

**The relationship of self-efficacy and engagement with formative and summative assessment
in Higher Education.**

Matti Kuhnert

s3976106

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

12

Supervisor: dr. A. Sarampalis

Second evaluator: B. Guetzkow, Msc

In collaboration with: Julia Wulf, Stella Fritzsche, Oliver Konradt, Lena Duiverman,

Tilman Mueller-Scholtz

February 3, 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

Across all educational institutions, assessment plays a central role. Summative assessment aims at testing whether students have successfully learned what was expected of them, while formative assessment is used to help students improve their performances by giving them feedback. A growing body of evidence suggests that the latter type of assessment is more beneficial for students' learning than the former. Furthermore, there is evidence associating the constructs self-efficacy and students' engagement to various desirable outcomes, such as higher academic achievement or greater well-being. However, only few studies have explored how both constructs (self-efficacy and engagement) vary depending on whether formative or summative assessment was employed in a higher education setting. The present study aimed at adding to this thin body of literature by employing an online questionnaire to bachelor psychology students from the Netherlands ($N = 210$). The study was explorative in nature and formulated no hypotheses. Repeated measures ANOVAs were conducted to investigate differences in means between the summative and formative assessment condition. The result showed that the participants were more engaged and higher in self-efficacy when participating in formative, rather than in summative assessment courses, with medium effect sizes. Although the study was correlational, the findings imply that students' experiences would benefit from the implementation of formative assessment methods in higher education.

Keywords: formative assessment, summative assessment, self-efficacy, engagement, higher education

The relationship of self-efficacy and engagement with formative and summative assessment in Higher Education.

One of the primary purposes of higher education institutions is to prepare their students for their future careers. Additionally, they also fulfil the function of accrediting that their students are qualified for certain jobs or further study related programmes. This happens mostly by teaching and examining content and skills that are perceived to be relevant. In both of these processes, assessment plays a crucial role. To instance, students' perceptions of how course content will be assessed influences what, how, and how much they study, which is also known as the backwash effect (Biggs, 1996; Gibbs & Simpson, 2005). It is therefore not only important how content and skills are being taught but also how they are assessed.

A common dichotomy of assessment types is summative and formative assessment. The former is used to evaluate whether students have successfully learned what they were expected to learn (Yorke, 2003). This is closely related to the previously mentioned function of institutions, namely ensuring that their students learned certain content or skills (Yorke, 2003). On the other hand, formative assessment involves processes aimed at helping students to improve their performances by giving them feedback about their previous work (Harlen & James, 1997; McDowell et al., 2009; Yorke, 2003). In the past thirty years, evidence accrued showing that feedback is one of the most influential factors for student learning (Black & Wiliam, 1998). Because of that, researchers shifted their attention from summative to formative assessment types (Rushton, 2005). Consequently, educational institutions for all age groups try to implement the latter and distance themselves from the former type of assessment (McDowell et al., 2009).

For formative assessment to be successful, both parties need to be active: the teacher needs to give relevant feedback about the student's performance, while the student needs to be

willing to use it (Yorke, 2003). Although in theory this concept is very different from summative assessment, a course's assessment might include summative, as well as formative assessment parts, to varying degrees (Yorke, 2003). For example, a course might include several assignments throughout the semester for which the students receive feedback, which is supposed to be used to improve the quality of their final assignment (formative assessment). However, they only pass the course if they submit the final assignment with sufficient quality (summative assessment). Because most educational institutions (especially in higher education) eventually reward students for their effort, it is often difficult to investigate courses that exclusively make use of formative assessment methods. The following review will therefore also include studies that investigated assessment types that were mostly (but not fully) formative in nature.

Furthermore, various assessment methods, such as peer-reviews, quizzes or oral feedback to assignments can be considered formative assessment. This creates challenges with generalising from one assessment method to the other, which is why it is difficult to gauge and quantify the effects of formative assessment on a general level. Besides studying the links between assessment types with student learning and achievement, associations with other concepts, such as self-efficacy have also been explored (e.g., den Boer et al., 2021).

Self-efficacy

Self-efficacy describes the "beliefs about one's ability to carry out certain actions required to attain a specific goal" (Hewstone et al., 2016, p. 196). Bandura (1997) argued that overcoming an obstacle requires the skills needed to do so, but also the belief that this is possible. Otherwise, the individual would not be capable of doing so or not motivated enough to try. Self-efficacy therefore works as a motivational factor and constitutes an important construct in research about education that has received considerable attention in the past decades (van Dinther et al., 2011).

There is ample evidence showing that self-efficacy is a reliable predictor of academic achievement in higher education (Ayllón et al., 2019; Richardson et al., 2012; van Dinther et al., 2011), while also being linked to other desirable outcomes, such as well-being (Boulton et al., 2019). For these reasons, institutions in higher education should consider which factors influence their students' self-efficacy.

With regards to the relationship between self-efficacy and formative assessment, the body of evidence is relatively scarce and contradictory. For example, den Boer et al. (2021) published results that showed no effect of formative assessment on student's self-efficacy. They conducted two field experiments with participants who took (up to) six tests throughout an undergraduate course. In one condition (summative assessment), the five best results would influence the final course grade by 30% (the final exam by 70%), in the other condition (formative assessment), the tests were not mandatory and did not affect the final course grade, meaning the final exam was the only determinant. The researchers (den Boer et al., 2021) did not find significant differences in student's sense of self-efficacy between the summative and formative assessment conditions. On the contrary, Lerdpornkulrat et al. (2019) conducted a quasi-experiment in which they compared two groups in an undergraduate information literacy course: the experimental group was given a rubric sheet for self-assessment in addition to feedback from their teacher, while the control group merely received feedback from their teacher. The results show that the experimental group had a significantly higher sense of self-efficacy than the control group with regards to information literacy skills on all dimensions: seeking/evaluating/processing information, as well as communicating and spreading (e.g., writing reports) it. Only few studies have investigated the links between self-efficacy and formative assessment methods in higher education. Particular studies that directly compare a summative and formative assessment

condition are rare. More research in this field is needed to understand how students' self-efficacy is linked to formative and summative assessment.

Engagement

Another important aspect for the academic development of students is their engagement. However, the construct is loosely defined (Kahn, 2014) and therefore difficult to measure (Boulton et al., 2019). In the present study, it is defined as "the active commitment and purposeful effort expended by students towards all aspects of their learning, including both formal and informal activities." (Boulton et al., 2019, p. 1). This rather broad definition was chosen because the present study seeks to explore the links between assessment types and engagement on a more general level. Engagement has been linked to various desirable outcomes, such as better academic performance (Salanova et al., 2010) or greater well-being (Boulton et al., 2019; Pascarella et al., 2010). For instance, Pascarella et al. (2010) conducted a longitudinal study in the United States that correlated certain forms of engagement with positive outcomes. The large sample of roughly 1400 to 2800 participants (depending on the scale) consisted of students in higher education from various institutions, such as universities or community colleges. The researchers found that "Active collaboration and learning" (e.g., interacting with peers) and "Student-faculty interaction" (e.g., contact with professors) were both positively correlated with "Effective reasoning and problem solving", as well as with the well-being of the participants. The medium effect sizes ranged from $r = .31$ to $r = .30$.

As previously mentioned, one issue with investigating student's engagement regarding formative assessment is that the former construct is loosely defined (Kahn, 2014) and therefore also differently operationalized. In this study, it is interpreted relatively broadly, entailing various dimensions, such as studying regularly or interacting with peers and professors. In

contrast to that, most of the published studies assess the participants' engagement with the respective formative assessment measure without exploring whether their general engagement with the course or programme changed as well. For instance, in online settings, engagement is often measured by assessing how much (and/or how many) students use formative assessment methods. To illustrate, Jacoby et al. (2014) conducted a study with 350 students who took part in a science related programme at a university in England that was meant to prepare them for traditional majors, such as Mathematics or Engineering. The formative assessment measure was embedded in an online learning environment in which the students were supposed to answer multiple choice questions. The more students used the measure, the more engaged the students in that programme were said to be. One might argue that it is evident that students become more engaged with the material when they complete tasks that are about that same material. However, it was not further explored whether the engagement of the participants with the programme on a more general level (e.g., attending lectures) varied as well.

Another example of this in an offline setting is illustrated by Han and Finkelstein (2013). Their study investigated the effects of remote controls (also known as "clickers") that students use in classrooms to answer questions posed by the person giving a lecture. Usually, the answers are subsequently shown and discussed immediately afterwards (Han & Finkelstein, 2013). In their study, the researchers compared two types of courses: one was defined by the fact that the professors used the clickers to grade students (summative assessment group), while the other group consisted of courses that were taught by professors who did not use them for grading purposes (formative assessment group). The students in the latter group were significantly more engaged with the lectures and its material than students in the summative assessment group. Similar to Jacoby et al. (2014), engagement in this study was also rather narrowly

operationalized, meaning as in whether the clickers help to actively engage the participants with the lectures. However, it might be valuable to explore whether desirable effects of formative assessment methods also go beyond their direct purpose. For example, do students tend to study more regularly when formative, rather than summative assessment methods are used? More research on this topic is needed to understand these mechanisms, especially in comparison to summative assessment.

The present study

The aim of the present study is to provide information on how to improve study programmes in higher education by focussing on how bachelor psychology students at the University of Groningen perceive formative and summative assessment methods. In contrast to most studies exploring the effects of formative assessment, the construct was investigated on a rather abstract level. This means that the focus was neither on a single formative assessment measure nor on a single course. Instead, it was operationalized in such a way that numerous types of assessment methods were seen as formative, as long as their direct purpose was to aid the students' learning process. Inferences might therefore be more applicable to various kinds of formative assessment and not only to certain sub-types of it. Because of their links to better academic performance (Ayllón et al., 2019; Richardson et al., 2012; Salanova et al., 2010; van Dinther et al., 2011) and greater well-being (Boulton et al., 2019; Pascarella et al., 2010), self-efficacy and student's engagement are the primary variables of interest. To explore their links with summative, as well as formative assessment, a questionnaire was administered to 210 psychology students. It featured a summative and a formative assessment condition and was used to compare how the participants' levels of self-efficacy and engagement varied depending on which assessment type was used. Additionally, the present study examines many types of the

participant's engagement with their programme, and not only with certain assessment methods, as it is common in the literature. For this reason, and because of the small body of evidence regarding self-efficacy and formative assessment in higher education, no hypotheses were formulated. The present study is therefore purely explorative in nature.

Methods

Participants

The participants were 210 bachelor psychology students from the University of Groningen. In that sample, 47 participants identified as male (22.3%), 161 as female (76.7%), one as "other", and one preferred not to say. Their mean age was 20.9 ± 5.0 years. Seventy-three participants were German (34.8%), 84 were Dutch (40.0%), and 53 had other nationalities (25.2%). For 156 participants (74.3%), this was their first programme at a university or college, while 54 (25.7%) indicated that they have studied in different programmes before. One hundred forty-three participants started their psychology bachelor's in 2021 (68.1%), eight in 2020 (3.8%), 27 in 2019 (13.0%), 26 in 2018 (12.4%), five in 2017 (2.4%), and one (0.3%) in 2016 or earlier.

Materials

To investigate the psychological states and behaviours of the participants, a self-constructed, online questionnaire was used (Qualtrics). Except otherwise noted, the answer options were always in the following five-point format: "Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly agree". The participants were provided with information about the study, an informed consent sheet and questions about their demographic characteristics.

They were also asked to give information about their general study habits, independent of

the assessment type. The six items of which the scale consisted of were self-constructed and focussed mostly on how the participant's study habits changed since leaving high school. One example is "I use the same study habits I have used in high school."

Furthermore, the present study investigated the participant's general procrastination behaviour (also independent of the assessment type) and possible reasons for that. The corresponding items for both scales were adapted from the Procrastination Assessment Scale - Students (Solomon & Rothblum, 1984). One scale measured the general procrastination behaviour and consisted of three items, such as "Procrastination on university activities is a problem for me.". The other scale, assessing possible reasons for procrastination, entailed nine items that were similar to "I tend to have too many other things to do". The answer options of the latter scale were adjusted to better suit the content of the items: "Not at all why I procrastinated", "Reflects a little", "Somewhat reflects", "Reflects a lot", "Definitely why I procrastinated".

The scales that are discussed next measured the respective constructs dependent on whether summative or formative assessment has been used (as opposed to the previously mentioned scales). This means that each scale was presented twice to the participants, once in the summative, and once in the formative assessment condition. Occasionally individual items were slightly altered to better suit the respective condition. For instance, a given item in the summative assessment condition might refer to exams, whereas the same item in the formative assessment condition might concern mandatory assignments.

The self-efficacy scale consisted of five items that were selected and adapted from the "Motivated Strategies for Learning Questionnaire" (Pintrich, 1991). Their content related to passing courses, understanding the course's material and mastering the skills that were taught. One example is "I am confident that I can understand the basic concepts taught.". All items are

displayed in Figure 1 in the results section.

The scale measuring engagement consisted of six items that concentrated mostly on the participants' interactions with their peers and professors, as well as on how much effort they put into certain courses. For instance, one item was "I regularly work with classmates on the material.". Some of the items were selected and adapted from a questionnaire measuring experiences of first-year university students in Australia (Krause & Coates, 2008), others were self-constructed, and one item was used from Schaufeli et al. (2002). All items are displayed in Figure 2 in the results section.

The "perceived sense of retention" scale entailed four self-constructed items. Their content related to remembering and understanding the general topic of courses, their central theories and concepts, as well as to the expectation of passing those courses. To illustrate, one item was "I tend to remember the general topic and learning goals in this type of course.".

The participant's learning approaches were measured with 10 items in the summative, and 11 items in the formative assessment part. The latter part included one extra item that was only concerning mandatory assignments, namely "The regular assignments help me structure.". The items were selected and adapted from the "Approaches and Study Skills Inventory for Students (*ASSIST*)" (Entwistle et al., 1997). Their content related to deep learning practices and the participant's satisfaction with their own learning approach. For instance, one item was: "It is important for me to follow arguments, or to see the reason behind contents.".

Additionally, the participant's procrastination behaviour in formative and summative assessment courses was measured. The scale included essentially the same three items that were also used to measure their tendency to procrastinate independent of the assessment type. However, they were slightly altered so that in the summative assessment condition, they related

to exams, whereas in the formative assessment condition, they concerned various mandatory tasks.

Another scale consisted of nine items that measured the participant's satisfaction with the respective type of course. The content focussed on various dimensions, such as whether the courses are interesting, enjoyable, well structured, or interactive. Seven of these items were adapted from the "Students Evaluation of Educational Quality" – Questionnaire (Coffey & Gibbs, 2001; Marsh, 1982), the remaining two were self-constructed. An example of one item is: "This type of course is challenging and interesting."

The participants were also asked to give information about their well-being with regards to certain courses. The content of the six self-constructed items concerned anxiety before exams, the stress level, and the workload. To illustrate, one item was: "The overall workload is too much."

Furthermore, the questionnaire asked the participants whether they cheated during exams or when working on assignments. To anonymise their answers, a coin toss method was used (Warner, 1965). The content of the items was inspired by the Academic Dishonesty Scale (Bashir & Bala, 2018). In both conditions, one item was about cheating in exams and asked the participants whether they engaged in any of the listed cheating behaviours (e.g., collaborating with others during an exam). In the summative assessment condition, the answer options were "Yes" and "No", while in the formative assessment condition, there was an additional "Not applicable" option in case the courses did not have exams. The latter condition also included one additional item that essentially asked the same question but with regards to assignments instead of exams. For this item, the answer options were "Yes" and "No". Before each item, the participants were asked to flip a coin by using the link to a website (*Just flip a coin*, 2010). In

case the coin flip resulted in tails, they were instructed to answer “Yes” regardless of whether this was true or not. If the coin showed heads, the participants were told to answer truthfully.

Procedure

The participants were recruited by the researchers via direct messages and messages in (large) group chats on social media (e.g., WhatsApp), as well as by using flyers. Additionally, first-year students had the opportunity to take part in this study in return for getting credits that were required to pass a mandatory course. The independent variable consisted of two levels: a formative and a summative assessment condition. The primary dependent variables of interest were self-efficacy and student engagement. The aim of this study was to investigate how the participant’s levels of self-efficacy and engagement vary depending on whether mandatory formative assessment types have been used or not. In the beginning of the questionnaire, the participants were given general information about the study, such as the involved researchers, or privacy concerns. Subsequently, it was necessary to consent to the conditions of the study before being able to answer the following part of the questionnaire. After that, the participants were asked for their gender, age, nationality, when they started their psychology programmes and whether it was their first college or university programme. The subsequent sections concerned general questions on study habits and procrastination behaviour, as well as possible reasons for procrastination, as described above. Then, the participants were informed that the rest of the questionnaire consisted out of two parts. One part assessed courses in which the final grade was only determined by an exam (or several partial exams) without any mandatory assignments. The other part concerned courses that included mandatory assignments (possibly in addition to an exam) throughout the courses that were supposed to aid the students’ learning or that were required to pass the course. The latter part was the formative assessment condition, while the

former part constituted the summative assessment condition. Subsequently, the participants were directed to the summative and formative assessment parts. The order of the parts was randomised. In the beginning of each part, the participants were asked to report their experiences with the specific type of course (summative or formative) in mind. Whenever there was a page break within one of the parts, the participants were reminded of what kind of courses they were supposed to be thinking of when answering the questionnaire. The order of the scales and questions was identical to the order in which they are discussed in the methods and results section of this paper. Prior to conducting the study, it was approved by the ethics committee of the University of Groningen.

Results

The responses of 46 participants were removed because they did not finish the questionnaire or did not give information about their demographics. Fourteen other responses were excluded since they were from the researchers to test the questionnaire. Additionally, responses of three participants were removed because they stated that they did not answer the questionnaire truthfully. Moreover, one response was excluded because the participant was part of the psychology minor and not bachelor programme. This resulted in a total sample size of 210. To test differences between the formative and summative conditions, as well as between students in different years, repeated measures ANOVAs were conducted. Since participants were recruited individually and there was no apparent reason or incentive for them to collaborate, the assumption of independent observations between subjects was assumed to be fulfilled. Moreover, the primary dependent variables self-efficacy and engagement deviated only slightly from normal distributions. Because of this and the large sample size, the ANOVAs were still conducted. To avoid chance capitalisation or a loss of statistical power, significance testing was

only done with the primary scales of interest, namely self-efficacy, and engagement. The alpha level of all conducted significance tests was five percent.

In order to measure the participant's general study habits, three items were combined, and the average was calculated. Higher scores indicated that the participants believed that since starting a university programme, their study habits have improved, possibly due to the help from the university or other students. The internal reliability of these three items was $\alpha = .39$. Their mean of it was 3.4 ($SD = 0.6$), indicating that the participants felt like their study habits had improved since starting a university programme, possibly because of the help from others. In the following, three single items were analysed, higher scores indicated that the participants agreed more strongly, lower scores that they rather disagreed, while the neutral condition ("Neither agree nor disagree") was coded as a three. The items were not included in the previous scale because their content does not relate to whether the participants thought that their study habits had improved since entering higher education. The mean score of 2.7 ($SD = 1.1$) of the item "I use the same study habits I have used in high school.", suggests that the participants felt a slight change in their study habits since leaving high school. The mean score of the item "I wish I could improve my study habits." was 3.8 ($SD = 1.0$), indicating that the participants felt a need to improve their study habits. Lastly, the mean of the item "I just memorize the material instead of trying to understand it." was 2.0 ($SD = 0.9$), suggesting that the participants tried to grasp the meaning of course material instead of merely memorising it.

General procrastination behaviour was measured by creating a variable that represented the average score of each participant on three items. Higher scores indicated more pronounced procrastination behaviour and a wish to decrease it. The mean score of this variable was 3.7 ($SD = 1.1$), the internal reliability $\alpha = .90$. Since a three was used to code the neutral condition

(“Neither agree nor disagree”), the data indicates that on average, the participants had a tendency to procrastinate and a wish to change this.

Moreover, the participants were asked to evaluate possible reasons for their procrastination behaviour (see Table 1). Higher scores indicated that the participants agreed more strongly, while lower scores suggested that they rather disagreed. The number three was coded as the neutral (“Neither agree nor disagree”) option.

Table 1

Possible reasons for procrastination behaviour

| Item | <i>M + SD</i> |
|--|---------------|
| I tend to have a hard time knowing what to study and what not to study. | 2.4 (1.1) |
| I tend to have too many other things to do. | 3.0 (1.1) |
| There tends to be some information I need to ask the professor, but I feel uncomfortable approaching them. | 1.7 (1.0) |
| I tend to be worried I get a bad grade. | 2.7 (1.4) |
| I really tend to dislike studying for exams. | 2.5 (1.1) |
| I tend to feel overwhelmed by the task. | 3.3 (1.3) |
| I tend to distrust myself to do a good job. | 2.7 (1.4) |
| I tend to lack the energy to begin studying. | 3.5 (1.3) |
| I tend to wait to see if the professor gives me some more information on the exam. | 1.9 (1.1) |

Note. List of single items of which the scale “Reasons for procrastination” consisted of.

The participants indicated that they knew what to study, while neither agreeing nor disagreeing as to whether they had too many other things to do. Also, most of them felt comfortable enough to approach the professor when they needed a piece of information. Moreover, the participants tended not to worry about receiving bad grades, while being fine with studying for exams.

However, the data suggests a tendency to feel overwhelmed by the given tasks and a lack of energy to start studying. Furthermore, merely a minority of participants procrastinated to receive

more information on the exam. Lastly, they tended to trust themselves to do a good job on given tasks.

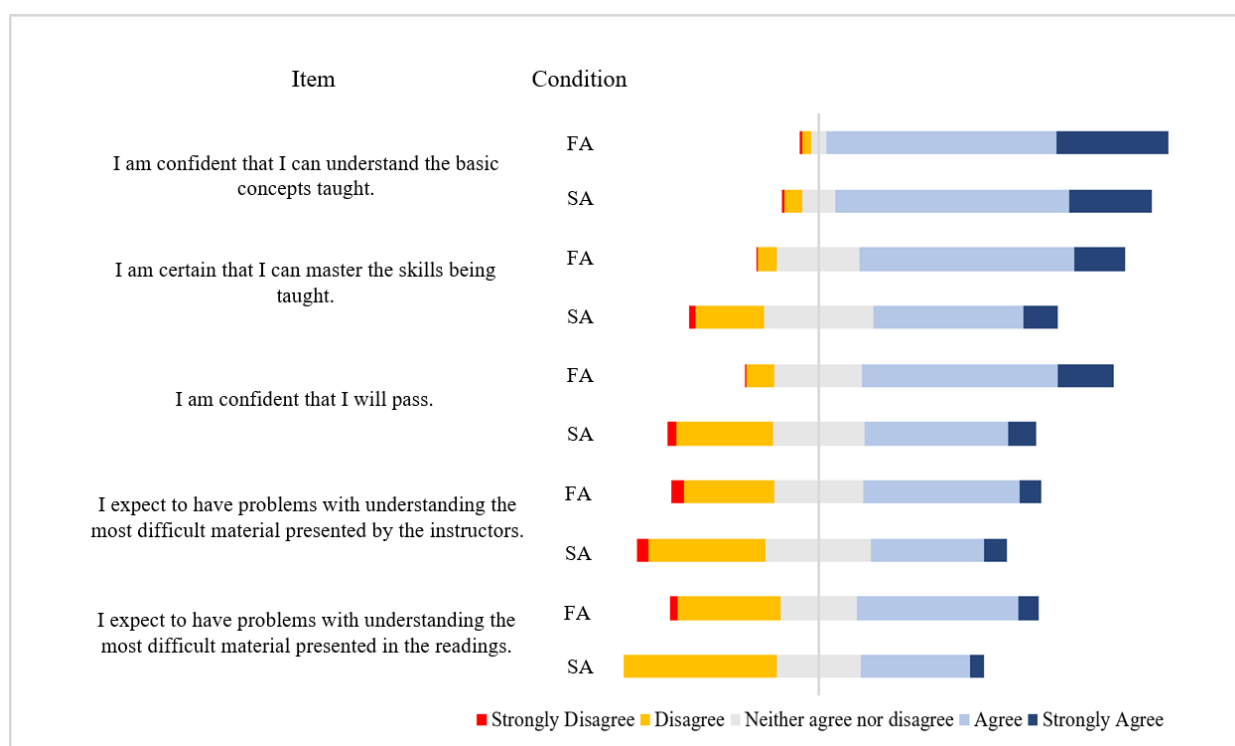
The following sections concern scales that were included in the summative, as well as in the formative assessment part. Hence, the respective items of which the scales consisted of were averaged in each part, resulting in two distinct variables that represented the participants' means on those items.

The self-efficacy scale consisted of five items, while two of those were reverse coded. Higher scores indicated higher self-efficacy. The internal reliability was $\alpha = .79$ in the summative, and $\alpha = .68$ in the formative assessment condition. The mean score in the former assessment condition was 3.3 ($SD = 0.7$), the one in the latter 3.6 ($SD = 0.6$). The participants thus reported being higher in self-efficacy when participating in formative, compared to summative assessment courses. A two-way repeated measures ANOVA revealed that this difference was statistically significant, $F(1, 208) = 47.38$, $MSE = 10.90$, $p < .001$, $d = .40$. Another aim of the present study was to investigate whether students from different year groups perceive summative and formative assessment types similarly. Because of that, the data set was split into two groups: first-year students ($n = 143$) and students from higher years combined ($n = 67$). The sample was not further divided into more year groups due to the (comparatively) small sample size of non-first-year students. The mean of the first-year students was 3.4 ($SD = 0.5$), the one of the non-first-year students 3.6 ($SD = 0.5$). The mean difference indicated that first-year students were lower in self-efficacy than students from higher years. The previously mentioned two-way repeated measures ANOVA showed that this difference was statistically significant, $F(1, 208) = 8.75$, $MSE = 4.84$, $p = .003$, $d = .40$. In contrast, the interaction effect between the assessment type and the study progress (meaning first-year student or not) was not statistically

significant, $F(1, 208) = 1.44$, $MSE = 0.33$, $p = .231$. In the following, the items of which the self-efficacy scale consisted of, were analysed and compared between the summative and formative assessment conditions. The exact percentages that are displayed in Figure 1 can be found in Appendix A.

Figure 1

Self-efficacy scores in formative and summative assessment condition



Note. FA = formative assessment; SA = summative assessment; The items “I expect to have problems with understanding the most difficult material presented in the readings.” and “I expect to have problems with understanding the most difficult material presented by the instructors.” were reverse coded.

In summative, as well as in formative assessment courses, almost all participants indicated that they were confident to understand the basic concepts taught. However, in the latter type of course, more participants reported that they were certainly able to master the skills that were taught. Similarly, in formative assessment courses, more participants indicated that they were confident to pass. Furthermore, in summative assessment courses, the participants indicated

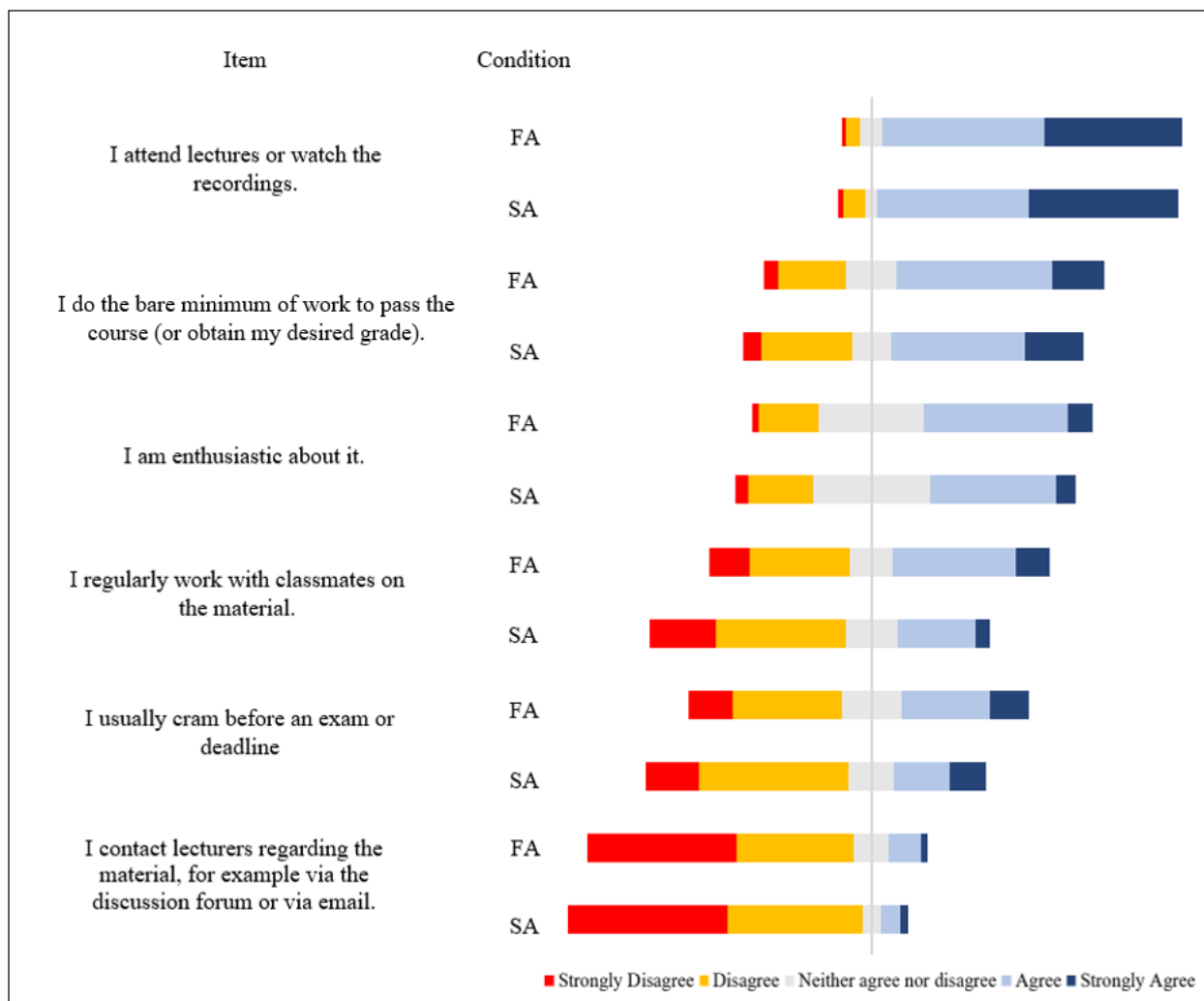
to expect having more problems with particularly difficult material presented by the instructors and in the readings.

The scale measuring student's engagement consisted of six items. Two of them were reverse coded. Higher scores indicated more engaged participants. The internal reliability was $\alpha = .56$ in the summative, and $\alpha = .48$ in the formative assessment condition. The mean in the former assessment condition was 3.0 ($SD = 0.6$), the one in the latter 3.2 ($SD = 0.6$). The data therefore suggested that the participants were more engaged in formative, rather than summative assessment courses. A two-way repeated measures ANOVA revealed that this difference was statistically significant, $F(1, 208) = 27.63$, $MSE = 4.24$, $p < .001$, $d = .35$. Similar to the analysis of self-efficacy, the effect of study year on engagement was also explored. The mean score of first-year students ($n = 143$) was 3.1 ($SD = 0.5$), the one from their peers from higher years ($n = 67$) 3.0 ($SD = 0.5$). The small difference in means suggested that the first-year students were slightly higher in self-efficacy than students from higher years. However, the aforementioned two-way repeated measures ANOVA revealed that this difference was non-significant, $F(1, 208) = 1.14$, $MSE = 0.60$, $p = .288$, $d = .20$. Furthermore, the interaction effect between the assessment type and the study year (first-year students or not) was also non-significant, $F(1, 208) = 2.36$, $MSE = 0.36$, $p = .126$.

The following section concerns the items of which the engagement scale consisted of. Figure 2 compares the participant's responses to individual items between the formative and summative assessment condition. The exact percentages that are displayed in Figure 2 can be found in Appendix B.

Figure 2

Engagement scores in formative and summative assessment condition



Note. FA = formative assessment; SA = summative assessment; The items “I usually cram before an exam” and “I do the bare minimum of work to pass the course (or obtain my desired grade).” were reverse coded.

In both conditions, the vast majority of participants indicated that they attended or watched the lectures, with almost no difference between summative and formative assessment courses. Furthermore, in the summative assessment condition, slightly more participants indicated that they only did the bare minimum of work to pass or obtain their desired grade.

However, the difference between both conditions was marginal. In contrast, in the formative assessment condition, more participants reported that they collaborated with classmates, compared to the summative assessment condition. Furthermore, in the latter condition, more participants stated that they tended to cram before exams, while slightly less participants reported being enthusiastic about the courses, relative to formative assessment classes. In both conditions, only a small minority of participants indicated to have contacted lecturers regarding the course material. However, slightly more did so in the summative, rather than the formative assessment condition.

Four items were used to measure the perceived sense of retention of the participants. Higher scores indicated a higher sense of perceived retention. In the summative assessment condition, the internal reliability was $\alpha = .68$, while being $\alpha = .59$ in the formative assessment condition. The mean in the formative assessment condition was 3.6 ($SD = 0.6$), the one in the summative assessment condition was 3.8 ($SD = 0.5$). The participants therefore reported that they felt like they remembered more from courses with formative, compared to courses with summative assessment.

The scale assessing “Learning approaches” was divided into two individual scales, one measured to what extent the participants engaged in deep learning, the other how satisfied they were with their learning approach in general. Furthermore, the part in the formative assessment condition included one extra item that was not included in the summative assessment condition.

The deep learning scale consisted of six items of which two were reverse coded. Higher scores indicated that the participants engaged more strongly in deep learning practices, such as finding out the meaning of given information, as opposed to merely memorising it. For this scale, Cronbach’s alpha was .59 in the summative, and .55 in the formative assessment condition. The

mean in the former condition was 3.2 ($SD = 0.8$), the one in the latter condition 3.4 ($SD = 0.6$). This suggested that in formative assessment courses, the participants tended to engage more strongly in deep learning practices, compared to summative assessment courses.

To measure how satisfied the participants were with their own learning approaches, four items constituted the corresponding scale. Two of those four items were reverse coded. Higher scores suggested that the participants tended to be more satisfied with their learning approach. The internal reliability was $\alpha = .77$ in the summative, and $\alpha = .72$ in the formative assessment condition. The mean in the former condition was 3.0 ($SD = 0.8$), the one in the latter condition was 3.2 ($SD = 0.8$). This indicated that in formative assessment courses, the participants were slightly more satisfied with their learning approaches than in summative assessment courses.

The item that was exclusively part of the formative assessment condition was: “The regular assignments help me structure.”. Higher scores indicated that the participants agreed more strongly with this item, while a three was used to code the neutral response (“Neither agree nor disagree”). The mean of this item was 3.8 ($SD = 1.0$), suggesting that a majority of the participants thought that regular assignments helped them to structure themselves.

To measure the participant’s procrastination behaviour, three items were combined and averaged. Higher scores indicated more pronounced procrastination behaviour and a wish to decrease it. The Cronbach’s alpha was .91 in the summative, and .90 in the formative assessment condition. The mean in the former assessment condition was 3.5 ($SD = 1.2$), the one in the latter 3.2 ($SD = 1.2$). The results therefore indicated that in summative assessment courses, the participants tended to procrastinate more, while having a stronger desire to change this, compared to formative assessment courses.

To investigate the participant's satisfaction, nine items were combined and averaged. Higher scores indicated greater satisfaction with the respective courses. The internal reliability was $\alpha = .79$ in the summative, and $\alpha = .78$ in the formative assessment condition. The mean in former assessment condition was 3.3 ($SD = .57$), the one in the latter was 3.8 ($SD = 0.5$). The data suggested that the participants were more satisfied with formative, rather than with summative assessment courses.

The participant's well-being was measured by combining and averaging six items, while five of those items were reverse coded. Higher scores indicated greater well-being, meaning that the participants felt more relaxed, less pressured in general, and also less anxious before exams. The Cronbach's alpha was .67 in both conditions. The mean in the summative assessment condition was 2.3 ($SD = 0.6$), the one in the formative assessment condition was 2.4 ($SD = 0.7$). Since the number three was used to code the neutral ("Neither agree nor disagree") answer, the data suggested that in both conditions, the participants tended to feel stressed and pressured, as well as anxious before exams. However, this was slightly more pronounced in the summative rather than the formative assessment condition. Furthermore, the well-being scale in the formative assessment part included one item that was not part of the summative assessment condition: "The mandatory assignments help me understand the course content.". The mean of this item was 3.9 ($SD = 0.8$), indicating that the participants considered mandatory assignments useful to better understand the course content.

Moreover, the cheating behaviour of the participants was investigated by using a randomised response method (Warner, 1965). As previously mentioned, the participants were asked to throw a coin, if heads came up, they were instructed to state the truth (either "Yes" or "No"), if tails came up, they were told to say "Yes" regardless of whether this was true or not.

Appendix C includes detailed information on how the actual numbers of cheaters were calculated. In the summative assessment condition, 10.5% of the participants reported that they cheated during an exam. In the formative assessment condition, 0% stated that they did so. Additionally, 6.2% of the participants indicated they cheated in the assignments in formative assessment courses. The data suggested that when participating summative assessment courses, the participants were more inclined to cheat, compared to when taking part in formative assessment courses.

Discussion

The goal of the present study was to give recommendations on how to improve programmes in higher education by reflecting on their assessment. It investigated differences in how bachelor psychology students perceived summative and formative assessment. The main variables of interest were students' engagement and their self-efficacy.

Self-efficacy

The results showed that the participants in this study were significantly higher in self-efficacy in formative, compared to summative assessment courses (medium effect size). This result is in line with the work of Lerdpornkulrat et al. (2019) who conducted a quasi-experiment and found that their additional formative assessment measure (rubrics for self-assessment) made their participants more self-efficacious. In contrast to that, den Boer et al. (2021) did not find a significant difference in self-efficacy between their formative and summative assessment condition. In their field-experiment, they compared two different groups with each other: for one group, several tests throughout the course contributed to the final grade (summative assessment), for the other group, the grade was only determined by the final exam (formative assessment).

One reason for why the previously mentioned results differ might be that formative assessment was operationalized differently. That is, in the present study, it can consist of various assessment types, such as quizzes or assignments throughout the course, as long as their direct purpose was to aid the students' learning. Meanwhile, in den Boer et al.'s (2021) study, the formative assessment method consisted of several tests throughout the course before a final exam. In Lerdpornkulrat et al.'s (2019) study, self-assessment with a rubric was considered the formative assessment method. Perhaps certain types of formative assessment affect self-efficacy differently than others do, which could explain the contradictory findings. Furthermore, all discussed studies (the present study; den Boer et al., 2021; Lerdpornkulrat et al., 2019) measured self-efficacy differently. To illustrate, self-efficacy in Lerdpornkulrat et al.'s study (2019) concerned information literacy skills, while in the present study, it related to academic achievement. Consequently, when comparing the results of different studies, one should be aware of the limited external validity due to the differences in formative assessment methods and the differences in how self-efficacy is defined and operationalized.

Besides investigating the effect of the assessment type on self-efficacy, the present study also explored how the study year in which the participants were in might have influenced their sense of self-efficacy. The results showed that first-year students (68.1%) scored significantly lower in self-efficacy than students from higher years (meaning all non-first-year students combined). The interaction effect between self-efficacy and the study year (first-year student or not) was non-significant. The latter means that the effect of the assessment type (summative and formative assessment) on self-efficacy was equal, regardless of whether the participants were first-year students or not. Conversely, the effect of study year on self-efficacy was also not affected by the assessment type. Reasons for the significant main effect (study year on self-

efficacy) might be that students from higher years are more experienced at studying than first-year students. Since these students have studied psychology for at least one year, they are more familiar with the assessment in this programme and because of this, might be more confident to pass courses and learn what they are supposed to learn (and more). This factor might be particularly important because a majority of the sample consisted of first-year students that just started their psychology programmes a couple of months prior to participating in this study (68.1%). This group might be more insecure about their abilities as to whether they are able to be successful at a university, especially if it is their first programme. Besides that, for many participants (74.8%), this was their first programme at a university or college which is why they might differ in certain characteristics, such as university experience, compared to students from higher years. Similarly, students who have participated in other programmes before are possibly more confident in their academic abilities because they are more experienced at for example, taking exams or because they are more familiar with studying independently. Thus, when generalizing the findings of the present study to other populations, one should consider that the majority of the sample consisted of first-year students (68.1%) and of students for whom this psychology bachelor was their first programme in higher education (74.8%).

Moreover, the participants felt more confident to pass formative, rather than summative assessment courses. In the former type of course, they also indicated having fewer problems understanding particularly difficult material presented by the instructors or in the readings. Furthermore, in formative assessment courses, more participants were certain that they could master the skills that were taught, compared to when participating in summative assessment courses. In both types of courses, most participants felt confident to understand the basic concepts taught, with only minor differences between summative and formative assessment.

Reasons for the mentioned differences might be that in formative assessment courses, assessment might occur more regularly than in summative assessment courses, in which knowledge is often tested by one final exam. Examples for this might be small tests throughout the course, as illustrated by den Boer et al.'s (2021) study. If such an assessment is employed, students are forced to think about the course material more regularly, compared to courses with only one final exam. This might help them understand more difficult course material better or to master the skills that are being taught. Consequently, students might be more confident to pass formative rather than summative assessment courses. This line of argumentation is in line with Bandura (1997) who explored various sources of self-efficacy and determined “enactive mastery experiences” as the most powerful. These experiences are characterised by mastering a skill, which leads to an increase in self-efficacy, while failing at doing so undermines it (Bandura, 1997). Since formative assessment aims at enhancing students’ future performances by giving them feedback (Harlen & James, 1997; McDowell et al., 2009; Yorke, 2003), it helps students to master skills. In contrast, summative assessment focusses on testing individuals’ knowledge or skills, without the direct intention of further improving them (Yorke, 2003). When students participate in formative assessment courses, they might therefore be more likely to experience enactive mastery experiences, which would increase their self-efficacy (Bandura, 1997).

Engagement

The second main variable of interest was student’s engagement, which was also significantly higher in formative than in summative assessment courses. This finding is in line with a study from Han & Finkelstein (2013) in which clickers (remote controls to answer questions in class) were used to answer questions that were posed during lectures. In case the system was used for grading purposes (summative assessment), students’ engagement with the

lecture and its material was lower than when it was used without grading purposes (formative assessment). Important to mention is that engagement has been differently operationalized; in Han & Finkelstein's (2013) study, it related more strongly to students being attentive in class, whereas in the present study, the focus was broader and mostly on apparent behavioural patterns like attending lectures, cramming, or collaborating with classmates. Additionally, and similar to the previously mentioned issue regarding self-efficacy, it is unclear which formative assessment types were used in the present study (e.g. Han & Finkelstein, 2013). These factors should be considered when comparing the findings of the present study and the ones from Han and Finkelstein (2013).

One explanation for higher levels of engagement in the present study might be that in formative assessment courses, the assessments took place more regularly than in summative assessment courses. For instance, if students are required to submit several small assignments throughout the course that build on the course's material, they are forced to engage it. This might influence them to watch the lectures or interact with their peers or professors because if they do not understand the material, they cannot complete the assignments.

In contrast to the results regarding self-efficacy, the participants' levels of engagement did not differ significantly depending on whether they were first-year students or not. Moreover, the interaction effect between the year in which the participants were in and the assessment type was non-significant. The latter means that the effect of the assessment type (summative and formative assessment) on the engagement scores was independent of whether the participants were first-year students or not. However, it also means that the (non-significant) effect of the study year on engagement was not dependent on the assessment type. One explanation for the non-significant main effect might be that students from higher years are higher in self-efficacy

than first-year students because they have more experience in higher education. However, both groups can be expected to have similar motives to be engaged with the programme, such as enjoying it or to pass courses, since there is no apparent reason for why this would be influenced by one's experience or age.

Furthermore, most participants reported that they attended the lectures or watched the recordings, while only a minority of them contacted lecturers regarding the material, with merely small differences between summative and formative assessment courses. Similarly, in formative assessment courses, slightly more participants felt enthusiastic about them, while slightly fewer reported to only put in the minimum amount of effort required to pass the courses. In contrast to that, substantially more participants stated that in formative assessment courses, they collaborated more frequently with classmates, while cramming less often before exams or deadlines.

Reasons for some of these results might be that in both types of courses, the students need to know the material that is presented in the lectures in order to pass the course, regardless of the assessment type, hence the negligible difference in whether the participants watched the lectures or not. Moreover, students in formative assessment courses might put in more effort than what is required to pass the respective course because they are more enthusiastic about them. An explanation for more collaboration in formative assessment courses might be that in some, students are required to work on their assignments in groups. Also possible is that more frequent assessments lead to more questions for the students that need to be answered in order to successfully complete them, which is why they might help each other out more often.

Future directions

One issue with the current body of evidence is that many studies regarding formative assessment and engagement, as well as formative assessment and self-efficacy were conducted in secondary schools, and not in higher education. Most students in higher education are older and more experienced than students in the school system, which could for example, affect their academic self-efficacy because they have overcome more academic obstacles yet, such as passing their final exams in secondary school. Also, students in higher education are expected to study more independently than in school, which might make them more engaged with their programmes because they themselves are responsible for acquiring certain skills or knowledge. For these reasons, more studies focussing on students in higher education are needed understand the links between formative and summative assessment and self-efficacy, as well as engagement.

Moreover, the comparisons of results from different studies regarding self-efficacy and engagement are often difficult because engagement is loosely defined, and both constructs (engagement and self-efficacy) can be operationalized in various ways. For example, engagement can be defined broadly, as in being engaged with the programme (e.g., Boulton et al., 2019) or narrowly, meaning being engaged with a certain assessment method, such as a virtual learning environment (e.g., Jacoby et al., 2014). Further exacerbating the problem is that in many studies, the items with which the constructs were measured were not made available (e.g., den Boer et al., 2021). Although it is unrealistic to expect all researchers to use the same instruments, more transparency would help to understand what exactly has been measured. Consequently, it would be easier to evaluate whether generalising certain findings is sensible or not.

Another issue with the current body of evidence is that only a few studies compare the outcomes of formative assessment with a control group (in the present study the summative

assessment condition) that does not experience the respective formative assessment measure. To evaluate the actual effect of formative assessment methods on certain outcomes, more studies using control groups are needed.

Conclusion

The present study compared courses that employed summative and formative assessment with regards to their links to students' self-efficacy and engagement. Both constructs were found to be significantly higher in formative, than in summative assessment courses, with medium effect sizes. Moreover, participants from higher years (meaning non-first-year students) were significantly higher in self-efficacy than first-year students. Conversely, no significant effect of the participants' study year on their levels of engagement was found. Moreover, there was no significant interaction effect between the study year and the assessment type, neither for self-efficacy nor for engagement. The findings complement the current body of evidence by directly comparing formative and summative assessment with each other. Moreover, by not determining particular formative assessment methods, the present study looks at formative assessment on a more abstract level than other studies do. For this reason, the inferences of this study relate more strongly to formative assessment in general and not only to certain sub-types of it. Similarly, students' engagement was also rather broadly defined, as in being engaged with the whole programme and not merely with a certain (formative) assessment method. The present study provides (correlational) evidence that formative assessment is associated with better outcomes regarding self-efficacy and engagement than summative assessment. Based on these results, institutions in higher education are advised to implement formative assessment in their curriculums.

References

- Ayllón, S., Alsina, Á., & Colomer, J. (2019). Teachers' involvement and students' self-efficacy: Keys to achievement in higher education. *PLOS ONE*, *14*(5), 1-11.
<https://doi.org/10.1371/journal.pone.0216865>
- Bandura, A. (1997). *Self-efficacy: The exercise of Control*. W. H. Freeman and Company.
- Bashir, H., & Bala, R. (2018). Development and validation of academic dishonesty scale (ADS): Presenting a multidimensional scale. *International Journal of Instruction*, *11*(2), 57-74.
- Biggs, J. (1996). Assessing Learning Quality: reconciling institutional, staff and educational demands. *Assessment & Evaluation in Higher Education*, *21*(1), 5-16.
<https://doi.org/10.1080/0260293960210101>
- 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>
- Boulton, C. A., Hughes, E., Kent, C., Smith, J. R., & Williams, H. T. P. (2019). Student engagement and wellbeing over time at a higher education institution. *PLOS ONE*, *14*(11), 1-20. <https://doi.org/10.1371/journal.pone.0225770>
- Coffey, M., & Gibbs, G. (2001). The Evaluation of the Student Evaluation of Educational Quality Questionnaire (SEEQ) in UK Higher Education. *Assessment & Evaluation in Higher Education*, *26*(1), 89-93. <https://doi.org/10.1080/02602930020022318>
- 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>

- Entwistle, N. J., McCune, V., & Tait, H. (1997). The approaches and study skills inventory for students (ASSIST). *Edinburgh: Centre for Research on Learning and Instruction, University of Edinburgh.*
- Gibbs, G., & Simpson, C. (2005). Conditions Under Which Assessment Supports Students' Learning. *Learning and Teaching in Higher Education*(1), 3-31.
- Han, J. H., & Finkelstein, A. (2013). Understanding the effects of professors' pedagogical development with Clicker Assessment and Feedback technologies and the impact on students' engagement and learning in higher education. *Computers & Education*, 65, 64-76. <https://doi.org/https://doi.org/10.1016/j.compedu.2013.02.002>
- Harlen, W., & James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379. <https://doi.org/10.1080/0969594970040304>
- Hewstone, M., Stroebe, W., & Jonas, K. (2016). *An Introduction to Social Psychology*. BPS Textbooks.
- Jacoby, C. J., Heugh, S., Bax, C., & Branford-White, C. (2014). Enhancing learning through formative assessment. *Innovations in Education and Teaching International*, 51(1), 72-83. <https://doi.org/10.1080/14703297.2013.771970>
- Just flip a coin*. (2010). <https://justflipacoin.com/>
- Kahn, P. E. (2014). Theorising student engagement in higher education. *British Educational Research Journal*, 40(6), 1005-1018. <https://doi.org/https://doi.org/10.1002/berj.3121>
- 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>
- Marsh, H. W. (1982). SEEQ: A reliable, valid and useful instrument for collecting student's evaluations of university teaching. *British journal of educational psychology*, 52(1), 77-95.
- McDowell, L., Sambell, K., & Davison, G. (2009). Assessment for learning: a brief history and review of terminology.
- Pascarella, E. T., Seifert, T. A., & Blaich, C. (2010). How Effective are the NSSE Benchmarks in Predicting Important Educational Outcomes? *Change: The Magazine of Higher Learning*, 42(1), 16-22. <https://doi.org/10.1080/00091380903449060>
- Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).
- 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/https://doi.apa.org/doi/10.1037/a0026838>
- Rushton, A. (2005). Formative assessment: a key to deep learning? *Med Teach*, 27(6), 509-513. <https://doi.org/10.1080/01421590500129159>
- Salanova, M., Schaufeli, W., Martínez, I., & Bresó, E. (2010). How obstacles and facilitators predict academic performance: the mediating role of study burnout and engagement. *Anxiety, Stress, & Coping*, 23(1), 53-70. <https://doi.org/10.1080/10615800802609965>

- 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>
- Solomon, L. J., & Rothblum, E. D. (1984). Academic procrastination: frequency and cognitive-behavioral correlates. *Journal of counseling psychology*, 31(4), 503. <https://doi.org/https://psycnet.apa.org/doi/10.1037/0022-0167.31.4.503>
- van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, 6(2), 95-108. <https://doi.org/https://doi.org/10.1016/j.edurev.2010.10.003>
- 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>
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher education*, 45(4), 477-501.

Appendix A

Table A1

Percentages of answers regarding self-efficacy scale in the summative assessment condition

| Item | Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly Agree |
|--------|-------------------|----------|----------------------------|-------|----------------|
| Item 1 | 2.4% | 26.2% | 24.8% | 39.0% | 7.6% |
| Item 2 | 3.8% | 29.5% | 22.9% | 41.4% | 2.4% |
| Item 3 | 0.5% | 4.8% | 9.0% | 63.3% | 22.4% |
| Item 4 | 6.2% | 30.5% | 28.6% | 31.9% | 2.9% |
| Item 5 | 1.9% | 18.6% | 29.5% | 41.0% | 9.0% |

Note. The content of the respective items is displayed below in Table A2.

Table A2

Self-efficacy items

| Item number | Content |
|-------------|--|
| 1 | I am confident that I will pass. |
| 2 | I expect to have problems with understanding the most difficult material presented in the readings. |
| 3 | I am confident that I can understand the basic concepts taught. |
| 4 | I expect to have problems with understanding the most difficult material presented by the instructors. |
| 5 | I am certain that I can master the skills being taught. |

Table A3*Percentages of answers regarding self-efficacy scale in the formative assessment condition*

| Item | Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly Agree |
|--------|-------------------|----------|----------------------------|-------|----------------|
| Item 1 | 0.5% | 7.6% | 23.8% | 52.9% | 15.2% |
| Item 2 | 5.7% | 43.8% | 20.5% | 28.1% | 1.9% |
| Item 3 | 0.5% | 2.4% | 4.3% | 62.4% | 30.5% |
| Item 4 | 5.7% | 42.4% | 24.3% | 24.3% | 3.3% |
| Item 5 | 0.5% | 5.2% | 22.4% | 58.1% | 13.8% |

Note. The content of the respective items is displayed above in Table A2.

Appendix B

Table B1

Percentages of answers regarding engagement scale in the summative assessment condition

| Item | Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly Agree |
|--------|-------------------|----------|----------------------------|-------|----------------|
| Item 1 | 3.8% | 19.0% | 34.3% | 37.1% | 5.7% |
| Item 2 | 17.1% | 39.5% | 11.4% | 26.7% | 5.2% |
| Item 3 | 19.5% | 38.1% | 15.2% | 22.9% | 4.3% |
| Item 4 | 11.0% | 16.2% | 13.3% | 43.8% | 15.7% |
| Item 5 | 1.4% | 6.7% | 3.3% | 44.8% | 43.8% |
| Item 6 | 47.1% | 39.5% | 5.2% | 5.7% | 2.4% |

Note. The content of the respective items is displayed below in Table B2.

Table B2

Engagement items

| Item number | Content |
|-------------|--|
| 1 | I am enthusiastic about it. |
| 2 | I do the bare minimum of work to pass the course (or obtain my desired grade). |
| 3 | I regularly work with classmates on the material. |
| 4 | I usually cram before an exam or deadline. |
| 5 | I attend lectures or watch the recordings. |
| 6 | I contact lecturers regarding the material, for example via the discussion forum or via email. |

Table B3

Percentages of answers regarding engagement scale in the formative assessment condition

| Item | Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly Agree |
|--------|-------------------|----------|----------------------------|-------|----------------|
| Item 1 | 1.9% | 17.6% | 31.0% | 42.4% | 7.1% |
| Item 2 | 15.2% | 45.7% | 14.8% | 20.0% | 4.3% |
| Item 3 | 11.9% | 29.5% | 12.4% | 36.2% | 10.0% |
| Item 4 | 11.4% | 26.2% | 17.6% | 31.9% | 12.9% |
| Item 5 | 1.0% | 4.3% | 6.7% | 47.6% | 40.5% |
| Item 6 | 43.8% | 34.3% | 10.5% | 9.5% | 1.9% |

Note. The content of the respective items is displayed above in Table B2.

Appendix C

Information on the calculations of the randomised response method

The percentage of participants who genuinely cheated was calculated by doubling the number of “No” responses and determining how many percent of the total amount of responses they constituted. Subsequently, this percentage was subtracted from 100% and the result represented the percentage of participants who indicated that they cheated. It was assumed that the chance of getting heads is equal to the chance of getting tails. For instance, in the summative assessment condition, 94 participants indicated “No” when being asked whether they cheated in exams. If that number is doubled, the result is 188, and 188 are 89.5% from 210. One hundred minus 89.5 are 10.5. Therefore, 10.5% of the participants reported that they cheated during exams in summative assessment courses. When calculating the percentage of participants who reported to have cheated in exams in formative assessment courses, the resulting percentage of was negative, possibly due to an unequal distribution of heads and tails. This result was considered to be 0%.