



Tests and academic cheating: do learning tasks influence cheating by way of negative evaluations?

Kristin Wenzel¹ · Marc-André Reinhard¹

Received: 22 August 2019 / Accepted: 9 April 2020 / Published online: 27 April 2020
© The Author(s) 2020

Abstract

Desirable difficulties like tests were often shown to increase long-term learning. However, due to the complexity and difficulty of such tasks, they are also argued to result in negative consequences like stress, anxiety, pressure, frustration, or negative evaluations. In other studies, such consequences were, in turn, often found to increase dishonest behaviour. Hence, the present work tests the assumptions that tests as difficult learning tasks, contrary to reading, lead to more negative evaluations of the learning situations, to more stress, and—directly and indirectly—to higher self-reported likelihoods of hypothetical cheating and to higher justifications for cheating. Thus, the learning situation itself, as well as negative consequences caused by the learning situation, is supposed to be linked to cheating. We conducted an online study in which participants read and imagined one of three hypothetical learning scenarios, either regarding one of two learning tests or a reading control task. Participants then rated negative consequences due to these scenarios, as well as likelihoods of cheating, and justifications for it, in a hypothetical examination. Our results showed no direct effects of the learning scenarios on likelihoods of hypothetical cheating or justifications. However, test scenarios were evaluated more negatively than the reading control scenario and these higher negative evaluations were in turn linked to higher likelihoods of own hypothetical cheating and to higher justifications. These findings indicate that tests as difficult learning tasks can indirectly influence cheating, at least in hypothetical scenarios. Future work should try to replicate and expand these results.

Keywords Tests · Cheating · Academic dishonesty · Desirable difficulties · Negative evaluations · Stress perceptions

✉ Kristin Wenzel
kristin.wenzel@uni-kassel.de

Marc-André Reinhard
reinhard@psychologie.uni-kassel.de

¹ Department of Psychology, University of Kassel, Holländische Straße 36-38, 34127 Kassel, Germany

1 Introduction

Challenging, difficult, and intentionally hindered learning tasks have often been shown to increase long-term learning outcomes compared to learning and processing that is fluent, easy, and simple, even though learners and lecturers normally assume the contrary (e.g., Bjork 1994; Bjork and Bjork 1992, 2011; Diemand-Yauman et al. 2011; Dobson and Linderholm 2015; Karpicke et al. 2009; Kornell et al. 2011). Previous work describes multiple incantations of such difficult learning tasks, for instance, *generation* (e.g., Bertsch et al. 2007), *distributed practice* (e.g., Cepeda et al. 2006), or *disfluency* (e.g., Diemand-Yauman et al. 2011). One of the most common desirable difficulties—and part of the present work—is, however, the application of learning tests or quizzes (also often called *testing effect*, *testing*, *learning/practice tests*, *test enhanced learning*, or *retrieval practice*): Taking a learning test on studied materials, after an initial study opportunity but before the final test or examination, increases retrieval of the learned information and enhances durable long-term learning as opposed to passively consumed and read materials (e.g., Adesope et al. 2017; Dobson and Linderholm 2015; McDaniel et al. 2007; Roediger and Karpicke 2006; Rowland 2014). These beneficial effects of tests were found in different settings (e.g., in laboratory, school, or university settings), for different ages (e.g., elementary school students, high school students, or university students), when using a broad array of materials or information (e.g., longer scientific textbook paragraphs, factual information, or lists of word-pairs like vocabulary), and when applying varying test question formats (e.g., multiple choice or short-answer questions inducing free recall, cued recall, or recognition; e.g., Adesope et al. 2017; Dobson and Linderholm 2015; Dunlosky et al. 2013; Roediger and Karpicke 2006; Rowland 2014).

Theoretically, the beneficial effects of desirable difficulties are attributed to stimulation of more elaborate cognitive processing, deeper semantic encoding, allocation of more resources, increased retention and transfer, strengthening of memory traces and associations, and anchoring of the learned information in long-term memory (e.g., Bjork 1994; Bjork and Bjork 1992, 2011; Dunlosky et al. 2013; McDaniel et al. 1988; McNamara et al. 1996; Roediger and Karpicke 2006; Rowland 2014). Higher applied (cognitive) effort during retrieval and processing, increased quality and depth of processing and encoding (induced by retrieval attempts), higher amounts of cognitive capacities and resources utilized during information processing and retrieval, higher effort needed to solve the tasks, as well as generally higher difficulty and effort induced by both the test and the underlying retrieval practice are especially valuable for the positive effects of desirable difficulties (e.g., Bertsch et al. 2007; Bjork and Bjork 1992; Karpicke and Roediger 2007; Rowland 2014; Tyler et al. 1979). Difficult tasks also reduce learners' existing overconfidence and their *illusion of competence*, which otherwise convey the mistaken assumption that read information is already internalized: The learning test and the—due to the test—reduced competence illusion also enhance meta-cognitive accuracy of the hitherto learning process, in turn triggering the allocation of more resources and deeper, more elaborate, and more

systematic processing (e.g., Alter et al. 2007; Bjork 1999; Mihalca et al. 2017; Pieger et al. 2016).

Nonetheless, although desirable difficulties are argued to be beneficial, they are also by definition demanding, complicated, and challenging. Thus, lecturers in particular often express concern about the applicability and effectiveness of such intentionally hindered learning tasks for every individual (e.g., Diemand-Yauman et al. 2011; Lipowsky et al. 2015). In line with these concerns of lecturers, researchers also proposed that desirable difficulties are only beneficial for those individuals who can handle the needed increased effort, extended thought, and more elaborated and deeper processing, and for those who can correctly retrieve information and overcome the posed challenge (e.g., Alter et al. 2013; Kaiser et al. 2018; Kornell et al. 2011; Oppenheimer and Alter 2014; Richland et al. 2005; Rowland 2014). This, however, may not prove possible for everyone: Previous studies, for instance, showed that special requirements like higher previous knowledge, higher working memory capacity, higher intelligence, and higher reading ability are relevant skills for desirable difficulties to actually increase learning outcomes (e.g., Lehmann et al. 2016; McDaniel et al. 2002; McNamara et al. 1996; Wenzel and Reinhard 2019a). Hence, it is argued that desirable difficulties are not beneficial for every learner.

Notably, apart from that, we assume that difficult tasks used in learning contexts could also result in further negative side-effects: For instance, difficult learning tasks can sometimes pose too much additional demand (regarding, for instance, cognitive capacities, processing capacities, cognitive effort, or working memory capacities) as well as too much cognitive load on the learner, especially concerning authentic and more complex tasks and high element interactivity information (this applies in particular to learners with less expertise; see e.g., Kalyuga et al. 2001; Roelle and Berthold 2017; Sweller and Chandler 1994; van Gog and Sweller 2015; Wenzel and Reinhard 2019a). Because desirable difficulties are hard to solve and challenge learners' competence illusion and overconfidence (e.g., Bjork 1999), they are also assumed to reduce self-efficiency and to increase negative emotions, pressure, or fear of failure: Empirically, difficult tasks in general trigger perceptions of threat or anxiety and experiencing difficulties or giving incorrect answers feeds negatively into learners' self-perceptions (e.g., O'Neil et al. 1969; Sarason and Sarason 1990; Schunk and Gaa 1981). Besides, performing poorly—which can happen while working on desirable difficulties—leads to experiencing stress (e.g., Sarason and Sarason 1990; Schunk and Gaa 1981). Students also perceived difficult learning tasks and tasks that required more time and effort—and thus more workload—as more stress-inducing than easier tasks (e.g., Kausar 2010). Moreover, a laboratory study showed that learning tests resulted in more experienced pressure compared to a re-reading control task, even controlling for participants' dispositional anxiety (Hinze and Rapp 2014). Tests with high-stakes—induced by stating that monetary rewards were dependent of individuals' test results—were perceived as even more pressuring than tests with low-stakes in which monetary rewards were independent of individuals' test results. Thus, individuals felt some pressure simply from taking learning tests. Additionally, high-stake learning tests led to more anxiety than did low-stake tests and also negatively influenced participants' attitudes and interests (Hinze and Rapp 2014). Fittingly, participants in a laboratory setting that learned with a test,

compared to students that learned through reading the same information, evaluated the learning situation as more negative and experienced more stress and anxiety, even controlling for individual differences like trait stress and trait anxiety (Study 2, Wenzel and Reinhard 2019b). Moreover, due to the increased effort and the challenge learners must overcome when working with desirable difficulties in their courses, they are also argued to feel treated unfairly by their lecturers, in particular because they normally believe easy and fluent learning to be more effective (e.g., Karpicke et al. 2009; Kornell et al. 2011). Hence, we assume that difficult learning tasks should feel especially hard, pointless, and unfair.

Most important, these just described negative consequences (like stress, anxiety, or feelings of unfairness) were in other literature often assumed to be related to deceptive behaviour like academic cheating (e.g., Agnew 1992; Houser et al. 2012; Wowra 2007; see also the following paragraphs). Hence, applying desirable difficulties as learning tasks in universities could, directly and indirectly, cause more academic cheating and increase justifications for such cheating.

1.1 Academic cheating

People generally value honesty, trustworthiness, and credibility (e.g., Geißler et al. 2013), which is why they often refuse to admit their own cheating—or at least underreport it. Nonetheless, cheating behaviour can be observed throughout our daily lives (e.g., DePaulo et al. 1996; Feldman et al. 2002) and specifically in academic contexts (e.g., Finn and Frone 2004; McCabe 2001; McCabe et al. 2001; Simha and Cullen 2012; Whitley 1998). In one American survey, for instance, 74% of the participating students reported having seriously cheated on at least one test, while over 30% admitted repetitive and serious cheating in tests and exams (McCabe 2001; see also: Simha and Cullen 2012; Whitley 1998; Wowra 2007). However, actual numbers of academic cheating may be even higher because previous studies found imbalances between what students reported and what teachers actually observed in terms of cheating behaviour (e.g., Naghdipour and Emeagwali 2013). Typical incantations of such cheating behaviour in academic contexts include using cheat sheets in exams, copying answers in tests, relying on inappropriate collaboration during exams, or plagiarism (e.g., Jensen et al. 2002; Simha and Cullen 2012; Whitley 1998).

In general, different theories regarding cheating and deception do exist, very common theories being economic models like the *rational choice theory* (e.g., Akers 1990; Becker 1968) or the *strain theory* (e.g., Agnew 1992; Agnew and White 1992; Carmichael and Piquero 2004).

1.1.1 The rational choice theory

The rational choice theory describes the assumption that individuals decide whether or not to cheat after assessing possible gains or costs of such behaviour. Hence, the expected utility due to a cost–benefit calculation is important (e.g., Becker 1968). Dishonesty is mostly shown when the (for instance) financial or social gains of cheating

outweigh the costs of such behaviours, such as feelings of guilt and immorality or facing the consequences of getting caught. Potential gains of cheating in tests would include getting better grades, achieving better results with less effort, or making a good impression on others. However, people are also motivated to maintain a positive self-concept depicting themselves as moral, trustworthy, and honest (e.g., Abeler et al. 2019; Fischbacher and Föllmi-Heusi 2013; Mazar et al. 2008; Shalvi et al. 2011). Thus, individuals show higher degrees of dishonest behaviour when they feel entitled, deserving, or justified to do so (e.g., Cameron et al. 2008; Campbell et al. 2004; Fida et al. 2018; Mazar et al. 2008; Shalvi et al. 2011, 2015). Individuals can feel justified or entitled to behave dishonestly when, for instance, they can excuse deviant behaviour through denying their own responsibility (e.g., by blaming external forces like excessive workload), through criticising those who are at the receiving end of their dishonesty (e.g., by blaming them as unfair or unethical), or through rationalizing/normalizing their cheating behaviour (e.g., by stating that everybody cheats; see e.g., Olafson et al. 2013).

1.1.2 The strain theory

The strain theory further assumes that criminal or dishonest behaviour is influenced by negative affective states that result from perceived strain, strainful experiences, or stressors. Strain thereby includes failing to achieve, or being denied achieving, positive outcomes (like good grades); expecting or actually experiencing negative stimuli; perceiving a disjunction between aspirations or expectations and actual achievements/rewards; and experiencing a disjunction between fair or just outcomes and actual outcomes (e.g., Agnew 1992). The resulting negative emotions can, for instance, be anger or anxiety (e.g., Carmichael and Piquero 2004). Researchers assume that, when faced with strains, stressors, or stressful situations, perceptions of frustration and unfairness arise, which in turn are crucial mechanisms for the link between strain and dishonest behaviour (e.g., Agnew 1992; Agnew and White 1992; Freiburger et al. 2017).

Instead of being contradictory theories, researchers today often propose that both theories together may explain dishonesty, cheating, and deviant behaviour. Negative emotions and strain can influence how rational choices are interpreted, thus influencing individuals' cost–benefit calculations: For instance, negative emotions can reduce individuals' concerns of getting caught, thereby reducing the costs of potential dishonesty; negative emotions can also increase individuals' justifications and rationalizations for their dishonest behaviour (e.g., Carmichael and Piquero 2004; Fida et al. 2015, 2018). In line with this, negative emotions induced by stressors can also increase individuals' perceptions of the importance of potential benefits or the importance of rewards gained by their deceptive behaviour (e.g., Carmichael and Piquero 2004; Fida et al. 2015, 2018).

1.2 Direct and indirect effects of tests as difficult learning tasks on academic cheating

Notably, the above described negative consequences of desirable difficulties in the learning context (e.g., stress, anxiety, perceptions of unfairness) fit the just presented theories explaining dishonesty and academic cheating. Thus, a direct relation between tests as an incantation of desirable difficulties and academic cheating, as well as an indirect relation between tests, thereby inflicted negative consequences (e.g., stress, anxiety, perceptions of unfairness), and academic cheating can be assumed.

For instance, worries about doing well in school, getting good grades, teachers' evaluations, and about the own performance compared to the performance of peers were positively correlated to cheating (e.g., Anderman et al. 1998). Thus, students often cheat to increase their performance and to make a good impression on others (e.g., Franklyn-Stokes and Newstead 1995; Newstead et al. 1996; Wowra 2007). Fear of not being able to succeed, an inability of keeping up with the assignments, lower self-efficiency, and fear of failure were also linked to more academic cheating and often reported as reasons for (past) cheating (e.g., Finn and Frone 2004; McCabe 1992; Schab 1991; Whitley 1998). Notably, as described before, we suppose that tests as difficult learning tasks increase such perceptions of performing poorly and fear of failure because they are difficult, hard to solve, and because they reduce learners' illusion of competence and reduce their overconfidence.

Test anxiety, social anxiety, and general anxiety were also positively correlated to (past) academic cheating (e.g., Rost and Wild 1994; Whitley 1998; Wowra 2007). Stress, parental pressure, pressure for good grades, and pressure in general were also often found to be linked to cheating or were reported as reasons and incentives for such dishonest behaviour (e.g., Brimble and Stevenson-Clarke 2005; Davis et al. 1992; Schab 1991; Whitley 1998). A study by Steininger et al. (1964) further showed that the more negative a (test) situation was perceived, the more anxiety-provoking it was; or, the more a test was perceived as difficult, the more cheating was considered as justified and the more participants reported that they would cheat. Negative emotions due to stressors—and we suppose learning tests to be acute stressors—were further correlated to more moral disengagement in the work context, and in turn to more justifications for deceptive or counterproductive work behaviour (e.g., Fida et al. 2015). Such moral disengagement and justifications thus increased deceptive or counterproductive work behaviour (e.g., Fida et al. 2015), which could also apply to deception in the academic context. Notably, as described before, tests and difficult learning tasks were shown to increase such negative emotions and perceptions of stress, anxiety, and pressure (e.g., Hinze and Rapp 2014; O'Neil et al. 1969; Study 2, Wenzel and Reinhard 2019b).

Furthermore, students' perceptions of the course or assessments as (too) difficult increased academic misconduct (e.g., Brimble and Stevenson-Clarke 2005; Freiburger et al. 2017), and the difficulty of the course was sometimes described as one reason to justify, rationalize, or neutralize cheating behaviour (e.g., Haines et al. 1986). In line with this, higher workload was also linked to more cheating (e.g., McCabe 1992; Whitley 1998). Similarly, participants who thought they had

indulged in more effort in a task felt more entitled and felt that they had earned good outcomes, like higher grades, which in turn led to more moral justifications (e.g., Hoffman and Spitzer 1985). Notably, tests as incantations of desirable difficulties are, even by definition, difficult and they logically increase learners' effort and workload. Thus, we suppose that these findings should also apply to tests as difficult learning tasks.

Another study showed that the more a test situation was perceived as pressuring and as uncomfortable, that is, the more it was perceived as a high-pressure situation, the more unfair the testing tool was perceived (Leiner et al. 2018). Because most learners often believe easier and fluent learning to be extremely effective (e.g., Karpicke et al. 2009; Kornell et al. 2011), we assume that they should perceive difficult tasks like tests and the increased effort they require as unfair and generally as negative, especially when they are forced to use tests in their university courses. In turn, students often reported that they would cheat more and that cheating was more justified when they perceived their teachers, the teaching practices, or the assessments as unfair and their schools as extremely competitive (e.g., Brimble and Stevenson-Clarke 2005; Calabrese and Cochran 1990; Finn and Frone 2004; LaBeff et al. 1990; McCabe 1992; Olafson et al. 2013; Whitley 1998). Fittingly, people who generally thought they were being treated unfairly were more inclined toward dishonesty (e.g., Houser et al. 2012) and perceived inequity was linked to more deceptive behaviour (e.g., Greenberg 1990).

1.3 The present research

In summary, the just described theoretical assumptions and the fitting empirical findings indicate that the application of tests as difficult learning tasks can directly or indirectly (via increasing negative consequences like perceptions of stress, anxiety, or feelings of unfairness) lead to more academic cheating. In more detail, difficult learning tests were argued to result in negative consequences like more stress, more negative perceptions and emotions, or more feelings of unfairness. These negative consequences were in turn often found to be linked to more cheating, more intentions to cheat, and to more justifications for cheating. Hence, the present work was conducted to test these theoretically derived direct and indirect effects of tests as difficult learning tasks on students' academic cheating.

Notably, there are to our knowledge neither studies exploring academic cheating as a result of tests as difficult learning tasks nor studies exploring academic cheating as a result of negative consequences like stress perceptions and negative situation evaluations caused by tests. Most of the existing studies regarding desirable difficulties focused on individual abilities or external factors serving as moderators or requirements for the described beneficial effects (see e.g., Adesope et al. 2017; Dobson and Linderholm 2015; Rowland 2014). However, the main focus was seldom on further negative consequences beyond reduced or restricted learning success and seldom on further triggered behaviour like cheating. We nonetheless argue that it is important to focus on these (new) assumptions because academic cheating can be seen as a widespread and problematic behaviour, even though students

themselves normally perceive cheating during an exam as having rather light consequences (because it is perceived as not directly harming others; e.g., Brimble and Stevenson-Clarke 2005; Marksteiner et al. 2013). For instance, due to cheating on a test, teachers cannot accurately grade students and can therefore not appropriately support their learning processes or help them to increase their skills (e.g., Reinhard et al. 2011). Students who enhance their performance through cheating can also gain an unfair and undeserved advantage compared to others, distort the performance succession in a class, increase competition, trigger peer cheating, and even normalize dishonest behaviour (e.g., Carrell et al. 2008; Fida et al. 2018; Gino et al. 2009; McCabe et al. 2001; Paternoster et al. 2013). Dishonesty in an academic setting is also often linked to further dishonesty in later workplaces (e.g., Nonis and Swift 2001). Due to these negative impacts of academic cheating and due to the lack of previous work, we think that it is relevant to investigate if the application of tests as difficult learning tasks directly or indirectly increases the probability of cheating before advising the usage of such learning tasks in universities. Hence, the present study uniquely contributes to the literature on desirable difficulties and to the literature on cheating behaviour.

To measure dishonest behaviour, researchers often use scenarios because these are assumed to accurately mirror emotions, intentions, and behaviours of individuals in different situations (e.g., Agnew 1992; Carmichael and Piquero 2004; Shu et al. 2011). Thus, we conducted an online study with the learning scenario condition (divided in one reading control scenario condition and two test scenario conditions) as the between-subjects variable. We further assessed individuals' negative evaluations of the learning scenarios as well as their stress perceptions in such imagined situations as two potential mediators. Self-reported likelihoods of hypothetical cheating and justifications for cheating served as our dependent variables.

1.4 Hypotheses

Due to the argumentations presented above, we assume the following hypotheses (see Fig. 1 for a conceptual diagram of the assumed relations): We suppose that both learning scenario conditions with tests lead to more negative evaluations of the learning situations (*Hypothesis 1*) and to higher stress perceptions (*Hypothesis 2*) than the reading control learning scenario condition. Both learning scenario conditions with tests are further assumed to directly lead to higher likelihoods of hypothetical cheating than the reading control learning scenario condition (*Hypothesis 3*). The negative evaluations of the learning situations are also hypothesized to be positively correlated to likelihoods of hypothetical cheating (*Hypothesis 4*). In line with this, stress perceptions are further assumed to be positively correlated to likelihoods of hypothetical cheating (*Hypothesis 5*). Moreover, we assume that both learning scenario conditions with tests directly lead to higher justifications for hypothetical cheating than the reading control learning scenario condition (*Hypothesis 6*). The negative evaluations of the learning situations are also hypothesized to be positively correlated to justifications for cheating (*Hypothesis 7*). In line with this, stress perceptions are assumed to be positively correlated to justifications for

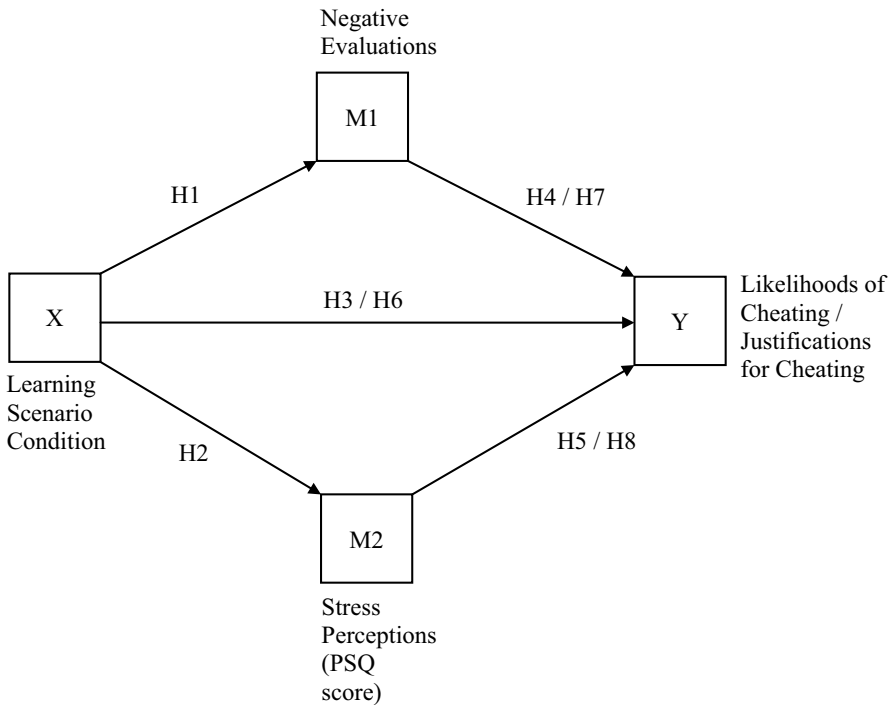


Fig. 1 Conceptual diagram of the assumed hypotheses. *Notes.* The learning scenario condition (X) includes a reading control scenario, a test with private results learning scenario, and a test with public results learning scenario

cheating (*Hypothesis 8*). Thus, apart from direct effects of the learning scenario condition on likelihoods of hypothetical cheating and on justifications for cheating, indirect effects via increases of the negative evaluations of the learning situations and via increases of stress perceptions are also assumed.

2 Methods

2.1 Participants

Power was set to .95 and sample size was calculated to detect a small to medium effect ($f=.20$). Using G*Power (Faul et al. 2009), a power analysis revealed a required sample size of $N=390$ to detect a significant effect (alpha level of .05), given there is a true effect. To test our hypotheses, we recruited an American online sample consisting of 458 participants, 53 of whom were excluded because they answered at least one of three attention-check questions incorrectly. Thus, our final sample consisted of $N=405$ participants from MTurk ($M_{age}=25.72$, $SD_{age}=6.65$, range=18–62, 48.4% female, 97.3% English native speakers,

all college or university students). Each participant was randomly assigned to one of the three learning scenario conditions: either the test with public results learning scenario condition ($n = 129$), the test with private results learning scenario condition ($n = 136$), or the reading control learning scenario condition ($n = 140$). Before starting the experiment, all participants had to provide their approval through reading and then agreeing to an informed consent (stating that they knew that their participation was completely voluntary and that they could withdraw at any time without explanation); participants also confirmed that they were at least 18 years old. The study was conducted in accordance with the Ethical Guidelines of the DGPs as well as the APA, and the project was approved by the Ethics Committee affiliated with the funding source. Participants received .60\$ for their participation.

2.2 Procedure and measures

The present work was conducted together with another study (concerning desirable difficulties, trait variables potentially linked to perceptions of such difficult learning situations, and by desirable difficulties caused stress experiences; Study 1, Wenzel and Reinhard 2019b). Our dependent variables assessing likelihoods of hypothetical cheating and justifications for cheating were assessed at the end of this other study.

At the beginning, participants read brief details about the study and then answered some questions regarding demographics, e.g., age, gender, and native language. Thereafter, different trait variables (e.g., trait test anxiety and trait stress) were assessed solely for the other study (Study 1, Wenzel and Reinhard 2019b; academic self-concept: Dickhäuser et al. 2002; PAF-E: Hoferichter et al. 2016; PSS: Cohen et al. 1983; SSS: Reeder et al. 1973). Although these dispositional variables may be related to the dependent variables of the present work, they will not be included in the analyses because dispositional variables were—unlike the direct and indirect effects of the learning situations—not the focus of the present study.

Participants were then randomly assigned to one of the three learning scenario conditions. As an example, the test with public results learning scenario condition, including the instructions, reads as follows:

This is a potential scenario that could happen in your daily life as a student. We would like to ask you to transport yourself in the situation, and to imagine it as strongly as you can. Imagine that you are a student in college and have lots of exams to write. During one of your majors your professor tries to increase your and your fellow students learning success, and enhance your chance to pass the exam. Therefore, half an hour before the end of every session you write an ungraded test, and answer multiple questions concerning the content of that session. Once the half an hour is up you can go home. Shortly following every session all students receive an e-mail with the matriculation numbers of everyone, and their test results, ranking from best to worst.

In the test with private results learning scenario condition, in which stakes should be perceived as even lower, participants read a slightly different scenario and were instructed to imagine that each student received the test results individually via

e-mail. In contrast, in the reading control learning scenario condition, the imagined process was that the professor would hand the students a summary of all relevant materials to read. See “Appendix A” for all three learning scenarios.

To follow, participants answered questions concerning their perceptions and evaluations of the imagined learning scenario, e.g., regarding difficulty, unfairness, inappropriateness, anger, or injustice. This concluded in an overall *negative evaluations of the learning situations* score using ten items ($\alpha = .89$; e.g., *How (un) just did you find the described and imagined way of learning in the situation?*, one (*extremely unjust*)—seven (*extremely just*)) on a seven-point Likert-like scale from one (lower scores) to seven (higher scores). Some of the items were reverse coded (e.g., participants were asked how fair they thought the learning in the scenario was). See “Appendix A” for a full list of all items, information about which items were recoded, and the complete scale labelling. We also added three—later not analysed—positive control items (e.g., asking for the perceived helpfulness or successfulness of such learning tasks) so that it was not completely clear that we wanted to assess an overall negative evaluations score. We added these positive control items because we wanted to avoid being too obvious, being potentially suggestive, or to unintentionally influence participants’ later responses. Participants were also asked about their situational *stress perceptions* in such an imagined learning scenario using the Perceived Stress Questionnaire (PSQ; Levenstein et al. 1993) that consists of 30 items ($\alpha = .95$; e.g., *You feel tense*) on a four-point Likert-like scale from one (*almost never*) to four (*usually*).

Subsequently, participants were told to again put themselves in the aforementioned scenario and to read the following statement regarding a hypothetical examination:

While preparing for the exam you took little notes and prepared a crib sheet you only wanted to use for your learning. Now imagine that you are in class with your fellow students writing the exam. Thinking about the answer to question number one you suddenly realize that the crib sheet you used to practice is still in your pocket.

Participants were then asked how likely it was for them to use the crib sheet to cheat on the exam (cheating item 1: *likelihoods own spontaneous cheating*) and how justifiable that was (cheating item 2: *justifications own spontaneous cheating*). Then, participants had to rate how likely it was for someone else to use the crib sheet to cheat on the exam (cheating item 3: *likelihoods others’ spontaneous cheating*) and how justifiable that was (cheating item 4: *justifications others’ spontaneous cheating*). Participants were then asked how likely it was for them to intentionally prepare a cheat sheet with the aim to use it during the exam (cheating item 5: *likelihoods own prepared cheating*) and how justifiable that was (cheating item 6: *justifications own prepared cheating*). They also reported how likely it was for someone else to intentionally prepare a cheat sheet with the aim to cheat during the exam (cheating item 7: *likelihoods others’ prepared cheating*) and how justifiable that was (cheating item 8: *justifications others’ prepared cheating*). These eight *cheating items*—four likelihoods items and four justifications items—were answered on a seven-point Likert-like scale from one (*not likely at all/not justifiable at all*) to seven (*extremely*

likely/extremely justifiable). See “Appendix A” for a full list of the items. In line with previous research (see e.g., Greene and Saxe 1992; Messick et al. 1985; Shu et al. 2011), we added items distinguishing between likelihoods and justifications for own hypothetical cheating behaviour and likelihoods and justifications for hypothetical cheating behaviour of other people. Because these cheating items were newly created for our study, we ran factorial analyses to test the underlying number of factors before testing our hypotheses. Regarding the four likelihoods of cheating items, the factor analysis yielded two factors: Factor 1 consisted of the two items regarding the likelihoods of own cheating (average score of the two items: *likelihoods own cheating*, $\alpha = .86$) and factor 2 consisted of the two items regarding the likelihoods of others’ cheating (average score of the two items: *likelihoods others’ cheating*, $\alpha = .84$). The second factor analysis was conducted with the four justification for cheating items and resulted in one factor (average score across the four items: *justifications for cheating*, $\alpha = .95$). A detailed description of the two factor analyses and the respective tables depicting the loadings of the factor analyses are available in “Appendix B”.

In the end, we measured general control variables (e.g., if participants had really imagined the read scenarios, if they understood the text, or how strongly they were able to put themselves in the learning scenarios). For instance, one item reads, *Did you understand the described scenario?*, and it was rated from one (*No, not at all*) to seven (*Yes, completely*). See “Appendix A” for a list of these items. We also included manipulation check questions regarding cheating (e.g., how important grades are for the participants, if they think they can improve their results through cheating, how likely it was to get caught in the imagined scenario, how likeable they rated the imagined lecturer in the scenario, and if they held negative or positive attitudes towards cheating). For instance, one item reads, *How likeable would you rate your professor?*, rated from one (*absolutely unlikeable*) to seven (*extremely likeable*). See “Appendix A” for a list of these items. These manipulation check questions were included to test for differences among participants in the three learning scenario conditions.

2.3 Statistical analyses

To test our hypotheses, we conducted three mediation analyses using PROCESS (Hayes 2018; model 4). Due to the factor analysis that yielded two factors for likelihoods of hypothetical cheating—one for own cheating behaviour and one for others’ cheating behaviour—we conducted two analyses to test the hypotheses that concern likelihoods of hypothetical cheating (e.g., predicting the influence of the learning scenario condition on likelihoods of hypothetical cheating as well as linkages between negative evaluations of learning situations and stress perceptions with likelihoods of cheating; Hypotheses 3, 4, and 5).

The first mediation analysis used likelihoods of own cheating (testing Hypotheses 1, 2, 3a, 4a, and 5a), the second mediation analysis used likelihoods of others’ cheating (testing Hypotheses 3b, 4b, and 5b), and the third mediation analysis used justifications for cheating (testing Hypotheses 6, 7, and 8) as the respective dependent

variable. All three mediation analyses used the learning scenario condition as the independent variable and participants' negative evaluations of the learning situations as well as participants' stress perceptions as two potential mediators. The learning scenario condition was dummy coded (X1: 1 = tests with private results learning scenario condition; X2: 1 = tests with public results learning scenario condition; reference category: reading control scenario condition). The mediator variables were z-standardized. To avoid unnecessary repetitions, only the description of the findings of the first mediation analysis will include the influence of the learning scenario condition on the two mediators, thus, on the negative evaluations of the learning situations and on participants' stress perceptions (testing Hypotheses 1 and 2).

3 Results

Neither participants' gender distribution nor their age differed among the three learning scenario conditions (both $ps \geq .230$). The general control variables and the manipulation check questions regarding cheating also did not differ among the three learning scenario conditions (all $ps \geq .091$). Only participants in the test with public results learning scenario condition rated the lecturer as more dislikeable than participants in the other two learning scenario conditions (both $ps \leq .001$).

The descriptive statistics of the negative evaluations of the learning situations, stress perceptions indicated by PSQ scores, likelihoods of own cheating, likelihoods of others' cheating, and justifications for cheating are presented in Table 1. Notably, likelihoods of others' hypothetical cheating were rated as significantly higher than likelihoods of own hypothetical cheating ($p < .001$).

The correlations among participants' negative evaluations of the learning situations, stress perceptions (PSQ), likelihoods of own hypothetical cheating, likelihoods of others' hypothetical cheating, and justifications for cheating are depicted in Table 2. Notably, the negative evaluations of the learning situations were significantly correlated to participants' likelihoods of own cheating ($r = .19, p < .001$) and to participants' justifications for cheating ($r = .16, p < .001$). The PSQ scores indicating stress perceptions were significantly correlated to likelihoods of others' cheating ($r = .14, p = .006$).

Table 1 Descriptive statistics of the negative evaluations of the learning situations, stress perceptions (PSQ), likelihoods of own cheating, likelihoods of others' cheating, and justifications for cheating

| Variables | M | SD | Range |
|---|------|------|-----------|
| Negative evaluations of the learning situations | 3.58 | 1.15 | 1.00–7.00 |
| Stress perceptions (PSQ) | 2.21 | .63 | 1.00–4.00 |
| Likelihoods own cheating | 2.29 | 1.55 | 1.00–7.00 |
| Likelihoods others' cheating | 4.21 | 1.46 | 1.00–7.00 |
| Justifications for cheating | 2.32 | 1.55 | 1.00–7.00 |

$N = 405$

Table 2 Correlations among the negative evaluations of the learning situations, stress perceptions (PSQ), likelihoods own cheating, likelihoods others' cheating, and justifications for cheating

| | 1. | 2. | 3. | 4. | 5. |
|--|-------|------------------|-------|-------|----|
| 1. Negative evaluations of the learning situations | 1 | | | | |
| 2. Stress perceptions (PSQ) | .50** | 1 | | | |
| 3. Likelihoods own cheating | .19** | .08 ⁺ | 1 | | |
| 4. Likelihoods others' cheating | .04 | .14** | .36** | 1 | |
| 5. Justifications for cheating | .16** | .08 | .86** | .33** | 1 |

⁺ $p < .10$, * $p < .05$, ** $p < .01$, Two-tailed. $N = 405$

3.1 Likelihoods own hypothetical cheating (Hypotheses 1, 2, 3a, 4a, and 5a)

Results of the first mediation analysis (see Fig. 2) showed that the learning scenario condition significantly predicted participants' negative evaluations of the learning situations (path a), $X1: B = .38, SE = .12, t(402) = 3.27, p = .001$; $X2: B = .79, SE = .12, t(402) = 6.66, p < .001$. In turn, the negative evaluations of the learning

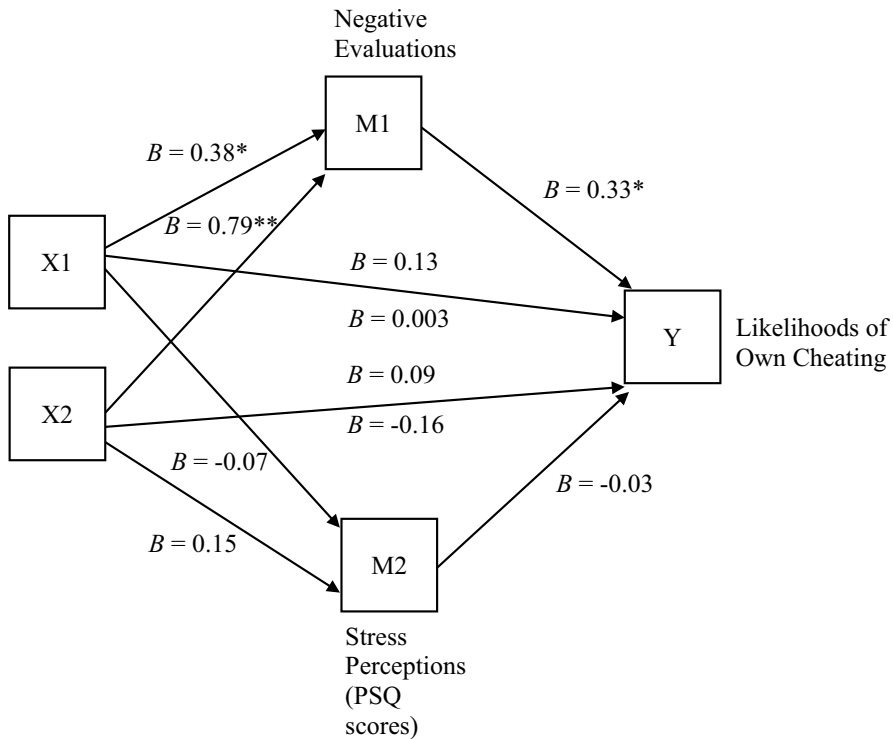


Fig. 2 First mediation analysis predicting likelihoods of own cheating. *Notes.* * $p < .05$, ** $p < .01$. The learning scenario condition was dummy coded ($X1: 1 = \text{tests with private results learning scenario condition}$; $X2: 1 = \text{tests with public results learning scenario condition}$; reference category: reading control scenario condition)

situations predicted participants' likelihoods of own hypothetical cheating (path b), $B = .33$, $SE = .09$, $t(400) = 3.61$, $p = .003$. The learning scenario condition did not significantly predict participants' stress perceptions indicated by their PSQ scores (path a), X1: $B = -.07$, $SE = .12$, $t(402) = -.61$, $p = .545$; X2: $B = .15$, $SE = .12$, $t(402) = 1.24$, $p = .216$. PSQ scores were also not linked to participants' likelihoods of own cheating (path b), $B = -.03$, $SE = .09$, $t(400) = .02$, $p = .706$. There was no significant direct effect (path c') of the learning scenario condition on likelihoods of own cheating, X1: $B = .003$, $SE = .19$, $t(400) = .02$, $p = .986$; X2: $B = -.16$, $SE = .20$, $t(400) = -.82$, $p = .411$. There was also no significant total effect (path c) of the learning scenario condition on likelihoods of own cheating, X1: $B = .13$, $SE = .19$, $t(402) = .71$, $p = .481$; X2: $B = .09$, $SE = .19$, $t(402) = .49$, $p = .626$. However, the results yielded significant indirect effects of the learning scenario condition via the negative evaluations of the learning situations on likelihoods of own hypothetical cheating (path a x path b), X1: $B = .13$, 95% CI [.037, .242]; X2: $B = .26$, 95% CI [.113, .423]. There were no indirect effects of the learning scenario condition via the PSQ scores, X1: $B = .002$, 95% CI [-.023, .030]; X2: $B = -.01$, 95% CI [-.050, .025].

These findings supported Hypothesis 1: Both learning scenarios including tests were, as assumed, evaluated more negatively than the reading control learning scenario. These negative evaluations included, for instance, higher perceptions of unfairness, strenuousness, and injustice, as well as higher feelings of anger. Unexpectedly, the learning scenario condition neither influenced participants' stress perceptions nor likelihoods of participants' own hypothetical cheating. Thus, Hypotheses 2 and 3a were not supported. In line with our assumptions—supporting Hypothesis 4a—negative evaluations of the learning situations were significantly and positively correlated to participants' own hypothetical cheating, indicating that higher negative evaluations were linked to higher likelihoods of own cheating. This indirect effect of the learning scenario condition on likelihoods of own hypothetical cheating (via increased negative evaluations of the learning situations) was significant. Hence, negative evaluations of the learning situations had a mediating effect. Contrary to Hypothesis 5a, stress perceptions were not significantly correlated to participants' likelihoods of own hypothetical cheating.

3.2 Likelihoods of others' hypothetical cheating (Hypotheses 3b, 4b, and 5b)

Results of the second mediation analysis (see Fig. 3) showed that the negative evaluations of the learning situations did not predict likelihoods of others' hypothetical cheating (path b), $B = -.07$, $SE = .09$, $t(400) = -.80$, $p = .425$. The PSQ score was, however, linked to participants' likelihoods of others' cheating (path b), $B = .23$, $SE = .08$, $t(400) = 2.74$, $p = .007$. There was again no significant direct effect (path c') of the learning scenario condition on likelihoods of others' cheating, X1: $B = .01$, $SE = .18$, $t(400) = .07$, $p = .946$; X2: $B = .11$, $SE = .19$, $t(400) = .58$, $p = .560$. There was also no significant total effect (path c) of the learning scenario condition on likelihoods of others' cheating, X1: $B = -.03$, $SE = .18$, $t(402) = -.18$, $p = .859$; X2: $B = .09$, $SE = .18$, $t(402) = .50$, $p = .615$.

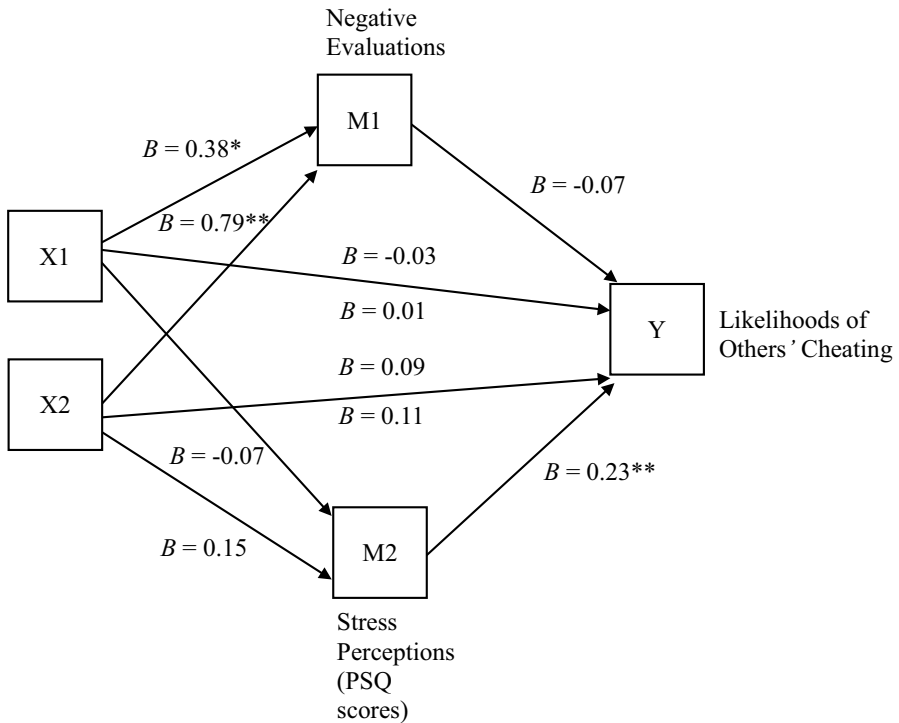


Fig. 3 Second mediation analysis predicting likelihoods of others' cheating. Notes. $*p < .05$, $**p < .01$. The learning scenario condition was dummy coded (X1: 1 = tests with private results learning scenario condition; X2: 1 = tests with public results learning scenario condition; reference category: reading control scenario condition)

Additionally, the findings yielded no significant indirect effects of the learning scenario condition via the negative evaluations of the learning situations on likelihoods of others' cheating (path a x path b), X1: $B = -.03$, 95% CI $[-.105, .044]$; X2: $B = -.06$, 95% CI $[-.197, .085]$. There were also no indirect effects of the learning scenario via the PSQ scores, X1: $B = -.02$, 95% CI $[-.086, .039]$; X2: $B = .04$, 95% CI $[-.021, .105]$.

Unexpectedly, the learning scenario condition did not influence likelihoods of others' hypothetical cheating. Thus, Hypothesis 3b could not be supported, indicating that the learning scenario had no effect on individuals' ratings of the probability of others' cheating in a hypothetical examination. Contrary to Hypothesis 4b, negative evaluations of the learning situations were not significantly linked to others' hypothetical cheating. Participants' stress perceptions were, however, significantly and positively correlated to likelihoods of others' hypothetical cheating, thus, supporting Hypothesis 5b. This indicated that higher stress perceptions were linked to higher ratings regarding likelihoods of others' hypothetical cheating behaviour. There were no indirect effects.

3.3 Justifications for hypothetical cheating (Hypotheses 6, 7, and 8)

Results of the third mediation analysis (see Fig. 4) showed that the negative evaluations of the learning situations significantly predicted justifications for cheating (path b), $B = .24, SE = .09, t(400) = 2.56, p = .011$. The PSQ scores indicating stress perceptions were not linked to justifications for cheating (path b), $B = -.003, SE = .09, t(400) = -.38, p = .970$. There was again no significant direct effect (path c') of the learning scenario condition on justifications for cheating, X1: $B = .02, SE = .19, t(400) = .13, p = .901$; X2: $B = .07, SE = .20, t(400) = .33, p = .743$. There was also no significant total effect (path c) of the learning scenario condition on justifications for cheating, X1: $B = .11, SE = .19, t(402) = .61, p = .542$; X2: $B = .25, SE = .19, t(402) = 1.33, p = .186$. However, the findings yielded significant indirect effects of the learning scenario condition via negative evaluations of the learning situations on justifications for cheating (path a x path b), X1: $B = .09, 95\% CI [.016, .188]$; X2: $B = .19, 95\% CI [.044, .341]$. There were no indirect effects of the learning scenario condition via the

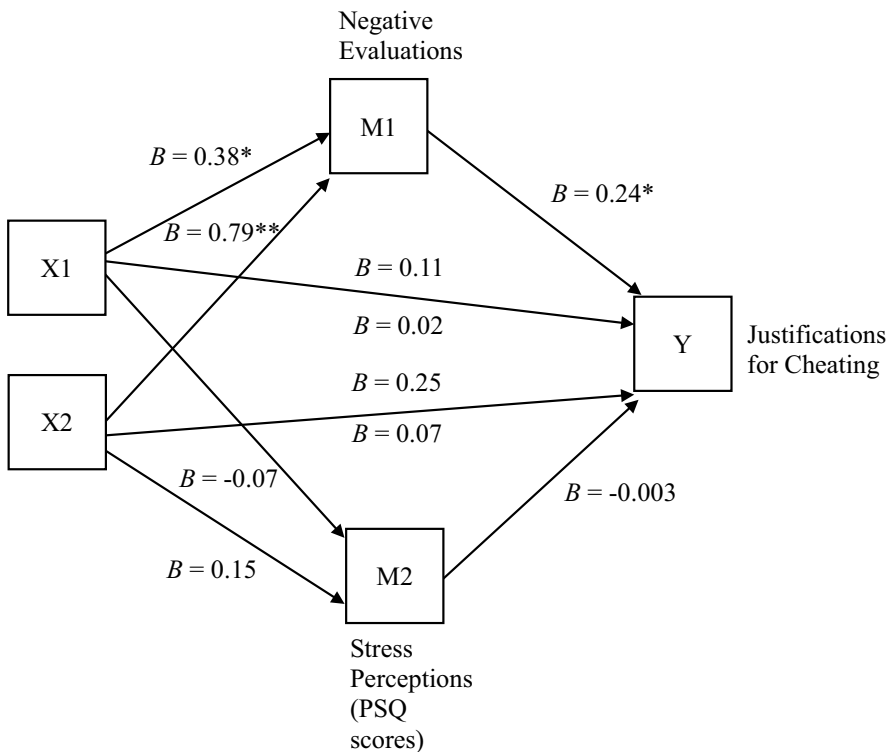


Fig. 4 Third mediation analysis predicting justifications for cheating. Notes. * $p < .05$, ** $p < .01$. The learning scenario condition was dummy coded (X1: 1 = tests with private results learning scenario condition; X2: 1 = tests with public results learning scenario condition; reference category: reading control scenario condition)

PSQ scores, X1: $B < .001$, 95% CI [$-.029, .026$]; X2: $B = -.001$, 95% CI [$-.042, .031$].

Contrary to Hypothesis 6, the learning scenario condition did not influence participants' justifications for hypothetical cheating. Thus, participants' ratings of justifications for hypothetical cheating were not dependent on whether participants had read scenarios including tests or including reading tasks. Negative evaluations of the learning situations were significantly and positively correlated to justifications for hypothetical cheating. This supported Hypothesis 7 and indicated that higher negative evaluations of the learning situations were linked to later higher justifications for cheating in the university context. The indirect effect of the learning scenario condition on justifications for hypothetical cheating (via increased negative evaluations of the learning situations) was also significant. Hence, negative evaluations of the learning situations had a mediating effect. Participants' stress perceptions were not correlated to justifications for hypothetical cheating. Thus, Hypothesis 8 was not supported.

4 Discussion

The aim of the present work was to test linkages among tests as difficult learning tasks, possible negative consequences of such difficult learning tasks like negative evaluations or stress perceptions, and hypothetical academic cheating. We assumed that learning scenarios including tests, as opposed to a control learning scenario including reading, directly and indirectly, lead to higher likelihoods of own and others' hypothetical cheating, as well as to higher justifications for such hypothetical cheating. The indirect effects should arise via increased negative evaluations of the learning situations and via increased stress perceptions due to the difficult test scenarios. Although ample research has focused on the application and effectiveness of tests as incantations of desirable difficulties (e.g., regarding potential moderators or boundary conditions; e.g., Adesope et al. 2017; Rowland 2014) and although previous studies showed that academic cheating has an abundance of negative impacts (e.g., regarding contagion effects of dishonesty through peers, relations between academic and workplace dishonesty, or validity of assessments and grading; e.g., Carrell et al. 2008; Gino et al. 2009; Nonis and Swift 2001; Reinhard et al. 2011), no research has—to our knowledge—previously tested our assumptions and hypotheses. Thus, our work using hypothetical scenarios uniquely contributes to the existing literature regarding cheating in the academic context and to the existing literature regarding tests as difficult learning tasks.

Our findings showed that although the learning scenario condition had neither direct effects on likelihoods of own and others' hypothetical cheating nor on justifications for cheating, it nonetheless indirectly affected likelihoods of own cheating and justifications for cheating through increasing participants' negative evaluations of the learning situations. Both imagined learning scenarios including tests were evaluated as significantly more negative than the learning control scenario including reading. These negative evaluations of the learning situations were in turn positively correlated with likelihoods of own hypothetical cheating and with justifications for

cheating, whereas participants' self-reported stress perceptions were only positively correlated to likelihoods of others' hypothetical cheating. In general, the cheating items had rather low mean scores, whereas likelihoods of others' hypothetical cheating were rated as significantly higher than likelihoods of own hypothetical cheating. This finding is in line with previous work showing that students often report lower frequencies of academic cheating compared to the amount that lecturers observed, and that students also report lower frequencies of their own dishonest behaviour compared to dishonest behaviour of their peers (e.g., Greene and Saxe 1992; Naghdipour and Emeagwali 2013). Additionally, students often perceive their own dishonest behaviour as less condemnable and less serious than the dishonest behaviour of their peers and generally believe that they are fairer than others (e.g., Greene and Saxe 1992; Messick et al. 1985). Thus, it could be that individuals underreport their own cheating behaviour (even in anonymous settings), likely because of the importance and value of norms like honesty and trustworthiness, the urge to maintain a positive self-concept, and the underlying social undesirability of dishonesty (e.g., Geißler et al. 2013; Mazar et al. 2008). Our factor analysis that yielded one factor regarding likelihoods of own hypothetical cheating and a second factor regarding likelihoods of others' hypothetical cheating further supported these findings. Interestingly, our factor analysis revealed only one factor underlying the four justifications for cheating items. Thus, our participants did not distinguish between justifications for own hypothetical cheating behaviour and justifications for others' hypothetical cheating behaviour (contrary to previously found differences between justifications for own and others' dishonesty, see e.g., Shu et al. 2011). In general, the rather low mean score of the justifications for cheating variable further indicates that participants rated hypothetical cheating in the presented scenarios as not justifiable, thus deeming academic cheating as ethically wrong. An explanation for the observed single factor could be that individuals normally try to maintain a positive self-concept and try to feel good or moral even when they cheat (e.g., Mazar et al. 2008): Therefore, they often compare their own behaviour with others' behaviour and, for instance, often believe that others cheated even more—and more severely—than they did (see e.g., Greene and Saxe 1992). This social comparison should, however, only increase individuals' perceptions of themselves as a better or more moral individual compared to others, if the justifications for their own and for others' behaviours are identically low-, because only then should the higher frequencies of others' dishonest behaviour compared to individuals' own less frequent dishonesty increase individuals' self-esteem and their moral self-concept. Moreover, it could also be possible that justifications for own cheating behaviour and justifications for others' cheating behaviour only significantly differ if the justifications ratings were rendered after individuals indulged in actual dishonest behaviour and not just in response to imagined hypothetical cheating.

Notably, our results were obtained even though participants did not really engage in an actual learning activity; they did not really take an exam with actual consequences for their everyday courses, but simply read and imagined scenarios and only self-reported hypothetical behaviour. Nonetheless, even such minimalistic operationalizations yielded significant effects. Thus, this indicates that actual learning in university settings, with real incentives to do well and with actual examinations

including opportunities of actual cheating behaviour, should lead to even stronger effects.

Our results partly fit the in the beginning described theoretical and empirical argumentations regarding negative consequences of desirable difficulties because the scenarios including tests were actually evaluated more negatively than the reading control scenario (e.g., Hinze and Rapp 2014; O’Neil et al. 1969). The observed indirect effects of learning scenarios with tests on own hypothetical cheating behaviour and justifications—via increased negative evaluations of the situations—were also in line with the in the Introduction presented theoretical and empirical argumentations regarding the emergence of cheating and dishonesty in academic contexts (e.g., Brimble and Stevenson-Clarke 2005; Steinger et al. 1964; Whitley 1998; Wowra 2007). Contrary to our assumptions and to literature described in the Introduction, there were neither effects of the learning scenario condition on participants’ stress perceptions nor direct effects of the learning scenario condition on the cheating variables. This could be due to our operationalizations and the application of hypothetical scenarios: It is possible that our scenarios were not strong enough to elicit actual affective responses as well as hypothetical cheating behaviour in only imagined situations (see also our discussion of limitations below).

Although not all our hypotheses were supported, it is still important to highlight that tests as difficult learning tasks can, at least indirectly and in scenarios, influence hypothetical cheating behaviour. Hence, lecturers thinking about applying tests as difficult learning tasks in their university courses should keep in mind that these can result in negative evaluations of the situations and can, indirectly, also result in increased likelihoods of cheating or justifications for cheating. Still, due to the explorative character of our work and because this is to our knowledge the first study testing possible effects of tests as difficult learning tasks on cheating, we suppose that it is too early for stating implications like advising against the usage of tests and desirable difficulties. Nonetheless, our work sheds light on this problematic issue, offering a valuable contributing to the literature regarding desirable difficulties as well as cheating.

4.1 Limitations and future research

There are also limitations of our study that we care to discuss. This includes, for instance, the applied learning scenarios: Although scenarios are often used in studies focusing on cheating behaviour (e.g., Agnew 1992; Carmichael and Piquero 2004; Shu et al. 2011), it is possible that the learning scenarios had no effects because they were too short, not detailed enough, framed as positive, or too low-stake. Still, we intentionally designed them to be preferably short, generalizable (e.g., regarding varying study paths or courses), and minimalistic (e.g., so as not to be suggestive or influencing). We additionally wanted to inquire if effects would arise even when using such simple operationalizations. However, the scenarios may further have been unable to adequately describe and convey the increased effort, difficulty, and cognitive processing triggered by desirable difficulties. The same applies to the short description of the hypothetical exam at the end of the semester, which also

could have been too short, too undifferentiated, or not detailed enough, because the short scenario did not actually describe features of the examination situation (e.g., regarding the importance of the exam, the existence of peers, or the topic of the exam). This could have reduced the transportability and imaginability of the scenario. Although our intention was to not prime or suggest responses due to more detailed descriptions of the hypothetical examination (e.g., by describing opportunities to cheat or the difficulty of the exam), it is possible that more details concerning the examination situation would have made the scenario more comprehensible, more realistic, and more transferable to participants' actual experiences and everyday lives. We may have then been unable to control how participants actually imagined the examination situation, which might have resulted in confounding variables that, in turn, could have influenced participants' answers. Another limitation is that regarding the negative evaluations of the learning situations and the cheating variables, we only observed correlations. Future work should also test causal relations. To do this, future studies could, for instance, directly manipulate the evaluations of the learning situations described in the scenarios, so that the test scenarios as well as the reading control scenario are respectively described as positive, negative, and neutral. This would make it possible to explore whether all conditions including test scenarios lead to higher hypothetical cheating and justifications, or whether only those scenarios that were described as negative would increase hypothetical cheating and justifications for cheating.

In line with the novelty of our research questions and their unique contributions to the cheating and education literature, one of the best aspects of the present work is that it is surely stimulating for further research. For instance, future studies could try to optimize our operationalizations, thus solving the limitations mentioned above, and generally try to replicate our findings using different samples (e.g., students from different countries), different desirable difficulties (e.g., generation or disfluency), or different negative consequences (e.g., negative affect, fear of failure, or feelings of pressure). More explicitly, future studies could also be conducted in laboratory settings or in actual classrooms, applying a real learning phase including actually learned information, so that real—and not only hypothetical—cheating behaviour can be observed. Moreover, future online studies should test our assumptions using different and more detailed scenarios that more adequately describe the learning situation, the learning materials, and the difficulty of the learning tasks. The description of the examination should also be longer and more detailed, for example describing the procedure of the exam, the applied questions, the presence of peers or lecturers, and precautions against cheating more realistic. We also solely presented the usage of cheat sheets in examinations as the incantation of cheating behaviour; however, a far wider range of such behaviour does exist and should also be examined (e.g., inappropriate collaboration during exams or plagiarism). Additionally, until now, we focused completely on situational variables but not on individual variables, whereas previous studies showed that multiple trait variables, individual characteristics, and individual differences (e.g., cognitive abilities, conscientiousness, learning-goal orientations, self-control, or self-efficacy) are simultaneously influential for (difficult) learning (e.g., for perceptions or effectiveness's) and for cheating behaviour and dishonesty (e.g., Bertrams and Englert 2014; de Bruin and Rudnick 2007;

Doménech-Betoret et al. 2017; Finn and Frone 2004; Giluk and Postlethwaite 2015; Ikeda et al. 2015; Koul 2012; Marcela 2015; Paulhus and Dubois 2015; Schunk 1996; Wenzel and Reinhard 2019a; Yu et al. 2017; see also “Appendix B” regarding correlations among our dependent variables and the assessed but not analysed trait variables). Thus, we argue that it is beneficial for future work to include the assessment of individual differences. Lastly, future research should of course also focus on reducing such direct and indirect negative consequences of tests as difficult learning tasks. Lecturers could, for instance, thoroughly explain the benefits of difficult learning to their students, reward them for their efforts, frame the difficulties as even more positive and low-stake, and adapt the difficulty of the tasks so that they are difficult enough to elicit beneficial effects but are not too difficult or overwhelming.

4.2 Conclusion

Summarizing, the present work shows that the application of tests as an incantation of desirable difficulties in the university context—although normally beneficial for long-term learning—can result in negative side effects: Learning scenarios including tests, in contrast to a reading control scenario, indirectly increased likelihoods of own hypothetical cheating and justifications for hypothetical cheating through increasing the negative evaluations of the imagined learning situations. Thus, this work serves as first evidence for the linkage among tests as difficult learning tasks, resulting negative consequences like negative evaluations or stress perceptions, and hypothetical cheating.

Acknowledgements Open Access funding provided by Projekt DEAL. We thank Chawwah Grünberg, Sonja Haverland, Antonia Mariß, Luisa Neufeld, Celina Stolz, Tobias Steppat, and Laura Wagner for their help in conducting/programming the study. We thank Sarah Tyrrell for proofreading the manuscript.

Funding This research was supported by a LOEWE grant from the Hessian Ministry for Science and the Arts entitled “desirable difficulties; intrinsic cognitive motivation and performance expectancies” awarded to Marc-André Reinhard.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study was conducted in accordance with the Ethical Guidelines of the German Association of Psychologists (DGPs) as well as the American Psychological Association (APA), and the project was approved by the Ethics Committee affiliated with the funding source. Additionally, by the time the data were acquired it was neither customary at Kassel University nor at most other German universities to seek ethics approval for simple studies on personality, attitudes, and hypothetical decisions. Thus, ethical approval was not required for this study in accordance with the national and institutional guidelines. However, the study exclusively makes use of anonymous questionnaires and no identifying information was obtained from participants. Moreover, every participant had to read and agree to two questions concerning their consent. They were also explicitly informed that all their data would be treated confidentially and that they could withdraw from the study at any given time, without explanation, by simply closing the internet browser.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Appendix A: Materials

Tests with public results learning scenario condition (including the instructions):

This is a potential scenario that could happen in your daily life as a student. We would like to ask you to transport yourself in the situation, and to imagine it as strongly as you can. Imagine that you are a student in college and have lots of exams to write. During one of your majors your professor tries to increase your and your fellow students learning success, and enhance your chance to pass the exam. Therefore, half an hour before the end of every session you write an ungraded test, and answer multiple questions concerning the content of that session. Once the half an hour is up you can go home. Shortly following every session all students receive an e-mail with the matriculation numbers of everyone, and their test results, ranking from best to worst.

Tests with private results learning scenario condition (including the instructions):

This is a potential scenario that could happen in your daily life as a student. We would like to ask you to transport yourself in the situation, and to imagine it as strongly as you can. Imagine that you are a student in college and have lots of exams to write. During one of your majors your professor tries to increase your and your fellow students learning success, and enhance your chance to pass the exam. Therefore, half an hour before the end of every session you write an ungraded test, and answer multiple questions concerning the content of that session. Once the half an hour is up you can go home. Shortly following every session you receive a private e-mail with your own test results.

Reading control learning scenario condition (including the instructions):

This is a potential scenario that could happen in your daily life as a student. We would like to ask you to transport yourself in the situation, and to imagine it as strongly as you can. Imagine that you are a student in college and have lots of exams to write. During one of your majors your professor tries to increase your and your fellow students learning success and enhance your chance to pass the exam. Therefore, half an hour before the end of every session your professor hands you a summary with all the relevant information of that session. In this time you read the materials. Once the half an hour is up you can go home.

Negative Evaluations of the Learning Situations (10 items, $\alpha = .89$): Instructions and items

“Please answer the following questions according to your imagined mood/perception/thoughts/feelings during the situation displayed in the scenario.”

“Concerning the imagined scenario, ...”

1. How strenuous did you find the described and imagined learning-situation? One (*not strenuous at all*)—seven (*extremely strenuous*)
2. How (un)just did you find the described and imagined way of learning in the situation? One (*extremely unjust*)—seven (*extremely just*), *recoded item
3. How difficult would you rate the learning in the described situation? One (*not difficult at all*)—seven (*extremely difficult*)
4. How fair or unfair would you rate the way of learning in such a situation? One (*extremely unfair*)—seven (*extremely fair*), *recoded item
5. How angry would you be if you were in such a situation and had to learn in such a manner? One (*not in the least bit angry*)—seven (*extremely angry*)
6. How relaxing would you rate such a learning-situation? One (*not relaxing at all*)—seven (*extremely relaxing*), *recoded item
7. How overstrained would you feel if you were in such a learning-situation? One (*not at all*)—seven (*totally*)
8. How annoyed would you feel if you were in such a learning-situation? One (*not at all*)—seven (*extremely*)
9. How uncertain would you feel if you had to learn in a way as described in the situation? One (*not at all*)—seven (*extremely*)
10. How inappropriate would you rate such a learning-situation? One (*not inappropriate at all*)—seven (*extremely inappropriate*)

Not analysed positive control items:

1. How attentive would you be in such a learning-situation? One (*not attentive at all*)—seven (*extremely attentive*)
2. How interesting did you find the described learning-situation? One (*not a bit interesting*)—seven (*extremely interesting*)
3. How helpful and successful would you rate such a learning-situation? One (*not helpful or successful at all*)—seven (*extremely helpful and successful*)

Cheating items:

1. How likely is it that you would use your crib sheet to cheat in the exam? One (*Not at all likely, you would never use the crib sheet*)—seven (*Extremely likely, you would use the crib sheet definitely*)
2. How justifiable is it for you to use your crib sheet to cheat in the exam? One (*Not justifiable at all*)—seven (*Extremely justifiable*)
3. How likely is it that someone else would use their crib sheet to cheat in the exam if they were in the same situation? One (*Not at all likely, other people would never*)

- use the crib sheet*)—seven (*Extremely likely, other people would use the crib sheet definitely*)
4. How justifiable is it for someone else to use their crib sheet to cheat in the exam? One (*Not justifiable at all*)—seven (*Extremely justifiable*)
 5. Furthermore, how likely is it that you would intentionally prepare a cheat sheet with the intention to use it in the exam? One (*Not at all likely, you would never prepare a cheat sheet*)—seven (*Extremely likely, you would definitely prepare a cheat sheet*)
 6. How justifiable is it for you to prepare a cheat sheet to cheat in the exam? One (*Not justifiable at all*)—seven (*Extremely justifiable*)
 7. How likely is it that someone else would intentionally prepare a cheat sheet to use it in the exam if they were in the same situation? One (*Not at all likely, other people would never prepare a cheat sheet*)—seven (*Extremely likely, other people would definitely prepare a cheat sheet*)
 8. How justifiable is it for someone else to prepare a cheat sheet to cheat in the exam? One (*Not justifiable at all*)—seven (*Extremely justifiable*)

Average across items 1 and 5: *likelihoods own cheating*, $\alpha = .86$

Average across items 3 and 7: *likelihoods others' cheating*, $\alpha = .84$

Average across items 2, 4, 6, and 8: *justifications for cheating*, $\alpha = .95$

General manipulation-check questions:

1. Have you really read and imagined the former scenario? *No, not at all/A little bit/Yes*
2. Did you understand the described scenario? One (*No, not at all*)—seven (*Yes, completely*)
3. Were you able to put yourself in the described scenario? One (*No, not at all*)—seven (*Yes, totally*)
4. Have you (in your daily life as a student) experienced situations similar to the ones described in the scenario? One (*No, never*)—seven (*Yes, multiple times*)

Manipulation-check questions concerning cheating:

“Given the former scenario, ...”

1. How likeable would you rate your professor? One (*absolutely unlikeable*)—seven (*extremely likeable*)
2. How important are good grades for you? One (*absolutely unimportant*)—seven (*extremely important*)
3. How much do you think you can improve the results of your exam through cheating? One (*no improvement at all*)—seven (*extremely high improvement*)
4. How likely is it that you (if you decided to cheat) would get caught? One (*absolutely unlikely*)—seven (*extremely likely*)
5. How would you rate the consequences of cheating if you would get caught? One (*absolutely no consequences*)—seven (*extremely severe consequences*)

“In general, ...”

6. How intense would you rate the pressure to perform during your study program? One (*not intense at all*)—seven (*very intense*)
7. Have you cheated in exams before? One (*no, never*)—seven (*yes, every time*)
8. Do you have negative or positive attitudes toward cheating during exams? One (*completely negative attitudes*)—seven (*completely positive attitudes*)

Appendix B: Further analyses

Factor analyses with Varimax rotation

We conducted two factor analyses to test the factor structure of the eight cheating items: Regarding the four likelihoods of cheating items, we conducted a factor analysis with varimax rotation. The eigen values and a scree plot yielded two factors explaining a total of 86.99% of the variance of all four likelihoods items (see Table 3). Factor 1 consisted of the two items regarding the likelihoods of own cheating (*likelihoods own cheating*, $\alpha = .86$; 58.98% of explained variance, eigen value = 2.36). Factor 2 consisted of the two items regarding the likelihoods of others' cheating (*likelihoods others' cheating*, $\alpha = .84$; 28.01% of explained variance, eigen value = 1.12). The second factor analysis with varimax rotation was conducted with the four justifications for cheating items. The eigen values and a scree plot yielded one factor explaining a total of 86.23% of the variance of all four items (*justifications for cheating*, $\alpha = .95$; eigen value = 3.45; see Table 4).

Table 3 Factor analysis likelihoods of hypothetical cheating variables

| Variables | Factor loading | |
|--|----------------|------|
| | 1 | 2 |
| Likelihoods own spontaneous cheating | .920 | .179 |
| Likelihoods others' spontaneous cheating | .171 | .912 |
| Likelihoods own prepared cheating | .924 | .161 |
| Likelihoods others' prepared cheating | .164 | .913 |

Table 4 Factor analysis justifications for cheating variables

| Variables | Factor loading |
|---|----------------|
| | 1 |
| Justifications own spontaneous cheating | .922 |
| Justifications others' spontaneous cheating | .938 |
| Justifications own prepared cheating | .936 |
| Justifications others' prepared cheating | .919 |

Table 5 Correlations among the negative evaluations of the learning situations, stress perceptions (PSQ), likelihoods own cheating, likelihoods others' cheating, justifications for cheating, trait test anxiety (PAF-E), trait stress scales (PSS, SSS), and academic self-concept

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. |
|--|--------|------------------|--------|------------------|--------|--------|--------|--------|----|
| 1. Negative evaluation of the learning situation | 1 | | | | | | | | |
| 2. Stress perceptions (PSQ) | .50** | 1 | | | | | | | |
| 3. Likelihoods own cheating | .19** | .08 ⁺ | 1 | | | | | | |
| 4. Likelihoods others' cheating | .04 | .14** | .36** | 1 | | | | | |
| 5. Justifications for cheating | .16** | .08 | .86** | .33** | 1 | | | | |
| 6. Trait test anxiety (PAF-E) | .28** | .56** | .14** | .14** | .11* | 1 | | | |
| 7. Trait stress (PSS) | .33** | .52** | .11* | .10 ⁺ | .07 | .65** | 1 | | |
| 8. Trait stress (SSS) | .31** | .54** | .08 | .07 | .05 | .60** | .67** | 1 | |
| 9. Academic self-concept | -.18** | -.27** | -.22** | -.05 | -.25** | -.43** | -.40** | -.29** | 1 |

⁺ $p < .10$, * $p < .05$, ** $p < .01$, Two-tailed. $N = 405$

We further conducted a MANCOVA predicting participants trait test anxiety (PAF-E), stress traits (PSS, SSS), and their academic self-concept using the learning scenario condition as the between-subjects variable to test for differences between participants in the three learning scenario conditions. The one-way MANCOVA yielded no significant multivariate main effect for the learning scenario condition, $F(8,800) = .50$, $p = .859$, $\eta_p^2 = .005$. Given the not significant overall test, the univariate main effects were not examined. We also conducted correlations among the negative evaluations of the learning situations, stress perceptions (PSQ), likelihoods own cheating, likelihoods others' cheating, and justifications for cheating with the trait variables (trait test anxiety, trait stress, and academic self-concept) that were solely assessed for the other study that was conducted together with this study (see Table 5).

References

- Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, *87*, 1115–1153. <https://doi.org/10.3982/ECTA14673>.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*, 659–701. <https://doi.org/10.3102/0034654316689306>.
- Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, *30*, 47–88. <https://doi.org/10.1111/j.1745-9125.1992.tb01093.x>.

- Agnew, R., & White, H. R. (1992). An empirical test of general strain theory. *Criminology*, 30(4), 475–500. <https://doi.org/10.1111/j.1745-9125.1992.tb01113.x>.
- Akers, R. L. (1990). Rational choice, deterrence, and social learning theory in criminology: The path not taken. *Journal of Criminal Law & Criminology*, 81, 653–676. <https://doi.org/10.2307/1143850>.
- Alter, A. L., Oppenheimer, D. M., & Epley, N. (2013). Disfluency prompts analytic thinking—But not always greater accuracy: Response to Thompson et al. (2013). *Cognition*, 128, 252–255. <https://doi.org/10.1016/j.cognition.2013.01.006>.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569. <https://doi.org/10.1037/0096-3445.136.4.569>.
- Anderman, E. M., Griesinger, T., & Westerfield, G. (1998). Motivation and cheating during early adolescence. *Journal of Educational Psychology*, 90, 84–93. <https://doi.org/10.1037/0022-0663.90.1.84>.
- Becker, G. S. (1968). Crime and punishment: An economic approach. In G. S. Becker & W. M. Landes (Eds.), *The economic dimensions of crime* (pp. 13–68). London: Palgrave Macmillan.
- Bertrams, A., & Englert, C. (2014). Test anxiety, self-control, and knowledge retrieval in secondary school students. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46, 165–170. <https://doi.org/10.1026/0049-8637/a000111>.
- Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition*, 35, 201–210. <https://doi.org/10.3758/BF03193441>.
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge: MIT Press.
- Bjork, R. A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application* (pp. 435–459). Cambridge: The MIT Press.
- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Boca Raton: CRC Press.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, 2, 59–68.
- Brimble, M., & Stevenson-Clarke, P. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *The Australian Educational Researcher*, 32, 19–44. <https://doi.org/10.1007/BF0321682>.
- Calabrese, R. L., & Cochran, J. T. (1990). The relationship of alienation to cheating among a sample of American adolescents. *Journal of Research & Development in Education*, 23(2), 65–72.
- Cameron, J. S., Miller, D., & Monin, B. (2008). Deservingness and unethical behavior in loss and gain frames. *Unpublished manuscript*. http://www.communicationcache.com/uploads/1/0/8/8/10887248/deservingness_and_unethical_behavior_in_loss_gain_frames.pdf.
- Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment*, 83, 29–45. https://doi.org/10.1207/s15327752jpa8301_04.
- Carmichael, S., & Piquero, A. R. (2004). Sanctions, perceived anger, and criminal offending. *Journal of Quantitative Criminology*, 20, 371–393. <https://doi.org/10.1007/s10940-004-5869-y>.
- Carrell, S. E., Malmstrom, F. V., & West, J. E. (2008). Peer effects in academic cheating. *Journal of Human Resources*, 43, 173–207. <https://doi.org/10.3368/jhr.43.1.173>.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*. <https://doi.org/10.2307/2136404>.
- Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, 19, 16–20. https://doi.org/10.1207/s15328023top1901_3.
- De Bruin, G. P., & Rudnick, H. (2007). Examining the cheats: The role of conscientiousness and excitement seeking in academic dishonesty. *South African Journal of Psychology*, 37, 153–164. <https://doi.org/10.1177/008124630703700111>.

- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*, 979–995. <https://doi.org/10.1037/0022-3514.70.5.979>.
- Dickhäuser, O., Schöne, C., Spinath, B., & Stiensmeier-Pelster, J. (2002). Die Skalen zum akademischen Selbstkonzept: Konstruktion und Überprüfung eines neuen Instrumentes [The academic self-concept scales: Construction and evaluation of a new instrument]. *Zeitschrift für differentielle und diagnostische Psychologie: ZDDP*, *23*, 393–405. <https://doi.org/10.1024/0170-1789.23.4.393>.
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, *118*, 111–115. <https://doi.org/10.1016/j.cognition.2010.09.012>.
- Dobson, J. L., & Linderholm, T. (2015). The effect of selected “desirable difficulties” on the ability to recall anatomy information. *Anatomical Sciences Education*, *8*, 395–403. <https://doi.org/10.1002/ase.1489>.
- Doménech-Betoret, F., Abellán-Roselló, L., & Gómez-Artiga, A. (2017). Self-efficacy, satisfaction, and academic achievement: The mediator role of Students’ expectancy-value beliefs. *Frontiers in Psychology*, *8*, 1193. <https://doi.org/10.3389/fpsyg.2017.01193>.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students’ learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*, 4–58. <https://doi.org/10.1177/1529100612453266>.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- Feldman, R. S., Forrest, J. A., & Happ, B. R. (2002). Self-presentation and verbal deception: Do self-presenters lie more? *Basic and Applied Social Psychology*, *24*, 163–170. https://doi.org/10.1207/S15324834BASP2402_8.
- Fida, R., Paciello, M., Tramontano, C., Fontaine, R. G., Barbaranelli, C., & Farnese, M. L. (2015). An integrative approach to understanding counterproductive work behavior: The roles of stressors, negative emotions, and moral disengagement. *Journal of Business Ethics*, *130*, 131–144. <https://doi.org/10.1007/s10551-014-2209-5>.
- Fida, R., Tramontano, C., Paciello, M., Ghezzi, V., & Barbaranelli, C. (2018). Understanding the interplay among regulatory self-efficacy, moral disengagement, and academic cheating behaviour during vocational education: A three-wave study. *Journal of Business Ethics*, *153*, 725–740. <https://doi.org/10.1007/s10551-016-3373-6>.
- Finn, K. V., & Frone, M. R. (2004). Academic performance and cheating: Moderating role of school identification and self-efficacy. *The Journal of Educational Research*, *97*, 115–121. <https://doi.org/10.3200/JOER.97.3.115-121>.
- Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise—An experimental study on cheating. *Journal of the European Economic Association*, *11*, 525–547. <https://doi.org/10.1111/jeea.12014>.
- Franklyn-Stokes, A., & Newstead, S. E. (1995). Undergraduate cheating: Who does what and why? *Studies in Higher Education*, *20*, 159–172. <https://doi.org/10.1080/03075079512331381673>.
- Freiburger, T. L., Romain, D. M., Randol, B. M., & Marcum, C. D. (2017). Cheating behaviors among undergraduate college students: Results from a factorial survey. *Journal of Criminal Justice Education*, *28*, 222–247. <https://doi.org/10.1080/10511253.2016.1203010>.
- Geißler, H., Schöpe, S., Klewes, J., Rauh, C., & von Alemann, U. (2013). *Wertestudie 2013: Wie groß ist die Kluft zwischen dem Volk und seinen Vertretern [Value Study 2013: How big is the gap between the people and their representatives]*. Köln: YouGov.
- Giluk, T. L., & Postlethwaite, B. E. (2015). Big five personality and academic dishonesty: A meta-analytic review. *Personality and Individual Differences*, *72*, 59–67. <https://doi.org/10.1016/j.paid.2014.08.027>.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, *20*, 393–398. <https://doi.org/10.1111/j.1467-9280.2009.02306.x>.
- Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*, 561–568. <https://doi.org/10.1037/0021-9010.75.5.561>.
- Greene, A. S., & Saxe, L. (1992). *Everybody (else) does it: Academic cheating* [Paper presentation]. Annual meeting of the Eastern Psychological Association, April 3–5, Boston, MA.

- Haines, V. J., Diekhoff, G. M., LaBeff, E. E., & Clark, R. E. (1986). College cheating: Immaturity, lack of commitment, and the neutralizing attitude. *Research in Higher Education*, 25, 342–354. <https://doi.org/10.1007/BF00992130>.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis* (2nd ed.). New York: The Guilford Press.
- Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, 28, 597–606. <https://doi.org/10.1002/acp.3032>.
- Hoferichter, F., Raufelder, D., Ringeisen, T., Rohrmann, S., & Bukowski, W. M. (2016). Assessing the multi-faceted nature of test anxiety among secondary school students: An English version of the German test anxiety questionnaire: PAF-E. *The Journal of Psychology*, 150, 450–468. <https://doi.org/10.1080/00223980.2015.1087374>.
- Hoffman, E., & Spitzer, M. L. (1985). Entitlements, rights, and fairness: An experimental examination of subjects' concepts of distributive justice. *The Journal of Legal Studies*, 14(2), 259–297.
- Houser, D., Vetter, S., & Winter, J. (2012). Fairness and cheating. *European Economic Review*, 56, 1645–1655. <https://doi.org/10.1016/j.euroecorev.2012.08.001>.
- Ikeda, K., Castel, A. D., & Murayama, K. (2015). Mastery-approach goals eliminate retrieval-induced forgetting: The role of achievement goals in memory inhibition. *Personality and Social Psychology Bulletin*, 41, 687–695. <https://doi.org/10.1177/0146167215575730>.
- Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, 27, 209–228. <https://doi.org/10.1006/ceps.2001.1088>.
- Kaiser, I., Mayer, J., & Malai, D. (2018). Self-generation in the context of inquiry-based learning. *Frontiers in Psychology*, 9, 2440. <https://doi.org/10.3389/fpsyg.2018.02440>.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579–588. <https://doi.org/10.1037/0022-0663.93.3.579>.
- Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, 17, 471–479. <https://doi.org/10.1080/09658210802647009>.
- Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 704–719. <https://doi.org/10.1037/0278-7393.33.4.704>.
- Kausar, R. (2010). Perceived stress, academic workloads and use of coping strategies by university students. *Journal of Behavioural Sciences*, 20(1), 31–45.
- Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, 22(6), 787–794. <https://doi.org/10.1177/0956797611407929>.
- Koul, R. (2012). Multiple motivational goals, values, and willingness to cheat. *International Journal of Educational Research*, 56, 1–9. <https://doi.org/10.1016/j.ijer.2012.10.002>.
- LaBeff, E. E., Clark, R. E., Haines, V. J., & Diekhoff, G. M. (1990). Situational ethics and college student cheating. *Sociological Inquiry*, 60(2), 190–198. <https://doi.org/10.1111/j.1475-682X.1990.tb00138.x>.
- Lehmann, J., Goussios, C., & Seufert, T. (2016). Working memory capacity and disfluency effect: An aptitude-treatment-interaction study. *Metacognition and Learning*, 11, 89–105. <https://doi.org/10.1007/s11409-015-9149-z>.
- Leiner, J. E. M., Scherndl, T., & Ortner, T. M. (2018). How do men and women perceive a high-stakes test situation? *Frontiers in Psychology*, 9, 2216. <https://doi.org/10.3389/fpsyg.2018.02216>.
- Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., et al. (1993). Development of the perceived stress questionnaire: A new tool for psychosomatic research. *Journal of Psychosomatic Research*, 37, 19–32. [https://doi.org/10.1016/0022-3999\(93\)90120-5](https://doi.org/10.1016/0022-3999(93)90120-5).
- Lipowsky, F., Richter, T., Borromeo-Ferri, R., Ebersbach, M., & Hänze, M. (2015). Wünschenswerte Erschwernisse beim Lernen [Desirable difficulties during learning]. *Schulpädagogik heute*, 6(11), 1–10.
- Marcela, V. (2015). Learning strategy, personality traits and academic achievement of university students. *Procedia-Social and Behavioral Sciences*, 174, 3473–3478. <https://doi.org/10.1016/j.sbspro.2015.01.1021>.

- Marksteiner, T., Reinhard, M. A., Lettau, F., & Dickhäuser, O. (2013). Bullying, cheating, deceiving: Teachers' perception of deceitful situations at school. *International Journal of Educational Psychology*, 2, 193–220. <https://doi.org/10.4471/ijep.2013.24>.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633–644. <https://doi.org/10.1509/jmkr.45.6.633>.
- McCabe, D. L. (1992). The influence of situational ethics on cheating among college students. *Sociological Inquiry*, 62, 365–374. <https://doi.org/10.1111/j.1475-682X.1992.tb00287.x>.
- McCabe, D. L. (2001). *Academic integrity—A research update*. College Station: Center for Academic Integrity.
- McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: A decade of research. *Ethics and Behavior*, 11, 219–232. https://doi.org/10.1207/S15327019EB1103_2.
- McDaniel, M. A., Hines, R. J., & Guynn, M. J. (2002). When text difficulty benefits less-skilled readers. *Journal of Memory and Language*, 46, 544–561. <https://doi.org/10.1006/jmla.2001.2819>.
- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14, 200–206. <https://doi.org/10.3758/BF03194052>.
- McDaniel, M. A., Waddill, P. J., & Einstein, G. O. (1988). A contextual account of the generation effect: A three-factor theory. *Journal of Memory and Language*, 27, 521–536. [https://doi.org/10.1016/0749-596X\(88\)90023-X](https://doi.org/10.1016/0749-596X(88)90023-X).
- McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14, 1–43. https://doi.org/10.1207/s1532690xci1401_1.
- Messick, D. M., Bloom, S., Boldizar, J. P., & Samuelson, C. D. (1985). Why we are fairer than others. *Journal of Experimental Social Psychology*, 21, 480–500. [https://doi.org/10.1016/0022-1031\(85\)90031-9](https://doi.org/10.1016/0022-1031(85)90031-9).
- Mihalca, L., Mengelkamp, C., & Schnotz, W. (2017). Accuracy of metacognitive judgments as a moderator of learner control effectiveness in problem-solving tasks. *Metacognition and Learning*, 12, 357–379. <https://doi.org/10.1007/s11409-017-9173-2>.
- Naghdipour, B., & Emeagwali, O. L. (2013). Students' justifications for academic dishonesty: Call for action. *Procedia-Social and Behavioral Sciences*, 83, 261–265. <https://doi.org/10.1016/j.sbspro.2013.06.051>.
- Newstead, S. E., Franklyn-Stokes, A., & Armstead, P. (1996). Individual differences in student cheating. *Journal of Educational Psychology*, 88, 229–241. <https://doi.org/10.1037/0022-0663.88.2.229>.
- Nonis, S., & Swift, C. O. (2001). An examination of the relationship between academic dishonesty and workplace dishonesty: A multicampus investigation. *Journal of Education for Business*, 77, 69–77. <https://doi.org/10.1080/08832320109599052>.
- Olafson, L., Schraw, G., Nadelson, L., Nadelson, S., & Kehrwald, N. (2013). Exploring the judgment-action gap: College students and academic dishonesty. *Ethics and Behavior*, 23, 148–162. <https://doi.org/10.1080/10508422.2012.714247>.
- O'Neil, J. H., Spielberger, C. D., & Hansen, D. N. (1969). Effects of state anxiety and task difficulty on computer-assisted learning. *Journal of Educational Psychology*, 60, 343–350. <https://doi.org/10.1037/h0028323>.
- Oppenheimer, D. M., & Alter, A. L. (2014). The search for moderators in disfluency research. *Applied Cognitive Psychology*, 28, 502–504. <https://doi.org/10.1002/acp.3023>.
- Paternoster, R., McGloin, J. M., Nguyen, H., & Thomas, K. J. (2013). The causal impact of exposure to deviant peers: An experimental investigation. *Journal of Research in Crime and Delinquency*, 50, 476–503. <https://doi.org/10.1177/0022427812444274>.
- Paulhus, D. L., & Dubois, P. J. (2015). The link between cognitive ability and scholastic cheating: A meta-analysis. *Review of General Psychology*, 19, 183–190. <https://doi.org/10.1037/gpr0000040>.
- Pieger, E., Mengelkamp, C., & Bannert, M. (2016). Metacognitive judgments and disfluency—Does disfluency lead to more accurate judgments, better control, and better performance? *Learning and Instruction*, 44, 31–40. <https://doi.org/10.1016/j.learninstruc.2016.01.012>.
- Reeder, L. G., Schrama, P. G. M., & Dirken, J. M. (1973). Stress and cardiovascular health: An international cooperative study—I. *Social Science and Medicine*, 1967(7), 573–584. [https://doi.org/10.1016/0037-7856\(73\)90026-7](https://doi.org/10.1016/0037-7856(73)90026-7).
- Reinhard, M. A., Dickhäuser, O., Marksteiner, T., & Sporer, S. L. (2011). The case of Pinocchio: Teachers' ability to detect deception. *Social Psychology of Education*, 14, 299–318. <https://doi.org/10.1007/s11218-010-9148-5>.

- Richland, L. E., Bjork, R. A., Finley, J. R., & Linn, M. C. (2005). Linking cognitive science to education: Generation and interleaving effects. In *Proceedings of the twenty-seventh annual conference of the Cognitive Science Society* (pp. 1850–1855). Mahwah: Erlbaum.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction*, *49*, 142–156. <https://doi.org/10.1016/j.learninstruc.2017.01.008>.
- Rost, D. H., & Wild, K. P. (1994). Cheating and achievement-avoidance at school: Components and assessment. *British Journal of Educational Psychology*, *64*, 119–132. <https://doi.org/10.1111/j.2044-8279.1994.tb01089.x>.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*. <https://doi.org/10.1037/a0037559>.
- Sarason, I. G., & Sarason, B. R. (1990). Test anxiety. *Handbook of social and evaluation anxiety* (pp. 475–495). Boston: Springer. https://doi.org/10.1007/978-1-4899-2504-6_16.
- Schab, F. (1991). Schooling without learning: Thirty years of cheating in high school. *Adolescence*, *26*(104), 839–847.
- Schunk, D. H. (1996). *Self-efficacy for learning and performance*. [Paper presentation]. American Educational Research Association, April 8–12, New York, NY.
- Schunk, D. H., & Gaa, J. P. (1981). Goal-setting influence on learning and self-evaluation. *The Journal of Classroom Interaction*, *16*(2), 38–44.
- Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, *115*, 181–190. <https://doi.org/10.1016/j.obhdp.2011.02.001>.
- Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications: Doing wrong and feeling moral. *Current Directions in Psychological Science*, *24*, 125–130. <https://doi.org/10.1177/0963721414553264>.
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, *37*, 330–349. <https://doi.org/10.1177/0146167211398138>.
- Simha, A., & Cullen, J. B. (2012). A comprehensive literature review on cheating. *International Journal of Cyber Ethics in Education (IJCEE)*, *2*, 24–44. <https://doi.org/10.4018/ijcee.2012100102>.
- Steininger, M., Johnson, R. E., & Kirts, D. K. (1964). Cheating on college examinations as a function of situationally aroused anxiety and hostility. *Journal of Educational Psychology*, *55*, 317–324. <https://doi.org/10.1037/h0042396>.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*, 185–233. https://doi.org/10.1207/s1532690xci1203_1.
- Tyler, S. W., Hertel, P. T., McCallum, M. C., & Ellis, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning and Memory*, *5*, 607. <https://doi.org/10.1037/0278-7393.5.6.607>.
- Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, *27*, 247–264. <https://doi.org/10.1007/s10648-015-9310-x>.
- Wenzel, K., & Reinhard, M. A. (2019a). Relatively unintelligent individuals do not benefit from intentionally hindered learning: The role of desirable difficulties. *Intelligence*, *77*, 101405. <https://doi.org/10.1016/j.intell.2019.101405>.
- Wenzel, K., & Reinhard, M. A. (2019b). Does the end justify the means? Learning tests lead to more negative evaluations and to more stress experiences. *Manuscript submitted for publication*.
- Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, *39*, 235–274. <https://doi.org/10.1023/A:1018724900565>.
- Wowra, S. A. (2007). Moral identities, social anxiety, and academic dishonesty among American college students. *Ethics and Behavior*, *17*, 303–321. <https://doi.org/10.1080/10508420701519312>.
- Yu, H., Glanzer, P. L., Sriram, R., Johnson, B. R., & Moore, B. (2017). What contributes to college students' cheating? A study of individual factors. *Ethics and Behavior*, *27*, 401–422. <https://doi.org/10.1080/10508422.2016.1169535>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kristin Wenzel is a Ph.D. student of Social Psychology at the Department of Psychology, University of Kassel, Germany. Her research interests include cognitive prerequisites and moderators of the positive effect of desirable difficulties as well as consequences of desirable difficulties. Additionally, her research focuses on deception and deception detection as well as on the belief in a just world.

Marc-André Reinhard is a Full Professor of Social Psychology at the Department of Psychology, University of Kassel, Germany. His research interests include cognitive and motivational moderators of the positive effect of desirable difficulties in learning. Moreover, his research focuses on the formation and consequences of performance expectancies.