Incremental Grading in Practice: First Experiences in Higher Education

Christian Köppe
c.koppe@uu.nl
Utrecht University
Utrecht, The Netherlands

Roald Verhoeff
r.p.verhoeff@uu.nl
Utrecht University
Utrecht, The Netherlands

Wouter van Joolingen
w.r.vanjoolingen@uu.nl
Utrecht University
Utrecht, The Netherlands

ABSTRACT

Incremental Grading is a student-driven assessment approach where students have the responsibility to grade their own work based on pre-defined assessment criteria, usually rubrics. The desired outcomes of Incremental Grading are higher self-assessment skills, higher ownership of learning, lower degree of procrastination, and a more distributed workload for teachers.

The approach has been described as a pattern language in previous work and has now been applied in two courses for academic teaching. In this experience report we evaluate the effectiveness of Incremental Grading in these courses, using a mixed-method approach. The results show that Incremental Grading has a positive impact on the self-assessment skills of students, can positively affect the quality of their work and consequently their final grades, and makes teacher's feedback more valuable.

KEYWORDS

Assessment, Self-Grading, Educational Patterns, Incremental Grading

1 INTRODUCTION

In higher education, just as in many other areas of education, it is common that students work individually or in groups on larger assignments over a period of several weeks to months. Examples are writing assignments such as an essay, research projects, or design projects. These assignments are often subject to some of the following challenges:

- low self-assessment skills - students have difficulties determining the quality of their work and what grades they can expect(see e.g. [1, 12]),
- formative vs summative assessment - students tend to ignore feedback given in summative assessment, as the focus is

more on the grade they got than on what they still can learn from it. However, feedback related to formative assessment is used for improvement but sometimes not seen connected to grading, making it less relevant for grade-oriented students (see e.g. [17]),
- procrastination - some students start too late with working on the assignment, usually leading to lower grades (see e.g. [15]),
- low ownership of learning - many students are reactive when working on assignments, instead of playing an active role in their learning they mainly do what they think is expected from them and improvements are only done when indicated as necessary by the teacher's feedback (see e.g. [11]), or
- high workload peaks of assessors for grading - usually after the final deadline larger pieces of work need to be graded in a relatively short time (see e.g. [4], needed extra staff for marking because of class size).

Incremental Grading is an assessment approach which is intended to address the above mentioned challenges in a holistic way. It was designed during development of a semester on Object-Oriented Software Engineering [9] and is described as a *pattern language* [7]. Such a pattern language consists of a collection of interrelated educational design patterns – basically well-known and well-described generalized good practices. The patterns consist of important practice elements such as context, the problem to be solved, forces which influence the problem, the (generalized) solutions, positive and negative consequences, and multiple concrete examples of the solution application. The advantage of using educational design patterns and such a pattern language is that this allows for context-specific applications and configurations of individual patterns. It does not require a complete application of the full approach.

In the next section we describe the core of the approach and also refer to the underlying patterns, highlighted in Small Caps and summarized in the appendix. The full description of the approach can be found in [7].

1.1 The Pattern Language of Incremental Grading

The core idea of Incremental Grading is that students assess their own work using pre-defined assessment criteria, usually in the form of Rubrics. The rubrics are based on a more general Assessment Criteria List and the results of a Criteria Refinement. Students perform regular Work Self-Assessments. Whenever a student (or group of students when working on group assignments) believes he or she achieved a certain level according to the quality descriptions

Table 1: Overview of Incremental Grading core patterns

Pattern Name	Summary
ASSESSMENT CRITERIA LIST [3]	Clearly communicate to students what the criteria for assessment are.
CRITERIA REFINEMENT [3]	Refine assessment criteria to a detailed level.
EMBRACE CORRECTION [2]	Give the students the chance to improve their work.
GO FOR GOLD [8]	Encourage the students to continue improving their work, even—or especially—when they already acquired a sufficient grade for it.
GRADE IT AGAIN, SAM [2]	Permit your students to change and re-submit an assignment for re-evaluation and re-grading, after you have graded it and provided feedback.
GRADING DASHBOARD [8]	Present the current status of the student's grading in an overview dashboard.
GRADING QUEUE [8]	Use a task management board (e.g. Kanban board) for handling grading requests in a timely and transparent manner.
REPAIR IT YOURSELF [10]	Let students correct their wrong or incorrect solutions, so that they understand better how to do it right.
RUBRIC [3]	Define levels of quality per criterion, rate each refined criterion on a sheet, and aggregate the mark.
WORK SELF-ASSESSMENT [8]	Assessment criteria are given (e.g. as RUBRICS) and students rate themselves using these criteria
STUDENT-DRIVEN GRADING [8]	Give students the responsibility for determining the quality of their work and what the grade for this (part of) work is. Let them justify and provide evidence for the determined quality and the corresponding grades. When accurate, then the students earn the grades.
THIS IS FEEDBACK [16]	For learners to act on feedback they first need to recognise when it has been given.

in the rubric for one or more elements of their work, they can request a self-determined grade per element using the accompanying rubric criterion for that element (STUDENT-DRIVEN GRADING). Per requested grade, they have to provide an underpinning for how they believe they have fulfilled the requirements. This underpinning is an essential part of the approach, as it requires students to acquire necessary self-assessment skills and to pro-actively engage in a fair grading process.

Teachers get notified of new grading requests via a GRADING QUEUE and then have to handle these requests in a timely way. If the quality description in the rubric corresponding to the requested grade, the actual work product and the provided underpinning do match, then the requested grade is given and added to an overview of all achievements (a GRADING DASHBOARD) so that both students and teachers have grip on the student's progress. If more than one grade is requested in a single grading request (e.g. for separate elements of a report), then the result can also be that only some of these grades are accepted if the other parts do not fulfill the required matching.

The students can request grades whenever and as often as they want (until the final deadline), on new work products and also on assignment elements that were previously graded and have been improved or corrected (EMBRACE CORRECTION, REPAIR IT YOURSELF and GRADE IT AGAIN, SAM). This way, the requests also serve as feedback (THIS IS FEEDBACK) and help the students with directing their own learning. Students can also be encouraged to improve their work even though they already achieved a passing grade (GO FOR GOLD). Optionally, there can be one or more interim deadlines where grading requests are required, so students *have to* perform self-assessments and self-grading.

Table 1 provides summaries of all core patterns of Incremental Grading. Figure 1 gives a visual overview of the core patterns and their relations.

1.2 Research Questions

This work is part of a larger research project with the goal of exploring the effects of the application of Incremental Grading as assessment approach in academic teaching courses. In this experience report we describe how Incremental Grading has been applied in two courses at the Faculty of Science at Utrecht University and evaluate its effectiveness regarding the impact on some of the above mentioned challenges.

In order to look more specifically at certain aspects of Incremental Grading, we formulated four research questions:

RQ1: To what extent does Incremental Grading affect students' self-assessment skills?

RQ2: How do students value the Incremental Grading approach?

RQ3: How do teachers value the Incremental Grading approach?

RQ4: How does the application of Incremental Grading affect the final assignment results?

Please note that in this first experience report we did not look at individual effects of the applied patterns, but instead looked at the approach as a whole. We therefore do not know which and how patterns and specific pattern implementation affected the outcomes. This will be part of future work.

In the next section we will describe the research method applied for answering the research questions. Section 3 comprises an overview of the two courses and how Incremental Grading was applied as assessment approach in these courses. The results of the application are presented next, followed by a discussion and conclusion, including a view on future work.

2 RESEARCH METHOD

This study is part of a larger design-based research project on Incremental Grading. The goal of this research was two-fold. Firstly, we wanted to validate the approach in its current design with respect to the expected positive effects. Secondly, we were also interested in gathering information that helps with improving the design of the Incremental Grading for future applications.

To achieve these goals, we applied a mixed-method/multiple-sources approach [6] in order to find answers to the research questions.

We examined students' perceptions of various aspects of Incremental Grading such as (perceived) self-assessment skills, effects on product quality and general appreciation of the approach using a questionnaire. It contained five closed questions with a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) and three open questions that were used to collect more qualitative
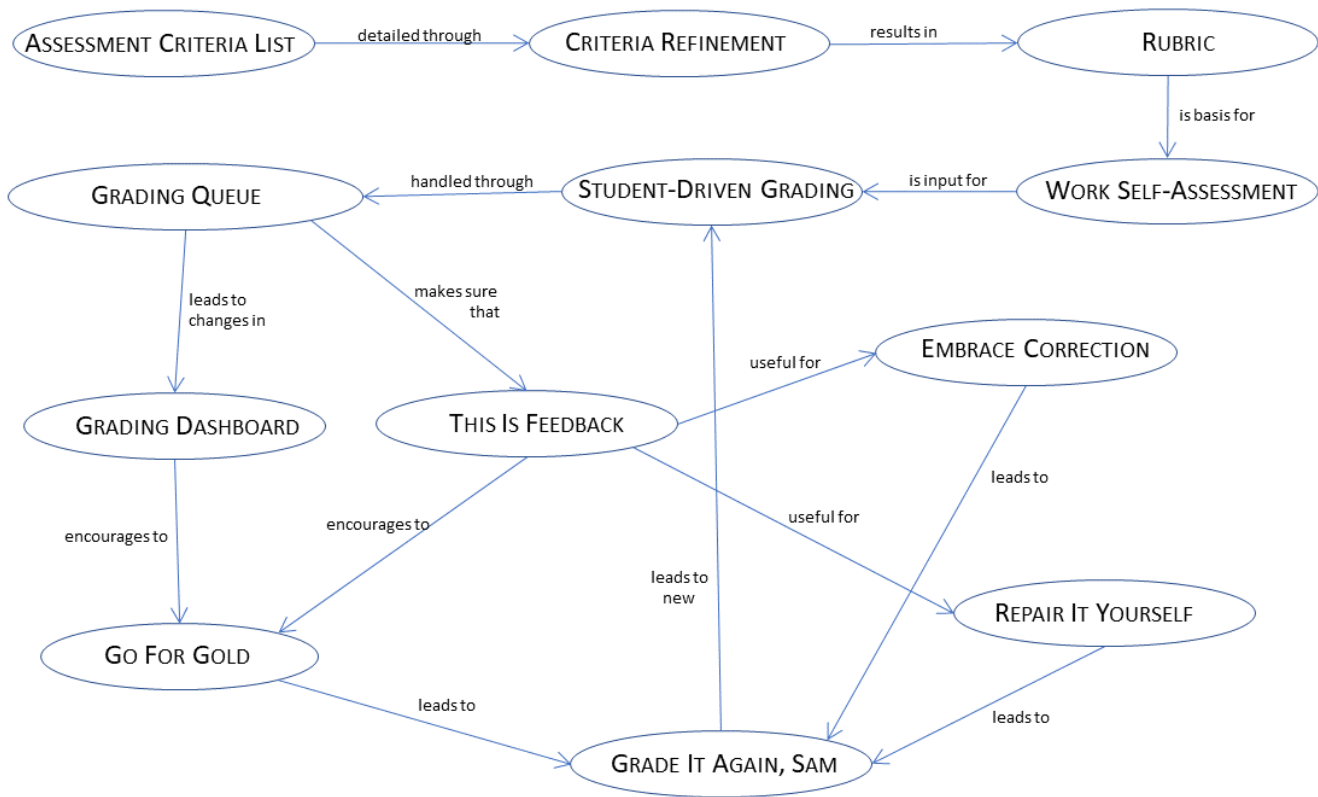
Figure 1: Overview of core elements (patterns) of Incremental Grading (adapted from [Köppe et al. 2019])

data. We applied provisional coding [14] for analysis of the student responses as the open questions were of explorative character and there was no existing framework. We established a start set of codes (see Figure 2) for both positive and negative aspects prior to the analysis. We based this set on concepts that were expected to come up in the answers and grouped it into four categories. When, during analysis, answers did not fit into the existing set, we extended the set of codes with these answers. The complete questionnaire is illustrated in appendix A.

We gathered quantitative data on how the students had applied Incremental Grading through an analysis of all grading requests. For this analysis, we extracted the data from the learning management system (LMS), including date of request, requested grade per rubric, and acceptance of grade request. These data were used for calculating various aggregated data, i.e. percentage of accepted grading requests (in general and per first and second half of assignment), percentage of unnecessary improvements, or the difference between handling grading requests after they have been declined by the teacher (improving work and request same grade *or* request lower grade). The results of the grading request analysis were also combined with the final assignment results of the students or student groups.

We gathered qualitative data on the teacher's point of view using semi-structured interviews. The interviews were transcribed and coded using the same approach as for the analysis of the open questions of the questionnaire.

Finally, we examined the open questions from the standard course evaluations. Answers which included information of relevance to the research questions were coded and included in the data set.

2.1 Variables

We defined the following variables which were used to gather data in order to answer the research questions.

RQ1: To what extent does Incremental Grading affect students' self-assessment skills?

- V1: percentage of accepted vs submitted grading requests per rubric and student/group in first and second half of all grading requests (independent of moment, e.g. if six grading requests were submitted, we looked at the first three as first half and the last three as second half; if there was an uneven number of requests, we omitted the middle one)
- V2: students' perceived improvement of their self-assessment skills

RQ2: How do students value the Incremental Grading approach?

- V3: general appreciation of the approach
- V4: most valued aspects of approach
- V5: least valued aspects of approach

Start set of codes

- Feedback
    - o F1: timing
    - o F2: quality
    - o F3: opportunity
- Structure/functioning
    - o S1: regular actions
    - o S2: motivation
    - o S3: self-directed learning
    - o S4: amount of work
    - o S5: guidance
- Self-assessment
    - o SA1: practicing self-assessment
    - o SA2: awareness of goal
    - o SA3: extra thinking over product
    - o SA4: having to be critical
    - o SA5: given criteria usable for evaluation
    - o SA6: priorities
    - o SA7: reasoning
    - o SA8: progress overview
- Impact
    - o I1: on quality
    - o I2: on marks

Figure 2: Start set of codes for analysis of open questions

RQ3: How do teachers value the Incremental Grading approach?

- V6: general appreciation of the approach
- V7: most valued aspects of approach
- V8: least valued aspects of approach

RQ4: How does the application of Incremental Grading affect the final assignment results?

- V9: characteristics of submitted grading requests that have a correlation with higher final results (such as many/few grading requests, many/few requested grades per grading request, working with increments etc.).
- V10: students' perceived effect of Incremental Grading application on the quality of their own work.

2.2 Data Collection

The following sources were used for gathering the data:

- The grading requests as stored in the used LMS, the relevant data such as number of requests, timing, and percentage of accepted grading requests per moment were extracted and analyzed by hand. (Vars V1, V6)
- A questionnaire on various aspects of Incremental Grading, taken at the end of the course (or shortly before). It contained 5 questions using a 5-point Likert scale and 3 open questions. The answers of the open questions were coded according to the criteria described below. The questionnaire can be found in appendix x. (Vars V2-V5, V7)

- Semi-structured interviews with lecturers after the course execution. (Vars V3-V5)
- Standard course evaluations where the students had the opportunity to provide feedback on the application of Incremental Grading. (Vars V3-V5)

The data were stored according to the guidelines of the Freudenthal Institute/Science Faculty of Utrecht University.

2.3 Ethical Considerations

We informed the students at the beginning of the courses that data on the application of Incremental Grading will be anonymously collected with the primary goal of improving the course. Furthermore, we informed them that the first author of this work also will use part of the anonymized data for research purposes. All students had the opportunity to opt-out of collection and use of their data.

Most students of the second course "Assessment and Evaluation" additionally provided an explicit informed consent on the usage of their anonymized data for research purposes. This gave us the possibility to combine the results of the questionnaire with usage data of Incremental Grading (such as number and quality of grading requests).

At the moment of data gathering, an ethics review board at the faculty was still in the process of being set up. Consequently, this project has not been approved by an ethics review board. However, the procedure as described above complied with the ethics standards of the Freudenthal Institute at that time.

3 APPLICATION CONTEXTS

Both courses were given at the Faculty of Science at Utrecht University as part of teacher education at Masters level. Both courses used Blackboard as LMS. Table 2 shows an overview of how the core patterns of Incremental Grading were applied in the two courses. It is notable that most elements of Incremental Grading were implemented in a similar way. However, there are some differences which might also have impact on the results: the first course made use of fixed interim deadlines and a separate assignment in the LMS per deadline, while the second course used one final deadline and one assignment in the LMS, allowing multiple grading requests as new attempts.

Both courses defined rubrics containing the three essential features as summarized in [5]: evaluative criteria, quality definitions for those criteria at particular levels, and a scoring strategy. They were used for *analytic grading*. This means that on each of a preset of criteria, qualitative judgements are made which afterwards get aggregated into a grade [13]. Figure 3 shows an exemplary exempt of the applied rubric.

3.1 Course: "Communicating Science with the Public"

The assignment in this course for which Incremental Grading was applied was the production of an educational video for secondary school, counting for 35% of the final grade. The students had to identify/define all relevant information such as aim, target group, topic, storyboard etc. and also produce the final video. The main focus was on the design process of the video and less on the final

Table 2: Overview of how the Incremental Grading patterns were implemented per course

Pattern	Implementation in course: "Communicating Science with the Public"	Implementation in course: "Assessment and Evaluation"
Assessment Criteria List & Criteria Refinement	7 weighted criteria	8 weighted criteria
Rubric	5 quality levels with corresponding grades (1,4,6,8,10)[1] per criterion	
Work Self-Assessment	Self-assessment encouraged by teacher, but regularly necessary due to interim deadlines	self-assessment was regularly encouraged by teacher
Student-Driven Grading	interim deadlines in weeks 2, 4, 6, 7, 8 and final deadline in week 9, handing in a grading request was required additional to product delivery, per deadline was a new assignment in Blackboard	only a final deadline in week 9 and no required interim deadlines, handing in grading requests was not required, but encouraged several times (in the beginning and during the course), one assignment in Blackboard, every grading request was a new attempt
Grading Queue	part of Blackboard's Grade Center (Needs Grading), not visible for students	
Embrace Correction	part of feedback given as response to grading requests	
Grading Dashboard	was only visible to students when looking at the feedback of their latest grading request in Blackboard, not easy to find for students	
Repair It Yourself	students were encouraged to use the feedback for improving their work	
Go For Gold	implicitly (by allowing improvements until the final deadline) and explicitly as part of feedback on grading requests	
Grade It Again, Sam	new grading requests were allowed for all assignment parts	

Figure 3: Exempt from rubric for video assignment in course "Communicating Science with the Public"

video artifact. The assignment duration was 10 weeks in total. Students had to work in groups of three. The course had two other assignments which were handled without Incremental Grading: writing a popular science text (counting for 55% of final grade) and developing a strategy for science communication in social media (10% of final grade).

3.2 Course: "Assessment and Evaluation"

All participants in this course were already working as teachers, this course was taken for teacher professional development. The assignment Incremental Grading was applied for was the design, execution, and evaluation of a larger summative assessment for a course they are giving at their own work/school. This was usually a written exam, a writing assignment or a larger practical. It counted for 50% of the final grade. The assignment duration was 10 weeks in total and was done individually. Apart from this assignment there also was a written exam at the end of the course, which counted for the other 50% of the final grade.

4 RESULTS

In this section, we present the results per course/assignment. Possible differences in the results per course and how these possibly relate to the ways Incremental Grading has been applied will be discussed in section 5.

4.1 Course: "Communicating Science with the Public"

The questionnaire was filled out by 29 students (out of 33 enrolled). Table 3 shows the results of the five closed questions, the coded answers of the three open questions are presented in the relevant variable sections. The standard course evaluation was done by 15 students.

We analyzed the characteristics of the submitted grading requests and the final grades with SPSS (version 26, IBM statistics). As the grading requests were handed in per group, the data-set was relatively small (N=11) and we did not find any significant correlations. However, we did see a few trends which might be worth exploring in future work. These are included in the relevant sections below.

Variable V1 (percentage accepted grading requests): A total of 56 grading requests were handed in by 11 groups. The average number of grading requests per group was 5. The average number of covered criteria dimensions of the rubric was 4. The acceptance rates of grading requests (per rubric/criterion) were 14% for the first half of the grading requests and 59% for the second half, the latter being significantly higher .

Variable V2 (students' perceived improvement of their self-assessment skills): 69% of the students (20/29) perceived that they were able to assess their own work based on the rubrics. 76% of the students (22/29) had the feeling that because of Incremental Grading their self-assessment ability had improved.

Table 3: Answers to first five questions in course: "Communicating Science with the Public", 5-point Likert scale with options strongly disagree (sd), disagree (d), neutral (n), agree (a), and strongly agree (sa); N=29

	sd	d	n	a	sa
Q1: Based on the rubrics, I was able to assess my own work well.	0	1	8	14	6
Q2: Through the self-grading approach, my ability to assess my own work has improved during the course.	0	4	3	21	1
Q3: I think that I would have achieved the same product quality without the self-grading approach.	2	13	9	3	2
Q4: Having to provide a motivation for the requested grade was unnecessary extra work.	0	11	10	6	2
Q5: I'd like to see this approach (self-grading with the possibility of improving the work) applied in other course assignments as well.	6	9	4	8	2

Students' responses in the questionnaire: Seven students out of 29 stated that they learned to be more critical on their own work, 3 students stated that they look at their product in a different way now, and 3 students learned that it helps to know expectations and aims. The following four responses on what they learned were each given twice: how to apply rubrics for their own work, to be more precise, to clearly distinguish between parts, and to identify missing or important parts.

Variable V3 (general appreciation of the approach by students): Only 34% of the students (10/29) would like to see the approach in other course assignments as well, 52% disagreed (15/29). This means that most of the students did not value the approach or were neutral about it.

Variable V4 (most valued aspects of approach by students): In the standard course evaluation, two students valued the amount of feedback which helped them learn and make progress, one student mentioned the interim feedback moments and one valued that it was well organized.

In the questionnaire, 28% of the students (8/29) saw the required motivation for the requested grade as unnecessary, 34% were neutral (10/29) and 38% disagreed (11/29, no strongly disagree). This means that only a small majority of students perceived the required underpinning as valuable for their progress.

Students' responses in the questionnaire on what they valued about the self-grading approach (as applied in this course): getting guidance (10 responses), having overview of their progress (7), having to be critical (7), being aware of the goal (6) and having to think more deeply about the product (5). They furthermore stated that they valued usable criteria for evaluation, having to provide a motivation, and that it has positive impact on quality and marks (each 3 responses). Two responses each valued priority setting and the regularity of actions.

Variable V5 (least valued aspects of approach by students): In the standard course evaluation, six students stated it required too much effort and time, also because this assignment only counted for 35% of the final grade. Two mentioned that the way the rubrics should be used was introduced too late and one student stated that there were too few moments to discuss grading with teacher directly.

Students' responses in the questionnaire on what they did not value about the approach (as applied in this course): by far most responses (16) mentioned the high amount of time/work it requires. Six responses described the underpinning as unnecessary work and three responses stated that the effort for the underpinning was comparatively high compared to the product. The structure and content for self-grading was unclear for six respondents (at least in the beginning). Four students found it difficult to determine the quality of their work. Two students each stated that there was too much guidance, the criteria were unclear, there was a repetition in the motivation, and the approach has no or a negative impact on the product quality.

Variable V6 (general appreciation of the approach by teacher): According to the teacher, Incremental Grading did fit the assignment and was valuable.

Variable V7 (most valued aspects of approach by teacher): Positive aspects for the teacher are clearer expectations after prior rework of rubrics, more objective assessments, as there was no comparing/ordering of groups, and also that there were no discussions on the acknowledged grades. The teacher also valued the positive impact on students: they are more focused on the right spots (more than in present instances of the course), they have to think more thoroughly because of the required underpinning, and they get more control on their own work and value that. She also mentioned that students try to get all points, working towards the highest possible grade, even though they recognize that it is a lot of work.

Variable V8 (least valued aspects of approach by teacher): The teacher stated that it was much work for her. Regarding the students she stated that some students found it very difficult (self-grading with underpinning), which leads to frustration and much work, especially to understand what is expected from them. Some students also used the rubrics as a kind of checklist, which might not leave much room for creativity. Finally, students felt that it was now too much work for an assignment which only counted for 35% of the final grade. This will be adjusted in the next instance of the course.

Variable V9 (characteristics of submitted grading requests that have a correlation with higher final results): As the questionnaire was filled in anonymously due to organizational reasons, we were not able to link the results to the grading request characteristics and the final grades. All student groups handed their first grading request in at a fixed deadline, therefore looking at the influence of the moment of first application on the final grades does not make sense.

In this course, the number of grading requests had no influence on the final grade (r=-0.151 for complete grading requests and r=-0,064 for number per rubric). The number of immediately approved grading requests (an indication for good self-assessment skills) might correlate with higher final grades (r=0.564, not significant).

The amount of unnecessary improvements (after a passing grade had been given for an element) seems to be higher with low performing groups (r=-0.487, not significant). Improving work and re-requesting the same or a higher grade tends to be in line with higher final grades (r=0.426, not significant).

Variable V10 (students' perceived effect on quality of their own work): 17% of the students (5/29) agreed that they would have achieved

the same product quality without self-grading, 31% were neutral (9/29), and 52% did disagree or strongly disagree (15/29). So half of the participants think that self-grading had a positive impact on the quality of their delivered products.

4.2 Course: "Assessment and Evaluation"

The questionnaire was filled in by 18 students out of 26 enrolled (N=18). Table 4 shows the results of the five closed questions, the coded answers of the three open questions are presented in the relevant variable sections. 15 of the 18 students finished the assignment. Besides that, five more students finished the assignment who haven't answered the questionnaire and haven't given the explicit informed consent. Therefore, the data of these students were only used in the aggregated data of the whole group. One student only answered questions 3, 4, 7 and 8 of the questionnaire as he hasn't applied Incremental Grading yet at the moment of filling in the questionnaire, so for all other questions N=17.

The standard course evaluation was done by 16 students.

We analyzed the characteristics of the submitted grading requests and the final grades with SPSS (version 26, IBM statistics). However, we did see a few trends which were not significant, but might be worth exploring in future work. These are also included in the relevant sections below.

Table 4: Answers to first five questions in course: "Assessment and Evaluation", 5-point Likert scale with options strongly disagree (sd), disagree (d), neutral (n), agree (a), and strongly agree (sa); N=17 for Q1, Q2, and Q5; N=18 for Q3 and Q4

	sd	d	n	a	sa
Q1: Based on the rubrics, I was able to assess my own work well.	0	1	4	7	5
Q2: Through the self-grading approach, my ability to assess my own work has improved during the course.	0	2	6	6	3
Q3: I think that I would have achieved the same product quality without the self-grading approach.	3	8	4	1	2
Q4: Having to provide a motivation for the requested grade was unnecessary extra work.	0	5	4	6	3
Q5: I'd like to see this approach (incremental grading with the possibility of improving the work) applied in other course assignments as well.	0	0	2	10	5

Variable V1 (percentage accepted grading requests): In total 65 grading requests were handed in by 18 groups. The average number of grading requests per student was 3.6 (with a minimum of 1 and a maximum of 9) and covered on average 3.6 criteria dimensions of the rubric. The acceptance rates of grading requests (per rubric/criterion) were 24% for the first half and with 59% significantly higher for the second half.

The analysis of the grading requests also shows that there is a correlation between starting early with handing in grading requests and having initial grading requests approved (-0.514 at 0.05 level). Formulated differently, the earlier they start the more they get correct the first time per rubric.

Variable V2 (students' perceived improvement of their self-assessment skills): 71% of the students (12/17) experienced that they were able to assess their own work based on the rubrics. 53% of the students

(9/17) had the feeling that because of Incremental Grading their self-assessment ability had improved, 35% were neutral (6/17), and 12% disagreed (2/17).

Variable V3 (general appreciation of the approach by students): 88% of the students (15/17) would like to see the approach in other course assignments as well, 12% were neutral (2/17) and no one disagreed. This means that the students of this course valued the approach.

The analysis of the grading requests also shows that there is a correlation between starting early with handing in grading requests and valuing the approach (r=-0.710 at 0.01 level). So, even most students did value the approach, the ones who started early valued it even more, an indication that making early experience with the process has a positive impact.

Variable V4 (most valued aspects of approach by students): In the standard course evaluation, 57% stated that there were (strongly) satisfied with the way Incremental Grading is performed, 21% disagreed. 93% stated that the assessment criteria were clear. Other students mentioned as positive aspects that Incremental Grading challenged them, is fine because everything was known upfront, they got regular feedback and the communication was good.

Students' responses in the questionnaire on what they valued about the approach (as applied in this course): the possibility for improvement (6 responses), the positive impact on quality and marks (5), a higher personal motivation (4), and the quality of feedback (3). Two times each mentioned were the feedback timing, the opportunity for getting feedback, and the removal of stress.

Variable V5 (least valued aspects of approach by students): In the standard course evaluation, one student mentioned that it was not clear what happens when your own assessment is too high and what (and if a) grade is given. Another student mentioned that applying Incremental Grading costs a lot of time.

Students' responses in the questionnaire on what they did not value about the approach (as applied in this course): three students mentioned providing the motivation was unnecessary. Two times each, students mentioned the missing of interim deadlines (as motivator), the high amount of work/time it requires, unclear criteria, a repetition when providing the motivation, and the activity of grading oneself. One student mentioned getting no mark when requested grade was too high as frustrating.

50% of the students (9/18) saw the required underpinning for the requested grade as unnecessary, 22% were neutral (4/18) and 28% disagreed (5/18, no strongly disagree). This means that most of the students did not perceive the required underpinning as valuable.

Variable V6 (general appreciation of the approach by teacher): The teacher stated in the interview that she is enthusiastic about the approach. An assignment like this (comprising different elements) fits well for Incremental Grading. She also stated that if more students would have used Incremental Grading, the peak at the end would have been smaller. Students need to be pulled into the process to make them enthusiast too, initial doubts about Incremental Grading changed often after having applied it.

Variable V7 (most valued aspects of approach by teacher): The teacher was positive about that students naturally used the feedback, that

there we no real peak moments besides a peak at the end (mainly caused by the students who did not apply Incremental Grading much) and that the work load was more or less evenly distributed. Furthermore, she felt that the quality of work has for some students strongly improved.

Positive aspects can also be found in what have been learned by the teacher when applying Incremental Grading. One of the biggest effects was on the used rubrics. During application it became clear that the rubrics had to be partially refined and that missing aspects, such as requiring that the end product consists of a coherent whole, need to be added. Weights will be added to the rubrics, as some elements turned out to be more important than others. Having students using the rubrics for underpinning of their grading requests also changed (or explicated) the interpretation of the criteria by the teacher.

Variable V8 (least valued aspects of approach by teacher): Negative effects, as stated by the teacher, were that feedback was mainly used as checklist and led only to concrete improvement, but less to more global quality improvements, making the end products more a collection of elements than a complete whole. She also saw disadvantages for the students who do not make use of Incremental Grading, as expectations from the teacher were for all the same, but these students were not aware of these expectations because of the missing feedback. Finally, it costs more time, as giving a lot of feedback also means to check often again if they have used the feedback. Besides that, a structure is needed such as reacting on each grading request within one week and it should not take more than one hour to handle it (depending on the size of the grading request). This means sufficient time needed for timely handling of the requests should be available.

Variable V9 (characteristics of submitted grading requests that have correlation with final results): Not unexpected, the number of declined grading requests correlates with lower final grades (r=-0.545 at 0.05 level). The same is valid for the percentage of declined grading requests (r=-0.710 at 0.01 level). Interestingly, an increase in correctness of grading requests does not have any correlation with the final grade (r=-0.008). The number of repeatedly declined grading requests for the same rubric elements correlates negatively with the final grade (r=-0.614 at 0.01 level). Also, students requesting a lower grade after an initial decline ended with lower final grades (r=-0.530 at 0.05 level).

Interestingly, there is no correlation between the moment of first grading request and the final grade (r=-0.042). This means that starting early does not directly impact the final grade.

The students who indicated that they would have achieved the same product quality without Incremental Grading did also submit the lowest numbers of grading requests (in general and per rubric, r=-0.542 and r=-0.493 respectively, both significant at 0.05 level). Interestingly, there is a trend that these students end with lower final grades (r=-0.475, not significant).

It is remarkable that the student with the most grading requests per rubric had all of them declined, while the average declining percentage is 49%. The reason for this is not clear, but the observation of the teacher was that this student "did not get the approach" and was not very open for feedback. However, if we interpret this

student as outlier and check for correlations again, then the numbers of grading requests (in general and per rubric) significantly correlate with the final grades (r=0.487 at 0.05 level and r=0.593 at 0.01 level, respectively). This shows that in most cases more usage of Incremental Grading (with a higher number of grading requests) correlates with higher final results.

Variable V10 (students' perceived effect on quality of their own work): 17% of the students (3/18) think that they would have achieved the same product quality without self-grading, 22% are neutral (4/18) and 61% disagree or strongly disagree (11/18). The majority perceived a positive effect on the quality of their own work.

5 DISCUSSION

In this section we discuss the results and answer the research questions. As we base the results only on experiences with two courses, the answers are more indicators of trends. To provide more validated answers, we will apply the approach in larger scale in future work.

RQ1: How does Incremental Grading affect the students' self-assessment skills?

In both courses, there's a significant difference of acceptance rates in the first and in the second half of the assignment period (14% vs 59% and 24% vs 59%). This shows that the percentage of approved grading requests improved from the first to the second half during both courses. In the course without interim deadlines it also shows that the earlier they start with handing in grading requests, the better they become in self-grading. The majority of students stated that they were able to self-assess their work well and that their self-assessment skills have improved.

Taking these data into account, the research question can be answered as follows: In our study Incremental Grading had a positive effect on the student's self-assessment skills. The correctness of self-assessments increased during the course which was also experienced by the students.

However, we looked only at the self-assessments of one assignment, handled by the same teacher. So a potential explanation is a learning effect specific for this assignment and that students learned how to satisfy their teacher when handing in a grading request. Future work has to show if Incremental Grading also increases self-assessment skills on long terms and independent of domain-specific assignments.

RQ2: How do students value the Incremental Grading approach?

Regarding the general appreciation of the approach by the students, we see clear differences between both courses. In the course with interim deadlines, only a third of the students were positive about Incremental Grading, compared to 90% in the other course. Feedback of the students showed that in the course with the interim deadlines, the effort they had to put into self-grading and underpinning was too high compared to the effort for the product itself. Furthermore, this product counted for only 35% of the final grade, showing a disproportion of required effort (high) and value for final grade (low). Potential reason is that students were working towards the

interim deadlines, but now additionally had to also hand in the grading request, so the time they spent shortly before the deadline on their request and underpinning was not available for working on the product.

In general, students valued that they got better guidance (aims and criteria are known upfront and clear) and that they had to be more critical about their own work. They also were positive about the feedback they got, both amount and quality. The chance to improve the work and (finally) the ability to earn high grades had, according to the students, a positive impact on quality and grades. The least valued aspects of Incremental Grading were the amount of work it did cost for the students. Some of them saw the required underpinning as unnecessary, not seeing the connection between their underpinning and the quality of feedback they get. The aspect of not getting a grade when the grading request was not (completely) acknowledged was mentioned only once.

Some students did not make use of Incremental Grading at all, some of these did also not finish the course. We did not ask specifically for the reason why they did not hand in a grading request, but assume that their initial view of the approach did not show any added value for them. It is notable that this was only the case in the course where there was no deadline at all, indicating that having at least one moment of required submission of a grading request might address this issue.

The research question can be answered as follows: Students generally value the different aspects of the approach and the approach in general too if no interim deadlines are applied. The biggest concern is the amount of time it costs for students.

RQ3: How do teachers value the Incremental Grading approach?

Both teachers value the approach for several reasons: the assessment criteria and their interpretations have improved, students are more focused on the task and make good use of teacher feedback. Some students also work harder to get high grades. The teachers also see some disadvantages, such as the work amount it costs them, even though this work is more distributed over the course. Some of these disadvantages can be lowered with implementation adjustments such as less interim deadlines and clear communication of what the teacher expects.

The research question can be answered as follows: Teachers generally value the applied aspects of the approach. The biggest concerns are the amount of time it costs to provide feedback and to handle the grading requests.

RQ4: How does the application of Incremental Grading have influence on the final assignment results?

When working with interim deadlines, it seems that the amount of grading requests has no influence on the final grades. When working with a final deadline only, then we can observe a connection between grading request quantity and final results. So in an implementation with no or only one interim deadline, more frequently requesting grades might connect to higher final grades. Potential explanations are the higher amount of feedback, but also the motivating effect of having finished elements with a certain grade. An

alternative potential explanation is that with interim deadlines, all students got feedback on multiple (at least 3) grading requests and therefore all benefit, leading to higher grades in general[2].

It seems that students with lower final results also tended to request lower grades after a grading request had been declined. On the other hand, students who did improve their work after a decline also achieved higher final results.

Students who indicated that Incremental Grading had no impact on their product quality also did not make much use of the approach and potentially also ended with lower final grades. This is in line with the teacher's experience that students who did use the approach also were more motivated and worked harder.

In general, a small majority (50% - 61%) also perceived a positive effect of Incremental Grading on the quality of the final products.

Summarized, the research question can be answered as follows: the number of grading requests relates to final grades in a positive way if no interim deadlines are applied. High performers tend to improve their work while low performers adjust the requested grades. Especially the more motivated students also perceive a positive effect on the quality of their products and in consequence also on their final grades.

Some of the results were not significant because the data-set was relatively small (especially for the first course). The obtained results are therefore merely an indication for potential trends and need further research with larger groups.

6   CONCLUSION AND FUTURE WORK

In this small study, we evaluated the effects of applying Incremental Grading as assessment approach for one assignment each in two courses in higher education. While most aspects of the implementation were comparable, some aspects such as interim deadlines differed. The results from both courses are indicators that Incremental Grading has positive effects on the self-assessment skills and final results. However, the difference in implementation per course also showed some contrasting results: students disliked the approach when multiple interim deadlines were present.

We suggest a few implementation specifics for future courses based on our experiences in this study:

- Use one interim deadline, relatively in the beginning of the assignment period. This way the students have to follow the process at least once early, get feedback on their self-grading and consequently experience the benefits.
- Assignments using Incremental Grading should be weighted realistically in a final grade, as the approach requires more and deeper thoughts by the students than simply submitting their learning artefacts.
- Providing examples of student work and having students practice to assess them using the rubrics might help them to get familiar with the rubrics and shows them how to translate the quality levels in the rubric to concrete learning artefacts.

Some elements of Incremental Grading were not applied due to technological issues, such as the grading dashboard and a grading

---

[2]The grades for this course assignment were higher than the previous years, but as various factors such as rubrics and assignment approach have been changed, the grades are not comparable

queue which is also visible for students. Future work with appropriate technological solutions should also look into the effects of these elements. Furthermore, the characteristics of the rubrics could be of influence on characteristics of submitted grading requests and final results. This will also be subject of future studies.

Also of interest are the effects of adding peer-assessment/peer-feedback to the approach as well as looking closer at the changed characteristic of teacher's workload.

ACKNOWLEDGMENTS

REFERENCES

[1] Heidi Andrade and Ying Du. 2005. Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation* 10, 1 (2005). https://scholarworks.umass.edu/pare/vol10/iss1/3

[2] Joe Bergin, Jutta Eckstein, Markus Völter, Marianna Sipos, Eugene Wallingford, Klaus Marquardt, Jane Chandler, Helen Sharp, and Mary Lynn Manns (Eds.). 2012. *Pedagogical Patterns: Advice for Educators.* Joseph Bergin Software Tools, New York, NY, USA. 230 pages.

[3] Joe Bergin, Christian Kohls, Christian Köppe, Yishay Mor, Michel Portier, Till Schümmer, and Steven Warburton. 2015. Assessment-Driven Course Design - Foundational Patterns. In *Proceedings of the 20th European Conference on Pattern Languages of Programs, EuroPLoP'15.* ACM, Irsee, Germany.

[4] Jaclyn Broadbent, Ernesto Panadero, and David Boud. 2018. Implementing summative assessment with a formative flavour: a case study in a large class. *Assessment & Evaluation in Higher Education* 43, 2 (feb 2018), 307–322. https://doi.org/10.1080/02602938.2017.1343455

[5] Phillip Dawson. 2017. Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education* 42, 3 (apr 2017), 347–360. https://doi.org/10.1080/02602938.2015.1111294

[6] Cristina B. Gibson. 2017. Elaboration, Generalization, Triangulation, and Interpretation: On Enhancing the Value of Mixed Method Research. *Organizational Research Methods* 20, 2 (apr 2017), 193–223. https://doi.org/10.1177/1094428116639133

[7] Christian Köppe, Mary Lynn Manns, and Rody Middelkoop. 2019. The Pattern Language of Incremental Grading. In *Proceedings of the 25th Conference on Pattern Languages of Programs, PLoP'18.* Hillside Group, Portland, OR, USA, 1–17.

[8] Christian Köppe, Mary Lynn Manns, and Rody Middelkoop. 2020. Educational Design Patterns for Student-Centered Assessments. In *Preprints of the 26th Conference on Pattern Languages of Programs, PLoP'19.* Hillside Group, Ottawa, Canada.

[9] Christian Köppe and Rody Middelkoop. 2020. On Using Hybrid Pedagogy as Guideline for Improving Assessment Design. In *Proceedings of the ECTEL workshop.* Delft, The Netherlands, 1–12.

[10] Christian Köppe, Ralph Niels, Robert Holwerda, Lars Tijsma, Niek Van Diepen, Koen Van Turnhout, and René Bakker. 2015. Flipped classroom patterns: designing valuable in-class meetings. In *Proceedings of the 20th European Conference on Pattern Languages of Programs - EuroPLoP '15*, Vol. 08-12-July. ACM Press, New York, New York, USA, 1–17. https://doi.org/10.1145/2855321.2855348

[11] Victor López-Pastor and Alvaro Sicilia-Camacho. 2017. Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment and Evaluation in Higher Education* 42, 1 (jan 2017), 77–97. https://doi.org/10.1080/02602938.2015.1083535

[12] Steven F. Raaijmakers, Martine Baars, Fred Paas, Jeroen J.G. van Merriënboer, and Tamara van Gog. 2019. Effects of self-assessment feedback on self-assessment and task-selection accuracy. *Metacognition and Learning* 14, 1 (mar 2019), 21–42. https://doi.org/10.1007/s11409-019-09189-5

[13] D. Royce Sadler. 2009. Indeterminacy in the use of preset criteria for assessment and grading. *Assessment and Evaluation in Higher Education* 34, 2 (2009), 159–179. https://doi.org/10.1080/02602930801956059

[14] Johnny Saldaña. 2013. *The Coding Manual for Qualitative Researchers* (second edi ed.). Sage publications. 282 pages.

[15] Piers Steel and Katrin B. Klingsieck. 2016. Academic Procrastination: Psychological Antecedents Revisited. *Australian Psychologist* 51, 1 (feb 2016), 36–46. https://doi.org/10.1111/ap.12173

[16] Steven Warburton, Joe Bergin, Christian Kohls, Christian Köppe, and Yishay Mor. 2016. Dialogical assessment patterns for learning from others. In *Proceedings of the 10th Travelling Conference on Pattern Languages of Programs - VikingPLoP '16.*

ACM Press, New York, New York, USA, 1–14. https://doi.org/10.1145/3022636.3022651

[17] Naomi E. Winstone and David Boud. 2020. The need to disentangle assessment and feedback in higher education. *Studies in Higher Education* (jun 2020), 1–12. https://doi.org/10.1080/03075079.2020.1779687

APPENDIX

INCREMENTAL GRADING QUESTIONNAIRE   **ID:**

In this course you had to grade the deliveries for the "Proeve van toetsing" yourself. As this was a new approach, we'd like to know how you experienced it. Please note: the following questions are only relevant for that part of the course.

Q1: Based on the rubrics, I was able to assess my own work well.

strongly disagree	disagree	neutral	agree	strongly agree
O	O	O	O	O

Q2: Through the Incremental Grading approach, my ability to assess my own work has improved during the course.

strongly disagree	disagree	neutral	agree	strongly agree
O	O	O	O	O

Q3: I think that I would have achieved the same product quality without Incremental Grading.

strongly disagree	disagree	neutral	agree	strongly agree
O	O	O	O	O

Q4: Having to provide a motivation for the requested grade was unnecessary extra work.

strongly disagree	disagree	neutral	agree	strongly agree
O	O	O	O	O

Q5: I'd like to see this approach (Incremental Grading with the possibility of improving the work) applied in other course assignments as well.

strongly disagree	disagree	neutral	agree	strongly agree
O	O	O	O	O

Q6: What I value the **most** about the Incremental Grading approach is:

Q7: What I value the **least** about the Incremental Grading approach is:

Q8: What I learned through the Incremental Grading approach is:

Figure 4: Questionnaire taken at the end of the course