}(S_i - 0)) + 1/2 \cdot (4 - S_i) = S_i - S_i$ leads to $\lambda_{Bi} = (4 - S_i)/S_i$. This measure of loss aversion is strictly decreasing in S_i and highly positively correlated with $1 - \beta_i$ leading to qualitatively similar regression results as $1 - \beta_i$ in all specifications. The regression results using λ_{Bi} can be obtained by the authors upon request. We interpret the finding that the risk measures derived from lottery series B have less predictive power for explaining answer behavior in the exam than those derived from lottery series A as evidence that students did not perceive lottery series B as having a quasi-mixed domain but a positive one with $r = 0$.

degree of constant relative risk aversion (CRRA).

We note that β_i is increasing in S_i, as follows from equation (4). While we use $1-\beta_i$ as regressor in our empirical analysis, we could use the cutoff values S_i interchangeably instead.

We combine the cutoffs of series A and B in order to derive a curvature-adjusted measure of loss aversion $\tilde{\lambda}_i$. For given β_i from series B and from the cutoff condition $0 = 1/2 \cdot 4^{\beta_i} + 1/2 \cdot (-\tilde{\lambda}_i)(-R_i)^{\beta_i}$, we obtain the degree of loss aversion of subject i

$$\tilde{\lambda}_i = \left(\frac{4}{|R_i|}\right)^{\beta_i}. \tag{5}$$

We note that, given β_i, $\tilde{\lambda}_i$ is increasing in R_i, as follows from equation (5). Only in this alternative specification, it is consistent with our identifying assumptions to consider loss and risk aversion as explanatory variables in the same regression.

We categorize the inverse measure of risk aversion from "risk-averse " to "risk-neutral or weakly risk-loving ",

$$\beta_i^c = \begin{cases} 0.25 \text{ "strongly risk-averse"}, & \text{if } \beta_i \leq 0.431; \\ 0.75 \text{ "risk-averse"}, & \text{if } \beta_i \in (0.431, 1); \\ 1.25 \text{ "risk-neutral or weakly risk-loving"}, & \text{if } \beta_i \geq 1. \end{cases}$$

Table 21 contains the descriptives of the mapping from cutoff values S in lottery series B into categories of risk aversion $1-\beta^c$. According to our categorization, students with cutoff value $S \geq 2$ are labelled "risk-neutral or weakly risk-averse" and those with $S < 1$ "strongly risk-averse". Students with intermediate values are labelled "risk-averse".

We also categorized the curvature-adjusted measure of loss aversion in three categories:

$$\tilde{\lambda}_i^c = \begin{cases} 1 \text{ "loss-neutral or weakly loss-averse"}, & \text{if } \tilde{\lambda}_i \leq 1.5; \\ 2 \text{ "weakly loss-averse"}, & \text{if } \tilde{\lambda}_i \in (1.5, 2]; \\ 3 \text{ "strongly loss-averse"}, & \text{if } \tilde{\lambda}_i > 2. \end{cases}$$

171 students are in category $\tilde{\lambda}_i^c = 1$, 255 in category $\tilde{\lambda}_i^c = 2$, and 220 category $\tilde{\lambda}_i^c = 3$. Table 22 reports correlations using the alternative risk preferences.

Table 21: Descriptive Statistics: Cutoffs in Lottery Series B and Risk Aversion Category

S	$1-\beta^c$		
	"risk-neutral or weakly risk-loving"	"risk-averse"	"strongly risk-averse"
	-0.25	0.25	0.75
0	0	0	29
0.4	0	0	9
0.8	0	0	37
1.2	0	160	0
1.6	0	264	0
2	59	0	0
2.4	88	0	0
Total	147	424	75

A choice in Lottery Series B is between a lottery with a 50-percent chance of winning 4 Euro and a 50-percent chance of winning zero, and a sure payment of S.

Table 22: Correlation between Alternative Risk Preferences and Other Variables

	Loss Av'n	Risk Av'n	Confidence	Cognitive Refl'n	Gender (F.)
Loss Aversion	1				
Risk Aversion	-0.0158	1			
Confidence	-0.119**	-0.145***	1		
Cognitive Refl'n	-0.137***	-0.129**	0.176***	1	
Gender (F.)	0.286***	0.174***	-0.364***	-0.307***	1

Table 22: Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

B.1 Cross-Section Regressions

The comparison of Table 23 vs. 4, Table 24 vs. 7, and Table 25 vs. 8, respectively shows that the alternative (or more precisely; curvature-adjusted) measure of loss aversion leads to only slightly less significant coefficients when the number of answered questions and propensity to gamble is the dependent variable. In most regressions, the risk aversion parameter is not statistically significant (an exception is that strong risk aversion negatively affects exam score). We also ran the regressions only with the inferred risk aversion dummies (but without the loss aversion dummies). In these regressions, which are available upon request, risk aversion shows no or low significance. We interpret our findings regarding the risk aversion dummies as evidence that at least in our context a lottery series with a positive domain is less well suited to elicit a measure of risk preferences that matters for exam choices (with mixed domain) than a lottery series with a mixed domain.

Regarding the number of correct answers loss-averse students continue to do worse (the level of significance changes in opposite direction for the two dummies) confirming the finding in the main text; compare Table 26 to Table 5. Risk aversion enters significantly with the expected sign for strongly risk-averse students. Also the finding regarding the effect of risk preferences on the number of correct answers conditional on answering are broadly in line with the finding in the main text (compare Table 27 to Table 6; in particular, row 4 of the respective tables). However, we lose the 10% significance of the dummy for loss-averse students.

Table 23: OLS Regression: Number of Answered Questions

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.561***	0.499***	0.504***	0.447***
	(0.171)	(0.167)	(0.170)	(0.167)
Loss Aversion		-1.290***		-1.246***
		(0.408)		(0.406)
Strong Loss Aversion		-1.263***		-1.221***
		(0.415)		(0.412)
Risk Aversion		0.358		0.469
		(0.378)		(0.375)
Strong Risk Aversion		-1.257*		-1.132*
		(0.651)		(0.648)
Confidence			0.340***	0.335***
			(0.102)	(0.106)
Gender (F.)	-2.026***	-1.657***	-1.600***	-1.263***
	(0.375)	(0.384)	(0.375)	(0.382)
Age	0.057	0.067	0.042	0.053
	(0.094)	(0.095)	(0.090)	(0.091)
Constant	21.916***	22.501***	25.738***	26.352***
	(2.101)	(2.181)	(1.807)	(1.911)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.3855	0.4041	0.3970	0.4147

Table 23: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 24: OLS Regression: Propensity to Gamble

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.098***	0.088**	0.087**	0.079**
	(0.035)	(0.035)	(0.035)	(0.035)
Loss Aversion		-0.311***		-0.304***
		(0.085)		(0.085)
Strong Loss Aversion		-0.308***		-0.300***
		(0.084)		(0.083)
Risk Aversion		0.105		0.127
		(0.079)		(0.078)
Strong Risk Aversion		-0.080		-0.056
		(0.124)		(0.123)
Confidence			0.064***	0.064***
			(0.021)	(0.021)
Gender (F.)	-0.304***	-0.229***	-0.224***	-0.155*
	(0.075)	(0.078)	(0.077)	(0.079)
Age	0.013	0.014	0.010	0.011
	(0.018)	(0.019)	(0.018)	(0.018)
Constant	-0.407	-0.270	0.231	0.374
	(0.410)	(0.430)	(0.360)	(0.383)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.2939	0.3140	0.3050	0.3246

Table 24: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 25: OLS Regression: Exam Score

	(1)	(2)	(3)	(4)
Cognitive Reflection	1.583***	1.488***	1.466***	1.383***
	(0.414)	(0.410)	(0.417)	(0.413)
Loss Aversion		-1.787		-1.670
		(1.115)		(1.115)
Strong Loss Aversion		-1.325		-1.221
		(1.128)		(1.128)
Risk Aversion		0.032		0.189
		(1.021)		(1.023)
Strong Risk Aversion		-2.933**		-2.733*
		(1.421)		(1.432)
Confidence			0.590**	0.568**
			(0.263)	(0.263)
Gender (F.)	-4.154***	-3.630***	-3.391***	-2.943***
	(0.876)	(0.912)	(0.918)	(0.951)
Age	-0.002	0.012	-0.023	-0.005
	(0.172)	(0.172)	(0.173)	(0.173)
Constant	58.697***	59.784***	67.337***	68.379***
	(4.131)	(4.392)	(3.627)	(3.964)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.3682	0.3762	0.3749	0.3821

Table 25: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 26: OLS Regression: Number of Correct Answers Unconditionally

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.715***	0.662***	0.657***	0.610***
	(0.175)	(0.172)	(0.175)	(0.172)
Loss Aversion		-1.026**		-0.972**
		(0.452)		(0.451)
Strong Loss Aversion		-0.863*		-0.814*
		(0.464)		(0.463)
Risk Aversion		0.130		0.219
		(0.417)		(0.417)
Strong Risk Aversion		-1.397**		-1.289**
		(0.614)		(0.616)
Confidence			0.310***	0.301***
			(0.109)	(0.110)
Gender (F.)	-2.060***	-1.762***	-1.664***	-1.402***
	(0.376)	(0.387)	(0.385)	(0.396)
Age	0.018	0.026	0.006	0.016
	(0.078)	(0.079)	(0.077)	(0.078)
Constant	17.204***	17.762***	21.358***	21.910***
	(1.856)	(1.959)	(1.602)	(1.746)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.4217	0.4335	0.4306	0.4415

Table 26: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 27: OLS Regression: Number of Correct Answers/ Questions Answered

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.013**	0.013**	0.013**	0.013**
	(0.005)	(0.005)	(0.005)	(0.005)
Loss Aversion		0.004		0.005
		(0.013)		(0.014)
Strong Loss Aversion		0.010		0.010
		(0.013)		(0.014)
Risk Aversion		-0.006		-0.005
		(0.012)		(0.012)
Strong Risk Aversion		-0.021		-0.020
		(0.018)		(0.018)
Confidence			0.003	0.002
			(0.003)	(0.003)
Gender (F.)	-0.021**	-0.022*	-0.018	-0.019
	(0.010)	(0.011)	(0.011)	(0.012)
Age	-0.001	-0.000	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.766***	0.765***	0.810***	0.808***
	(0.051)	(0.053)	(0.045)	(0.048)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	646	646	645	645
R square	0.1329	0.1356	0.1344	0.1371

Table 27: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

B.2 Panel Data Estimation

We consider the results of the panel regressions when the alternative measure of loss aversion is used. Comparing Table 28 vs. 10, Table 29 vs. 11, Table 30 vs. 12 we find our qualitative findings confirmed: Loss aversion negatively affects the likelihood to answer and more-loss-averse students are less likely to get the correct answer, predicted both by the causal and the spurious channel. There is no significant effect on giving the correct answer conditionally on answering; see Table 30. Moreover, risk aversion is not significant in most regressions (not significant or significant only at the 10%).

Doing the exogenous sample split, in the subsample of students in other fields of study, we find a positive effect of loss aversion on giving the correct answer conditionally on answering; risk aversion is not significant. The positive effect of loss aversion is significant at the 10% level for strongly loss-averse students; see Table 31 and compare to the corresponding Table 14 in the main text. Our take-away from this alternative specification is that the findings in the main text regarding the exogenous sample split are robust (despite the drop in statistic significance).

Table 28: Random-Effect Logit Regression: Answer a Question

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.286***	0.263***	0.267***	0.247***
	(0.058)	(0.057)	(0.057)	(0.057)
Loss Aversion		-0.470***		-0.461***
		(0.150)		(0.149)
Strong Loss Aversion		-0.506***		-0.491***
		(0.147)		(0.145)
Risk Aversion		0.162		0.208
		(0.135)		(0.134)
Strong Risk Aversion		-0.191		-0.138
		(0.189)		(0.190)
Confidence			0.106***	0.107***
			(0.034)	(0.035)
Gender (F.)	-0.667***	-0.556***	-0.532***	-0.431***
	(0.120)	(0.119)	(0.122)	(0.120)
Time Micro	-0.704***	-0.704***	-0.702***	-0.702***
	(0.098)	(0.098)	(0.098)	(0.098)
Time Macro	-1.350***	-1.350***	-1.347***	-1.348***
	(0.093)	(0.093)	(0.093)	(0.093)
Cognitive Reflection × Micro	-0.132***	-0.132***	-0.132***	-0.132***
	(0.027)	(0.027)	(0.028)	(0.027)
Constant	2.115***	2.405***	2.102***	2.346***
	(0.268)	(0.294)	(0.269)	(0.295)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	18,734	18,734	18,705	18,705

Table 28 : Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 29: Random-Effect Logit Regression: Unconditionally Correct Answer

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.199***	0.189***	0.189***	0.179***
	(0.033)	(0.033)	(0.033)	(0.033)
Loss Aversion		-0.205**		-0.197**
		(0.087)		(0.087)
Strong Loss Aversion		-0.185**		-0.176**
		(0.088)		(0.087)
Risk Aversion		0.024		0.043
		(0.079)		(0.079)
Strong Risk Aversion		-0.216*		-0.193*
		(0.110)		(0.111)
Confidence			0.054***	0.053***
			(0.020)	(0.020)
Gender (F.)	-0.369***	-0.312***	-0.299***	-0.249***
	(0.069)	(0.069)	(0.070)	(0.071)
Time Micro	-1.013***	-1.013***	-1.017***	-1.017***
	(0.069)	(0.069)	(0.069)	(0.069)
Time Macro	-1.063***	-1.063***	-1.064***	-1.064***
	(0.076)	(0.076)	(0.076)	(0.076)
Cognitive Reflection × Micro	-0.134***	-0.133***	-0.133***	-0.133***
	(0.020)	(0.020)	(0.020)	(0.020)
Constant	0.911***	1.064***	0.910***	1.042***
	(0.166)	(0.179)	(0.168)	(0.182)
N. Obs.	18,734	18,734	18,705	18,705

Table 29: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 30: Random-Effect Logit Regression: Correct Answer, Conditionally On Answering

	(1)	(2)	(3)	(4)
Cognitive Reflection	0.142***	0.139***	0.138***	0.136***
	(0.035)	(0.035)	(0.036)	(0.036)
Loss Aversion		-0.007		-0.002
		(0.090)		(0.090)
Strong Loss Aversion		0.020		0.024
		(0.089)		(0.089)
Risk Aversion		-0.017		-0.014
		(0.083)		(0.083)
Strong Risk Aversion		-0.129		-0.124
		(0.116)		(0.118)
Confidence			0.014	0.013
			(0.021)	(0.021)
Gender (F.)	-0.143**	-0.138*	-0.124*	-0.121
	(0.069)	(0.073)	(0.074)	(0.077)
Time Micro	-1.081***	-1.081***	-1.089***	-1.089***
	(0.084)	(0.084)	(0.084)	(0.084)
Time Macro	-0.632***	-0.632***	-0.637***	-0.637***
	(0.101)	(0.101)	(0.101)	(0.101)
Cognitive Reflection × Micro	-0.095***	-0.095***	-0.094***	-0.094***
	(0.024)	(0.024)	(0.024)	(0.024)
Constant	1.575***	1.598***	1.579***	1.597***
	(0.170)	(0.189)	(0.171)	(0.190)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	15,501	15,501	15,473	15,473

Table 30: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

Table 31: Random-Effect Logit Regression: Correct Answer, Conditionally On Answering by Subsamples

	Other Fields:		Business Adm'n and Econ:	
	(1)	(2)	(3)	(4)
Cognitive Reflection	0.152***	0.166***	0.137***	0.145***
	(0.054)	(0.055)	(0.047)	(0.046)
Loss Aversion		0.174		-0.209*
		(0.126)		(0.124)
Strong Loss Aversion		0.208*		-0.220*
		(0.120)		(0.123)
Risk Aversion		-0.173		0.136
		(0.130)		(0.104)
Strong Risk Aversion		-0.239		-0.007
		(0.147)		(0.204)
Gender (F.)	-0.198*	-0.209**	-0.097	-0.041
	(0.103)	(0.107)	(0.092)	(0.097)
Time Micro	-1.037***	-1.034***	-1.113***	-1.113***
	(0.120)	(0.119)	(0.120)	(0.120)
Time Macro	-0.648***	-0.644***	-0.619***	-0.619***
	(0.152)	(0.152)	(0.132)	(0.132)
Cognitive Reflection $\times$ Micro	-0.132***	-0.132***	-0.071**	-0.071**
	(0.038)	(0.038)	(0.032)	(0.032)
Constant	1.615***	1.577***	1.856***	1.896***
	(0.226)	(0.267)	(0.178)	(0.205)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	6,212	6,212	9,289	9,289

Table 31: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

C Alternative Specification of the Threshold above which a Student Answers a Question

As an alternative approach to derive a threshold above which a student answers a question k, in this appendix, we apply the concept of expectation-based loss aversion by Kőszegi and Rabin (2006, 2007). Similar to the threshold p^* derived in the main text, this alternative threshold p^{**} is increasing in λ.

An expectation-based loss-averse student with success probability p_k who plans to answer and a degree of loss aversion λ uses her expected score from answering question k as her reference point r_k which is equal to $p_k \cdot 3 + (1 - p_k) \cdot 0 = 3p_k$. Her gain-loss utility is derived as follows. With probability p_k, she gives the correct answer to question k and gets 3 points. She will therefore experience a gain of $3 - r_k = 3(1 - p_k)$. With probability $1 - p_k$, her answer turns out to be wrong and she gets 0. She will therefore suffer a loss of $\lambda \cdot (0 - r_k) = -3p_k\lambda$. Her expected gain-loss utility then equals $p_k \cdot 3(1 - p_k) + (1 - p_k) \cdot (-3p_k\lambda)$ which simplifies to $-3p_k(1 - p_k)(\lambda - 1)$. Her expected total utility additionally includes the expected value of answering question k, $3p_k$. Her expected total utility of not answering question k equals 1. To obtain the threshold, we have to set the expected total utility of answering question k (with a weight $\eta = 1$ on gain-loss utility relative to consumption utility) equal to that of not answering the question,

$$3p_k - 3p_k(1 - p_k)(\lambda - 1) = 1.$$

From this equation, the threshold p^{**} can be derived as a function of λ,

$$p^{**}(\lambda) \equiv \frac{2}{\sqrt{3\left(\lambda(3\lambda - 8) + 8\right)} - 3(\lambda - 2)},$$

which takes values in the interval $[1/3, 1]$ for $\lambda \geq 1$. The slope is equal to

$$\frac{dp^{**}(\lambda)}{d\lambda} = \frac{\lambda + \sqrt{3}\sqrt{\lambda(3\lambda - 8) + 8} - 4}{2\sqrt{3}(\lambda - 1)^2\sqrt{\lambda(3\lambda - 8) + 8}}$$

The threshold $p^{**}(\lambda)$ is strictly increasing in λ because $dp^{**}(\lambda)/d\lambda$ is a hyperbola with $dp^{**}(\lambda)/d\lambda > 0$.

D Additional Regressions

Table 32: Random-Effect Logit Regression: Answer a Question by Subsamples

	Other Fields:		Business Adm'n and Econ:	
	(1)	(2)	(3)	(4)
Cognitive Reflection	0.275***	0.203***	0.291***	0.277***
	(0.070)	(0.068)	(0.098)	(0.098)
Loss Aversion		-0.711***		-0.411*
		(0.187)		(0.230)
Strong Loss Aversion		-0.700***		-0.591**
		(0.205)		(0.265)
Confidence		0.110***		0.080
		(0.038)		(0.062)
Gender (F.)	-0.565***	-0.318**	-0.780***	-0.512***
	(0.151)	(0.149)	(0.191)	(0.196)
Time Micro	-0.840***	-0.838***	-0.456***	-0.456***
	(0.117)	(0.117)	(0.173)	(0.173)
Time Macro	-1.636***	-1.633***	-0.613***	-0.613***
	(0.109)	(0.109)	(0.180)	(0.180)
Cognitive Reflection × Micro	-0.134***	-0.135***	-0.088*	-0.088*
	(0.035)	(0.035)	(0.046)	(0.046)
Constant	2.185***	2.798***	3.050***	3.394***
	(0.291)	(0.324)	(0.344)	(0.393)
Field Fixed Effects	Yes	Yes	Yes	Yes
N. Obs.	8,642	8,613	10,092	10,092

Table 32: Standard errors are in parentheses. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

E Instructions

UNIVERSITÄT
MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Dear Participant,

First of all, we would like to thank you very much for participating in this experiment!
In total, participation in this experiment does not last longer than 60 minutes.
All data collected will be treated anonymously. Therefore we would like to ask you not to mark the questionnaires by name.

The experiment consists of personal questions, test questions, questions about your risk behaviour and questions about your participation in various lotteries. For the lottery questions, a monetary payment will be made. You will receive 6 Euro for your participation in this experiment. In the lottery part you can win or lose up to 4 Euro in addition, i.e. you receive between 2 and 10 Euro.

Please do not talk to other participants during the whole experiment, do not look at other participants' sheets and do not use any electronic devices. Please fill in the form yourself. The experiment takes place under similar conditions to writing a written exam. If you violate these fair play rules, we will immediately collect your sheet and exclude you from the experiment and all payouts.

What's your student ID? We need your student ID to clearly identify you when you collect your monetary payment after the experiment. During the evaluation of this study, all your data will be anonymized so that no conclusions can be drawn about your person. We will not associate your decisions with your name.

Student ID ________________

please turn the page...

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

UNIVERSITÄT MANNHEIM

Personal questions:

1. Do you agree with the anonymous evaluation of your data?

   yes ☐ (please fill in the remaining questions)

   no ☐ (please put your pen aside and sit still without disturbing your fellow students)

2. How old are you?

   ___ years

3. Please tick your gender:

   f ☐ m ☐

4. What semester are you in?

   ___ Semester

5. What is your main field of study?

   ___

6. What is the name of the place (city or municipality) where you obtained your high school degree?

   ___

7. What is your mother tongue?

   ___

please turn the page...

UNIVERSITÄT MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Introductory questions I:

8. A racket and a ball together cost 1.10 Euro. The racket costs 1 Euro more than the ball. How much costs the ball?

   ___ Euro

9. If 5 machines take 5 minutes to produce 5 parts, how long does 100 machines take to produce 100 parts?

   ___ minutes

10. A carpet of water lilies grows on a lake. Every day this carpet doubles in size. If it takes 48 days for the water lily carpet to cover the whole lake, how long would it take for this carpet to cover half the lake?

    ___ days

please turn the page...

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Questions about behaviour I:

For each of the following statements, indicate the **likelihood** that you would engage in such activity or conduct. Please use the following scale from **1 to 5**:

11. ... camping in the wilderness far away from civilization and campsites?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

12. ...follow a tornado in a car to take dramatic pictures?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

13. ...risking a day's income in a poker game?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

14. ... invest 5% of your annual income in a very speculative stock?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

15. ...don't buckle your seatbelt in a car?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

16. ...go home at night alone through an unsafe part of town?

highly unlikely ☐1 ☐2 ☐3 ☐4 ☐5 very likely

please turn the page...

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

UNIVERSITÄT MANNHEIM

Introductory questions II:

17. What's the name of the author of William Tell?

☐ a) Johann Wolfgang von Goethe ☐ b) Friedrich Schiller

☐ c) Friedrich Hölderlin ☐ (d) Theodor Fontane

18. What year did Albert Einstein die?

☐ a) 1955 ☐ b) 1947 ☐ c) 1961 ☐ d) 1938

19. How many inhabitants does the Saarland (federal state) have?

☐ a) 2,132,000 ☐ b) 1,670,000 ☐ c) 1,037,000 ☐ d) 890,000

20. How big is the distance between earth and sun in "astronomical units"?

☐ a) 587 ☐ b) 1 ☐ c) 4553 ☐ d) 14

21. Which urban area has the largest population?

☐ (a) Shanghai ☐ (b) Istanbul ☐ (c) Los Angeles ☐ (d) Moscow

22. How many of the last five questions do you think you answered correctly?

__ % (0% to 100%)

23. How many questions do you think the other participants answered correctly on average?

__ % (0% to 100%)

please turn the page...

UNIVERSITÄT MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Decision about lottery participation:

You will now be presented with several lotteries, each of which you can either play or not play. The lotteries differ in the amount you can lose. By the end of this part, one lottery will be randomly selected and played in order to determine your payoff. You have a budget of 6 euros. You can win or lose a maximum of 4 Euros, i.e. you will receive a payout of between 10 and 2 Euros.

Here is a brief **example:**

Lottery Series Z:

Gain **4,00 Euro** Probability of winning **50%.**

Loss **see below** Probability of losing **50%**

Please make a cross each time you want to play the Z Series lottery!

	X	X	☐	☐	☐	☐
loss	-0,60 Euro	-1,20 Euro	-1,80 Euro	-2,40 Euro	-3,00 Euro	-4,00 Euro

- ➔ The player would play a Series Z lottery up to a loss of -1.20 euros.
- ➔ If the lottery with a loss of -0.60 Euro is randomly selected for being paid out, the player would win 4 Euro with 50% of chance, or lose 0.60 Euro with 50% of chance.
- ➔ A 50% probability, for example, corresponds to the probability of getting a 1, 2, or 3 when rolling the dice.
- ➔ If the lottery with a loss of -1.80 Euros is randomly selected for being paid out, the player would not play and would not win or lose anything.

If anything is still unclear to you, please contact the experimenter now.

please turn the page...

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

UNIVERSITÄT MANNHEIM

Once you have understood the instructions, make the following lottery choices as described in the example above. After all the questionnaires have been collected, a lottery series and one lottery within this series will be randomly selected and then played. You will be paid the **week after next week** according to your choices and the realized payoffs in the lecture.

Lottery Series A:

Gain 4**,00 Euro** Probability of winning **50%**

Loss **see below** Probability of losing **50%**

Please make a cross each time you want to play the respective Series A lottery!

	☐	☐	☐	☐	☐	☐
loss	-0,60 Euro	-1,20 Euro	-1,80 Euro	-2,40 Euro	-3,00 Euro	-4,00 Euro

Question 24-29.

please turn the page...

UNIVERSITÄT MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Now choose between **Lottery B** and a **safe payment**. In **Lottery B,** you can win 4 Euros or 0 Euros, but lose nothing. The probability is 50% in each case.

Lottery B:

Gain **4,00 Euro** Probability **50%**

Gain **0,00 Euro** Probability **50%**

Safe payment:

Payment **A (see right column in the table)**

Please choose between Lottery B and the Safe payment in each of the lines 1-6. Please make exactly one cross in each line!

Table	Lottery B	Safe payment
Line 1	Lottery B ☐	A=0,40 Euro ☐
Line 2	Lottery B ☐	A=0,80 Euro ☐
Line 3	Lottery B ☐	A=1,20 Euro ☐
Line 4	Lottery B ☐	A=1,60 Euro ☐
Line 5	Lottery B ☐	A=2,00 Euro ☐
Line 6	Lottery B ☐	A=2,40 Euro ☐

Question 30-35.

please turn the page...

UNIVERSITÄT MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Questions about behaviour II:

36. After every decision I've made, I wonder what would have happened if I'd made a different decision.

not applicable at all ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 fully applicable

37. When I make a decision, I try to find out afterwards what the other alternatives would have led to.

not applicable at all ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 fully applicable

38. Even a good decision is a failure if it turns out that another option would have been better.

not applicable at all ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 fully applicable

39. When I think about my life, missed opportunities often come to mind.

not applicable at all ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 fully applicable

40. Once I have decided, I do not question that decision.

not applicable at all ☐1 ☐2 ☐3 ☐4 ☐5 ☐6 ☐7 fully applicable

please turn the page...

UNIVERSITÄT
MANNHEIM

Faculty of Law and Economics
- Department of Economics -
Prof. Dr. Martin Peitz

Do you have any comments on this experiment?

You have now completed the questionnaire for the experiment. Please wait until your sheet is picked up. Thank you for your patience!

Please put your pen away and keep quiet.

please turn the page...