Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 19-2021

Christoph Bühren and Marvin Gabriel

Performing Best When It Matters the Most: Evidence from Professional Handball

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers

Coordination: Bernd Hayo • Philipps-University Marburg
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Performing best when it matters the most: Evidence from professional handball

Christoph Bühren, Clausthal University of Technology, Department of Economics, Julius-Albert-Str. 2, 38678 Clausthal-Zellerfeld, Germany, email: christoph.buehren@tu-clausthal.de

Marvin Gabriel, University of Kassel, Department of Economics, Nora-Platiel-Str. 4, 34127 Kassel, Germany, email: marvin.gabriel@hotmail.de

Abstract

We analyze the impact of psychological pressure on individual performance with handball penalties thrown in the decisive stage vs. the rest of the game. Contrary to the phenomenon of choking under pressure, we observe that most of the analyzed players perform best when it matters the most. The positive effect of pressure on performance is especially pronounced when the score is level or when the thrower's team is lagging. We control for gender and psychological traits assessed with a survey. Female players score with a higher probability than male players in our sample. The positive impact of pressure is not significantly higher for female players.

JEL: D 91
Keywords: Performance under pressure; sports data; psychological traits; survey

1. Introduction

The nature of work in a globalized and digital economy is characterized by increased competitive pressure, and the corona-crisis accelerated this trend (Burke and Ng, 2006; Schwarzmüller et al., 2018; Kraus et al., 2020). While some employees can cope with psychological pressure and retain their typical performance level or even perform better, others "choke" under pressure and perceive stress (Vijahav, 2017). Sports economics, interpreted as behavioral economics in sports, analyzes very similar phenomena with rich data sets trying to explain the heterogeneous impact of psychological pressure on performance (Bühren & Steinberg, 2019). Choking under pressure (CUP) describes performance decrements at particularly important moments in competition compared to the performance that a person can usually achieve (Ötting et al., 2020). This phenomenon is especially pronounced in crunch time, the final minutes of tight games in professional team sports (Toma, 2017).

What factors make some athletes more and some less inclined to CUP? Psychological concepts can help to answer this question. Several studies analyze personality traits related to job satisfaction, motivation, and performance (e.g., Wright et al., 2004, Dudley et al., 2006; Lin et al., 2014). In this context, the concept of the core self-evaluation (CSE) has received much attention – Judge and Bono (2001) as well as Judge and Kammeyer-Mueller (2011), e.g., observe positive correlations of CSE with job satisfaction. Bühren and Steinberg (2019) were the first to analyze the effect of psychological traits on performance under pressure in sequential tournaments. In their tennis field experiment, they use an additional survey that addresses the individual self-assessment categories according to Judge et al. (1997). The authors find interaction effects of psychological traits on performance. Whereas subjects with low self-esteem face a first-mover advantage, subjects with high self-esteem perform better as a second-mover. Clarke et al. (2020) examine golfers' and archers' personality traits connected with choking under pressure. They use a questionnaire with items on fear of negative evaluation, doubts of action, and non-display of imperfection. Comparing these results to self-reported CUP, they can classify over 70% of choking athletes based on their psychological traits.

The remainder of the paper is structured as follows. Section 2 reviews relevant literature on performance under pressure and psychological traits related to performance. Section 3 explains our empirical approach, describes the survey we use to measure psychological traits, and derives our hypotheses. Section 4 presents the results, and Section 5 discusses them and concludes.

2. Related literature

2.1 Choking under pressure vs. performing best when it matters the most

The findings of Ariely et al. (2009) indicate that the desire to perform especially well when the stakes are high increases performance in effort-based tasks (such as clicking keys on a computer as fast as possible) but decreases performance in skill-based tasks (such as solving math problems). Harb-Wu and Krumer (2019) confirm this observation for biathletes' skiing (effort-based) and shooting performance (skill-based) in front of a supportive home audience. Social facilitation theory suggests a home advantage for relatively simple effort tasks (Ötting et al., 2020). In contrast, Baumeister and Steinhilber (1984) were the first to observe a paradoxical home disadvantage for relatively complex skill tasks (see also Wallace et al., 2005). They explain the disadvantage with monitoring pressure induced by a supportive audience. Similarly, Dohmen (2008) find that, ceteris paribus, soccer players miss penalties more often if they are from the home team. Performance decrements during skill tasks in high-pressure situations (compared to the typical performance in lower-pressure situations) have two opposing explanations. (1) Either people are distracted by the pressure, and their performance drops because they do not fully concentrate on the specific task (Sanders & Walia, 2012). (2) Or they focus too much on the task execution in high-pressure situations, and their performance drops because their automatic processes are disturbed (Baumeister, 1984).

Wilson et al. (2009a,b) show that high levels of anxiety impair the shot success of penalty kicks in soccer and free-throws in basketball. Performance decrements due to suboptimal stress levels are likely to be observed at the end of tight competitions. Jordet et al. (2007) analyze penalty shootouts in soccer and find that the probability to miss a penalty increases towards the end of the shootout. This is confirmed by Toma (2017), who observes that basketball players' free-throw percentage drops in the last 30 seconds of tight games in the National Basketball Association (NBA), especially when their team is lagging. Similarly, Cao et al. (2011) find evidence for CUP from the free-throw line in the last minute of tight NBA games, especially for weaker free throw shooters and if the first of a pair of two free throws was missed. Complementing this result, Worthy et al. (2009) observe a drop in free throw percentages in crunch time if teams lag or lead by one point, but not if the score is level. Analyzing the Professional Golf Association (PGA) Tour, Wells and Skowronski (2012) find that the average performance in the final fourth round is significantly lower compared to the third round.

Analyzing CUP in archery, Bucciol and Castagnetti (2020) observe that the players' performance significantly drops in tiebreaks, especially for women in the most prestigious tournament. Likewise, Dilmaghani (2020) finds that time pressure in chess leads to

underperformance, especially for female elite players. Yet, Cohen-Zada et al. (2017) find that male tennis players choke more likely under pressure than female tennis players. They argue that for men it seems to be too important to win (see also Niederle & Vesterlundt, 2007), resulting in too high cortisol levels. In a recent real effort experiment, Booth & Nolen (2021) show that women's effort is not dependent on the degree of pressure but men's effort increases under tournament compared to piece-rate incentives – especially when playing against women.

In contrast to CUP, the attentional control theory by Eysenck et al. (2007) suggests that high-pressure does not yield worse performance if individuals compensate for the effect of anxiety with effort or increased use of processing resources. Krumer (2020) argues that ability is the main determinant for performance under pressure. In the most critical moment of club matches in soccer, the penalty shoot-out, he observes that teams of a higher league usually win against teams from a lower league, especially in the finals. Similarly, González-Díaz et al. (2012) find that the best tennis players are successful because their performance is especially good in key moments of the game – these players are able to "perform best when it matters the most". In the same vein, Jetter and Walker (2015) observe that the top 4 tennis players win tie-breaks and matches at the most important tournaments (Grand Slams) with a higher probability than at less important tournaments – the "clutch-player" effect. And Klaassen and Magnus (2014) calculate that the top 50 players retain their typical performance level at important points of the match, whereas lower-ranked tennis players underperform in these situations.

Taken together, CUP is likely to occur in skill tasks towards the end of tight competitions when it is very important for players to perform well. However, performance under pressure depends on the task and individual characteristics. For instance, Mesagno et al. (2019) observe that dominant left-handers experience less CUP than dominant right-handers in American football. And Ötting et al. (2020) do not find evidence for CUP in professional darts reasoning that darts players are used to high-pressure situations.

2.2 Personality traits concerning performance under pressure

In addition to the question of whether and when performance declines, remains constant, or even increases under pressure, the following section establishes a connection between performance under pressure and personality traits as possible predictors. Otten (2009) tries to model "choking vs. clutch performance" based on self-confidence and anxiety. Clarke et al. (2020) argue that the fear of negative evaluations by outsiders, the display of one's imperfections, the concern over mistakes, general anxiety, and too high expectations could be responsible for CUP. Likewise, we focus on the core self-evaluation according to Judge et al.

(1997). CSE is a fundamental evaluation construct of a person about him- or herself, including all competencies, abilities, and the external world (Judge et al., 2003; Erez & Judge, 2001; Chang et al., 2012). CSE encompasses the four personality traits self-esteem, self-efficacy, emotional stability, and locus of control (Judge et al., 1997; Judge et al., 2003). Persons with a high CSE level are characterized by a positive assessment of themselves. One advantage of the CSE scale is that this generalized self-concept is not as complex as e.g. the Big 5 personality traits. And many studies show a robust relationship between CSE and the Big 5 – in particular, they find strong correlations with extraversion as well as conscientiousness (Chang et al., 2012; Judge et al., 2002; Robins et al., 2001; Farmer et al., 2001).

Some studies reveal that CSE is associated with performance and job satisfaction (e.g., Judge and Bono, 2001; Judge et al. 2003). Judge and Kammeyer-Mueller (2011) argue that people with strong CSE are more satisfied with their work because they are more optimistic about challenges. And they perform better at work because CSE strengthens task motivation and persistence concerning goals (Judge et al., 1998; Erez & Judge, 2001; Judge & Ilies, 2002). Likewise, Kirmani et al. (2019) find that the connection between CSE and performance is based on achievement, power, and affiliation (McClelland,1961), which is confirmed by Boon et al. (2011). Furthermore, Judge et al. (2000) observe a positive effect of CSE on job satisfaction. The meta-analysis of Judge and Bono (2001) verifies that all four personality traits of CSE are associated with job performance and satisfaction. Chang et al. (2012) argue that CSE reduces counterproductive behavior at work and enhances organizational citizenship behavior. Moreover, CSE is related to mental health with a positive effect on stress management (Kammeyer-Mueller et al., 2009).

However, higher self-evaluation does not necessarily lead to better work performance. For example, men usually rate themselves higher than women, especially concerning self-esteem (Bleidorn et al. 2016; Robins et al., 2002; Kling et al., 1999). Yet they do not perform better than women (Niederle and Vesterlund, 2011; Soll & Klayman, 2004; Beyer, 1990). This suggests that men are more inclined to overconfidence. Croson & Gneezy (2009) argue that confidence relates to risk taking. Whereas Hardies et al. (2013) cannot find this relationship with students and auditors, Broihanne et al. (2014) confirm that finance professionals' overconfidence leads to increased risk taking. In the next subsection, we review the role of risk taking under pressure.

2.3 Individual willingness to take risks in pressure situations

The willingness to take risks describes the fundamental preference of individuals to force or avoid risky behavior options (Kam, 2012). Grund et al. (2013) observe that lagging NBA teams

inefficiently increase their risk taking by attempting too many 3-point instead of 2-point shots. Lehman and Hahn (2013) support this result calculating that lagging football teams too often "go for it" on the fourth down.[1] However, the results of Klaassen and Magnus (2014) and Paserman (2010) suggest that most tennis players' performance drops in crucial stages of the game because they inefficiently *reduce* their risk taking. The findings of Klaassen and Magnus (2014) and Paserman (2007) suggest that women choke more under pressure than men by winning fewer points on serve and making relatively more unforced errors at important points, respectively. However, Paserman's (2010) model suggests that especially men's performance decreases under pressure because they play these points too safely.

Thus, the relationship between risk taking and CUP is likely to be gender-specific. Women are typically more risk-averse than men (Charness & Gneezy, 2012; Eckel & Grossman, 2008), although some studies find no differences (see the meta-study of Byrnes et al., 1999). Decisions under uncertainty are based on emotions (Croson & Gneezy, 2009), and women tend to be more affected by emotions than men (Harshman & Paivio, 1987). Women feel more anxiety and nervousness compared to men when they are confronted with negative prospects (Fujita et al., 1991). Moreover, men express more self-confidence than women – especially in competitions (Niederle & Vesterlund, 2011; Soll & Klayman, 2004; Beyer, 1990).

Moreover, the score of the game is likely to interfere with the relationship between risk taking and CUP. While CUP seems to be associated with less risk taking, there is evidence that loss aversion induces higher risk taking of lagging teams and individuals: At the PGA Tour, Pope and Schweitzer (2011) observe that bogey putts (one stroke above the reference "par") are ceteris paribus faster than birdie putts (one stroke below par). In line with prospect theory (Kahneman & Tversky, 1979), this finding speaks for risk aversion in the domain of gains and risk taking in the domain of losses. Similarly, Elmore and Urbaczewski (2021) identify a natural experiment in professional golf supporting the reference-dependence of players' behavior. The US Golf Association changed the par rating, and thus the reference point, for two holes at US Open tournaments from par 5 to par 4. Although the difficulty of the holes stayed the same, golfers needed ceteris paribus fewer strokes for the holes with the new par rating.

[1] The fourth down is the last chance of the attacking team to pass 10 yards (with a play) in order to get four new trials. Alternatively, the attacking team can choose the less risky variants of kicking the ball into the opponent's field (punting) or trying to kick a field goal.

3. Design of the study

3.1 Empirical approach and survey

We measure performance under pressure with the shot success of penalties in the crunch time of professional handball games and compare it to the shot success of penalties thrown in the rest of the game. Similar to the penalty kick in soccer, the seven-meter-throw is a one on one situation between the thrower and the goalkeeper (Apestiguea & Palacios-Huerta, 2010). All other players are direct observers of this situation and cannot intervene. As the probability of a goal being scored is much higher than the goalkeeper's chance to save the seven-meter, the psychological pressure lies on the thrower's shoulders. Sufficient time for reflection before the throw and the responsibility for the potential success or failure of the entire team increases this pressure – especially in deciding situations of the game (Bühren & Krabel, 2019).

A synonym for the deciding phase of professional team ball games is crunch time (Christmann et al., 2018). Depending on the ball game, it occurs in the final minutes or seconds when both teams still have the chance to win and every move might turn the game into a success or defeat. In contrast, the final moments of already decided games are called garbage time (Ertug & Maoret, 2020). We categorize a penalty to be thrown during crunch time in the last 15 minutes of a handball game if the thrower's team is a maximum of four goals behind and three goals ahead. Thus, when scoring the penalty, the difference between the teams' scores is a maximum of four goals. During the second last minute of the game, we allow the score difference in crunch time to be a maximum of three goals, and during the last minute a maximum of two goals.

The official homepages of the German Handball Federation, the German Handball Bundesliga, and the Women's Handball Bundesliga records all seven-meter throws (www.dhb.de, https://www.liquimoly-hbl.de/de/, http://www.hbf-info.de/o.red.r/home.html). In addition to the game data collected for the penalty throws, we use individual data of the thrower available on the same homepages. Beyond that, we assess the players' CES via a questionnaire according to Judge et al. (2003). Originally, Judge et al. (1997) used CSE as a disposition to explain job satisfaction (Judge et al., 1998). Then, Judge et al. (2003) developed the CSE scale to measure CSE as one construct. Although Gardner and Pierce (2010) argue that self-esteem, self-efficacy, emotional stability, and internal control conviction are better measured separately, there are high correlations between the four traits (on average approx. r=0.6). Moreover, factor analyses extract only one common factor (Judge et al. 1998; Judge et al. 2000; Erez & Judge, 2001; Judge et al. 2002; Brückner, 2020). And this factor has incremental validity concerning the four separate traits (Judge et al. 2002; Judge et al. 2003; Brückner, 2020). Judge et al. (2003) use a questionnaire with 12 items (see appendix) from a pool of 65

items. These items are assessed on a 5-level Likert scale (from strong disagreement to strong agreement) and yield very good internal consistency (Cronbach's alpha > 0.8), retest reliability (r=0.81), and convergent validity (Judge et al. 2003). For native German speakers, the questionnaire of Brückner (2020) is used, who modified the version of Heilmann and Jonas (2010). The German version of the CES scale by Brückner (2020) shows the same very good quality criteria as the original by Judge et al. (2003) and the earlier German translation by Heilmann and Jonas (2010).

Beyond CSE, we measure the individual willingness to take risks because risk taking influences performance in general and specifically under pressure (see section 2.3). We use the short scale of Beierlein et al. (2015). The 7-level Likert scale ranges from not willing to take risks to very willing to take risks. The short-scale has very satisfactory quality criteria (Beierlein et al., 2015). In experimental economics, the standard risk measures are incentivized lotteries á la Holt and Laury (2002). However, the recent study of Arslan et al. (2020) suggests an even higher external validity for survey-based stated risk preferences.

Following Clarke et al. (2020), we also ask whether the players have ever experienced a dramatic drop in performance that is beyond their control. This variable is likely to be negatively correlated to the CSE scale because relatively more self-confident people are expected to express fewer experiences with CUP – or these experiences might rather cause a person to be less self-confident. The self-assessment serves as a control variable for CUP observed in our handball dataset.

3.2 Variables

In the focus of interest is the dependent variable *shot success*. It takes the values 1 if the thrower scores the seven-meter-throw, and 0 otherwise. We try to explain the shot success with the independent individual and competition-related variables described below and summarized in Table 1.

3.2.1 Individual variables

Gender takes the value 0 for females and 1 for males. *Age* is measured in years on December 31st, 2019, which corresponds to the middle of the season 2019/2020. *Throwing hand* takes the value 0 for left and 1 for right (see also Mesagno et al., 2019). *League* indicates if the individual plays in the first, second, or third league. Similar to Krumer (2020), we use this variable as a measure for a general handball ability.

The players' *CSE* is measured from 1 (strong disagreement) to 5 (strong agreement) with th12 items taken from Judge et al. (2003) for the English version of our survey and the translation by Brückner (2020) for the German version, respectively. Our Cronbach's alpha for the CSE scale is 0.83. We measure the players' *risk taking* from 1 (not willing to take risks at all) to 7 (very willing to take risks) with the single item from Beierlein et al. (2015). Finally, *Self-report choking* takes the value 1 if the player reports having experienced a drop in performance under pressure that has been beyond their control, and 0 if the player denies having experienced CUP before (Clarke, 2020).

3.2.2 Competition-related variables

Location indicates if the penalty is taken in front of a home audience (0) or at the opponent's court (1). *Crunch time* defines both the time of a game and a certain score difference between the respective teams (Toma, 2017). It takes the value 1 in the last 15 minutes of a tight game (maximum score difference of 4, in the second last minute of 3, and the last minute of 2), and 0 otherwise. Finally, the categorical variable score distinguishes between a draw (0) and whether the thrower's team is lagging (1) or leading (2) before the seven-meter.

Table 1: Overview of the variables

	Dependent variable	Individual variables							Competition related variables		
Variable	**Shot success**	**Gender**	**Age**	**Throwing hand**	**League**	**CSE**	**Risk taking**	**Self-report choking**	**Location**	**Crunch time**	**Score**
Specification	0 = No 1 = Yes	0 = Female 1 = Male		0 = Left 1 = Right	1 = 1st League 2 = 2nd League 3 = 3rd League	1 - 5	1 - 7	0 = No 1 = Yes	0 = Home 1 = Away	0 = No 1 = Yes	0 = Draw 1 = Lagging 2 = Leading
Scale	Nominal	Nominal	Ratio	Nominal	Ordinal	Ordinal	Ordinal	Nominal	Nominal	Nominal	Nominal

3.3 Hypotheses

The following hypotheses are based on the discussed results of the related literature. Based on the findings of Baumeister (1984), Jordet et al. (2007), Wells and Skowronski (2012), Toma (2017), and Cao et al. (2011), we hypothesize to find evidence for CUP.

Hypothesis 1: The shot success of penalties in crunch time is smaller than in the rest of the game.

The findings of Cao et al. (2011), González-Díaz et al. (2012), Klaassen and Magnus (2014), and Krumer (2020) suggest that in very important moments of a competition weaker players,

in our case players of a lower league, are more likely to suffer a drop in performance than stronger players.

Hypothesis 2: Weaker players are more inclined to performance decreases in crunch time than stronger players.

The meta-analysis of Judge et al. (1997) shows that CES is positively associated with job performance (see also Judge & Bono, 2001 and Chang et al., 2012). We hypothesize that this relationship also holds in our setting and argue that better performance reduces the probability of CUP.

Hypothesis 3: Higher scores in CES result in higher shot success in crunch time.

Prospect theory suggests that risk preferences are especially important during crunch time – when the outcome of the penalty is likely to shift the prospect from the loss to the gain domain or vice versa (Kahneman & Tversky, 1979; Pope & Schweitzer, 2011). According to Grund et al. (2013) as well as Lehman and Hahn (2013) increased risk taking of lagging teams impedes their performance.

Hypothesis 4: The individual willingness to take risks reduces the shot success in crunch time. This effect is more pronounced if the player's team is lagging rather than leading.

Toma (2017) observes CUP for players of lagging teams. Likewise, Worthy et al. (2009) expected to find CUP for players of lagging but not of leading teams.

Hypothesis 5: The effect of crunch time is more pronounced if the player's team is lagging rather than leading.

Baumeister and Steinhilber (1984), Wallace et al. (2005), Dohmen (2008), and Harb-Wu and Krumer (2019) find that players choke more often in their home stadiums.

Hypothesis 6: The effect of crunch time is more pronounced if the team is playing at home rather than away.

Whereas Klaassen and Magnus (2014) find that women are more inclined to CUP than men, Cohen-Zada et al. (2017) obtain the opposite result. Toma (2017) observes no differences in CUP from the free-throw line by gender.

Hypothesis 7: The effect of crunch time is the same for women and men.

4. Results

4.1 Descriptive statistics

From 5506 penalties of the season 2019/2020 in the first three handball leagues in Germany, we know the psychological traits of the 236 throwers who answered our survey.[2] In our sample, 25 male and 18 female players threw 645 and 537 penalties in the first leagues, 22 male and 29 female players 696 and 672 penalties in the second leagues, and 53 male and 89 female players 1323 and 1633 in the third leagues. We distributed our online-survey (programmed with LamaPoll) in German and English to 146 German handball clubs – every club from the first and second league and every club from two out of four divisions of the third league. Around 60% (86 clubs) agreed to distribute the survey to their seven-meter throwers, and 54% (236 out of 440) of them answered the survey.

Table 2 summarizes the descriptive statistics of all the variables that we assess from the handball games and our survey. The mean value of the dependent variable shot success is 77%, which corresponds to a miss at every fourth or fifth seven-meter. We observe fewer penalties by men (48%) than by women. The average age in our sample is 25.49 years. The youngest player was 16 years old in the middle of the 2019/2020 season and the oldest 38. The thrower's right hand is typically the dominant throwing hand (in 77%). Most of the penalties of our data set are from the third league – the mean of the league variable is 2.32. 12% of the penalties are thrown in crunch time, and most of the penalties are thrown when the player's team is lagging (47% vs. 38% when the team is leading and 14% when the score is level). On average, the players have a relatively high CSE of 4.02 (on a scale from 1 to 5) and assess their risk taking with 4.92 (on a scale from 1 to 7). Only 30% of the participants reported having personally experienced CUP.

[2] In Germany, the first three leagues of Olympic disciplines, like handball, are categorized as professional sports. Whereas handball is a full-time job for a large majority in the first league and most of the players in the second league, it is usually a part-time job for players in the third league.

Table 2: Descriptive statistics

	Variables	Mean	SD	Min	Max
Dependent variable					
1	Shot success	0.77	0.42	0	1
Individual variables					
2	Gender	0.48	0.50	0	1
3	Age	25.49	4.56	16	38
4	Throwing hand	0.77	0.42	0	1
5	League	2.32	0.81	1	3
6	CSE	4.02	0.45	1.83	5
7	Risk taking	4.92	1.06	2	7
8	Self-report choking	0.30	0.46	0	1
Competition-related variables					
9	Location	0.50	0.50	0	1
10	Crunch time	0.12	0.33	0	1
11	Score	1.24	0.68	0	2

Note: N = 5506

Figure 1 displays the shot success in crunch time vs. the rest of the game by gender, league, and the competition-related variables. It can be seen that for all variables, except for leading, the scoring percentage of the penalties is higher in crunch time compared to the rest of the game. We observe the highest shot success for women in crunch time (84.74%) and the lowest for men during the rest of the game (75.53%).

Figure 1: Shot success by crunch time, gender, league, and game-related variables

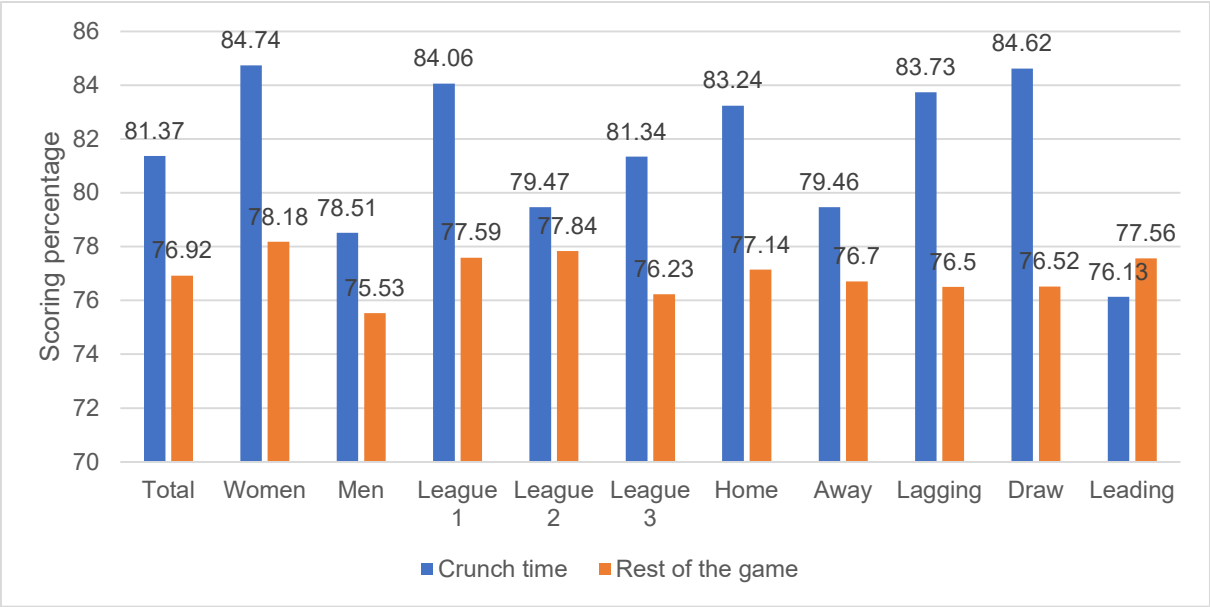

Table 3 shows the pairwise correlations of our variables. There are very small correlations between the competition-related variables and the individual variables. The highest Pearson correlation coefficient in our dataset is r=-0.39 between league and age. In the first leagues for men and women, the penalties are thrown by more experienced players than in the second and third leagues. In our data set, their average age is 26.47, 24.45, and 23.51 respectively. Furthermore, the first league throwers are more specialized – on average, they threw 27.49 penalties in league 1, 26.82 in league 2, and 20.82 in league 3 in the season 2019/2020. Men's teams are older than women's teams (25.03 vs. 23.68). Moreover, CSE and gender are significantly associated – male players assessed their CSE level higher (on average 4.16) than women (3.90). This is in line with Bleidorn et al. (2016), who find that men have higher self-esteem than women. Self-report choking correlates negatively with CSE: the higher the level of CSE, the lower the probability that a player reports having experienced CUP before. Additionally, age correlates with CSE (r=0.15): the older the subjects, the higher their self-evaluation. This is consistent with previous research showing that individual self-evaluation increases until middle adulthood before it gradually starts to decrease (Bleidorn et al. 2016; Robins et al. 2002; Kling et al. 1999). Finally, the players' self-assessment on risk taking is higher if they are male (in line with Croson & Gneezy, 2009, and Eckel & Grossman, 2008).

Table 3: Pairwise correlations

	Variables	1	2	3	4	5	6	7	8	9	10	11
1	Shot success	1										
2	Gender	-0.04**	1									
3	Age	0.01	0.24**	1								
4	Throwing hand	-0.01	-0.12**	-0.17**	1							
5	League	-0.02	-0.08**	-0.39**	0.16**	1						
6	CSE	-0.02	0.29**	0.15**	0.07**	-0.02	1					
7	Risk taking	-0.01	0.15**	0.11**	-0.14**	-0.25**	0.16**	1				
8	Self-report choking	-0.01	-0.08**	0.13**	-0.12**	-0.12**	-0.27**	0.14**	1			
9	Location	-0.01	0.02	-0.04**	0.00	0.02	0.04**	0.01	-0.06**	1		
10	Crunch time	0.04**	0.04**	0.01	0.01	-0.01	0.01	0.00	-0.02	0.00	1	
11	Score	0.00	-0.03*	-0.01	-0.01	0.02	0.01	0.01	-0.02	-0.06**	-0.05**	1

Notes: N = 5506; significance levels: ** $p < 0.01$; * $p < 0.05$

Table 4: Marginal effects of probit regressions

	Model 1		Model 2		Model 3		Model 4	
Dependent variable								
Shot success	ME	SD	ME	SD	ME	SD	ME	SD
Individual variables								
Gender	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)
Age	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Throwing hand	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
League	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
CSE	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)
Risk taking	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)
Self-report choking	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Competition-related variables								
Location	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Crunch time	0.05*	(0.02)	0.06	(0.06)	0.01	(0.18)	0.01	(0.09)
Lagging	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)
Leading	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)
Interaction terms								
League*Crunch time			0.00	(0.02)				
CSE*Crunch time					0.01	(0.04)		
Risk taking*Crunch time							0.01	(0.02)
Lagging*Crunch time								
LeadingCrunch time								
Location*Crunch time								
Gender*Crunch time								

Number of obs	5506	5506	5506	5506
Wald chi2	17.81	17.90	18.37	18.49
Prob > chi2	0.0860	0.1188	0.1049	0.1015
Log pseudolikelihood	-2928.21	-2928.20	-2928.18	-2928.09
Pseudo R2	0.0034	0.0034	0.0034	0.0035

Notes: Marginal effects in the ME columns; standard deviations in parentheses in the SD columns; standard deviations clustered at player level; reference category for score: draw; significance level: *** p < 0.001; ** p < 0.01; * p < 0.05

	Modell 5		Modell 6		Modell 7		Modell 8		Modell 9	
Dependent variable										
Shot success	ME	SD	ME	SD	ME	SD	ME	SD	ME	SD
Individual variables										
Gender	-0.03**	(0.01)	-0.03**	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)
Age	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Throwing hand	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
League	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
CSE	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)
Risk taking	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)
Self-report choking	-0.02	(0.01)	-0.02	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Competition-related variables										
Location	-0.01	(0.01)	-0.01	(0.01)	-0.00	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Crunch time	0.02	(0.03)	0.08**	(0.03)	0.07**	(0.02)	0.07*	(0.03)	0.12**	(0.03)
Lagging	-0.01	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)
Leading	-0.01	(0.02)	0.01	(0.02)	-0.00	(0.02)	-0.00	(0.02)	0.01	(0.02)

Interaction terms					
League*Crunch time					
CSE*Crunch time					
Risk taking*Crunch time					
Lagging*Crunch time	0.06 (0.03)				
Leading*Crunch time		-0.09* (0.04)			-0.10** (0.04)
Location*Crunch time			- 0.03 (0.03)		
Gender*Crunch time				-0.04 (0.03)	-0.05 (0.04)
Number of obs	5506	5506	5506	5506	5506
Wald chi2	20.40	23.32	22.08	20.10	27.53
Prob > chi2	0.0598	0.0252	0.0367	0.0653	0.0105
Log pseudolikelihood	-2926.58	-2924.74	-2927.66	-2927.48	-2923.69
Pseudo R2	0.0040	0.0046	0.0036	0.0037	0.0050

Notes: Marginal effects in the ME columns; standard deviations in parentheses in the SD columns; standard deviations clustered at player level; reference category for score: draw; significance level: *** p < 0.001; ** p < 0.01; * p < 0.05

4.2 Multivariate analysis

Table 4 presents the marginal effects of probit regressions trying to explain the shot success of handball penalties. Model 1 includes all the variables discussed in 4.1. Similar to Gonzáles-Díaz et al. (2012) and Ötting et al. (2020), our results indicate clutch performance rather than CUP. Our subjects are 5 percentage points more likely to hit the seven-meter if it is thrown in crunch time. This effect is not only statistically, but also economically significant. Especially in the men's leagues, a lot of games are decided by just one goal difference (around 15% in our sample, the score difference of nearly half of our games is less than 4 goals). By definition, a penalty in crunch time is decisive for the game. The final score difference in the leagues we analyzed in the season 2019/2020 was 4.80 on average; if there was a penalty in crunch time: 2.57. Therefore, it is very likely that the effect size of a miss or hit in this crucial situation is a lower boundary for the effect on the team's winning probability – especially if the outcome creates momentum (Lehman and Hahn, 2013). Our finding suggests that the analyzed handball players perform best when it matters the most. Thus, we can reject Hypothesis 1. The increased pressure of crunch time seems to induce more focus and concentration of the thrower rather than anxiety. According to attentional control theory (Eysenck et al., 2007), they appear to use more processing resources in the decisive throws. Moreover, Model 1 shows that women score the penalties, ceteris paribus, with a 3 percentage point higher probability than men. Although men more often select themselves into competitions than women, they do not necessarily perform better in competitive environments (Niederle & Vesterlund, 2011; Soll & Klayman, 2004; Cohen-Zada et al., 2007). Our data rather suggest the opposite. For the other control variables, we do not find any significant direct effect on shot success.

To analyze the impact of psychological pressure on performance deeper, Models 2 – 9 include interaction terms with our treatment variable crunch time. We follow Ai and Norton (2013) in correcting potentially biased interaction effects, their standard errors, and p-values in the nonlinear probit model. According to Klaassen and Magnus (2014) as well as González-Díaz et al. (2012), especially weaker players are likely to choke under pressure. Model 2 adds the interaction effect between crunch time and league. League represents the general performance level of a player. Yet, it has no direct effect on the shot success and does not interact with crunch time. As a result, we find no evidence supporting Hypothesis 2. Our data suggest that weaker players do not perform worse in crunch time. It could be that the league measures the potential to perform well under pressure too roughly. Therefore, we include the interaction of CSE and crunch time in Model 3. We do not find support for Hypothesis 3 that more positive self-evaluation leads to better performance in general and specifically in crunch time.

In Model 4, the interaction between risk taking and crunch time is not significant. The general propensity to take risks does not seem to influence the shot success of penalties, neither in crunch time nor in the rest of the game. If we consider the interaction with lagging in this analysis, the marginal effect of the three-way interaction is 0.01 and not significant (p=0.087). If we instead include the interaction with leading, the marginal effect is -0.02 and significant (p=0.020). Thus, higher risk taking in crunch time only reduces shot success if the team is leading. This speaks against Hypothesis 4. However, we do not have data on specific risk taking of the penalties, similar to the speed (Pope and Schweitzer, 2011) or the style[3] of the throw. Models 5 and 6 examine the interaction effects between crunch time with lagging and leading, respectively. In Model 5, lagging increases the relationship between crunch time and shot success by 6 percentage points, but not significantly (p=0.081). The main effect of crunch time is not significant in this model. In Model 6, however, the main effect of crunch time increases to 8 percentage points and is highly significant. The significant interaction effect of -9 percentage points between crunch time and leading means that the positive effect of crunch time on shot success is contradicted if the thrower's team is leading. Thus, Model 6 supports the descriptive statistics presented in Figure 1. These results confirm Hypothesis 5. Figure 2 illustrates the interaction between crunch time and leading.

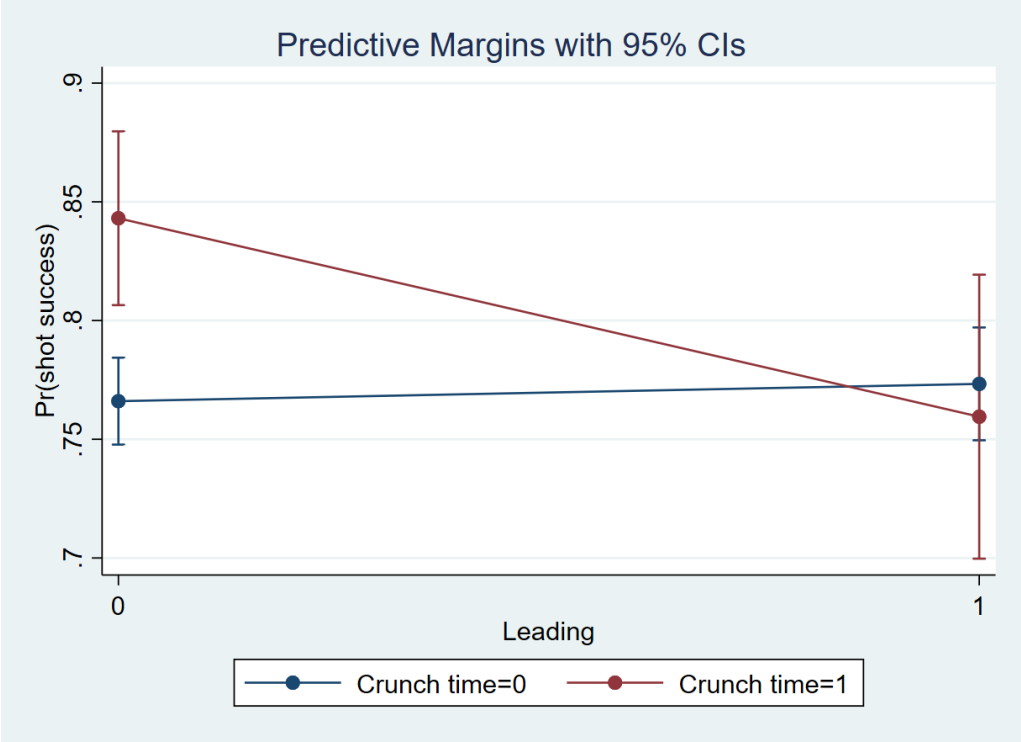

Figure 2: Interaction effect of *Crunch time* and *Leading* (Model 6, Table 4) according to Williams (2012)

[3] A twister or a lob would be a risky throw. Likewise, Bozhinov and Grote (2018) use the jump serve as a proxy for risk taking in volleyball.

Model 7 tests if the effect of crunch time is especially relevant for home teams. The location of the game (home or away) does not directly influence the shot success of penalties, and the interaction effect between crunch time and location is not significant, which speaks against Hypothesis 6. In Model 8, we analyze the interaction effect between crunch time and gender on shot success. These are the only two independent variables that have significant main effects on the seven-meter scoring (see Model 1). Both main effects remain significant in Model 8. The interaction effect is negative (-0.04), but not significant (p=0.307). Both genders perform better in crunch time, and we do not find evidence that either men or women can cope better with psychological pressure. This finding is different from Klaassen and Magnus (2014) and Cohen-Zada et al. (2017) but in line with Toma (2017), confirming Hypothesis 7.

Model 9 extends Model 1 by the interaction effects from models 6 and 8. While women still have a 3 percentage points higher chance of scoring the seven-meter in this model, the main effect of crunch time in Model 9 increases to 12 percentage points ($p<0.001$). The interaction effect between crunch time and leading is -10 percentage points in this model ($p<0.01$) in this model. Thus, Model 9 confirms models 1, 6, and 7. To further check the robustness of our results, we restricted crunch time to the last 5 minutes only (instead of 15 minutes) and rerun our analyses – without changing anything else of the crunch time definition. The results in the appendix show that we can replicate our main findings with the restricted crunch time. However, only 4% (instead of 12%) of the analyzed penalties are thrown in the last 5 minutes.

5. Conclusion and discussion

The primary goal of our study was to gain insights into the performance of professional handball players in situations of particular pressure. Therefore, we compared the shot success of decisive penalties to the shot success of less important penalties. Our findings support the idea of performing best when it matters the most instead of choking under pressure. The positive effect of pressure was especially relevant when the thrower's team was not leading. On average, women performed better than men in our sample. However, pressure affected the players' performance not significantly different by gender.

Additionally, we surveyed the players' psychological traits that are supposed to influence behavior under pressure. The results show no significant interaction effects between these traits and psychological pressure on shot success. The reason could be that our subjects are selected to be regularly responsible for throwing penalties for their team. Thus, they are experienced and trained to perform well under psychological pressure. Whereas 63% of golfers and archers reported CUP experiences in the study by Clarke et al. (2020), only 30% remembered personal CUP in our sample. Furthermore, the average CSE of our players is

higher compared to the results of Brückner (2020), Heilmann and Jonas (2010), and Judge et al. (2003). Thus, our findings seem to be particularly relevant for specialists who are used to execute a familiar task under high psychological pressure, taking responsibility for their team.

There would be a selection bias if participants of our survey systematically deviate from the full sample of penalty throwers in the 2019/2020 season in terms of behavior under pressure. To check this, we compare the shot success in our sample to the season averages of the penalty throwers in 2019/2020 who did not take part in our survey.

Table 5: Comparison of penalty shot success of survey participants and non-participants

	Shot success in %							
	Overall	Crunch time	Rest of the game	Home	Away	Behind	Draw	Lead
Participants	77.46	81.13	76.95	77.89	77.03	77.30	77.44	77.66
Non-participants	74.86	76.67	74.57	75.03	74.68	74.47	74.74	75.35

Table 5 shows that the scoring probability of our subjects' penalties in crunch time is higher than of non-participants. However, this difference is not significant according to a two-sided t-test (p=0.310). The overall shot success of our participants is significantly higher than of non-participants (p=0.01). Yet if we only analyze players who threw at least five penalties in the 2019/2020 season, the difference is no longer significant (p=0.228).

Another potential limitation of our empirical approach concerns the definition of crunch time. This treatment variable should distinguish decisive stages of the game (high-pressure situations) from conventional stages of the game (lower-pressure situations). Finding clear-cut definition criteria for crunch time is not easy. For the analysis, we need to set unambiguous and preferably objective rules indicating if the penalty is thrown in a decisive stage of the game or not. There could be game- or player-specific heterogeneity not covered by our definition, or by our control variables, that make players perceive a penalty as a high-pressure situation although it is not thrown in crunch time. Similarly, there might be penalties in crunch time that players perceive to throw under low pressure. Thus, we conducted robustness checks which confirm that our main findings are robust against different definitions of crunch time (see appendix for the results in which we restrict crunch time to 5 minutes). Furthermore, we included a self-assessment of CUP which is consistent with our definition of crunch time. During conventional game stages, the shot success of players who stated having experienced CUP before is the same as the shot success of players without CUP experiences (77%). Yet they distinguish each other in the expected direction in crunch time: 78% with CUP experiences vs. 83% without. The marginal effect of this self-assessment on shot success in

a variant of Model 1 (Table 4) only for crunch time penalties is 6 percentage points, but not significant (p=0.101).

While objective game-specific control variables and player characteristics are easy to assess, it is difficult to measure the most interesting variables in our context, the psychological traits related to performance under pressure. Future studies should also consider different psychological control variables, e.g. Clarke et al. (2020) surveyed the fear of negative evaluation and the desire for perfect self-presentation in addition to the Big 5 and CSE. Furthermore, there could be psychological effects specific to the situation of the game that are not captured by our control variables. For instance, Lehman and Hahn (2013) show that negative momentum leads to less risk taking.

Finally, it can be discussed whether a seven-meter in handball is an appropriate unit of observation to analyze individual performance under pressure. The fact that the opposing goalkeeper influences the thrower's performance might bias our results. Free-throws in basketball, putts in golf, and dart throws may be easier to analyze in this respect. However, similar to the argument of Apesteguia and Palacios-Huerta (2010) in soccer, the thrower's performance is the main determinant of the shot success of a handball penalty – the average scoring rate is nearly 80% in our sample. Furthermore, on average, the influence of the goalkeeper levels out across treatments (during crunch time and the rest of the game). And one of the reasons why we study seven-meter throws is that the thrower's anxiety in crunch time may be especially high because he or she directly looks into the opponent's eyes before the whistle blows.[4]

[4] For soccer shoot-outs, Jordet and Hartman (2008) observe that players are more likely to look away from the goalkeeper when preparing kicks that lead to a loss if they are missed.

References

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123-129.

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *The Review of Economic Studies*, 76(2), 451-469.

Arslan, R. C., Brümmer, M., Dohmen, T., Drewelies, J., Hertwig, R., & Wagner, G. G. (2020). How people know their risk preference. *Nature Scientific Reports*, *10*(1), 1-14.

Apesteguia, J., & Palacios-Huerta, I. (2010). Psychological pressure in competitive environments: Evidence from a randomized natural experiment. *American Economic Review*, *100*(5), 2548-64.

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46(3), 610-620.

Baumeister, R. F., & Steinhilber, A. (1984). Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships. *Journal of Personality and Social Psychology*, 47(1), 85-93.

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2015). Kurzskala zur Erfassung der Risikobereitschaft (R-1), [Short-scale for measuring risk taking], *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS).* https://doi.org/10.6102/zis236.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology,* 59(5), 960-970.

Bleidorn, W., Arslan, R. C., Denissen, J. J. A., Rentfrow, P. J., Gebauer, J. E., Potter, J., Gosling, S. D. (2016). Age and gender differences in self-esteem – A cross-cultural window. *Journal of Personality and Social Psychology*, 111(3), 396-410.

Bono, J. E., & Judge, T. A. (2003). Core self-evaluations: A review of the trait and its role in job satisfaction and job performance. *European Journal of Personality*, 17(S1), 5-18.

Boon, C. T., Low, D. F. L., Lim, S. Y., Ng, X. Q., & Wong, P. P. (2011). *The Influence of Core Self-Evaluation (CSE) on Employees' Motivation in Continuing Education*. London: Sage.

Booth, A. L., & Nolen, P. J. (2021). *Gender and Psychological Pressure in Competitive Environments*. IZA Discussion Paper No. 14174. http://ftp.iza.org/dp14174.pdf.

Bozhinov, V., & Grote, N. (2018). *Performance under Pressure on the Court: Evidence from Professional Volleyball*. Available at SSRN 3312153. http://dx.doi.org/10.2139/ssrn.3312153.

Brückner, J.-P. (2020). *ACSES – eine deutschsprachige allgemeine Core Self-Evaluations-Skala*, [ACSES – A German generalized core-self-evaluation scale], Discussion Paper, Christian- Albrechts- Universität zu Kiel. https://macau.uni-kiel.de/receive/macau_mods_00000405.

Bucciol, A., & Castagnetti, A. (2020). Choking under pressure in archery. *Journal of Behavioral and Experimental Economics*, 89, 101581.

Burke, R. J., & Ng, E. (2006). The changing nature of work and organizations: Implications for human resource management. *Human Resource Management Review*, 16(2), 86-94.

Bühren, C., & Krabel, S. (2019). Human performance after success and failure: evidence from the NBA. *The International Journal of Human Resource Management*, 1-24.

Bühren, C., & Steinberg, P. J. (2019). The impact of psychological traits on performance in sequential tournaments: Evidence from a tennis field experiment. *Journal of Economic Psychology*, 72, 12-29.

Broihanne, M. H., Merli, M., & Roger, P. (2014). Overconfidence, risk perception and the risk-taking behavior of finance professionals. *Finance Research Letters*, 11(2), 64-73.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., & Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during Covid-19. *Economics Letters*, 198, 109664.

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, 125(3), 367-383.

Cao, Z., Price, J., & Stone, D. (2011). Performance under pressure in the NBA. *Journal of Sports Economics*, 12(3), 231-252.

Chang, C.-H., Ferris, D. L., Johnson, R. E., Rosen, C. C., & Tan, J. A. (2012). Core self-evaluations: A review and evaluation of the literature. *Journal of Management*, 38(1), 81-128.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50-58.

Christmann, J., Akamphuber, M., Müllenbach, A. L., & Güllich, A. (2018). Crunch time in the NBA–The effectiveness of different play types in the endgame of close matches in professional basketball. *International Journal of Sports Science & Coaching*, 13(6), 1090-1099.

Clarke, P., Sheffield, D., & Akehurst, S. (2020). Personality predictors of yips and choking susceptibility. *Frontiers in Psychology*, Article 2784. https://doi.org/10.3389/fpsyg.2019.02784.

Cohen-Zada, D., Krumer, A., Rosenboim, M., & Shapir, O. M. (2017). Choking under pressure and gender: Evidence from professional tennis. *Journal of Economic Psychology*, 61, 176-190.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.

Dilmaghani, M. (2020). Gender differences in performance under time constraint: Evidence from chess tournaments. *Journal of Behavioral and Experimental Economics*, 89, 101505.

Dohmen, T. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, 65(3/4), 636-653.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology*, 91(1), 40.

Eckel, C., & Grossman, P. (2008). Men, women and risk aversion: Experimental evidence. In: C. R. Plott, & V. L. Smith (Eds.), *Handbook of Experimental Economics Results*, 1061-1073. Amsterdam: Elsevier.

Elmore, R., & Urbaczewski, A. (2021). Loss aversion in professional golf. Journal of Sports Economics, 22 (2), 1-16. https://doi.org/10.1177/1527002520967403.

Erez, A., & Judge, T. A. (2001). Relationship of core self-evaluations to goal setting, motivation, and performance. *Journal of Applied Psychology*, 86(6), 1270-1279.

Ertug, G., & Maoret, M. (2020). Do Coaches in the National Basketball Association Actually Display Racial Bias? A Replication and Extension. *Academy of Management Discoveries*, 6(2), 206-234.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, 7(2), 336.

Farmer, R. F., Jarvis, L. L., Berent, M. K., & Corbett, A. (2001). Contributions to global self-esteem: The role of importance attached to self-concepts associated with the five-factor model. *Journal of Research in Personality*, 35(4), 483-499.

Fujita, F., Diener, E., & Sandvik, E. (1991). Gender differences in negative affect and well-being: The case for emotional intensity. *Journal of Personality and Social Psychology*, 61(3), 427-434.

Gardner, D. G., & Pierce, J. L. (2010). The core self-evaluation scale: Further construct validation evidence. *Educational and Psychological Measurement*, 70(2), 291-304.

González-Díaz, J., Gossner, O., & Rogers, B. W. (2012). Performing best when it matters most: Evidence from professional tennis. *Journal of Economic Behavior & Organization*, 84(3), 767-781.

Grund, C., Höcker, J., & Zimmermann, S. (2013). Incidence and consequences of risk-taking behavior in tournaments - Evidence from the NBA. *Economic Inquiry*, 51(2), 1489-1501.

Harb-Wu, K., & Krumer, A. (2019). Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, 166, 246-262.

Hardies, K., Breesch, D., & Branson, J. (2013). Gender differences in overconfidence and risk taking: Do self-selection and socialization matter?. *Economics Letters*, 118(3), 442-444.

Harshman, R. A., & Paivio, A. (1987). Paradoxical sex differences in self-reported imagery. *Canadian Journal of Psychology*, 41(3), 287-302.

Heilmann, T., & Jonas, K. (2010). Validation of a German-language core self-evaluations scale. *Social Behavior and Personality*, 38(2), 209-226.

Hickman, D. C., & Metz, N. E. (2015). The impact of pressure on performance: Evidence from the PGA TOUR. *Journal of Economic Behavior & Organization*, 116, 319-330.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644-1655.

Jetter, M., & Walker, J. K. (2015). Game, set, and match: Do women and men perform differently in competitive situations?. *Journal of Economic Behavior & Organization*, 119, 96-108.

Jordet, G., & Hartman, E. (2008). Avoidance motivation and choking under pressure in soccer penalty shootouts. *Journal of Sport and Exercise Psychology*, 30(4), 450-457.

Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. A. P. M. (2007). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *Journal of Sports Sciences*, 25(2), 121-129.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits – self-esteem, generalized self-efficacy, locus of control, and emotional stability – with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, 86(1), 80-92.

Judge, T. A., Bono, J. E., & Locke, E. A. (2000). Personality and job satisfaction: The mediating role of job characteristics. *Journal of Applied Psychology*, 85(2), 237-249.

Judge, T. A., Erez, A., & Bono, J. E. (1998). The power of being positive: The relation between positive self-concept and job performance. *Human Performance*, 11(2/3), 167-187.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, 83(3), 693-710.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology*, 56(2), 303-331.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology*, 87(4), 797-810.

Judge, T. A., & Kammeyer-Mueller, J. D. (2011). Implications of core self-evaluations for a changing organizational context. *Human Resource Management Review*, 21(4), 331-341.

Judge, T. A., Locke, E. A., & Durham, C. C. (1997). The dispositional causes of job satisfaction: A core evaluations approach. *Research in Organizational Behavior*, 19, 151-188.

Judge, T. A., van Vianen, A. E. M., & de Pater, I. E. (2004). Emotional stability, core self-evaluations, and job outcomes: A review of the evidence and an agenda for future research. *Human Performance*, 17(3), 325-346.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.

Kam, C. D. (2012). Risk attitudes and political participation. *American Journal of Political Science*, 56(4), 817-836.

Kammeyer-Mueller, J. D., Judge, T. A., & Scott, B. A. (2009). The Role of Core Self-Evaluations in the Coping Process. *Journal of Applied Psychology*, 94(1), 177-195.

Kirmani, S. S., Attiq, S., Bakari, H., & Irfan, M. (2019). Role of core self evaluation and acquired motivations in employee task performance. *Pakistan Journal of Psychological Research*, 34(2), 401-418.

Klaassen, F., & Magnus, J.R. (2014). *Analyzing Wimbledon: The Power of Statistics*. New York: Oxford University Press.

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin*. 125(4), 470-500.

Kraus, S., Clauss, T., Breier, M., Gast, J., Zardini, A., & Tiberius, V. (2020). The economics of COVID-19: initial empirical evidence on how family firms in five European countries cope with the corona crisis. forthcoming: *International Journal of Entrepreneurial Behavior & Research*.

Krumer, A. (2020). Pressure versus ability: Evidence from penalty shoot-outs between teams from different divisions. *Journal of Behavioral and Experimental Economics*, 89, 101578.

Lehman, D. W., Hahn, J. (2013). Momentum and organizational risk taking: Evidence from the National Football League. *Management Science*, 59(4), 852-868.

Lin, S., Lamond, D., Yang, C. L., & Hwang, M. (2014). Personality traits and simultaneous reciprocal influences between job performance and job satisfaction. *Chinese Management Studies*, 8(1), 6-26.

Mesagno, C., Garvey, J., Tibbert, S. J., & Gröpel, P. (2019). An investigation into handedness and choking under pressure in sport. *Research Quarterly for Exercise and Sport*, 90(2), 217-226.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, 122(3), 1067-1101.

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, 3(1), 601-630.

Otten, M. (2009). Choking vs. clutch performance: A study of sport performance under pressure. *Journal of Sport and Exercise Psychology*, 31(5), 583-601.

Ötting, M., Deutscher, C., Schneemann, S., Langrock, R., Gehrmann, S., & Scholten, H. (2020). Performance under pressure in skill tasks: An analysis of professional darts. *PloS ONE*, 15(2). https://doi.org/10.1371/journal.pone.0228870.

Page, K., & Page, L. (2010). Alone against the crowd: Individual differences in referees' ability to cope under pressure. *Journal of Economic Psychology*, 31(2), 192-199.

Paserman, M. D. (2007). *Gender differences in performance in competitive environments: Evidence from professional tennis players*. IZA Discussion Paper (2834). http://ftp.iza.org/dp2834.pdf.

Paserman, M. (2010). *Gender differences in performance in competitive environments: Evidence from professional tennis players*. Mimeo. http://people.bu.edu/paserman/papers/Paserman_Tennis_January2010.pdf.

Pope, D. G., & Schweitzer, M. E. (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 101(1), 129-157.

Robins, R. W., Tracy, J. L., Trzesniewski, K. H., Potter, J., & Gosling, S. D. (2001). Personality correlates of self-esteem. *Journal of Research in Personality*, 35(4), 463-482.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology and Aging*, 17(3), 423-434.

Sanders, S., & Walia, B. (2012). Shirking and "choking" under incentive-based pressure: A behavioral economic theory of performance production. *Economics Letters*, 116(3), 363-366.

Schwarzmüller, T., Brosi, P., Duman, D., & Welpe, I. M. (2018). How does the digital transformation affect organizations? Key themes of change in work design and leadership. *MREV Management Revue*, 29(2), 114-138.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 299-314.

Toma, M. (2017). Missed shots at the free-throw line: Analyzing the determinants of choking under pressure. *Journal of Sports Economics*, 18(6), 539-559.

Vijayan, M. (2017). Impact of job stress on employees' job performance in Aavin, Coimbatore. *Journal of Organisation and Human Behaviour*, 6(3), 21.

Wells, B., & Skowronski, J. (2012). Evidence of choking under pressure on the PGA Tour. *Basic and Applied Social Psychology*, 34(2), 175-182.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, 12(2), 308-331.

Wilson, M. R., Wood, G., & Vine, S. J. (2009a). Anxiety, attentional control, and performance impairment in penalty kicks. *Journal of Sport and Exercise Psychology*, 31(6), 761-775.

Wilson, M. R., Vine, S. J., & Wood, G. (2009b). The influence of anxiety on visual attentional control in basketball free throw shooting. *Journal of Sport and Exercise Psychology*, 31(2), 152-168.

Worthy, D. A., Markman, A. B., & Maddox, W. T. (2009). Choking and excelling at the free throw line. *The International Journal of Creativity and Problem Solving*, 19, 53-58.

Wright, T. A., Cropanzano, R., & Meyer, D. G. (2004). State and trait correlates of job performance: A tale of two perspectives. *Journal of Business and Psychology*, 18(3), 365-383.

Appendix

Robustness check with crunch time restricted to 5 minutes

Figure A1: Shot success by crunch time (restricted to 5 minutes), gender, league, and game-related variables

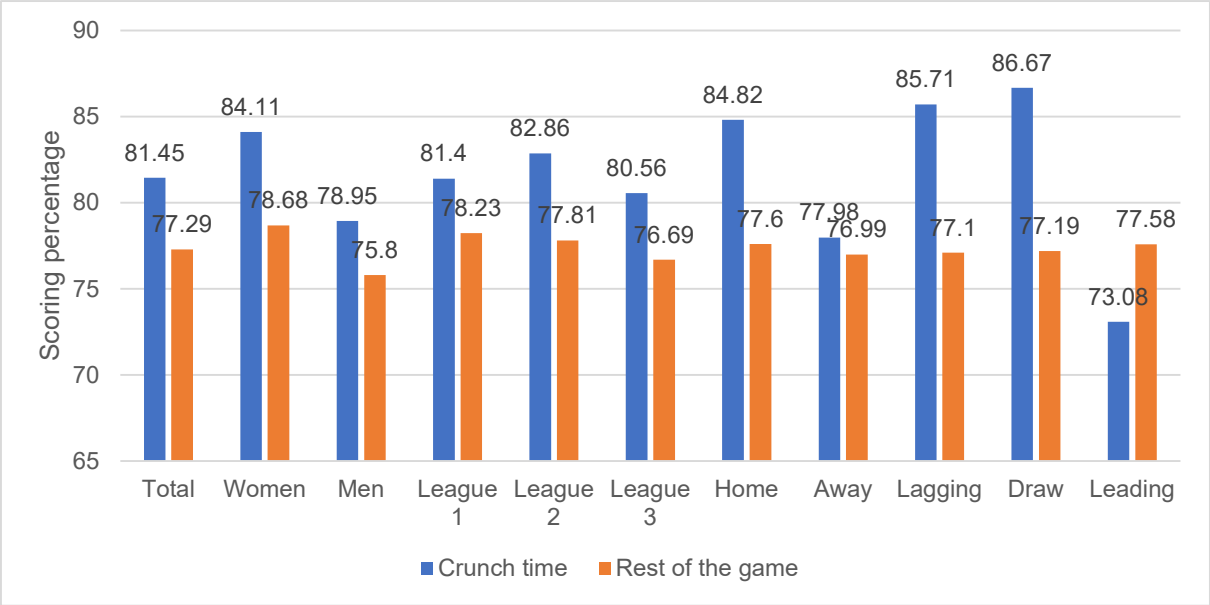

Table A1: Marginal effects of probit regressions with crunch time restricted to 5 minutes

	Model 1		Model 2		Model 3		Model 4	
Dependent variable								
Shot success	ME	SD	ME	SD	ME	SD	ME	SD
Individual variables								
Gender	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)
Age	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Throwing hand	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
League	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
CSE	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)
Risk taking	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)
Self-report choking	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Competition-related variables								
Location	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Crunch time	0.05	(0.03)	0.04	(0.04)	-0.26	(0.30)	-0.12	(0.13)
Lagging	0.00	(0.02)	0.00	(0.02)	0.00	(0.02)	0.00	(0.02)
Leading	0.00	(0.02)	0.00	(0.02)	0.00	(0.02)	0.01	(0.02)
Interaction terms								
League*Crunch time			0.00	(0.04)				
CSE*Crunch time					0.07	(0.08)		
Risk taking*Crunch time							0.03	(0.03)
Lagging*Crunch time								
LeadingCrunch time								
Location*Crunch time								

Gender*Crunch time				
Number of obs	5506	5506	5506	5506
Wald chi2	15.51	15.52	15.87	18.65
Prob > chi2	0.0860	0.2145	0.1974	0.0974
Log pseudolikelihood	-2930.80	-2930.79	-2930.27	-2930.18
Pseudo R2	0.0025	0.0025	0.0027	0.0027

Notes: Marginal effects in the ME columns; standard deviations in parentheses in the SD columns; standard deviations clustered at player level; reference category for score: draw; significance level: *** p < 0.001; ** p < 0.01; * p < 0.05

	Modell 5		Modell 6		Modell 7		Modell 8		Modell 9	
Dependent variable										
Shot success	ME	SD	ME	SD	ME	SD	ME	SD	ME	SD
Individual variables										
Gender	-0.03*	(0.02)	-0.03**	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)	-0.03*	(0.01)
Age	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Throwing hand	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
League	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
CSE	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)	-0.01	(0.02)
Risk taking	-0.00	(0.01)	-0.00	(0.01)	-0.00	(0.01)	- 0.00	(0.01)	-0.00	(0.01)
Self-report choking	-0.02	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Competition-related variables										
Location	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
Crunch time	0.01	(0.04)	0.10*	(0.04)	0.08*	(0.04)	0.06	(0.05)	0.12*	(0.06)
Lagging	-0.01	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)
Leading	-0.01	(0.02)	-0.00	(0.02)	-0.00	(0.02)	-0.00	(0.02)	0.00	(0.02)

Interaction terms										
League*Crunch time										
CSE*Crunch time										
Risk taking*Crunch time										
Lagging*Crunch time	0.09	(0.06)								
Leading*Crunch time			-0.14*	(0.07)					-0.15*	(0.07)
Location*Crunch time					- 0.07	(0.06)				
Gender*Crunch time							-0.03	(0.01)	-0.04	(0.06)
Number of obs	5506		5506		5506		5506		5506	
Wald chi2	18.65		23.21		17.06		15.83		23.58	
Prob > chi2	0.0973		0.0260		0.1474		0.1993		0.0352	
Log pseudolikelihood	-2929.63		-2928.08		-2930.08		-2930.69		-2927.91	
Pseudo R2	0.0029		0.0035		0.0028		0.0026		0.0035	

Notes:Marginal effects in the ME columns; standard deviations in parentheses in the SD columns; standard deviations clustered at player level; reference category for score: draw; significance level: *** p < 0.001; ** p < 0.01; * p < 0.05

Survey

CSE scale (Judge et al., 2003)

Please rate the following 12 questions on a scale of 1-5

1 = strongly disagree 2 = disagree 3 = neutral 4 = agree 5 = strongly agree

	1	2	3	4	5
1) I am confident I get the success I deserve in life.	◯	◯	◯	◯	◯
2) Sometimes I feel depressed.	◯	◯	◯	◯	◯
3) When I try, I generally succeed.	◯	◯	◯	◯	◯
4) Sometimes when I fail I feel worthless.	◯	◯	◯	◯	◯
5) I complete tasks successfully.	◯	◯	◯	◯	◯
6) Sometimes, I do not feel in control of my work.	◯	◯	◯	◯	◯
7) Overall, I am satisfied with myself.	◯	◯	◯	◯	◯
8) I am filled with doubts about my competence.	◯	◯	◯	◯	◯
9) I determine what will happen in my life.	◯	◯	◯	◯	◯
10) I do not feel in control of my success in my career.	◯	◯	◯	◯	◯
11) I am capable of coping with most of my problems.	◯	◯	◯	◯	◯
12) There are times when things look pretty bleak and hopeless to me.	◯	◯	◯	◯	◯

Willingness to take risks - (Beierlein et al., 2015):

Please rate the following question on a scale of 1-7.

1 = not at all willing to take risks to **7 = very willing to take risks**

	1	2	3	4	5	6	7
How do you see yourself - how willing are you in general to take risk?	◯	◯	◯	◯	◯	◯	◯

Self-report Choking (Clarke et al., 2020):

Have you ever experienced a dramatic drop in performance that has been out of your control?

◯ Yes ◯ No