



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 629

Grit Trumps Talent? An experimental approach

Leonie Gerhards and Christina Gravert

October 2015

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Grit Trumps Talent? An experimental approach*

Leonie Gerhards^{†1} and Christina Gravert^{‡2}

¹Aarhus University

²University of Gothenburg

September 2015

“It’s not that I’m so smart, it’s just that I stay with problems longer.” – Albert Einstein

Abstract

Perseverance to accomplish long-term goals, also known as grit, is a crucial determinant for success in life. In the present study we introduce an innovative laboratory design to elicit grit in an incentivized and controlled way. Subjects work on a computerized task to solve anagrams. By observing their decision not to shirk, we measure their grittiness experimentally. We find that the original questionnaire measure of grit developed by Duckworth et al. (2007) is significantly correlated with our new experimental measure – even when controlling for ability and a questionnaire measure of self-control. Moreover, subjects’ earnings increase in their experimentally elicited grit.

Keywords: Grit, perseverance, laboratory experiment, real-effort task

JEL codes: C91, D03, M50, J24

*We are grateful for helpful comments from Michele Bélot, Teodora Boneva, Tor Eriksson, Nabanita Datta Gupta, Alexander Koch and Bertil Tungodden as well as seminar and conference participants at University of Aarhus, University of Amsterdam, University of Copenhagen, University of Gothenburg, University of Munich, University of Nottingham, the Copenhagen Network of Experimental Economics, the Meeting of German Economists Abroad 2014 and the Nordic Conference in Behavioral and Experimental Economics 2015. This research was supported by funding from the Danish Council for Independent Research — Social Sciences (Det Frie Forskningsråd — Samfund og Erhverv), grant number 12-124835 and from the Swedish Research Council (Vetenskapsrådet) (2009-33472-67049-56).

[†]Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark; lgerhards@econ.au.dk, Tel.: +45 871 65229

[‡]Department of Economics, University of Gothenburg, Vasagatan 1, S-41124 Göteborg, Sweden; christina.gravert@economics.gu.se, Tel.: +46 (0)31 786 5252

1 Introduction

In 1921 Lewis Terman started to select a group of gifted children and followed them in a longitudinal study (Terman & Oden 1947). Although all of them had extraordinarily high IQs (they averaged at 150), he observed a high variance in their level of success in life. In particular, Terman found that the more successful individuals exhibited higher rates of perseverance, goal orientation and will-power than the less successful ones. Similarly, other researchers have noticed positive correlations between zeal, hard labor and success, independently of initial talent (Galton 1892, Ericsson & Charness 1994). Economists and psychologists alike acknowledge that there is still a large variation in success once cognitive skills such as intelligence, test scores and basic demographics have been controlled for (Almlund et al. 2011).

In this paper we focus on “grit”, one of the non-cognitive skills that has been shown to be linked to success. It is defined as perseverance and passion to accomplish long-term or higher-order goals in the face of challenges and setbacks (Duckworth et al. 2007). In a number of studies Duckworth and co-authors find that grit, elicited in questionnaires, is a reliable predictor for pupils’ success in the national spelling bee, retention rates of West Point Military Academy cadets and Grade Point Averages (GPA) among Ivy League undergraduates (Duckworth et al. 2007, Eskreis-Winkler et al. 2014).

However, by design, these questionnaire studies rely only on self-reported information, a methodology that has some well known drawbacks. Answers to survey questions can be distorted if survey respondents misunderstand statements or interpret terms or descriptions of situations differently than intended by the researcher. Further biases might arise from experimenter demand effects or the respondents’ desire to adhere to a supposed standard. To address these concerns, a large strand of literature has evolved, comparing the predictive validity of survey responses to experimental measures. Examples of experimentally validated questionnaires are, for instance, trust, altruism, competitiveness and risk and time preferences (see for instance Glaeser et al. 2000 and Becker et al. 2012). Motivated by this line of reasoning, we developed an experimental measure of grit which is less strongly subject to biases and experimenter demand effects than the survey measure. While it is easy to fake grit in a questionnaire, it is much harder to fake it in a real effort task. Furthermore, compared to the previous survey studies on grit and perseverance conducted in the field, our controlled laboratory environment provides a way to exclude confounding factors such as peer effects, and it allows us to control for individual ability.

We elicit grit in an experimental task that involves solving anagrams, i.e., rearranging letters to form new words. Subjects have the opportunity to either skip individual anagrams or switch to solving easier anagrams altogether if they want to avoid working on the harder ones. Both shirking behaviors come at a cost and provide a measure of how easily subjects give up when the task is difficult or when they become bored.

We supplement our experimental task with an extensive questionnaire, including the Short Grit Scale developed by Duckworth & Quinn (2009) and the Brief Self-Control Scale (Tangney et al. 2004) to elicit subjects’ self-reported general grit and self-control abilities. The self control measure is in particular aimed at capturing the ability to avoid momentary temptation, a concept related, but not identical to grit.

We assume the survey measure of grit to be positively correlated with perseverance on the

task. Said differently, we expect the self-reported grit measure to be negatively correlated with shirking on the task. In line with this conjecture, we indeed find that higher self-reported grit correlates with less skipping on the experimental task – even when controlling for task-related ability and the questionnaire measure of self-control. The correlation between self-reported grit and switching to easy anagrams is negative too. But it is not statistically significant at conventional levels.

Moreover, similar to Duckworth et al. (2007), who observe a positive correlation between their survey measure of grit and different success measures such as GPA and ranking in spelling bees, our experimental measures of grit are significantly linked to subjects’ earnings in our experimental study. Attempting to work on hard (instead of easy) anagrams significantly increases subjects’ payoffs. Skipping individual hard anagrams significantly decreases payoffs, also when controlling for ability and performance.

Lastly, our study complements recent economic and psychological studies that investigate if and how cognitive and non-cognitive skills interact with each other. Most of these studies link measures for ability and Big Five personality traits to various forms of success. In particular interesting for us are findings on the personality factor conscientiousness, of which grit and perseverance are sub-traits (Almlund et al. 2011). Chamorro-Premuzic & Arceche (2008) and Moutafi et al. (2005), for instance, show that when it comes to academic outcomes, such as GPA or logic and verbal reasoning test scores, conscientiousness and intelligence can act as substitutes. Duckworth et al. (2007) focus on the non-cognitive trait grit and find that while grit is not positively related to their measure of ability (i.e. Scholastic Assessment Test (SAT) scores), both ability and grit positively predict the academic success (i.e. cumulative GPA) of undergraduates at an American elite university. Moreover, they observe a significant negative correlation of grit and SAT scores, which leads them to conclude that “among relatively intelligent individuals, those who are less bright than their peers compensate by working harder and with more determination” (p. 1093), an interpretation that was similarly proposed by Moutafi et al. (2005). The empirical results so far hence suggests that grit can be used as a compensation mechanism or a substitute for lack of talent or IQ. However, in the experiment at hand, we do not observe significant (negative) interactions between our empirically elicited grit and ability measures, which points to the conclusion that grit and ability can also act as complements rather than substitutes in certain tasks.

2 Experimental Design

2.1 The task

To rule out that the subjects’ effort is mainly driven by their intrinsic motivation to work on the task, we implement a rather tedious computerized real-effort task. In particular, we opted for an “anagram solving task” in which subjects have to rearrange the letters of English words to form new ones.¹ We introduce two levels of difficulty: “easy anagrams” consist of 3 to 4 letters, “hard anagrams” comprise 5 to 7 letters.

¹Most of the subjects in our sample are Danish undergraduates. They had on average 9 years of English at school with 3 years being the minimum.

The experiment starts with a five minute practice round in which subjects are only presented with easy anagrams. Performance in this part is not monetarily incentivized. It allows subjects to familiarize themselves with the experimental task. In the main part of the experiment on the other hand, subjects are paid according to their performance and free to choose between solving hard and easy anagrams. For each correctly solved hard anagram subjects earn DKK 5.00². For each solved easy anagram subjects earn DKK 0.50.

Each anagram is presented for up to 90 seconds. If a subject does not manage to solve it correctly within this time frame, a new anagram will be displayed. However, when working on hard anagrams, subjects can also “skip” individual anagrams at a cost of DKK 3.00. Then they do not have to wait until the end of the 90 seconds, but a new anagram is generated immediately.

The main part of the experiment is divided into two identical halves of 30 minutes each. The first half mainly serves to measure subjects’ ability on the task. The second half is the working period of interest in our later data analysis. Both halves are partitioned into 10 sub-periods of 3 minutes. In the first sub-period of each half subjects have to work on hard anagrams. We chose this set-up in order to stress the default character of working on hard anagrams and to make sure that the subjects get to know the level of difficulty of the hard anagrams. At the beginning of each of the following nine sub-periods subjects can choose to “stay with the hard anagrams” or to “switch to the easy anagrams” for the coming three minutes.³ Thus, we allow subjects to go back to solving hard anagrams after having switched to easy anagrams in one sub-period (and vice versa). If subjects decide to switch to the easy anagrams at the beginning of a new sub-period, they also have to bear a cost of DKK 3. The total cost of switching to easy anagrams for one sub-period hence consists of the explicit switching cost and the implicit cost of reduced earning opportunities.

Table 1: Payoffs in DKK

Action	Payoff
Solving a hard anagram	5.00
Solving an easy anagram	0.50
Skipping a hard anagram	-3.00
Switching to easy anagrams	-3.00

Table 1 gives an overview of the provided incentives. We chose this payoff structure to make switching to easy anagrams clearly monetarily unattractive. Subjects should be able to see without any formal calculation that even for less able individuals switching to easy anagrams is not optimal. The choice can only be rationalized by a desire to avoid working hard. And indeed, comparing the subjects’ maximum three minute sub-period earnings when working on hard anagrams to the respective three sub-period earnings when working on easy anagrams, we find that only 1 out of our 62 subjects would have been monetarily better off working only on easy anagrams⁴. Similarly, considering the individual mean sub-period

²Note that DKK 1.00 corresponds to approximately EUR 0.13.

³See Figure B.1 in the Online Appendix for a screen shot of this decision stage.

⁴That subject’s maximum sub-period earnings when working on hard anagrams is DKK 0.00, whereas it

earnings when working on hard and easy anagrams, only 2 out of our 62 subjects would have been monetarily better off working only on easy anagrams⁵. For the vast majority of our subjects, the payoff maximizing choice was hence to work on the hard task.

In both halves of the main part, an information box on the subjects' computer screen continuously informs them about how many hard anagrams they have already skipped since the beginning of the current half (1 or 2) and how many hard anagrams they have skipped in the current sub-period⁶. In sub-periods in which subjects work on easy anagrams, no such information is displayed.

At the end of the first half subjects receive feedback about their productivity and pay-off in that half. Furthermore, we ask them (without previous announcement) to make an incentivized guess which performance quintile they belong to⁷. In particular, we ask them to make this guess based on their number of correctly solved hard anagrams. We reward the correct answer with DKK 20. Afterwards their computer screens display their actual performance quintile, before the second half commences.

To sum up, the two halves of the main part of the experiment are identical in their set-up. We conjecture, however, that subjects' intrinsic motivation to do well on the task is higher at the beginning of the experiment. Moreover, we assume in line with previous studies that grit will have a stronger influence the longer a work task lasts. Therefore, we will focus on the second half when we experimentally elicit subjects' grit, and we will use the number of correctly solved hard anagrams in the first half as our proxy of subjects' ability (and motivation) in the task.

2.2 Questionnaire

Besides performing on the task, we also ask the subjects to fill out a short questionnaire. This comprises two parts: in one part we ask them how they perceived working on the task; in the other part we elicit a number of non-cognitive skills and personality traits through non-incentivized survey questions. In particular, we administer the Short-Grit-Scale (Duckworth & Quinn 2009), which comprises eight items that are ranked on a 5 point scale and measure the tendency to sustain effort and stay focused on interests over a long period of time. Since previous studies have found grit to be correlated with self-control (see for instance Duckworth et al. 2007), we also elicit this non-cognitive skill using Tangney et al.'s (2004) Brief Self-Control Scale. It contains 13 items endorsed on the same 5 point scale as the grit measure. Self-control refers to the voluntary regulation of behavioral, emotional, and attentional impulses. Although being related, it focuses on momentary pleasure/pain trade-off, while grit captures a longer term concept.

We supplemented these survey questions by questions on general time (patience) and risk preferences as for instance used in the German SOEP (Becker et al. 2012) and a general

is DKK 4.50 when working on easy anagrams.

⁵Those subjects' mean sub-period earnings when working on hard anagrams are DKK -2.00 and DKK 0.47, whereas their mean sub-period earnings when working on easy anagrams are DKK 0.67 and DKK 1.29.

⁶See Figure B.2 in the Online Appendix for a screenshot of this performance stage.

⁷In one of our sessions only 6 subjects participated. Therefore we asked them to guess which performance tertile they belonged to instead.

question on optimism. The full list of questionnaire items is available in Online Appendix C⁸.

A valid concern when running survey questionnaires at the end of an experiment is that the outcome of the experiment might influence the answers given in the questionnaire. To address this point, we randomize the order of the experiment and the survey part of the questionnaire at the session level. The task-related questions were always asked at the end of the experiment.

As is evident from Table 2, the sequence of task and questionnaire in the experiment does not significantly affect the elicited variables. From the insignificant Mann-Whitney ranksum test results we conclude that our performance and questionnaire measures are mutually independent and hence not subject to experimenter demand effects or framing. Also, whether the task was perceived as challenging, enjoyable or exhausting is independent of the order of task and questionnaire. Consequently, we pool all observations for the following analysis.

Table 2: Summary statistics (mean values and standard deviations) for selected variables

	All observations	Questionnaire last	Questionnaire first	Ranksum test results
<i>Performance in the practice round and main part</i>				
Performance in practice round	9.113 (6.621)	8.763 (6.973)	9.667 (6.127)	$p = 0.357$
Correctly solved hard anagrams	33.839 (26.220)	32.368 (26.836)	36.167 (25.605)	$p = 0.427$
Correctly solved easy anagrams	42.403 (64.401)	47.447 (65.329)	34.417 (63.446)	$p = 0.481$
Skipped hard anagrams	7.935 (12.716)	7.684 (13.563)	8.333 (11.518)	$p = 0.634$
Switches to easy anagrams	4.452 (5.861)	5.053 (6.177)	3.500 (5.308)	$p = 0.382$
<i>Questionnaire measures</i>				
Grit	3.063 (0.632)	3.079 (0.666)	3.036 (0.586)	$p = 0.988$
Self-control	3.077 (0.647)	3.107 (0.678)	3.029 (0.606)	$p = 1.000$
Narcissism	0.511 (0.268)	0.508 (0.259)	0.514 (0.283)	$p = 0.824$
General optimism	7.258 (2.495)	7.421 (2.274)	7.000 (2.844)	$p = 0.678$
General patience	5.548 (2.815)	5.579 (2.786)	5.500 (2.919)	$p = 0.827$
General risk attitudes	5.145 (2.679)	5.368 (2.804)	4.792 (2.484)	$p = 0.480$
Perceived task as a challenge	4.065 (0.787)	4.026 (0.822)	4.125 (0.741)	$p = 0.639$
Exerted effort on task	4.339 (0.848)	4.368 (0.852)	4.292 (0.859)	$p = 0.671$
Enjoyed working on task	3.113 (1.282)	3.237 (1.344)	2.917 (1.176)	$p = 0.303$
Perceived task as exhausting	3.532 (1.004)	3.500 (0.980)	3.583 (1.060)	$p = 0.880$
Expected performance quintile	3.194 (1.252)	3.395 (1.242)	2.875 (1.227)	$p = 0.134$
Observations	62	38	24	

2.3 Procedure

We ran the experimental sessions in Spring and Summer 2014 at the Cognition and Behavior (Cobe) Lab at Aarhus University. The subjects were recruited via the laboratory’s online recruiting website from a subject pool of mostly undergraduate students from all faculties. In total 62 subjects participated in three “questionnaire first” and four “questionnaire last” sessions, comprising 24 and 38 subjects, respectively.

At the beginning of each session consent forms and detailed instructions about the experimental task were distributed to the subjects (available in Online Appendix D). The subjects had 10 minutes to read the instructions. Afterwards the experimenter asked if there were any questions or if anyone needed more time to read. When all subjects declined, the experiment (programmed in *z-Tree* (Fischbacher 2007)) started, and all further instructions were provided on the computer screens. At the end of each session either part 1 or part 2 was randomly selected for payments. Average earnings for the 90 minutes sessions amounted

⁸We elicited a number of further characteristics such as narcissism, altruism etc. These measures have no effect on the behavior in this experimental task and are used in another project (Gerhards & Gravert 2015).

to DKK 120, including the reward for a correct guess about one’s own performance rank and a DKK 40 show-up fee. The payments were directly transferred to the subjects’ bank accounts.

3 Results

Table 2 reveals that subjects differ considerably in their performance on the task in the main part of the experiment. On average, subjects solve 34 hard anagrams with a standard deviation of 26; and 42 easy anagrams with a standard deviation (s.d.) of 64. With regard to our experimental grit measures, we observe that subjects skip on average 8 hard anagrams (s.d. 13) and switch 4 times to easy anagrams (s.d. 6) during the main part of the experiment. Below we start our analysis by presenting correlations between the self-reported measure of grit and our experimental outcomes. Next we show how our experimental measure of grit impacts other experimental outcomes.

3.1 The self-reported measure of grit

The two performance measures that we are particularly interested in are the number of hard anagrams skipped and the decisions to switch to easy anagrams. We conjecture that both shirking behaviors are manifestations of subjects’ (non-) grittiness. We complement the analysis by an examination of two additional performance measures. The first being the number of hard anagrams solved (conditional on the fact that a subject actually worked on hard anagrams in the given sub-periods), the second being earnings. The latter represents the most aggregated individual performance measure since earnings depend both on the difficulty and the total number of correctly solved anagrams as well as on the amount of skipping. Figure A.1 in Appendix A shows the development of subjects’ shirking behavior and performance over the two halves of the experiment. The figures and related Wilcoxon signed-rank tests show no clear time trends⁹. From this we conclude that subjects do not exhibit learning or tiring effects over the course of the experiment that we need to account for.

In the regressions in Table 3 we examine in how far self-reported grit affects these performance measures in the second half of the experiment. In the odd-numbered specifications, we regress subjects’ earnings (model (1)), the number of anagrams they solve in sub-periods, in which they work on hard anagrams (model (3)), the decision to switch to easy anagrams (model (5)) and the number of anagrams skipped when working on hard anagrams in a given

⁹In the first half, subjects solve on average 1.989 anagrams (s.d. 1.400) during the sub-periods in which they work on hard anagrams. In the second half this number amounts to 1.914 anagrams (s.d. 1.246, $p = 0.537$). Similarly, conditional on working on hard anagrams, subjects skip on average 0.533 hard anagrams (s.d. 0.663) in the first and 0.492 anagrams (s.d. 0.848) in the second half ($= 0.178$). The mean total number of switches to easy anagrams is 2.355 (s.d. 2.674) and thus only slightly higher in the first, than in the second half (mean = 2.097, s.d. 3.505, $p = 0.062$). The latter test result amounts to $p = 0.192$ if one excludes the decisions from the second sub-periods in part 1 (and hence also part 2), which can apparently be considered an outlier. The relatively high desire to switch to easy in that sub-period could be explained by the subjects’ desire to check the level of difficulty after having been forced to work on hard anagrams in the first sub-period.

Table 3: Grit and ability effects (in the second half of the main part)

	Earnings		Hard anagrams solved		Switch to easy		Hard anagrams skipped	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Self-reported grit	0.415 (0.504)	0.464 (0.473)	-0.109 (0.126)	-0.114 (0.134)	-0.070 (0.174)	-0.104 (0.163)	-0.323*** (0.102)	-0.299*** (0.096)
Solved hard in 1st half	3.604*** (0.428)	3.536*** (0.410)	0.924*** (0.112)	0.919*** (0.119)	-0.539** (0.272)	-0.537** (0.269)	0.413*** (0.110)	0.433*** (0.129)
Grit x Solved hard in 1st half		0.418 (0.287)		0.026 (0.106)		-0.066 (0.149)		-0.123 (0.092)
Constant	5.989*** (0.992)	6.111*** (0.976)	1.791*** (0.203)	1.798*** (0.196)	-1.044*** (0.354)	-1.059*** (0.360)	0.912*** (0.261)	0.878*** (0.253)
R^2	0.243	0.246	0.330	0.330			0.272	0.290
Pseudo R^2					0.106	0.107		
Number of clusters	62	62	62	62	62	62	62	62
Observations	620	620	490	490	558	558	490	490

Regression 1 and 2: Pooled OLS, Dep. variable: Earnings; Regression 3 and 4: Pooled OLS, Dep. variable: Number of solved hard anagrams (given worked on hard); Regression 5 and 6: Pooled Probit, Dep. variable: Decision to switch to easy anagrams; Regression 7 and 8: Pooled OLS, Dep. variable: Number of of skipped hard anagrams (given worked on hard). Cluster-robust standard errors are given in parentheses: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Note that all continuous independent variables are standardized. All regressions include dummy variables for session weeks and sub-periods.

sub-period (model (7)) on task-related ability and self-reported grit. We measure the former by the number of solved anagrams per sub-period that they worked on hard anagrams in the first half and elicit the latter in our questionnaire. The small and insignificant Spearman rank correlation coefficient of 0.076 ($p = 0.559$) confirms that these two variables measure fundamentally different factors. In the even-numbered specifications we run similar regressions in which we additionally include an interaction term of self-reported grit and our ability variable. Note we standardize the continuous independent variables in all regressions in this paper. On the one hand this helps us avoid potential problems of multicollinearity. On the other hand, it facilitates comparisons of the sizes of coefficients.

As becomes evident from the first four specifications in Table 3, the questionnaire measure of grit is neither significantly correlated with earnings, nor with the number of hard anagrams solved. Both dependent variables are mainly explained by individual ability. The coefficients of “hard anagrams solved in part 1” are large and highly significant in all these regressions. Moreover, the effects of ability and self-reported grit on the dependent variables neither significantly reinforce, nor weaken each other, as indicated by the non-significant and small interaction terms in models (2) and (4).

Also the decision to switch to easy anagrams depends mainly on the subjects’ ability (see specification (5)). Self-reported grit has a negative impact on switching, but it is small and statistically insignificant. Specification (7), however, reveals that a subject’s self-reported grit significantly reduces the amount of hard anagrams skipped. Gritty individuals stay the course if they encounter a difficult anagram and skip significantly less often. We further find that subjects who solved more hard anagrams correctly in the first half are more likely to skip in the second half. An explanation for this initially counter-intuitive result could be that high ability subjects prefer to skip individual anagrams to which they cannot immediately find an answer at the cost of DKK 3.00 in order to take the opportunity to solve the next anagram and earn DKK 5.00.

The coefficients of the interaction term of self-reported grit and ability are neither significant in specification (6) nor in specification (8). This suggests that no significant interdependencies exist between these two variables. Self-reported grit hence explains shirking

irrespective of subjects' ability.

In the first three specifications of Table 4 we investigate which other personal characteristics besides self-reported grit and ability can explain subjects' skipping behavior. As becomes evident, gender has only a marginally significant effect on skipping. Women skip slightly less than men. Also age and the questionnaire measures of patience, optimism and risk attitudes have no stable significant effect on skipping. Interestingly, however, the correlation between the questionnaire measure of grit and skipping behavior remains significant once one controls for subjects' self-control. The coefficient for self-control is comparably smaller and neither significant in specification (2) nor (3). So while self-control and grit are related traits – as shown by their high correlation in previous studies (for instance Duckworth et al. 2007), and in our study where the Spearman's rank correlation between the (non-standardized) questionnaire measures of self-control and self-reported grit is high and statistically significant ($\rho = .669, p < .01$) – self-reported grit is the better predictor of skipping behavior. Specification (3) furthermore suggests that a subjects' overconfidence with respect to his or her performance (as elicited after the first half) is negatively related to skipping in a magnitude that lies between self-reported grit and individual ability.

The question remains whether a subject's self-reported grittiness does not only reduce skipping of hard anagrams but ultimately also increases payoffs once one controls for additional characteristics. The regressions in the second half of Table 4 show that this is not the case. Instead, it is mainly ability, measured as the number of solved anagrams per rounds worked on hard anagrams in the first half of the experiment, that affects individual earnings. The corresponding coefficients stands out both in magnitude and significance. The realized payoffs furthermore significantly decrease in the subjects' age and increase in their general patience, as elicited in the questionnaire. We continue the discussion of the effect of grit on payoffs in the following section.

3.2 The experimental measure of grit

Woody Allan famously said “80 percent of success is showing up” (Safire 1989). “Showing up” can be understood as a subject's decision to try to work on hard anagrams rather than choosing the easy ones which require very little effort. We test his statement empirically by looking at how much our subjects' payoffs vary depending on their decision to try. While we cannot entirely agree with Allan's statement, column (1) in Table 5 reveals that subjects' earnings in the second half of the experiment depend to a significant extent on their decision to work on hard anagrams (or not), even when controlling for their ability. This is captured by the number of hard anagrams solved correctly in the first half and has itself a significant effect on earnings too. In our data set 42 out of 62 subjects never worked on easy anagram in the second half of the experiment (68 percent). When we look at the average profit earned in that part of the experiment (excluding show up fee and reward for a correctly guessed performance quintile), we find that these subjects who “show up” in every round earn on average more than twice as much as subjects who, at least once, take the easy way out by switching to easy anagrams (DKK 97 vs. DKK 40).

In column (2) we focus on our second experimental grit measure, i.e., subjects' skipping behavior. In this specification, we regress earnings in the second half of the experiment on the

Table 4: Characteristics that affect the number of skipped anagrams and payoffs (in the second half of the main part)

	Hard anagrams skipped			Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)
Self-reported grit	-0.313*** (0.096)	-0.293** (0.119)	-0.204** (0.098)	0.469 (0.480)	0.605 (0.655)	0.513 (0.649)
Solved hard in 1st half	0.395*** (0.115)	0.375*** (0.125)	0.324*** (0.099)	3.596*** (0.429)	3.820*** (0.399)	3.971*** (0.379)
Female	-0.198* (0.117)	-0.175 (0.107)	-0.173* (0.103)	-0.298 (0.473)	-0.344 (0.461)	-0.319 (0.460)
Age	-0.036 (0.074)	-0.041 (0.070)	-0.073 (0.077)	-0.699* (0.377)	-0.733** (0.366)	-0.749** (0.374)
Self-reported self-control		0.052 (0.125)	0.059 (0.109)		-0.338 (0.423)	-0.390 (0.430)
General patience		-0.071 (0.087)	-0.041 (0.084)		0.883** (0.402)	0.919** (0.412)
General optimism		-0.155 (0.100)	-0.221** (0.107)		-0.273 (0.541)	-0.188 (0.524)
General risk attitudes		0.057 (0.108)	0.110 (0.101)		-0.121 (0.492)	-0.189 (0.478)
Overconfidence			-0.287** (0.121)			0.432 (0.404)
Constant	0.949*** (0.257)	0.941*** (0.236)	0.946*** (0.227)	5.994*** (1.057)	5.923*** (1.035)	5.883*** (1.033)
R^2	0.307	0.335	0.384	0.251	0.261	0.263
Number of clusters	62	62	62	62	62	62
Observations	490	490	490	620	620	620

Pooled OLS regressions; Models (1) - (3): Dependent variable: The number of skipped anagrams (given worked on hard anagrams). Models (4) - (6): Dependent variable: Payoffs. Cluster-robust standard errors are given in parentheses: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Note that all continuous independent variables are standardized. Moreover, all regressions include dummy variables for session weeks and sub-periods.

Table 5: What explains subjects' earnings?

	(1)	(2)	(3)	(4)	(5)
Easy sub-periods in 2nd half	-18.632*** (4.413)			-4.156** (1.843)	-3.444 (2.101)
Skipped hard in 2nd half		-0.675 (5.329)	-19.064*** (2.020)	-17.987*** (2.005)	-17.024*** (2.106)
Solved hard in 2nd half			53.747*** (2.615)	51.043*** (2.792)	52.487*** (2.965)
Solved hard in 1st half	30.484*** (4.538)	36.772*** (5.346)	5.771** (2.361)	5.928** (2.278)	5.875** (2.298)
Easy sub-periods in 2nd half x Solved hard in 1st half					1.211 (2.127)
Skipped hard in 2nd half x Solved hard in 1st half					-2.727 (1.998)
Constant	78.365*** (7.450)	79.508*** (8.712)	78.669*** (2.960)	78.456*** (2.856)	78.902*** (2.850)
R^2	0.641	0.525	0.946	0.951	0.954
Observations	62	62	62	62	62

OLS regressions; Dependent variable: Payoffs in the second half of the main part. Standard errors are given in parentheses: * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Note that all continuous independent variables are standardized. Moreover, all regressions include dummy variables for session weeks.

number of skipped hard anagrams and control for the subjects' ability. The latter variable again has a significant effect on earnings. Skipping, however, only significantly negatively affects subjects' earnings once we control for the contemporaneous number of solved hard anagrams (see specification (3)). This could potentially be explained by the fact that not only subjects who want to shirk skip anagrams, but also those subjects, who skip individual anagrams in order to continue working on the next displayed word. This effect can only be factored in by controlling for the number of solved anagrams.

In column (4) we add both of our experimental grit measures to the regression model. The size and significance of the coefficient of skipping remains largely unaffected compared to model (3). The coefficient of switching, however, decreases both in size and significance compared to column (1). Since we use standardized right-hand side variables in our two experimental grit measures, we can conclude that skipping has a stronger effect on earnings than switching has.

Lastly, in column (5), we add to the regression model interaction terms between our experimental grit measures and ability. Since both of their coefficients are insignificant, we conclude that grit and ability must not necessarily act as substitutes, as hypothesized by earlier studies (see for instance Duckworth et al. (2007) and Moutafi et al. (2005)). In our experimental setup, they act as complements. Interestingly, this holds true both when using the survey measure of grit (see specification (2) in Table 3) as well as when applying our experimental measures of grit (see specification (5) in Table 5) in the analysis of subjects' earnings.

All above described findings are robust to the inclusion of the self-reported grit measure in the regressions from Table 5. It neither changes the above reported findings, nor does its coefficient ever turn out to be significant itself. We report the corresponding regressions in Table A.1 in Appendix A.

4 Discussion and Conclusion

This paper develops and tests an experimental real-effort task which elicits grit in an incentivized decision making setting rather than by using a self-report scale in a questionnaire. We find a strong correlation between skipping behavior in our experimental task and low levels of perseverance on the grit scale developed by Duckworth et al. (2007) and Duckworth & Quinn (2009).

In our experiment we are, moreover, able to separate self-reported grit from raw ability on the task. Both self-reported grit and task-specific ability have statistically significant effects on (not) shirking. The questionnaire measure of grit does, however, not explain subjects' earnings in the experimental task. Only our experimental measures of grit, that is not switching to an easier task and not skipping individual hard sub-tasks, are significantly correlated with experimental earnings. Subjects who attempt to solve hard anagrams earn significantly more than subjects who sometimes slack off by solving easy anagrams. Also, skipping individual hard anagrams has a clear negative effect on payoff even when controlling for initial ability and current performance.

This paper provides a basis for future work on the experimental investigation of grit and perseverance. Interesting further avenues are, for instance, further investigations of the

question of under which circumstances grit and ability act as substitutes or complements. Moreover, both from a scientific and from a practitioner’s perspective, it seems worthwhile to study whether grit can be learned or at least imitated by peers and whether and to which extent grit can be influenced by training or incentives (for the former question see, Gerhards & Gravert (2015), for the latter see Alan et al. (2015)). The answers to these questions can have important policy implications for both the educational and employment sector.

References

- Alan, S., Boneva, T. & Ertac, S. (2015), ‘Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit’, *Working Paper* pp. –.
- Almlund, M., Duckworth, A. L., Heckman, J. J. & Kautz, T. D. (2011), ‘Personality Psychology and Economics’, (16822).
- Becker, A., Deckers, T., Dohmen, T., Falk, A. & Kosse, F. (2012), ‘The Relationship Between Economic Preferences and Psychological Personality Measures’, *IZA Discussion Paper* .
- Chamorro-Premuzic, T. & Arteche, A. (2008), ‘Intellectual competence and academic performance: Preliminary validation of a model’, *Intelligence* **36**(6), 564–573.
- Duckworth, A. L., Peterson, C., Matthews, M. D. & Kelly, D. R. (2007), ‘Grit: Perseverance and Passion for Long-term Goals’, *Journal of Personality and Social Psychology* **92**(6), 1087–1101.
- Duckworth, A. L. & Quinn, P. D. (2009), ‘Development and Validation of the Short Grit Scale (Grit-S)’, *Journal of Personality Assessment* **91**(2), 166–174.
- Ericsson, K. A. & Charness, N. (1994), ‘Expert Performance: Its Structure and Acquisition’, *American Psychologist* **49**, 725747.
- Eskreis-Winkler, L., Shulman, E. P., Beal, S. A. & Duckworth, A. L. (2014), ‘The Grit Effect: Predicting Retention in the Military, the Workplace, School and Marriage’, *Frontiers in Psychology* **5**, 36.
- Fischbacher, U. (2007), ‘z-Tree: Zurich Toolbox for Ready-made Economic Experiments’, *Experimental Economics* **10**(2), 171–178.
- Galton, F. (1892), *Hereditary Genius: An Inquiry into its Laws and Consequences*, Macmillan, London.
- Gerhards, L. & Gravert, C. (2015), ‘Because of you I did not give up How peers affect perseverance’, *Working Paper* .
- Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. & Soutter, C. L. (2000), ‘Measuring trust’, *The Quarterly Journal of Economics* **115**(3), 811–846.

- Moutafi, J., Furnham, A. & Paltiel, L. (2005), 'Can Personality Factors Predict Intelligence?', *Personality and Individual Differences* **38**(5), 1021–1033.
- Safire, W. (1989), 'Woody Allen's Secret is Out', *The Herald-Journal*. p. A10.
- Tangney, J., Baumeister, R. & Boone, A. (2004), 'High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success', *Journal of Personality* **72**(2), 271–322.
- Terman, L. M. & Oden, M. H. (1947), *The Gifted Child Grows up: Twenty-five Years' Follow-up of a Superior Group*, Stanford University Press, Oxford, England.

A Appendix

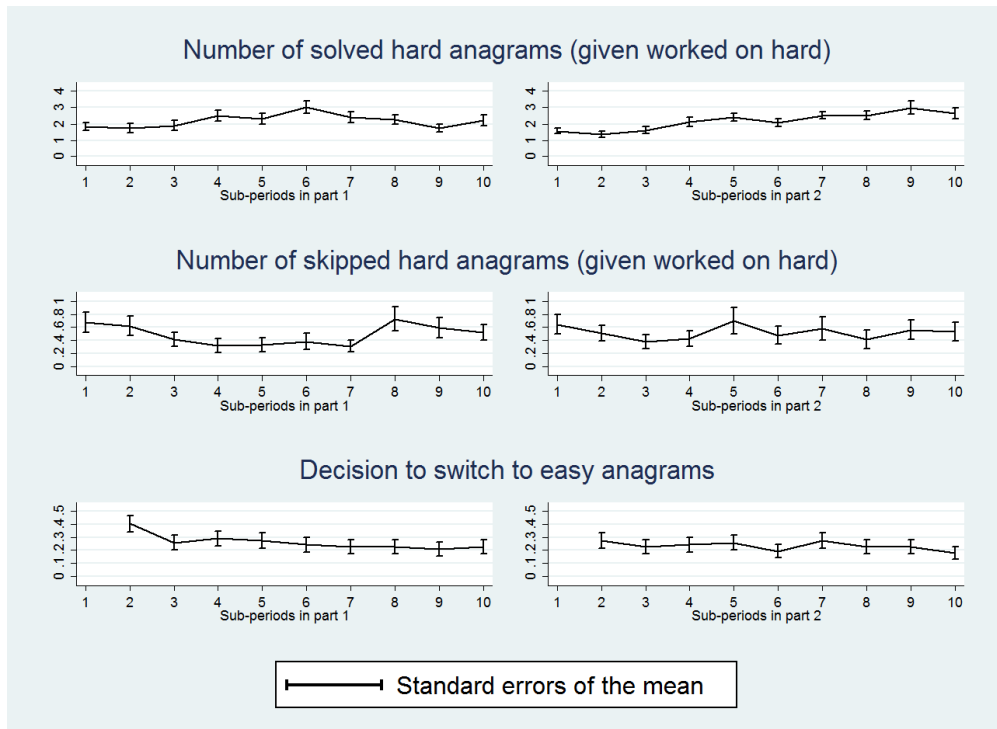


Figure A.1: Outcome variables over time

Table A.1: What explains subjects' earnings?

	(1)	(2)	(3)	(4)	(5)
Easy sub-periods in 2nd half	-18.446*** (4.442)			-4.078** (1.829)	-3.222 (2.094)
Skipped hard in 2nd half		1.159 (5.773)	-17.976*** (2.150)	-16.971*** (2.122)	-16.179*** (2.191)
Solved hard in 2nd half			53.554*** (2.597)	50.909*** (2.771)	52.413*** (2.946)
Solved hard in 1st half	30.192*** (4.579)	35.555*** (5.553)	5.205** (2.376)	5.386** (2.293)	5.282** (2.328)
Self-reported grit	3.049 (4.387)	4.560 (5.432)	2.542 (1.828)	2.420 (1.764)	2.299 (1.775)
Easy sub-periods in 2nd half x Solved hard in 1st half					1.537 (2.128)
Skipped hard in 2nd half x Solved hard in 1st half					-2.375 (2.004)
Constant	78.525*** (7.488)	79.189*** (8.744)	78.494*** (2.938)	78.293*** (2.835)	78.694*** (2.836)
R^2	0.645	0.532	0.948	0.953	0.955
Observations	62	62	62	62	62

OLS regressions; Dependent variable: Payoffs in the second half of the main part. Standard errors are given in parentheses:
 * $p < 0.10$ ** $p < 0.05$ *** $p < 0.01$. Note that all continuous independent variables are standardized. Moreover, all regressions include dummy variables for session weeks.

Online appendix

B Screenshots of the experiment

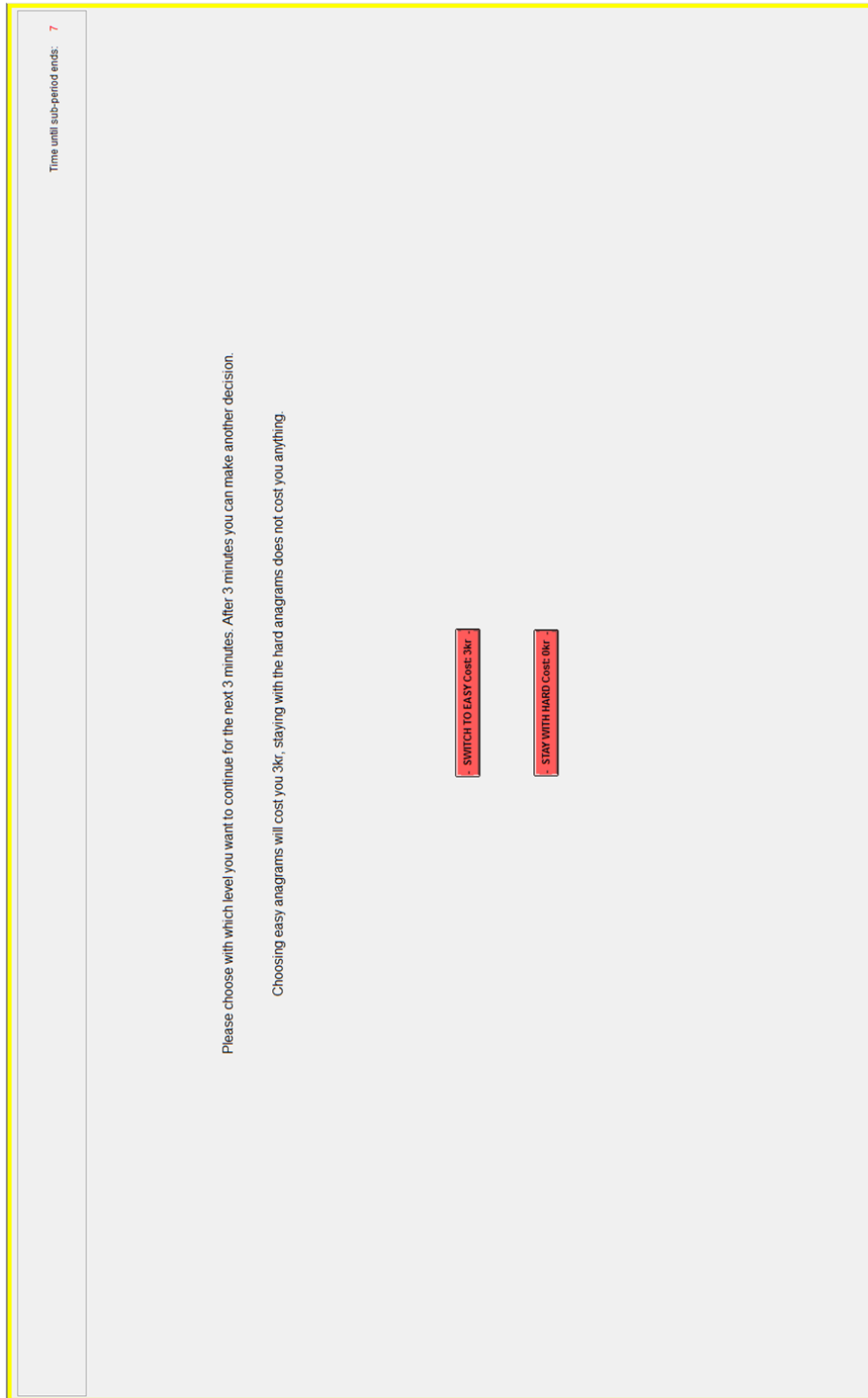


Figure B.1: Decision Stage

Time until sub-period ends: 170

Your total number of skipped anagrams in part 1 until the last sub-period (incl.):

0

Your number of skipped anagrams in the current sub-period:

1

This is the starting word: present

Please enter the anagram here:

Remaining time for this anagram: 87

Part 1: Sub-period 1 of 10

Figure B.2: Hard Anagram Stage

C Questionnaire

Question sets 2-9 were asked either at the beginning of the experiment or at the end. Question sets 1 and 10 were always asked at the end of the experiment. The headlines for each set of questions were neutral in the questionnaire.

1. Perception of the task

- (a) How much did you enjoy working on the task?
- (b) How challenging did you perceive the task?
- (c) How much effort did you provide during the task?
- (d) How exhausting did you perceive the task?

2. General questions

- (a) Please tell us your age
- (b) What is your gender?
- (c) What is your field of study?
- (d) How many years of English did you have in school?
- (e) How often do you play scrabble or solve crossword puzzles?

3. Questions on time and risk preferences and optimism (Scored from 0-10)

- (a) Are you generally an impatient person or someone who always shows great patience?
- (b) Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks?
- (c) Are you generally an optimistic person or do you expect things to go wrong?

4. Grit Scale (Scored from 1-5)

- (a) New ideas and projects sometimes distract me from previous ones.
- (b) Setbacks dont discourage me.
- (c) I have been obsessed with a certain idea or project for a short time but later lost interest.
- (d) I am a hard worker.
- (e) I often set a goal but later choose to pursue a different one
- (f) I have difficulty maintaining my focus on projects that take more than a few months to complete.
- (g) I finish whatever I begin.
- (h) I am diligent.

5. Brief-Self-Control-Scale (Scored 1-5)

- (a) I am good at resisting temptation.
- (b) I have a hard time breaking bad habits.
- (c) I am lazy.
- (d) I say inappropriate things.
- (e) I do certain things that are bad for me, if they are fun.
- (f) I refuse things that are bad for me.
- (g) I wish I had more self-discipline
- (h) People would say that I have iron self- discipline.
- (i) Pleasure and fun sometimes keep me from getting work done.
- (j) I have trouble concentrating.
- (k) I am able to work effectively toward long-term goals.
- (l) Sometimes I cant stop myself from doing something, even if I know it is wrong.
- (m) I often act without thinking through all the alternatives.

6. Rotter's 4-Item Locus of Control Scale (Choose A or B and on 2 levels)

- (a) A. What happens to me is my own doing.
- (b) B. Sometimes I feel that I don't have enough control over the direction my life is taking.
- (c) A. When I make plans, I am almost certain that I can make them work.
- (d) B. It is not always wise to plan too far ahead because many things turn out to be a matter of good or bad fortune.
- (e) A. In my case getting what I want has little or nothing to do with luck
- (f) B. Many times we might just as well decide what to do by flipping a coin.
- (g) A. Many times I feel that I have little influence over the things that happen to me.
- (h) B. It is impossible for me to believe that chance or luck plays an important role in my life.

7. Narcissism Scale (Choose A or B)

- (a) A. I know that I am good because everybody keeps telling me so.
- (b) B. When people compliment me I sometimes get embarrassed.
- (c) A. I like having authority over people.
- (d) B. I don't mind following orders.
- (e) A. I really like to be the center of attention.

- (f) B. It makes me uncomfortable to be the center of attention.
- (g) A. People always seem to recognize my authority
- (h) B. Being an authority doesn't mean that much to me
- (i) A. I find it easy to manipulate people.
- (j) B. I don't like it when I find myself manipulating people.
- (k) A. I have a natural talent for influencing people.
- (l) B. I am not good at influencing people.
- (m) A. I see myself as a good leader.
- (n) B. I am not sure if I would make a good leader.
- (o) A. I would prefer to be a leader.
- (p) B. It makes little difference to me whether I am a leader or not.
- (q) A. I am a born leader.
- (r) B. Leadership is a quality that takes a long time to develop

8. Altruism

- (a) How would you assess your willingness to share with others without expecting anything in return, for example your willingness to give to charity?
- (b) Imagine the following situation: You unexpectedly receive 7000 kr. How much of this would you donate to charity? Give a value between 0 and 7000.
- (c) How well does the following statement describe you as a person? I do not understand why people spend their lifetime fighting for a cause that is not directly beneficial for them.

9. Personal Questions

- (a) Do you smoke cigarettes?
- (b) Do you exercise regularly (at least once a week or more)?
- (c) Are you on a sports team?
- (d) How many siblings do you have?
- (e) How many younger siblings do you have?
- (f) Do you currently volunteer anywhere?
- (g) Do you have any official function in any student association or club?
- (h) Do you usually study in a study group?
- (i) If you study in a study group are you the person who encourages others do better or does studying with others rather help you to be better?

10. Last question

- (a) What do you think this experiment was about?

D Instructions

Welcome and thank you for taking part in our experiment.

In this experiment you can, depending on the decisions you make, earn money in addition to the 40kr you receive as a show-up fee. Your earnings will be added to your show up fee and transferred to your Nem-ID account after the session.

You are not allowed to communicate during the session. If you have questions, please raise your hand. A member of the research team will come to you and answer them in private. Violation of this rule will lead to the exclusion from the session and all payments.

Please take your time to read the instructions carefully before the start of the experiment!

The whole experiment will take about 1.5 hours.

Instructions

The experiment consists of 4 parts. The first part consists of a questionnaire. You will receive the instructions for the subsequent parts of the experiment after you have filled out the questionnaire.

Instructions

The experiment consists of 4 parts. A practice round of 5 minutes and two payment-relevant parts of 30 minutes each where you will be asked to work on a task. The last part consists of a questionnaire. You will get paid, if you complete all 4 parts.

Description of the task

You will be asked to solve anagrams consisting of 5 to 7 letters. An anagram is a word formed by reordering the letters of another word.

Example of an anagram:

section -> notices

For each solved anagram you will earn 5kr. You have 90 seconds to work on each anagram. You can only move to the next anagram by entering a correct solution and clicking the **“Submit”** button. If you submit a wrong answer, the anagram will remain the same until you solve it or the time runs out. Submitting a wrong answer is not penalized. If you have not solved the anagram within the time limit a new anagram will be generated at no cost.

You have the possibility to skip an anagram, if you do not know the answer. If you click the **“Skip”** button a new anagram will be generated and **3kr will be deducted from your total earnings.**

The two payment-relevant parts of 30 minutes are divided into 10 sub-periods of 3 minutes each. After every sub-period you have the opportunity to switch to easy anagrams with 3 to 4 letters by clicking the **“Switch to Easy”** button. **For each solved easy anagram you will earn 0.5kr.** As with the hard anagrams you have 90 seconds to solve it until the time runs out and a new anagram will be generated. When **switching from hard to easy anagrams 3kr will be deducted,** as this is equivalent to skipping a hard anagram. Stay with the hard anagrams by clicking the **“Stay with Hard”** button leads to no deduction.

Anagrams might have more than one solution. We will accept any correct solution.

While you work on the task several pieces of information will be displayed on the screen: A countdown for the 3 minutes sub-period that you currently work in, a countdown for the current anagram and the number of anagrams you have skipped in the current sub-period of the experiment. Please note that as you approach the end of a three minute sub-period, the sub-period countdown determines how much time you have left to work on the displayed anagram (and not the countdown for the current anagram).

In the first 5 minutes of the experiment you have the chance to familiarize yourself with the task by solving easy anagrams in a **practice round which is not payment relevant!** After 5 minutes you will be told that the payment relevant part begins.

At the end of the second payment-relevant part, either the first or the second part will be randomly selected to calculate your payment. Your earnings will depend on your performance in the chosen part. It is therefore in your best interest to always act as if it is the part that counts.

Payoff example:

Consider the case that a person solved 30 hard and 10 easy anagrams, skipped 2 hard anagrams and switched 2 times to the easy anagrams in the part of the experiment that was randomly selected for payment.

The person hence earns $30 \cdot 5kr + 10 \cdot 0,50kr - 2 \cdot 3kr - 2 \cdot 3kr = 150kr + 5kr - 6kr - 6kr = 143kr$ additional to the 40kr show-up fee.

Please note: Those sample numbers are randomly chosen and must not necessarily correspond to actual behavior of previous participants in this experiment!