

**University Student Perceptions: A Comparison Between Formative and Summative  
Assessment Based on Self-Efficacy and Academic Misconduct**

Julia Wulf

S3771466

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Group number 12

Supervisor: dr. A. Sarampalis

Second evaluator: B. Gützkow, Msc

In collaboration with: Lena Duiverman, Stelle Fritzsche, Oliver Konradt, Matti Konstantin

Kuhnert, Tilman Müller-Scholtz.

Date: January 17, 2022

*A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned.*

### **Abstract**

This study examines how students' perceptions of formative assessments and summative assessments at the University of Groningen differ in terms of their self-efficacy and frequency of cheating. It was hypothesized that formative assessments, such as assignments and exams, would be associated with higher self-efficacy and fewer cheating behavior than in summative exams. Further, it was hypothesized that higher self-efficacy would lead to less cheating. To test these hypotheses, an online study was conducted in which psychology undergraduate students reported their self-efficacy and frequency of cheating in courses using formative and summative assessments. A randomized response method was used for the cheating items. The results indicated that students had significantly higher self-efficacy in courses using formative assessments than summative ones. Also, results suggested that students cheated less in formative assignments and exams than in summative exams. However, results were inconclusive whether high self-efficacy would also lead to fewer cheating. The study provides valuable insight into the students' perceptions on the formative and summative assessment methods based on their self-efficacy and cheating behavior.

*Keywords:* formative assessment, summative assessment, student perception, self-efficacy, academic misconduct

## **University Student Perceptions: A Comparison Between Formative and Summative Assessment Based on Self-Efficacy and Academic Misconduct**

As the Council of Europe stated, higher education serves the purposes of preparing students “for sustainable employment, personal development, preparing students for active citizenship, and creating a broad advanced knowledge base and stimulating research and innovation.” (Education, 2014, p. 7). However, for universities worldwide it is seldom possible to fully live up to these aims considering the challenges they face these days. For example, one of those challenges is a decline in funding per student at Dutch universities and while the number of students is growing, Dutch universities struggle with increasing workload, shortage of staff and hence rising student-teacher ratios in study programs (*De dalende rijksbijdrage en de druk op onderwijs en onderzoek*, 2021). To still manage the increasing numbers of students these days and to ensure a comparability of students’ performances in a standardized way, the choice of assessment at universities is often of summative nature such as multiple-choice exams (Qu & Zhang, 2013). To explain, summative assessment is the graded evaluation of students’ performance at the end of a study module reflecting the extent to which students achieved a required learning objective (Yorke, 2003). Another type of assessment is formative assessment that has been loosely defined as “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, pp. 7-8).

In direct comparisons with summative assessment, formative assessment seems to be experienced more positively by students. For example, Struyven et al. (2005) reviewed 36 empirical studies on students’ experiences with evaluations and assessments at universities and found that students perceived alternative assessments, such as formative assessment, as fairer because “it appeared to measure qualities, skills and competences which would be

valuable in contexts other than the immediate context of assessment” (Struyven et al., 2005, p. 333). Additionally, Weurlander et al. (2012) conducted an exploratory study that asked medical students about their experiences of different forms of assessments. The students described that formative assessment motivated them to study, made them aware of their own learning and stimulated their learning process e.g., learning when having to explain a learned construct to others. Also, Lerdpornkulrat et al. (2019) found self-efficacy scores of first-year students in an formalized Information Literacy Skills course in Thailand to be significantly higher than in the control group. Hence, formative assessment appears to produce more positive learning experiences for students.

One university that takes the education of their students’ seriously is the Dutch University of Groningen, describing their mission as a higher education institution as follows: “Students learn best when they process new information in such a way that it makes sense to them, when their learning is contextualized” (*Making Connections* 2021, p. 7). Even though the University recognizes the importance of the applicability of new knowledge for the learning experience, their Faculty of Behavioral and Social Sciences predominantly uses summative assessment as an assessment method although it has been known to not transfer information well to other contexts (Resnick & Resnick, 1992). This seems contradictory since students reported their learning experience to have been more positive in formative assessment than in traditional, summative assessment types (Lerdpornkulrat et al., 2019; Struyven et al., 2005; Weurlander et al., 2012). Therefore, this study aims to investigate whether the undergraduate psychology students at the University of Groningen also experience formative assessments to improve their study experience. The ultimate goal would be to give an according recommendation to the University of Groningen in regard to assessment forms. Self-efficacy was used as an indicator of the learning experience as students’ high self-efficacy beliefs have been linked to enhanced commitment, decision-

making, motivation, and goal setting (Linnenbrink & Pintrich, 2003; Murdock & Anderman, 2006). Further, cheating behavior was explored as it could serve as an indicator whether students enjoy their learning experience. That is, Simkin and McLeod (2010) asked 158 students for their reasons to cheat and found that the most important reason was to pass the course to move on. This suggests that students aimed at completing the courses regardless of the learning opportunity indicating a poor learning experience. Thus, altogether, the aim of the study was to investigate the students' perceptions at the University of Groningen on the different assessment forms, formative and summative, based on their self-efficacy and frequency of cheating behavior.

### **Advantages and Disadvantages of Assessment Types**

#### ***Summative Assessment***

Summative assessment has been the choice of assessment since the need for norm-referenced and fair evaluation emerged beginning at the 19<sup>th</sup> century (Butler, 2018). Nowadays, multiple choice exams dominate higher education as they can reduce teacher bias in grading and can be administered to a large number of students at the same time while still being considerably inexpensive to administer (Qu & Zhang, 2013). Universally, it is also used to determine professional or academic aptitude because grade point averages of summative exams are often used as a benchmark in the world of work (Starovoytova & Namango, 2016).

Additionally, multiple choice exams have shown consistently that they enhance recall by providing retrieval cues and functions as additional study time because students are again confronted with the study material as discussed in a review by Marsh et al. (2007). However, Struyven et al. (2005) found students in multiple choice exams to adopt a superficial study approach and a deep learning approach in assignments. Although the students also indicated to favor multiple choice exams as the assessment is perceived to be more straightforward and accessible (Struyven et al., 2005) this was not supported by Harlen and Crick (2002). They

conducted a systematic review of 19 studies on the impact of summative assessment on the motivation of students aged four to 19 and found that students had generally negative experiences: They disliked high-stake exams, experienced high exam anxiety and especially low achievers, who did not or barely achieve the required objectives at the end of the study module, were at risk for lower self-esteem in summative assessment compared to other forms. Considering the literature, summative assessment has obvious structural advantages, but more research is necessary to establish which assessment form students experience to facilitate their study experience as will be done at the University of Groningen.

### ***Formative Assessment***

The alternative to summative assessment is formative assessment and its purpose is to create a reciprocal relationship between student and teacher: On the one hand enhancing the learning of students by providing them with feedback and on the other hand providing the teacher with valuable information regarding the students' learning in order to provide them with a constructive learning structure (Andrade, 2010). Effective feedback by the teacher is central and was found to be an important predictor for effective learning (Dixson & Worrell, 2016; McCarthy, 2017; Rakoczy et al., 2019). Hattie and Timperley (2007) reviewed 12 meta-analyses on the effect of feedback in the classroom and found a large average effect size of ( $d = 0.79$ ) indicating high practical relevance. In addition, feedback was also to found have an indirect effect on self-efficacy as reported by van Dinther et al. (2014). They conducted a study with 138 first-year undergraduate students at a Dutch university investigating the relationships between student perceptions of formative assessment and self-efficacy and their influence on the outcomes of their learning. The results indicated that students experience higher self-efficacy if formative assessment was perceived as being practice-oriented and the feedback was focused on improving their learning. However, the results of studies such as cited above need to be viewed critically as formative assessment suffers from a definitional

problem, meaning that there is no coherent conceptualization of the construct (e.g., Bennett, 2011). That means that formative assessment is so vaguely defined that it may be operationalized differently depending on the researcher, possibly restricting its effectiveness and comparability. As a result, the effectiveness of formative assessment is questionable when higher education uses their formatively inspired assessments to contribute to a final grade. Hence, it is necessary to further investigate whether students at the University of Groningen indeed experience formative assessment to enhance their study progress.

### **Self-efficacy**

Based on the above-mentioned literature, self-efficacy seems to be enhanced in formative assessment (van Dinther et al., 2014) and is hence important to explore in this context. Self-efficacy is part of the social cognitive theory (Bandura, 1997) and can be defined as the “beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p. 3). It has shown to have positive effects on the performance of students as a meta-analysis exploring the psychological antecedents of university students’ performance, found self-efficacy to be the strongest predictor (Richardson et al., 2012). More positive outcomes were found by Freire et al. (2019) who found that self-efficacy positively predicted effective coping with stress in 1402 undergraduate students in Spain as well as Dogan (2015) who identified self-efficacy to contribute significantly to academic performance in middle and high school students in Turkey. Regarding the effect of formative assessment on self-efficacy, Panadero et al. (2017) conducted a comprehensive meta-analysis of 19 studies from various countries around the world and found that encouraging students to self-assessment practices, a form of formative assessment, predicted higher self-efficacy with a large effect size ( $d = 0.73$ ). Based on the existing literature, self-efficacy seems to have an overall positive impact on students’ study



experience, and it needs to be investigated whether the effect that formative assessment can lead to higher self-efficacy can be also replicated at the University of Groningen.

### **Cheating**

As described above, cheating may be an indicator for how students experience their study progress. Academic cheating may be defined as the deliberate effort to misrepresent knowledge or wrongfully taking credit for resources by engaging in deceitful actions (Eckstein & Unesco, 2003). This includes cheating on written work such as submitting someone else's work, fabricating or falsifying data and copying work without citing the author, and cheating on tests such as copying answers, using non-permitted notes and paying someone to do the work for oneself, also known as contract cheating (Eckstein & Unesco, 2003). Among the main reasons to cheat are the pressure to attain good grades, a lack of interest in the course or the material, a lack of preparation time (Cizek, 1999; McCabe, 1999; Starovoytova & Namango, 2016), and peers finding cheating acceptable (McCabe, 2016; Papp & Wertz, 2009). An additional factor linked to cheating is a lack of clarity of what is defined as cheating behavior (Keener et al., 2019) resulting from misunderstandings between student and university (Bisping et al., 2008). Moreover, demographics and personality characteristics (Kisamore et al., 2007), motivational factors such as outcome expectations, achievement goals and self-efficacy (Anderman & Koenka, 2017) could be linked to cheating. Cheating behavior at universities has been found to continue after graduating e.g., in the work environment (Harding et al., 2004; Swift & Nonis, 1998) and can cause serious repercussions for the cheater as it can end careers and for the institutions that are compromised in their integrity and efficiency (Ogilvie & Stewart, 2010). Hence, it is important to reduce cheating also at the University of Groningen to ensure the integrity of the students' performances.

The literature investigating the whether the type of assessment is related to cheating is scarce and contradictory. For example, Gardner et al. (1988) identified a mean cheating rate

of 50.8% on assignments in psychology first-year students, whereas Karlins et al. (1988) only found 3% of undergraduate students to be cheating in an assignment. Research on plagiarism, which is a form of cheating, only exerted in assignments, has also revealed contradictory results. On the one hand, López Puga (2014) questioned 65 undergraduate psychology students regarding their beliefs of plagiarism and found about 60% to admit to engaging in plagiarism, while on the other hand, Stephens et al. (2007) only found about 33.3% of undergraduate students from the US admitting to it. In conclusion, because cheating is a vital indicator of the study experience of students and the literature on the effect of assessment type on cheating is scarce and ambiguous, it needs to be investigated further.

### **Cheating and Self-efficacy**

Both constructs of self-efficacy and cheating have also been linked to each other: Tas and Tekkaya (2010) investigated this relationship in Turkish elementary students via self-reports and found that students with high levels of self-efficacy reported less cheating than those with low levels of self-efficacy. Additionally, Ogilvie and Stewart (2010) conducted a situational analysis of academic misconduct and self-efficacy using undergraduate students from Australia and found low self-efficacy to be a predictor of higher intentions to engage in plagiarism. Since the literature suggests that self-efficacy beliefs in students are linked to cheating behavior, the current study attempts to replicate the findings with students at the University of Groningen.

### **This Study**

In the current study, we administered an exploratory survey to psychology undergraduate students from the University of Groningen in order to assess self-efficacy beliefs and cheating behaviors in courses that used formative assessment methods i.e., assignments with the purpose of increasing learning and courses that used summative assessment courses i.e., a final exam. The educational value of formative and summative

assessment will be compared by determining the differences between reported cheating behavior and self-efficacy, between the assessment types. Based on the literature, it is hypothesized that:

H1: In comparison to summative assessment, self-efficacy beliefs of students are significantly higher in formative assessment.

H2: In comparison to summative assessment, students engaging in formative assessment show lower cheating behavior.

H3: High self-efficacy beliefs will be negatively associated with cheating behavior in students.

## **Method**

### **Participants**

Participants were all either current or previous students from the psychology bachelor programme at the University of Groningen. Out of a total of 259 responses, 45 were removed as they did not finish the questionnaire and three responses were removed indicating that they did not answer the questions truthfully. In the end, 211 responses were used in the main analysis. The sample consisted of 144 first-year students and 67 higher-year students, among which eight started in 2020, 27 started in 2019, 26 started in 2018, five started in 2017 and one participant started in 2015. Altogether the sample had a mean age of 20.83 ( $SD = 2.14$ ). Due to the unequal distribution of first-year students and higher-year students, Table 1 provides a detailed overview of the distribution within the different starting years. For 156 participants, the psychology bachelor at the University of Groningen was their first university programme, while 55 were already enrolled in other programmes. In total, 210 participants reported their gender: 162 females (76.8%), 47 males (22.3%), one participant indicating “other” (0.5%), while one participant indicated “would rather not say” (0.5%). Of all participants, 85 were Dutch (40.3%), 73 were German (34.6%) and 53 indicated to have other

nationalities (25.1%). Those consisted of 28 different nationalities. The distribution of gender and nationalities reflected the general student population at the psychology bachelor programme at the University of Groningen. Their participation was voluntary, and consent was given. Also, ethical approval was given by the Ethics Committee of the Faculty of Behavioral and Social Sciences at the University of Groningen.

**Table 1**

*Mean Age, Gender Ratio and Enrollment Ratio of the First-year and Higher-Year Students*

Student starting year	First-year students	Higher-year students
$M_{age}$	19.77	22.19
$M_{SD}$	1.91	1.88
Gender ratio (% female)	81.94	65.67
First enrollment (% yes)	74.31	73.13

*Note:* First-year students include starting year 2021, Higher-year students include starting years 2020, 2019, 2018, 2017, 2015.

## **Materials**

The study consisted of an online questionnaire using the online software Qualtrics. As no generalization to other universities or years was intended, reliability or validity indices of the items were of no concern. All items were formulated as statements and participants indicated their agreement or disagreement along a 5-point scale, namely “strongly disagree”, “disagree”, “neither agree nor disagree”, “agree”, and “strongly agree”. The different constructs were ordered from general to more specific e.g., general procrastination at the beginning of the survey and specific procrastination behavior in the sections of formative and summative courses. Within the course type conditions, the constructs did not follow a specific order. Only cheating behavior was put at the end due to its different assessment style.

## **Demographics**

Participants' demographic information was assessed using five items including participants' gender, age, nationality, the year in which they started their psychology bachelor programme and whether programme was their first university programme, respectively. For that, participants were either able to choose from precast answer options or were asked to indicate the correct answer if the option did not apply e.g., "What is your nationality: Dutch, German or Other (please indicate)." The purpose of those questions was to determine whether the sample was representative of students studying psychology at the University of Groningen.

### ***General Study Habits***

The general study habits of participants since starting the psychology bachelor at the University of Groningen were assessed using six items. Study habits referred to routinely performed actions encompassing time management, learning techniques, and frequency of studying that are performed by students in order to memorize the material. The study habits could either be effective or ineffective. The items were designed specifically for this questionnaire and were not structured in a specific manner. The first three items aimed at exploring whether and how the participants felt that their study habits had improved since enrolling in the programme e.g., through other students or the help of the University. For example: "I feel like my study habits have improved since enrolling for this programme". An indication of "strongly agree" marked that an improvement in their study habits took place since enrolling at the University of Groningen, while "strongly disagree" indicated no improvement. The following three items aimed at investigating whether the participants wished to improve their study habits, employed the same study habits as in high school and focused on the sole memorization of the material instead of trying to understand it. A response of "strongly disagree" indicated an improvement in study habits whereby "strongly

agree” indicated no improvement. These items were reversed coded for the calculation of internal reliability.

### ***Procrastination***

All items assessing procrastination behavior were adapted from the Procrastination Assessment Scale Students (PASS; Solomon & Rothblum, 1984). The original PASS encompasses 44 items with good internal reliability ( $\alpha = .80$ ) and good concurrent validity. The term procrastination refers to the behavior of delaying a task while being aware of its possible negative consequences. No specific structure of the items was applied nor was reverse coding used.

Three items assessed general procrastination behavior that was independent from a specific course type and procrastination behavior depending on formative and summative course type. The items aimed at investigating to what extent the participants procrastinate on study related tasks and whether they find it problematic. For example: “I want to decrease my tendency to procrastinate on university activities.” An indication of “strongly agree” implied that the procrastination behavior is high and/or problematic.

The following nine items assessing possible reasons for the participants’ procrastination behavior independent of course type included insecurities e.g., being worried about getting a bad grade or distrusting the self, overcoming structural issues such as waiting for additional information on the exam and having a hard time knowing what to study, or other personal issues such as disliking studying for exams. The answer options for these items were phrased differently, namely “not at all reflects why I procrastinated”, “reflects a little”, “somewhat reflects”, “reflects a lot” and “definitely reflects why I procrastinated.” An indication of the first response option assumed that the stated reason was not of relevance while the last assumed the statement to be accurate.

### ***Self-efficacy***

The five items used for assessing self-efficacy were used for both formative and summative courses, and were inspired and adapted from the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991). By answering these items, participants indicated to which extent they believed in their own capacities to be successful in their studies, such as understanding the material. The order of the items did not follow a specific structure. Out of the five items, two were phrased in a reversed manner asking whether the participants expected to have problems with understanding the most difficult material in the readings or presented by instructors. Participants indicated with “strongly agree” that their self-efficacy was low, whereas “strongly disagree” stated the opposite. The residual items assessed their confidence in passing, understanding basic concepts, and mastering skills.

### ***Engagement***

Engagement was measured using six items in both formative and summative courses, which were to some extent inspired and adapted from items created by Krause and Coates (2008) and Schaufeli et al. (2002). In this study, engagement was defined as the students’ drive towards learning marked by commitment and persistence. The order of the items was arbitrary. The two reverse coded items assessed the approach of studying for an exam in the stated course type e.g., “I do the bare minimum of work to pass the course.” If participants strongly agreed with these statements, it indicated low engagement, while the opposite answer option indicated high engagement. The remaining items focused on the whether the participants work with classmates on the material, attend lectures or watch recordings, contact with lecturers regarding material, or are enthusiastic in the stated type of course. For these items, low engagement was indicated by the answer option “strongly disagree” and high engagement by “strongly agree.”

### ***Perceived Retention***

The four items assessing retention of the course material were identical for formative and summative courses. In this study, the term retention evaluated the students' ability to retain and recall study related information effectively. Retention is usually assessed using direct tests, however, to gain some insight into how students perceive their retention in specific course types, four items based on learning modes were developed by Simpson et al. (1994) that determine the depth of processing and the degree of integration of knowledge. The different items asked for the participants' ability to remember and explain general topics or learning goals, central concepts, and theories in the specific course types. Also, independently from retention, they were asked about their grade, namely with "generally receive a high grade in a course like this". An indication of "strongly disagree" indicated a high ability to retain study information in this type of course, while "strongly disagree" suggested low retention.

### ***Learning Approaches***

The learning approaches were assessed with 10 items, all used in formative and summative courses. For the former course type, one additional course type specific item was added. All items were to some extent based on the Approaches and Study Skills Inventory for Students (Entwistle & Tait, 2013). The construct of learning approaches described the strategies used by students to execute learning tasks. No specific structure was used, and four items were reversed coded. Those assessed the students' doubts with the study material of the specific course type, e.g., wondering about the relevance of the study work, and the students' study approach in this type of course, whereby an indication of "strongly agree" demonstrated a superficial learning approach, while the opposite suggested a deep learning approach. The remaining items described the extent to which students interact and invest effort to understand the study material e.g., studying to master the material or studying regularly. The one additional item for formative courses was "The regular assignments help me structure." If



participants strongly agreed with the statement, it was an indication of a deep learning approach, while strongly disagreeing indicated a superficial learning approach.

### ***Satisfaction***

A total of nine items were used to assess students' satisfaction in formative and summative courses and were adapted from the "Students Evaluation of Educational Quality" Questionnaire (Marsh, 2011). The construct of satisfaction was defined as the students' contentment with the formative and summative courses. No items were phrased in reverse and they followed no specific structure. For example, the items assessed the clarity of the course aims, helpfulness of feedback, quality of communication of approaching deadlines and clarity of assessment requirements. A response of "strongly agree" indicated high satisfaction with the specific course type while "strongly disagree" indicated low satisfaction.

### ***Well-Being***

The first six items assessing the well-being of students within a specific type of course were used in both formative and summative courses. For formative courses, an additional two items were added. All were self-constructed and did not follow a specific order. For this study, well-being was defined as the impact of study related tasks on the students' mental health. The additional two items for formative courses were also reverse coded as they asked whether there are times where there is nothing to study for and whether the mandatory assignments help understand the course content. If participants responded with "strongly agree", high well-being was assumed while the opposite was the case with a response of "strongly disagree." The remaining items evaluated the experienced stress regarding workload, studying, keeping up with courses and examinations e.g., "I feel anxious before an exam." For these items, an indication of "strongly agree" indicated low well-being, while "strongly disagree" indicated high well-being in this type of course.

### ***Cheating***

The assessment of cheating behavior used a randomized response method, namely a coin toss method. It was first introduced by Warner (1965) and has been shown to provide a framework in which participants indicated significantly more embarrassing behavior, e.g., cheating, than in direct surveys (Pitsch et al., 2017). Cheating behavior was investigated in exams and assignments in formative courses and exams in summative ones. The introduction text for the method was self-constructed and stated:

For the next questions, a coin toss method (please find detailed information down below) is used to ensure that your answers to this question are fully anonymous. Please use this webpage (*Just flip a coin*, 2010) to flip a coin before answering each question and answer the question according to the outcome of the coin toss. If the coin comes up heads, then answer the question truthfully; if it comes up tails, just say 'yes' no matter what you would have answered. Follow this link (Pitsch et al., 2017) for more information on the coin toss method.

The items themselves included lists of specific cheating behaviors that were inspired by the Academic Dishonesty Scale (Bashir & Bala, 2018). Cheating in exams included behaviors such as the usage of prohibited things such as notes, copying answers, paying someone else to take the exam, and collaborating with others during the exam. Cheating in assignments included behaviors such as receiving help, using resourcing without citing the author, copying answers from someone who had already done the assignment and letting someone else complete the assignment. The participants were asked to indicate whether they had ever done any of the stated behaviors in this specific assessment. The “no” answers represented true responses for those who got “head” and the “yes” responses represented partly true responses from the coin flip outcome of “head” and partly from “tail” as those were instructed to indicate “yes” regardless of the truth.

## **Design and Procedure**

The present study used a correlation design while the variable of interest was the general cheating behavior, measured by the coin toss method, allowing for a comparison of cheating behavior in both summative and formative assessment. These scores, in turn, were compared to the self-efficacy scores, giving an indication about the mediating relationship.

The first-year students were recruited via SONA, a website on which first-year students of the psychology bachelor programme are obliged to participate in scientific research in exchange for credit. The higher-year students were recruited via convenience and snowball sampling through advertisement messages in WhatsApp year-groups and private messages in personal WhatsApp chats. Thereby, no inducement such as money or credit was given.

Via a link, all participants were directed to the Qualtrics website where the questionnaire was presented. They were first provided with an information form that included details regarding the background and purpose of the study, their rights of participating, possible benefits and risk of participating, storage and handling of data and contact information of the researchers. Before giving their consent, participants were encouraged to ask questions before during and after participating. After that, they were asked to provide their details regarding their demographics, their experiences with their general study habits, and procrastination behaviors.

The following questions were divided into two parts, one assessing experiences with courses using formative assessment and the other using summative assessment. In both parts similar questions were used and, if necessary, adapted to the assessment type. The participants were randomly presented with either one of those parts first, followed by the second part, such that all participants completed both parts. Before each part, an introduction scenario, depending on the assessment type, was presented. Within the parts, participants were asked

about their specific study habits and procrastination behaviors in such courses, as well as their well-being, perceived retention, and satisfaction.

In the end, participants were asked in both parts about their general cheating behavior during either exams and/or assignments with the coin toss method. A link for further details on this method was included, allowing participants to receive more information. Additionally, participants were instructed to follow a link to a website generating coin tosses before answering the questions and to subsequently answer the question according to the outcome of the coin toss that was either “yes” or “no”. If the coin came up “heads”, they were asked to answer truthfully. In contrast, if the coin came up tails, they were asked to answer “yes” regardless of the truth. The participants were able to skip the question by clicking on the arrow for the next page.

After the completion of the questionnaire, participants were asked to indicate whether they responded truthfully and followed the instructions of the coin toss method. Additionally, they had the opportunity to leave a comment and to receive more information about the study results by emailing the researcher.

## **Results**

The students’ perceptions on their self-efficacy and cheating behavior in assessments and exams in the different university assessment methods, formative and summative, at the University of Groningen, were analyzed with the statistical software platform SPSS.

### **General Student Experiences: Study Habits and Procrastination**

Since the six single items assessing study habits had low internal reliability ( $\alpha = .403$ ), the items were not averaged and merged, but viewed individually. For all items, a value of 5 indicated complete agreement with the given statement, while a value of 1 indicated complete disagreement. Based on the means (and standard deviations) presented in Table 2, it is visible that participants generally felt that their study habits had improved since enrolling at the

University of Groningen. In line with that, participants generally disagreed that their study habits stayed the same as in high school. Also, participants indicated that students and the University itself facilitated an improvement of study habits since their enrollment.

Additionally, participants generally wished to improve their study habits. Moreover, they indicated that they do not just memorize the material presented instead of trying to understand it since starting higher education.

**Table 2**

*Means and Standard Deviations of General Study Habits*

Item	<i>M</i>	<i>SD</i>
Study_Habits_Improved_Since_Enrollment	3.78	0.81
Study_Habits_Other_Students_Helped_Improve	3.15	1.08
Study_Habits_University_Information_Helped_Improve	3.36	0.97
Study_Habits_Wish_To_Improve	3.76	1.01
Study_Habits_Same_As_Highschool	2.73	1.16
Study_Habits_Just_Memorize_Material	2.03	0.86

*Note.* All Scores were measured with 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), 5 (strongly agree).

Due to the high internal reliability of the three items assessing the participants' general procrastination ( $\alpha = .898$ ), their scores were averaged and merged into one variable. While a high value (5) indicated that the participants generally procrastinate and indicate that this is a problem, a low value (1) indicated no problem with procrastination. The averaged scores of the three items had a mean of 3.67 ( $SD = 1.09$ ), suggesting that most participants seem to procrastinate to a problematic extent. In Table 3 the means and standard deviations of the reasons are displayed. The three top reasons for procrastinating are lacking energy to begin studying, overwhelmedness by the task and the tendency of having too many other things to

do. The three least common reasons are difficulty to approach the professor for necessary information, waiting for information from the professor and the tendency of not knowing what to study.

**Table 3**

*Means and Standard Deviations of Reasons for General Procrastination*

Item	<i>M</i>	<i>SD</i>
Procrastination_Reason_Hard_Time_Knowing_What_To_Study	2.35	1.13
Procrastination_Reason_Too_Many_Other_Things_To_Do	3.02	1.14
Procrastination_Reason_Things_To_Ask_Professor	1.70	1.02
Procrastination_Reason_Reason_Worried_To_Get_Bad_Grade	2.73	1.40
Procrastination_Reason_Dislike_Studying	2.48	1.06
Procrastination_Reason_Overwhelmed_By_Task	3.30	1.30
Procrastination_Reason_Distrust_Myself_To_Do_Good_Job	2.69	1.42
Procrastination_Reason_Lacking_Energy	3.45	1.26
Procrastination_Reason_Waiting_For_Information_Professor	1.93	1.09

*Note.* All Scores were measured with 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), 5 (strongly agree).

### **Comparison of Student Experiences in Formative and Summative Courses**

As the scales assessing various student experiences had generally high internal consistency, the respective scores were averaged and merged into new variables.

#### ***Engagement***

The first scale assessed the participants' engagement in formative and summative courses. A high value (5) indicated great engagement in the respective courses, whereas low engagement corresponded to a low value (1). The mean in courses with formative assessment was 3.16 and 2.96 in courses with summative assessment. The difference of the means

suggested that participants were slightly more engaged in courses with formative assessment than in ones using summative assessment.

### ***Retention***

Participants were also asked about their ability to retain study material in formative and summative courses. While a level of 1 indicated a low ability to retain study material, a level of 5 indicated a high ability of retention. A difference in the means was notable as the mean in formative courses was 3.82 and the mean in summative courses was 3.54. This indicated that participants seemed to retain more of the material in courses with formative assessment than in summative assessment.

### ***Learning Approaches***

For the scale learning approaches, a value of 1 indicated a surface learning approach, while a value of 5 indicated deep learning. Whereas the averaged scores in the formative courses yielded a mean of 3.45, the scores in summative courses corresponded to a mean of 3.24, suggesting that participants adopted more deep learning approaches in the formative courses compared to summative courses.

### ***Procrastination***

Participants were asked about their tendency to procrastinate in courses using formative or summative assessment. A high level of procrastination corresponded to high values (5), while a low level of procrastination was indicated by low values (1). The average procrastination mean in formative courses was 3.20 and 3.51 in summative courses. This suggests that the amount of procrastinating was higher in summative than in formative courses.

### ***Satisfaction***

The participants' satisfaction with formative and summative courses. Low satisfaction was indicated by a value of 1 and high satisfaction corresponded to a value of 5. As the scores

in formative courses produced a higher mean ( $M = 3.79$ ) than in summative courses ( $M = 3.29$ ), participants seemed to be more satisfied in courses that adopt formative assessment as compared to summative assessment.

### ***Well-Being***

The participants' well-being in formative and summative courses. A value of 1 corresponded to lower stress while a value of 5 corresponded to higher stress levels in the respective courses. The formative well-being averages were 3.36 while the summative averages were 3.48, indicating that the amount of stress in courses with formative assessment was lower compared to the summative assessment.

Based on the last few paragraphs one can see that the direction of the scores was consistent across all six scales assessing student experiences, indicating that indeed there is a difference between the students' experiences in courses using formative and summative courses. Table 4 provides an overview of the means, standard deviations, and Cronbach's alphas of the six variables assessing student experiences in courses using formative and summative assessment.

**Table 4**

*Means, Standard Deviations and Cronbach's Alphas of the Averaged Scores of Six Variables Assessing Student Experiences in Formative and Summative Courses*

Variables	Formative				Summative			
	<i>n</i>	$\alpha$	<i>M</i>	<i>SD</i>	<i>n</i>	$\alpha$	<i>M</i>	<i>SD</i>
Engagement	6	0.48 <sup>a</sup>	3.16	0.57	6	0.56	2.96	0.60
Retention	4	0.59	3.82	0.50	4	0.69	3.54	0.65
Learning approaches	11	0.70	3.45	0.49	10	0.74	3.24	0.55
Procrastination	3	0.89	3.20	1.18	3	0.91	3.51	1.19
Satisfaction	9	0.77	3.79	0.51	9	0.79	3.29	0.57



Well-being	7	0.69	3.36	0.59	6	0.65	3.48	0.64
------------	---	------	------	------	---	------	------	------

*Note.* All Scores were measured with 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), 5 (strongly agree).

<sup>a</sup> Even though the internal consistency was low ( $\alpha < .05$ ), it was averaged and merged regardless because of its closeness to the cutoff.

### ***Self-efficacy***

As the self-efficacy items had high internal consistency in both the formative courses ( $\alpha = 0.68$ ) and summative courses ( $\alpha = 0.80$ ), one new variable was created for both course types by merging and averaging the scores of the five self-efficacy items. High values (5) of the scores indicated higher senses of self-efficacy while lower values (1) indicated lower senses of self-efficacy. Overall, the average score of the merged self-efficacy scale in formative courses was 3.64 ( $SD = 0.57$ ), whereas the mean in summative courses was 3.31 ( $SD = 0.71$ ). This suggests that participants on average had higher self-efficacy in courses with formative assessments than in courses using summative assessments.

The differences between the two course types become clearer when comparing some of the individual subscale items. For example, around 72% of the participants in formative courses felt that they expect to master the skills being taught in the courses, indicated by responding agree (3) or highly agree (4) to that question, while in summative courses only 50% felt that way. Similarly, in formative courses, around 48% of the participants expected experiencing no problems with understanding the material of the lecture, while only around 37% expected the same in summative courses. This was indicated by responding strongly disagree (1) or disagree (2). Another variation became visible when the participants were asked about their confidence of understanding basic concepts: While 93% of the participants in formative courses agreed (4) or strongly agreed (5) to that statement, only 85% of participants did so in summative courses. An overview of the descriptive statistics of the five

items assessing self-efficacy in formative and summative courses is displayed in Table 5. Again, the direction of the scores is consistent across the items and suggests a difference between the experience of formative and summative courses regarding the sense of self-efficacy. A more detailed overview of the distribution of responses in the self-efficacy items can be found in Figure 1 for summative courses and in Figure 2 for formative courses.

**Table 5**

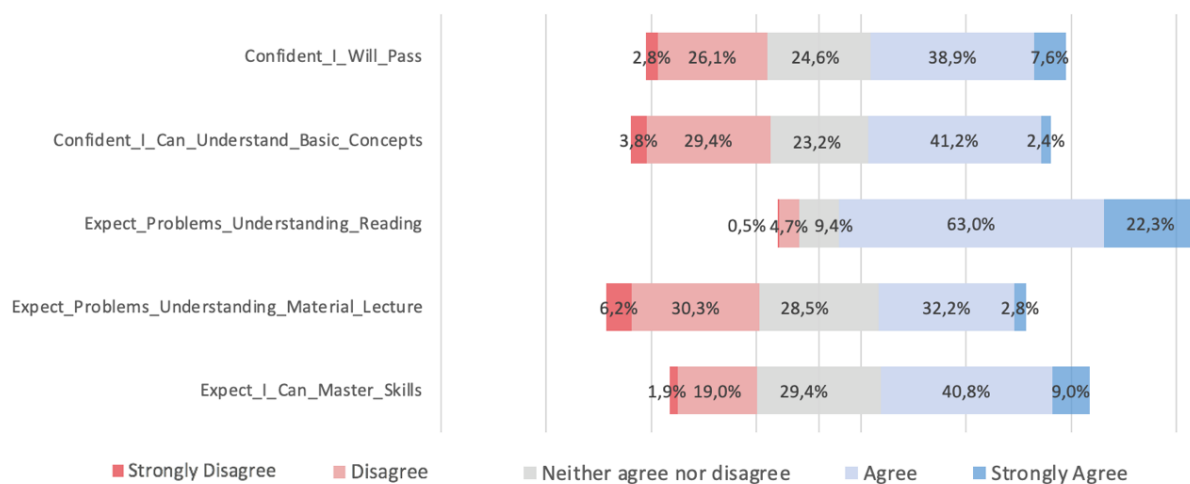
*Mean and Standard Deviation of Self-Efficacy Items in Formative and Summative Courses*

Items	Formative		Summative	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Confident_I_Will_Pass	3.74	0.83	3.22	1.00
Confident_I_Can_Understand_Basic_Concepts	4.20	0.67	4.02	0.74
Expect_Problems_Understanding_Reading	2.76	9.86	3.09	0.97
Expect_Problems_Understanding_Material_Lecture	2.77	0.99	2.95	0.99
Expect_I_Can_Master_Skills	3.79	0.76	3.36	0.95

*Note.* All Scores were measured with 1 (strongly disagree), 2 (disagree), 3 (neither agree nor disagree), 4 (agree), 5 (strongly agree).

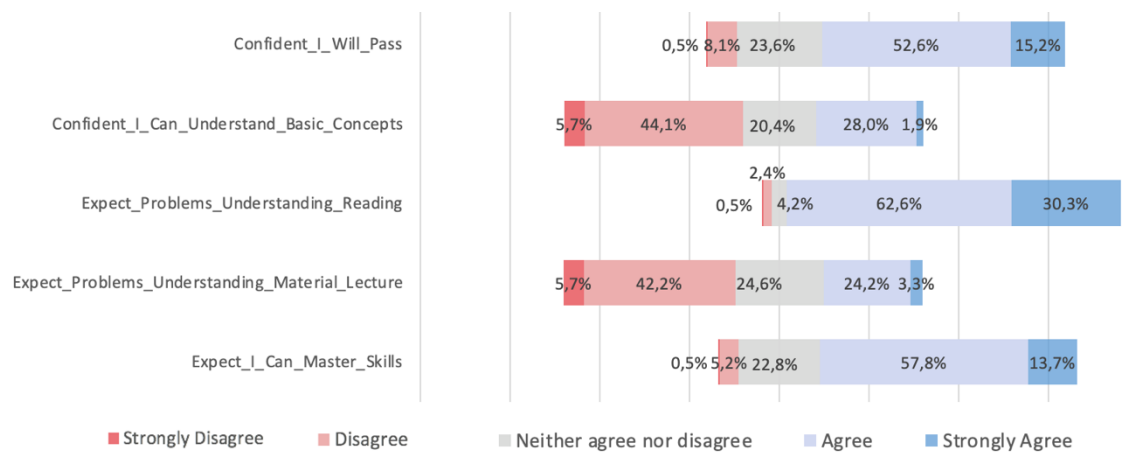
**Figure 1**

*Distribution of Self-Efficacy Responses in the Five Individual Items in Summative Courses*



**Figure 2**

*Distribution of Self-Efficacy Responses in the Five Individual Items in Formative Courses*



### ***Cheating***

Participants were asked about their cheating behavior by using a coin toss method whereby the “yes” responses included those who were instructed to answer “yes” regardless of the truth (“tail”) but also those who were instructed to respond truthfully (“head”). The “no” responses reflect the score for those who only got ‘head’ and were instructed to answer truthfully. Hence, the “no” responses displayed the true score for half of the total responses. This was under the assumption that the participants had an equal chance of 50% of getting either “heads” or “tail”.

To determine the actual percentage of cheaters, the number of “no” responses (e.g., 94) was doubled (e.g., 188) and the percentage of the total number of responses (e.g., 211) determined (89.1%). Then, the percentage of “no” answers (89.1%) was subtracted from 100% to obtain the percentage of actual cheaters (10.9%).

Hence, in summative courses approximately 10.9% of the participants admitted to cheating in some form during exams. In contrast, 6.2% indicated to have cheated in some form during assignments in formative courses when using the same calculation method as described previously. However, the online coin toss website did not seem to produce a fair

50% distribution of “head” or “tail” when the participants were asked about exams in formative courses. This resulted in fewer participants getting “tail” and, therefore, led to a negative percentage of -4% when using the above-mentioned calculation method. Thus, it was treated as 0% in this paper. Hence, about 0% indicated to have cheated in exams in formative courses. Looking closer at these results, 6.2% indicated to have cheated in assignments in the formative condition when also exams were administered. In those exams, 0% indicated to have cheated in some form. This was in contrast to exams without additional assignments, in which about 10.9% of participants indicated to have cheated in some form. Since this difference seems great, it can be assumed that cheating in formative assessment was lower than in the summative assessment. However, the result has to be viewed with caution as no significance test was conducted. Hence, the second hypothesis is only partly supported. Table 6 provides an overview of the unprocessed frequencies and percentages of cheating using the coin toss method.

**Table 6**

*Frequency and Percentage of Cheating in Formative and Summative Courses with Coin Toss Method*

Items	Formative				Summative	
	Exams		Assignments		Exams	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Yes	96	45.5	112	53.1	117	55.5
No	114	54	99	46.9	94	44.5
Total	210	99.5	211	100	211	100

*Note.* The “yes” responses include the responses of the coin toss method “tail” responses.

### **Self-efficacy: Difference Between Summative and Formative Courses**

Prior to the inferential analyses of the self-efficacy scores, the assumption of normality was checked. For that, the Shapiro-Wilk test was used for both summative and formative assessment and indicated a significant violation of normality,  $W = 0.983, p = .011$ ;  $W = 0.972, p = .001$ , respectively, at a  $p < .05$  threshold for statistical significance. However, normality was considered robust due to the large sample size ( $n = 211$ ; Agresti, 2018). The assumption of independence between subjects was considered met. The data was also examined for influential points. Given that all participants answered the questions of both the formative and summative part, the scores were not considered independent from one another. Hence, a paired samples t-test was conducted to account for the dependence.

The above-mentioned comparison of the self-efficacy means of formative and summative courses revealed a difference of 0.33 between the two means, giving reason to test the significance of the difference. For the analysis, the five items assessing self-efficacy were merged into one variable by averaging the scores. That was done for the self-efficacy items in the formative and summative part. Consistent with the first hypothesis, the paired-sample t-test suggested that there was a significant difference between the self-efficacy scores of formative and summative courses ( $t(210) = -7.0, p < .001, d = -.482$ ) with medium effect size (Cohen, 1992). This result suggested that the participants' self-efficacy was higher in courses with formative assessments compared to summative assessments.

### **Influence of Self-Efficacy on Cheating**

Since the variable assessing cheating behavior was dichotomous and the proportion of actual cheaters had to be calculated manually, the continuous self-efficacy scales of formative and summative courses needed to be transformed into a dichotomous variable by splitting it into groups allowing for a comparison of the scores. Hence, the variables consisting of the averaged five items scores assessing self-efficacy were split into high and low self-efficacy. All values below 3, including responses of disagree (2) or strongly disagree (1), were

considered as low self-efficacious beliefs, and all values above 3, indicated by agree (4) and strongly agree (5), were considered high. After that, a crosstab was created displaying the frequencies of the cheating responses in formative courses with exams and assignments and summative courses with exams, visible in Table 7.

Based on the manual calculation of true cheating scores, one can say that 12.5% of participants with low self-efficacy admitted to cheating in courses using summative exams. In comparison, approximately 9.92% of the participants with high self-efficacy admitted to cheating. The difference of the results was not further tested for significance as the study design precluded the possibility. Thus, the descriptives indicated that participants with high self-efficacy cheated less on summative exams than participants with low self-efficacy. The same calculation was done for the assessment methods in formative courses. Surprisingly, only about 2.56% of the participants with low self-efficacy indicated to have cheated in formative assignments, in contrast to 6.98% of the participants with high self-efficacy. No comparison could be made between the two groups of self-efficacy formative courses and formative exams because the coin toss website did not seem to produce an equal 50% chance of “heads” or “tails”, meaning that less people than anticipated indicated to have cheated. The calculations in both groups were treated hence as 0%. The third hypothesis that participants with high self-efficacy cheat less than those with low self-efficacy was partly supported based on the direction of scores for the summative exam but not for the formative assessments.

**Table 7**

*Frequency and Percentage of Cheating in Formative and Summative Courses Across High and Low Self-Efficacy*

Items	Formative		Summative	
	Exams	Assignments	Exams	
	Low	High	Low	High

Yes	18	78	20	92	45	72
No	21	93	19	80	35	59
Total	39	171	39	172	80	131

*Note.* The “yes” responses include the responses of the coin toss method “tail” responses.

Low = average self-efficacy response below value of 3 (including disagree (2) and highly disagree (1)); High = average self-efficacy response above value of 3 (including agree (4) and highly agree (5)).

## **Discussion**

In the present study we investigated how psychology undergraduate students at the University of Groningen perceived their self-efficacy and cheating behavior in university assignments or exams to be different between formative and summative courses. We hypothesized that students’ self-efficacy would be significantly higher in courses using formative assessment than compared to courses using summative ones. Similarly, it was hypothesized that students would cheat less in assessments in formative courses than compared to summative ones. The last hypothesis stated that those students having high self-efficacy would also cheat less. Overall, all hypotheses were supported or partly supported.

### **Difference Between Formative and Summative Courses**

#### ***Self-efficacy***

In the first hypothesis, we assumed that students would have higher self-efficacy in courses using formative assessment compared to using summative assessment. The findings of the present study supported this hypothesis as the average self-efficacy score in formative courses was significantly higher than in summative courses with a medium effect size ( $d = .482$ ). This indicated that the participants experienced their self-efficacy during formative courses to be significantly higher than in summative courses to a medium extent. This may be explained by the nature of formative assessment, as it, when conducted correctly, focuses on

the acquisition of skills such as monitoring or self-reflection. These skills not only strengthen the sense of being in control of the learning process but also increase awareness once students successfully learn something new. Reflecting on moments of success may feed into a cycle of positive thoughts and more successes that ultimately strengthen the belief in their own capacities. Support of this assumption was found in a comprehensive meta-analysis by Panadero et al. (2017) who reviewed 19 studies from around the world and, based on their results, concluded that implementing self-assessment interventions that aiming to increase skills such as monitoring or self-evaluation in students, had a positive effect on their self-efficacy with a large effect size ( $d = 0.73$ ). This indicates that the skills and mechanisms involved in performing formative assessment indeed increase the feeling of self-efficacy.

Interestingly, the result of the present study is inconsistent with the literature investigating the influence of formative assessment on self-efficacy scores. For example, den Boer et al. (2021) questioned mechanical engineering students at a Dutch University of Applied Sciences in a self-report about their self-efficacy in formative and summative courses and did not find significant. Similarly, Yin et al. (2008) compared self-efficacy scores of middle school students in a small, randomized experiment, where also no significant differences were found in pre- and post-questionnaires. Their inconsistency with the result of the present study may be their lack of focus on monitoring or self-evaluation skills.

Another possible explanation for the mixed results may be the lack of control concerning the way feedback was provided to students. That is, in both studies it was not further described how students received feedback to their formative assessments. For example, whether they were solely given a sheet with correct answers or were their responses individually corrected, discussed and their strengths and weaknesses for the next study session implemented. The way feedback is delivered to students can make a difference in the effectiveness of formative assessment and hence their self-efficacy and ultimately learning



experience. The importance of feedback was summarized by Hattie and Timperley (2007). They stated that feedback is only effective if it stimulates student's motivation to further engage with the task, enhances self-efficacy, and is generally meaningful in a way that it is "effective in reducing the discrepancy between the current understandings and what is desired" (Hattie & Timperley, 2007, p. 86). Since the above-mentioned studies did not elaborate on how their feedback was provided, it is unclear whether these goals or standards were adhered to and may have diminished the effectiveness of formative assessment on the students' self-efficacy and learning experience.

Given that the present study did also not specify or control for the way feedback was provided to students, it is surprising that the result still showed a significant difference in students' self-efficacy depending on the type of assessment. One reason for that may be a misunderstanding of the definition of formative assessment in the present study. That is, the sample of this study consisted mainly of first-year students that may have reported higher self-efficacy in the questionnaire because they possibly used a course as point of reference for formative courses that, in fact, was not formative. To explain, only two first-year courses applied formative assessment strategies, namely Statistics 1a, which includes weekly homework assignments and individualized feedback, and Academic Skills, which teaches skills such as writing and critical thinking. The latter is fundamentally different to the former and may seem to be a formative course due to the assessment being assignments. However, it is not since the assignments are graded and altogether make up the final course grade, hence, making it a summative assessment. Since the academic skills course is mostly experienced quite positively by students, it is possible that some first-year students have mistaken the course as formative, influencing their responses for formative courses to be more positive and leading to unrepresentative results. In order to investigate whether the first-year students influenced the results, a speculative analysis was performed.

For that, a repeated measures analysis of variance (RM-ANOVA) was conducted and revealed a significant effect for the starting year of the psychology bachelor ( $F(1, 5.06) = 9.108, p < .005, \eta^2 = .042$ ). This indicates that there was indeed a difference in the first-year students' self-efficacy and the higher-year students' self-efficacy. This result is only speculative and may be used as an indicator that the disproportionately distribution of first-year students had an impact on the results. Hence, the results for the first hypothesis have to be viewed with caution.

### ***Cheating***

In the second hypothesis, we stated that students would cheat less in assignments and exams that are part of formative courses in comparison to summative ones. The results of this study supported the hypothesis as fewer students at the University of Groningen cheated in formative final exams and assignments than in summative exams. Interestingly, in comparison to summative final exams, the amount of cheating in formative assignments was already lower but practically non-existent in formative final exams. This indicates that students cheated less in formative assessment methods, i.e., assignments and exams, compared to exams in summative courses. One possibility for the general lower frequency of cheating in formative courses than in summative courses is the possibility that students interacted more consistently with the study material throughout the study module in the formative assignments. That may have led the students to store some of the new information in their memory and felt no need to cheat in the final exam as they felt more prepared. However, this result must be viewed with caution as we were not able to correlate participants' individual cheating scores to their self-efficacy scores. That is, the method of randomized responses allowed participants to be honest with their cheating behavior on average but not individually. Therefore, only indirect comparisons of students' group scores were possible. Although the current study design precludes a causal statement as it is no experimental design, the

descriptive results of the present study are an early indication that students cheat less in formative than in summative courses.

The literature concerning a direct comparison between formative and summative assessment is also scarce and inconclusive due to variations in definitions of cheating, different types of cheating, different implementation strategies of formative and summative assessment, study subject, location (e.g., online or offline) and biases in reports (e.g., surveillance bias or social desirability bias). Richardson et al. (2012) partly acknowledged this issue in a literature review in which coursework (i.e., assignments) and examinations were compared. Due to the lack of research in this subject, the different components for formative assessment, such as assignments, and summative assessment, such as exams, need to be investigated individually for each of the possible cheating behaviors. For instance, the results concerning literature on assignments are mixed, as Gardner et al. (1988) found that nearly all (98%) of first-year psychology students cheated on assignments at least once in two experiments. In contrast, Karlins et al. (1988) only found 3% of the undergraduate students cheating in the course principles of management. Even though research on cheating in assignments was scarce, more literature could be found when looking for the type of cheating behavior that can be performed in assignments e.g., plagiarism. When reviewing literature on estimates of plagiarism, it revealed again mixed results: On the one hand, López Puga (2014) found undergraduate psychology students admitting to plagiarism ranging from 62% to 57%, while Stephens et al. (2007), on the other hand, only found about 31% of undergraduate US students. None of them used a randomization method. The latter results overtop what was found in the present study. One explanation for the lower cheating frequency in this study is that students had to self-report their cheating behavior and may have been, despite the randomization method, hesitant to answer truthfully. They also might have been ashamed of their behavior and answered in a socially desirable way of not having cheated. Also, since the

present study asked the students to indicate specific cheating behaviors in specific courses in retrospect, their memory of those behaviors may have been biased. In conclusion, the results of the present study are an early indication that students cheat less in formative courses than in summative ones, implying that universities that solely do summative assessments, would possibly benefit from adding formative assessments as they may improve the study experience of students.

### **Self-efficacy and Cheating**

In the third hypothesis, we assumed that high self-efficacy beliefs would be negatively associated with cheating behavior. The results indicated that participants with high self-efficacy reported less cheating in summative final exams than those with low cheating behavior. This result was consistent with the hypothesis. However, we found the opposite for formative assignments: Participants with high self-efficacy cheated more than those with low self-efficacy. This result was not in line with the hypothesis. The different results may be explained by the differing nature of cheating behaviors in exams and assignments. That is, high self-efficacious students may feel less need to cheat in final exams as their positive beliefs about their capacities prevents them from cheating. Also, exams are proctored by supervisors, and being detected with cheating in the presence of others would also be embarrassing. In contrast, assignments are written without a proctor, and participants with high self-efficacy may be more prone to engaging in plagiarism or using notes from others. Another explanation may be the small sample size in the individual groups after splitting self-efficacy into two groups so that personal preferences of courses or teachers served as a confound when the participants reported their experiences. For example, if a participant used a favored course as their point of reference for formative courses in which their self-efficacy was also high, their responses would have reflected that. Similar to the second hypothesis, the study design precluded the possibility of testing the significance of the mean differences of

cheating in high and low efficacious participants. Hence, the results have to be viewed with caution.

The literature on cheating behavior supported the study's finding that high self-efficacy leads to less cheating in summative final exams but was inconsistent with the finding concerning formative assignments. For example, Pavlin-Bernardić et al. (2017) investigated the relationship between self-efficacy and academic cheating in second and third grade students from six different high schools in Croatia based on self-report and found that high self-efficacy was associated with less cheating. A similar result was recently found in Polish university students asking them via self-report to indicate their self-efficacy and cheating behavior (Baran & Jonason, 2020). The findings suggested again that those students with high self-efficacy reported less cheating. The ambiguous results with the findings of the present study may be explained by the assessment types. That is, the above-mentioned studies did not specify the type of assessment in which participants cheated e.g., exams or assignments in formative or summative courses. This may be a confound. Also, it is unclear whether participants referred to online or offline assessment when answering the course specific questions. That is, due to the Covid-19 pandemic, assessment for the higher-year students suddenly shifted from offline to online assessment, making it challenging for teachers to ensure quality teaching and may have had an influence on their willingness to cheat. In conclusion, the descriptive results are a first indication that students who are high in self-efficacy report less cheating in summative exams and more cheating in formative assignments. However, due to the study design that does not allow for a significance test of the differences, this is only speculative.

### **Future Directions**

Based on the findings of the present study, formative assessment seems superior to summative assessment in increasing the students' senses of self-efficacy and leading to lower

cheating behavior. Future research could compare whether cheating behavior is reduced in final exams if formative assessments were used throughout the course module. Also, it could be investigated whether high self-efficacy does indeed lead to more cheating behavior in formative assessments.

### **Conclusion**

The present study adds value to the existing literature on the students' perception at on formative and summative assessment based on their self-efficacy and cheating behavior. In particular, in comparison to summative assessment, formative assessments seem to increase self-efficacy beliefs of students at the University of Groningen and descriptive data suggests that this type of assessment is associated with less cheating behavior. The results enhance the understanding of how undergraduate psychology students at the University of Groningen experience assessments and can be used to create assessments that strengthen students' self-efficacy and to make their learning experience worthwhile.

## References

- Agresti, A. (2017). *Statistical Methods for the Social Sciences* (5th ed.). Pearson.
- Anderman, E. M., & Koenka, A. C. (2017). The Relation between Academic Motivation and Cheating. *Theory Into Practice, 56*(2), 95-102.
- Andrade, H. (2010). Summing up and moving forward: Key challenges and future directions for research and development in formative assessment. *Handbook of formative assessment, 344-351*.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co.
- Baran, L., & Jonason, P. (2020). Academic dishonesty among university students: The roles of the psychopathy, motivation, and self-efficacy. *PLOS ONE, 15*, e0238141.  
<https://doi.org/10.1371/journal.pone.0238141>
- Bashir, H., & Bala, R. (2018). Development and Validation of Academic Dishonesty Scale (ADS): Presenting a Multidimensional Scale. *International Journal of Instruction, 11*, 57-74. <https://doi.org/10.12973/iji.2018.1125a>
- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.  
<https://doi.org/10.1080/0969594X.2010.513678>
- Bisping, T. O., Patron, H., & Roskelley, K. (2008). Modeling Academic Dishonesty: The Role of Student Perceptions and Misconduct Type. *The Journal of Economic Education, 39*(1), 4-21.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.  
<https://doi.org/10.1080/0969595980050102>

- Butler, A. C. (2018). Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition*, 7(3), 323-331. <https://doi.org/https://doi.org/10.1016/j.jarmac.2018.07.002>
- Cizek, G. J. (1999). *Handbook of educational policy*. Academic Press.  
<http://site.ebrary.com/id/10186362>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- den Boer, A. W., Verkoeijen, P. P., & Heijltjes, A. E. (2021). Comparing Formative and Summative Cumulative Assessment: Two Field Experiments in an Applied University Engineering Course. *Psychology Learning & Teaching*, 20(1), 128-143.  
<https://doi.org/10.1177/1475725720971946>
- De dalende rijksbijdrage en de druk op onderwijs en onderzoek*. (2021, August 16). Universiteiten van Nederland. Retrieved January 13, 2022, from <https://www.universiteitenvannederland.nl/dalende-rijksbijdrage.html>
- Dixson, D. D., & Worrell, F. C. (2016). Formative and Summative Assessment in the Classroom. *Theory Into Practice*, 55(2), 153-159.  
<https://doi.org/10.1080/00405841.2016.1148989>
- Dogan, U. (2015). Student Engagement, Academic Self-efficacy, and Academic Motivation as Predictors of Academic Performance. *The Anthropologist*, 20(3), 553-561.  
<https://doi.org/10.1080/09720073.2015.11891759>
- Eckstein, M., & Unesco, I. (2003). Combating academic fraud: towards a culture of integrity. [http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/\[in=epidoc1.in\]/?t2000=018777/\(100\)](http://lst-iiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=018777/(100)).



- Education, E. A. o. I. i. H. (2014). *Professional Higher Education in Europe: Characteristics, Practices examples and National differences*. K. I. Centre.  
[https://www.eurashe.eu/library/mission-phe/PHE\\_in\\_Europe\\_Oct2014.pdf](https://www.eurashe.eu/library/mission-phe/PHE_in_Europe_Oct2014.pdf)
- Entwistle, N., & Tait, H. (2013). Approaches and Study Skills Inventory for Students (ASSIST) (incorporating the Revised Approaches to Studying Inventory - RASI).
- Freire, C., Ferradás, M. d. M., Núñez, J. C., Valle, A., & Vallejo, G. (2019). Eudaimonic Well-Being and Coping with Stress in University Students: The Mediating/Moderating Role of Self-Efficacy. *International Journal of Environmental Research and Public Health*, 16(1), 48. <https://www.mdpi.com/1660-4601/16/1/48>
- Gardner, W. M., Roper, J. T., Gonzalez, C. C., & Simpson, R. G. (1988). Analysis of Cheating on Academic Assignments. *The Psychological Record*, 38(4), 543-555.  
<https://doi.org/10.1007/BF03395046>
- Harding, T. S., Passow, H. J., Carpenter, D. D., & Finelli, C. J. (2004). An examination of the relationship between academic dishonesty and professional behavior. *IEEE Antennas and Propagation Magazine*, 46(5), 133-138.  
<https://doi.org/10.1109/MAP.2004.1388860>
- Harlen, W., & Crick, R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. [http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/\[in=epidoc1.in\]/?t2000=021193/\(100\)](http://lst-iiiep.iiep-unesco.org/cgi-bin/wwwi32.exe/[in=epidoc1.in]/?t2000=021193/(100)).
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Just flip a coin*. (2010). Retrieved January 11 from <https://justflipacoin.com/>
- Karlins, M., Michaels, C., & Podlogar, S. (1988). An Empirical Investigation of Actual Cheating in a Large Sample of Undergraduates. *Research in Higher Education*, 29(4), 359-364.

- Keener, T. A., Galvez Peralta, M., Smith, M., Swager, L., Ingles, J., Wen, S., & Barbier, M. (2019). Student and faculty perceptions: appropriate consequences of lapses in academic integrity in health sciences education. *BMC Medical Education*, *19*(1), 209. <https://doi.org/10.1186/s12909-019-1645-4>
- Kisamore, J. L., Stone, T. H., & Jawahar, I. M. (2007). Academic integrity: The relationship between individual and situational factors on misconduct contemplations. *Journal of Business Ethics*, *75*(4), 381-394. <https://doi.org/10.1007/s10551-006-9260-9>
- Krause, K. L., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education*, *33*(5), 493-505. <https://doi.org/10.1080/02602930701698892>
- Lerdpornkulrat, T., Poondej, C., Koul, R., Khiawrod, G., & Prasertsirikul, P. (2019). The Positive Effect of Intrinsic Feedback on Motivational Engagement and Self-Efficacy in Information Literacy. *Journal of Psychoeducational Assessment*, *37*(4), 421-434. <https://doi.org/10.1177/0734282917747423>
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *19*(2), 119-137. <https://doi.org/10.1080/10573560308223>
- López Puga, J. (2014). Analyzing and reducing plagiarism at university. *Journal of Education and Psychology*, *7*, 131-140. <https://doi.org/10.1989/ejep.v7i2.186>
- Making Connections* (2021). (Strategisch Plan 2021-2026, Issue. <https://www.rug.nl/about-ug/policy-and-strategy/strategic-plan/strategisch-plan-eng-2021.pdf>
- Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*(2), 194-199. <https://doi.org/10.3758/BF03194051>

- Marsh, H. (2011). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. *British Journal of Educational Psychology*, 52, 77-95. <https://doi.org/10.1111/j.2044-8279.1982.tb02505.x>
- McCabe, D. (2016). Cheating and Honor: Lessons from a Long-Term Research Project. In T. Bretag (Ed.), *Handbook of Academic Integrity* (pp. 187-198). Springer Singapore. [https://doi.org/10.1007/978-981-287-098-8\\_35](https://doi.org/10.1007/978-981-287-098-8_35)
- McCabe, D. L. (1999). Academic dishonesty among high school students. *Adolescence*, 34(136), 681-687.
- McCarthy, J. (2017). Enhancing feedback in higher education: Students' attitudes towards online and in-class formative assessment feedback models. *Active Learning in Higher Education*, 18, 127-141. <https://doi.org/10.1177/1469787417707615>
- Murdock, T. B., & Anderman, E. M. (2006). Motivational Perspectives on Student Cheating: Toward an Integrated Model of Academic Dishonesty. *Educational Psychologist*, 41(3), 129-145. [https://doi.org/10.1207/s15326985ep4103\\_1](https://doi.org/10.1207/s15326985ep4103_1)
- Ogilvie, J., & Stewart, A. (2010). The Integration of Rational Choice and Self-Efficacy Theories: A Situational Analysis of Student Misconduct. *The Australian & New Zealand journal of criminology*, 43(1), 130-155.
- Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74-98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Papp, R., & Wertz, M. (2009). To Pass At Any Cost: Addressing Academic Integrity Violations.
- Pavlin-Bernardić, N., Rovani, D., & Pavlović, J. (2017). Academic Cheating in Mathematics Classes: A Motivational Perspective. *Ethics & Behavior*, 27(6), 486-501. <https://doi.org/10.1080/10508422.2016.1265891>

- Pintrich, P., Smith, D., Duncan, T., & McKeachie, W. (1991). A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor, Michigan, 48109*, 1259.
- Pitsch, W., Feth, S., Frenger, M., & Schmelzeisen, P. (2017). *Cheater Detection for Randomized Response-Techniques. Derivation, Analyses and Application*.
- Qu, W., & Zhang, C. (2013). The Analysis of Summative Assessment and Formative Assessment and Their Roles in College English Assessment System. *Journal of Language Teaching and Research*, 4. <https://doi.org/10.4304/jltr.4.2.335-339>
- Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, 154-165. <https://doi.org/10.1016/j.learninstruc.2018.01.004>
- Resnick, L., & Resnick, D. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. [https://doi.org/10.1007/978-94-011-2968-8\\_3](https://doi.org/10.1007/978-94-011-2968-8_3)
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin*, 138(2), 353-387. <https://doi.org/10.1037/a0026838>
- Schaufeli, W. B., Salanova, M., González-romá, V., & Bakker, A. B. (2002). The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *Journal of Happiness Studies*, 3(1), 71-92. <https://doi.org/10.1023/A:1015630930326>
- Simkin, M., & McLeod, A. (2010). Why Do College Students Cheat? *Journal of Business Ethics*, 94, 441-453. <https://doi.org/10.1007/s10551-009-0275-x>

- Simpson, M. L., Olejnik, S., Tam, A. Y.-W., & Supattathum, S. (1994). Elaborative verbal rehearsals and college students' cognitive performance. *Journal of Educational Psychology, 86*(2), 267-278. <https://doi.org/10.1037/0022-0663.86.2.267>
- Solomon, L., & Rothblum, E. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology, 31*, 503-509. <https://doi.org/10.1037/0022-0167.31.4.503>
- Starovoytova, D., & Namango, S. (2016). Factors Affecting Cheating-Behavior at Undergraduate-Engineering. 7.
- Stephens, J. M., Young, M. F., & Calabrese, T. (2007). Does Moral Judgment Go Offline When Students Are Online? A Comparative Analysis of Undergraduates' Beliefs and Behaviors Related to Conventional and Digital Cheating. *Ethics & Behavior, 17*(3), 233-254. <https://doi.org/10.1080/10508420701519197>
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education, 30*, 331-347. <https://doi.org/10.1080/02602930500099102>
- Swift, C. O., & Nonis, S. (1998). When No One Is Watching: Cheating Behavior on Projects and Assignments. *Marketing Education Review, 8*(1), 27-36. <https://doi.org/10.1080/10528008.1998.11488616>
- Tas, Y., & Tekkaya, C. (2010). Personal and Contextual Factors Associated With Students' Cheating in Science. *The Journal of Experimental Education, 78*(4), 440-463.
- van Dinther, M., Dochy, F., Segers, M., & Braeken, J. (2014). Student perceptions of assessment and student self-efficacy in competence-based education. *Educational Studies, 40*(3), 330-351. <https://doi.org/10.1080/03055698.2014.898577>

Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309), 63-69.

<https://doi.org/10.1080/01621459.1965.10480775>

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, 37(6), 747-

760. <https://doi.org/10.1080/02602938.2011.572153>

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M.,

Tomita, M. K., & Young, D. B. (2008). On the Impact of Formative Assessment on Student Motivation, Achievement, and Conceptual Change. *Applied Measurement in Education*, 21(4), 335-359.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.

<https://doi.org/10.1023/A:1023967026413>