# Enhancing learning with inspectable student models



**DOI:**
10.1016/j.chb.2020.106276

**Document status and date:**
Published: 01/06/2020

**Document Version:**
Publisher's PDF, also known as Version of record

**Document license:**
Taverne

Computers in Human Behavior 107 (2020) 106276

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

# Enhancing learning with inspectable student models: Worth the effort?

Sietske Tacoma [a,*], Corine Geurts [a], Bert Slof [a], Johan Jeuring [a,b], Paul Drijvers [a]

[a] Utrecht University, Utrecht, the Netherlands
[b] Open University of the Netherlands, Heerlen, the Netherlands

ARTICLE INFO

*Keywords:*
Feedback-seeking behavior
Higher education
Inspectable student model
Log file analysis
Statistics education

ABSTRACT

In electronic learning environments, information about a student's performance can be provided to the student in the form of an inspectable student model. While relatively easy to implement, little is known about whether students use the feedback provided by such models and whether they benefit from it. In this study, the use of inspectable student models in an introductory university statistics course by 599 first-year social science students was monitored. Research questions focused on whether students sought feedback from the student models, which decisions for subsequent study steps they made, and how this feedback seeking and decision making related to results on their statistics exams. Results showed a large variety among students in feedback-seeking and decision-making behavior. Lower student model scores seemed to encourage students to practice more on the same topic and higher scores seemed to evoke the decision to move to a different topic. Viewing frequency and amount of variety in decision making were positively related to exam results, even when controlling for total time students worked. These findings imply that inspectable student models can be a valuable addition to electronic learning environments and suggest that more intensive use of inspectable student models may contribute to learning.

## 1. Introduction

University education puts high demands on students in taking responsibility for their learning (Krause & Coates, 2008; Torenbeek, Jansen, & Suhre, 2013). A potentially effective way to support them in doing so is to offer formative assessment opportunities: assessment of their performance aimed at improving the learning process prior to grading (Birenbaum et al., 2015; Timmers, Braber-van den Broek, & Van den Berg, 2013). Whereas many educators and researchers advocate the potential of formative assessment for learning, sound empirical evidence for this is lacking (Hendriks, 2014). Scarce (Robinson, Myran, Strauss, & Reed, 2014) and ineffective (Bennett, 2011; Heitink, Van der Kleij, Veldkamp, Schildkamp, & Kippers, 2016) implementations of formative assessment in educational settings are regularly voiced explanations for this lack. To reach its full potential, formative assessment should be a cyclical process (Gikandi, Morrow, & Davis, 2011). Besides gathering information about student performance, two other elements are part of such formative assessment cycles, namely providing tailored feedback on performance and deciding on actions to enhance learning based on the provided feedback (Antoniou & James, 2014; Black & Wiliam, 2012). Whereas educational practitioners gather a lot of assessment data (e.g., Tempelaar, Rienties, & Giesbers, 2015), they often experience difficulties in proving tailored feedback and determining how their students make use of it. To address this, more insight into the interplay between the provided feedback and students' feedback-seeking and decision-making behavior is needed. The current study addresses this by implementing and examining formative assessment cycles – by means of inspectable student models – in an electronic learning environment in the context of a university statistics course.

For statistics education, the use of formative assessment – e.g., self-tests – has been recommended by several authors (Carver et al., 2016; Tishkovskaya & Lancaster, 2012). The low-stake assessment setting might support students in reducing statistics anxiety (Chew & Dillon, 2014) and procrastination (Onwuegbuzie, 2004), two factors that often result in lower grades for statistics (Paechter, Macher, Martskvishvili, Wimmer, & Papousek, 2017). By conducting self-tests, students have the opportunity to gain insight into their current mastery of the study domain (Dirkx, Kester, & Kirschner, 2014). For the case of statistics, this study domain involves a large number of abstract concepts (Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007). Hence, the tailored feedback element of the formative assessment cycle should support students in gaining insight into their mastery of these various concepts. A promising operationalization of feedback in this respect is the inspectable student model, that can be offered to students within

* Corresponding author. Freudenthal Institute, Utrecht University, P.O. Box 85170, 3508, AD, Utrecht, the Netherlands.
*E-mail address:* s.g.tacoma@uu.nl (S. Tacoma).

https://doi.org/10.1016/j.chb.2020.106276
Received 17 May 2019; Received in revised form 8 January 2020; Accepted 27 January 2020
Available online 30 January 2020

electronic learning environments. In this study, we examine students' use of inspectable student models, i.e., whether and how students consult inspectable student models and make decisions on actions after consultation, and its effect on students' performance on a statistics exam.

## 2. Inspectable student models in electronic learning environments

Electronic learning environments are gaining in popularity for realizing formative assessment in education (Van der Kleij, Timmers, & Eggen, 2011). Due to technological advancements (e.g., open source, interactive visualizations, learning analytics) implementing such tools in educational practices nowadays requires less money and effort than in the past, and these advancements also provide more opportunities for integrating the complete formative assessment cycle in the educational design. Electronic learning environments have the advantage that information about student performance is usually automatically captured and stored (e.g., log files) by means of a student model: a representation of a student's current mastery of important topics in the study domain (Herder, Sosnovsky, & Dimitrova, 2017). A visualization (e.g., figure, table) of the student model that students can consult – an *inspectable student model* – can serve as the tailored feedback element in the formative assessment cycle. Enriching electronic learning environments with inspectable student models has the potential to foster student learning in two ways. First, inspectable student models provide an overview of the important topics in the domain, which can support students in understanding the domain structure (Mitrovic & Martin, 2007). Second, inspectable student models provide an estimate of the student's current mastery of the topics included in the model. Low estimates for topics might stimulate students to exert more effort and practice on these topics. Furthermore, when estimates conflict with a student's own perception of his or her mastery level, the student is more likely to consider further practice (Bull & Kay, 2007; Long & Aleven, 2011). Hence, enriching electronic learning environments with inspectable student models is an added service, which could support students in deciding on appropriate subsequent actions (e.g., selecting additional practice tasks).

Earlier studies revealed that students value the presence of inspectable student models in weekly homework sets (Mitrovic & Martin, 2007; Tacoma, Sosnovsky, Boon, Jeuring, & Drijvers, 2018). To our knowledge, previous research focused on the effects of inspectable student models combined with either (1) task selection adapted by the electronic learning environment based on the content of the student model (e.g., Brusilovsky, Sosnovsky, & Yudelson, 2009), or (2) monitoring of and feedback on task selection by the student (Mitrovic & Martin, 2007). Hence, the potential of integrating inspectable student models as an added service, to strengthen the formative assessment cycle through tailored feedback while leaving control over task selection fully with students, has not been studied extensively. Thus, it remains unclear how this added service affects student learning, a knowledge gap that this paper aims to fill.

For feedback to affect student learning, students need to actively seek for it, process it and decide which, if any, subsequent actions to carry out (Timmers et al., 2013). Various factors, such as motivation and accessibility of feedback information, may influence whether and how students engage in such behavior. To better understand these factors and, more specifically, how providing inspectable student models might foster student learning, more insight into students' feedback-seeking and decision-making behavior is required.

## 3. Feedback-seeking and decision-making behavior

Feedback-seeking behavior has been defined as the proactive search for feedback information in one's environment (Ashford & Cummings, 1983). Although inspectable student models are intended to foster student learning, there is no guarantee that students will engage in a proactive search for the feedback the student models provide, especially when this is not a mandatory learning activity. For a student to exhibit feedback-seeking behavior, the assumed values should outweigh the assumed costs (Anseel, Beatty, Shen, Lievens, & Sackett, 2015; Ashford & Cummings, 1983). In the context of the present study, students should see the value of inspecting the student model as well as undertaking subsequent actions based on the provided feedback. According to Anseel, Lievens, and Levy (2007) students might value feedback for different motives, namely: self-assessment (i.e., knowing how well one is doing), self-improvement (i.e., acquiring a higher mastery level), self-enhancement (i.e., coping with stress and anxiety), and self-verification (i.e., maintaining consistency between self-conceptions and new self-relevant information).

Especially students with strong self-improvement motives are more inclined to exhibit feedback-seeking behavior when they value the tool's potential for their learning process. The self-improvement value is particularly relevant when a student considers appropriate subsequent study steps, for example immediately after completing an initial set of practice tasks (Gikandi et al., 2011). Whether and how the provided feedback affects a student's decision making at such moments depends on several factors, such as perceived usefulness of the feedback (Harks, Rakoczy, Hattie, Besser, & Klieme, 2014) and the student's desire and intention to respond to the feedback (Kinicki, Prussia, Wu, & McKee-Ryan, 2004). In the context of this study, feedback indicating that current mastery is below the expected standards could lead to more practice and more feedback-seeking behavior (Hattie & Yates, 2014; VandeWalle & Cummings, 1997). If, however, the perceived costs of exposing one's uncertainty and need for help outweigh the student's value of self-improvement, such feedback might also lead to less feedback-seeking behavior, to avoid loss of face and ego costs of repeated negative feedback (Abraham, Burnett, & Morrison, 2006; Timmers et al., 2013). Yet, for inspectable student models these costs are relatively low compared to seeking feedback from a tutor or peer (Timmers et al., 2013). Receiving feedback indicating that the current mastery level is above the expected standard can also have diverse effects on both practice and subsequent feedback-seeking behavior. Students will only be inclined to practice more and exhibit more feedback-seeking behavior when they expect that the additional time investment will result in a gain in mastery level.

Previous studies on feedback-seeking behavior revealed no strong relationship between feedback-seeking behavior and performance (Anseel et al., 2015). When one attaches a high value to the feedback, one is inclined to proactively seek for (additional) feedback (Morrison & Cummings, 1992; Tuckey, Brewer, & Williamson, 2002). However, more feedback-seeking behavior does not automatically result in better performance such as a higher mastery level (Ang, Cummings, Straub, & Earley, 1993; Ashford & Black, 1996). Similarly, a review by Crommelinck and Anseel (2013) questioned the implicit assumption that feedback seeking is positively associated with performance, since most of the studies pay little empirical attention to the question whether and how feedback-seeking behavior affects performance. Consequently, a more in-depth understanding of the factors that explain whether and how feedback seeking leads to better performance is needed.

To this end, the current study examines the interplay between students' use of inspectable student models, i.e., their feedback-seeking and decision-making behavior, and their exam grades for the case of a university statistics course. The study is guided by three research questions:

RQ1: How do first-year university students in social science seek feedback from inspectable student models in an introductory statistics course?

RQ2: How does feedback from inspectable student models inform these students' decisions about subsequent actions?

RQ3: How do these students' feedback-seeking and decision-making behavior relate to performance on a statistics exam?

## 4. Materials and methods

### 4.1. Participants

Participants were 599 first-year university students who were enrolled in an introductory Methods and Statistics course at a Dutch research university. To be eligible for enrollment at this university, students needed to have followed a pre-university track in secondary education or at a university of applied sciences, which means that these students belonged to the top 20% of students their age. The course was mandatory for all bachelor's degree programs in the social sciences. The students were informed about this study and were asked for their consent. Of the 1025 students who were enrolled in the course, 599 made use of the electronic learning environment and gave consent for the use of their work and exam results for this study. Of the 599 students, 77% was female and 23% was male. Their ages varied between 17 and 43 years ($M = 19.5$, $SD = 2.2$).

### 4.2. Description of the course and the electronic learning environment

The Methods and Statistics course was an eight-week course in which new methods and statistical concepts were introduced in week 1, 2, 4, 5 and 7. Intermediate exams were administered in week 3 and in week 6, and the final exam was administered in week 8. Learning objectives of the course were outlined in a course manual. In the weeks in which new concepts were introduced, a lecture on these concepts was given and students were offered online homework sets on the statistical topics. Students could choose to work on these homework sets at home or in lab sessions supervised by teachers. Tasks from the homework sets and their relations with the learning objectives were discussed in weekly discussion sessions.

The electronic learning environment in which the homework sets were made available was the Digital Mathematics Environment (DME, see Drijvers, Boon, Doorman, Bokhove, & Tacoma, 2013). Tasks in the homework sets addressed, for example, selecting appropriate measures of center and spread for given variables, or carrying out hypothesis tests for given situations and samples. Students received immediate feedback on the correctness of their answers, but the correct answer itself was not provided to students. Students could attempt answering tasks until they found the correct answer. A typical task from the first homework set is displayed in Fig. 1. The tasks were designed by a team of teachers in the university's Methods and Statistics department.

In the weeks prior to the intermediate and final exams, extra practice sets were provided in the DME, allowing students to prepare for the exams. The extra practice sets contained between six and eleven new practice tasks on all topics covered so far. All homework and extra practice sets remained available for the students until the end of the course period. All interactions of the students in the DME were logged.

### 4.3. Design and implementation of the inspectable student models

The DME was enriched with an inspectable student model for each homework set. Fig. 2 shows two examples of an inspectable student model for the first homework set. Each student model contained a list of important topics in the homework set, grouped into two or three categories. The number of topics per category varied between two and seven. Most tasks in the homework sets were connected to the topic(s) they were related to. Lists of topics, connections between tasks and topics and the tasks themselves were optimized informed by findings by Tacoma et al. (2018), based on the same course in the previous academic year. In particular, this previous study showed that some tasks served a useful function in the homework set (such as introducing a new topic), but were not appropriate for informing student models, and hence should

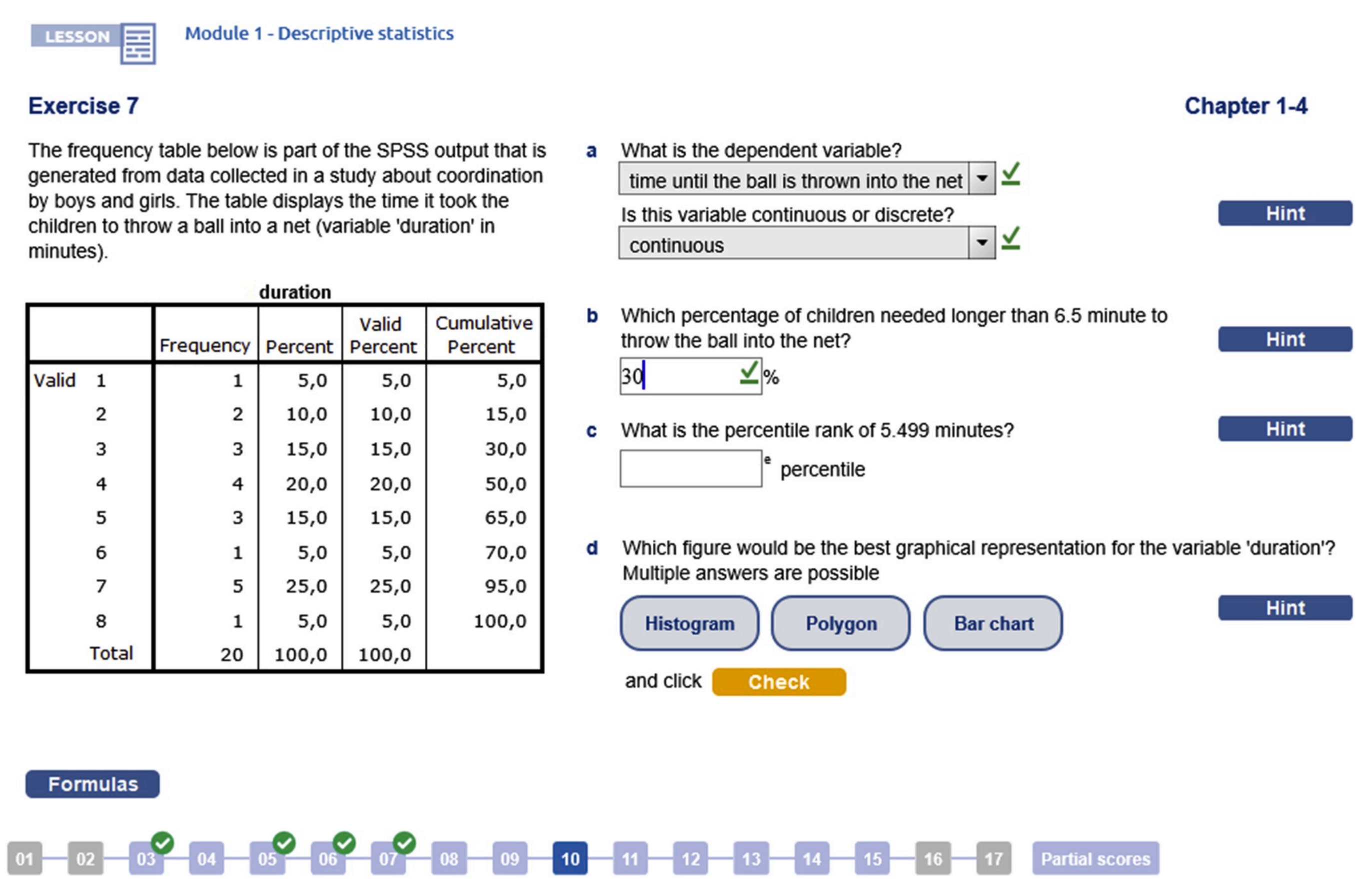


**Fig. 1.** Example of a practice task in the first homework set in the DME.

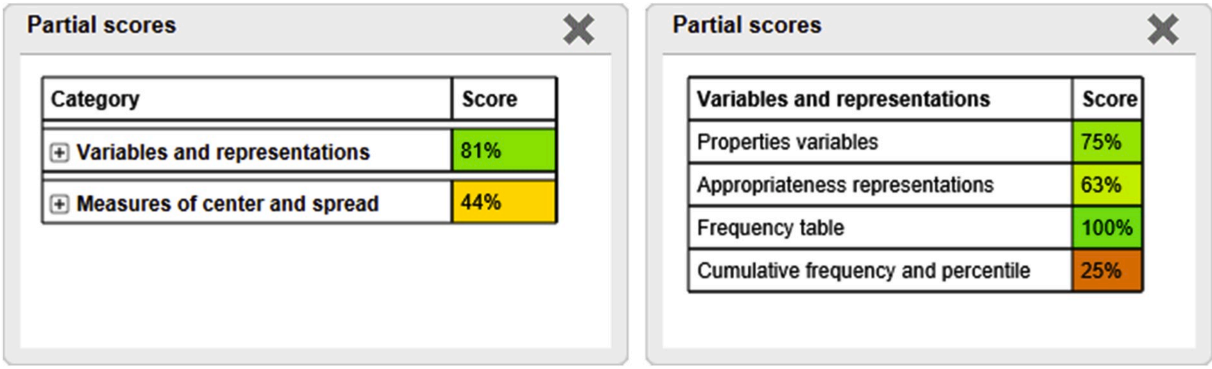


**Fig. 2.** Student models for the first homework set when a student has worked on several categories (left) or on one category only (right).

not be connected to any topic. Furthermore, new tasks were added to address topics that had been underrepresented in the previous year, and a number of multiple-choice tasks that had been found to offer too few opportunities to learn from (i.e., asking students to select one out of only two options) were redesigned.

Scores in the student models were calculated based on the student's correct and incorrect attempts on the tasks in the homework set: for each task a task score was calculated as the number of correct attempts on that task (usually 1) divided by the total number of attempts the student made on the task. Topic scores were calculated as the mean task score of all tasks that were connected to the topic and for which the student had made at least one attempt. Category scores were a weighted average of topic scores, weighted by the number of tasks connected to the topic.

Students could access the inspectable student model for a homework set by clicking on the button "Partial scores" (bottom right corner in Fig. 1). On the final page of each homework set this service was explicitly mentioned to students, with the suggestion to use the student model to select topics for further practice. When students opened the student model, only the categories and category scores were shown (Fig. 2, left). Students could use the plus-buttons to view the topics in each category and their scores on these topics. Only categories that the student had worked on were shown and if a student had only worked on one category yet, this category was shown folded out immediately (Fig. 2, right).

On the first page of the extra practice sets, students received instruction that they could either choose to work on all extra practice tasks, or to make a selection based on their inspectable student models. Links to the homework sets were included, so that students could easily access the student models for the different homework sets. In each extra practice set, the first page also contained an overview indicating which extra practice tasks addressed which topics. This enabled students to select extra practice tasks for topics that needed their attention.

### 4.4. Data collection

Data for this study consisted of log files of the students' work on homework and extra practice sets in the DME, including logs of student model views. Additionally, students' grades for the final exam were collected. The possibility to log student's actions in electronic learning environments provides an opportunity to monitor meticulously what students do with inspectable student models that are provided to them. For each student model view, the DME logged the time of opening and closing the student model, the corresponding homework set, current student model scores for all topics and categories in the student model, and categories that the student opened (if any). After the end of the course period, log files were exported from the DME. Logs from students who did not give consent were deleted and all other logs were rendered anonymous. Exam results were rendered anonymous as well, using the same key to enable connecting them to the students' use of the DME. The final exam lasted 2 h and consisted of 30 4-option multiple-choice items: 14 about methods and 16 about statistics. Only the students' results on the statistics items were included in this study. For these 16 items, Cronbach's α was .60, which seems an appropriate value for an exam consisting of relatively few items that assess a wide range of topics (e.g., normal distribution, confidence intervals, hypothesis testing) within the domain of statistics (Taber, 2018). An example question is:

> It was investigated whether in the 2010 elections politicians who were active on Twitter received more preference votes than their colleagues who were not active on Twitter. The report mentioned both a *p*-value (.001) as well as the effect size ($d = .01$). What is the correct conclusion when testing with $\alpha = 1\%$?

Multiple-choice options for this item were (a) The result is not significant and the effect is small; (b) The result is significant, but the effect is small; (c) The result is significant and the effect is large; (d) The result is not significant, but the effect is large.

### 4.5. Data analysis

To answer RQ1 on feedback-seeking behavior, the logged information was used to describe how often, how long and in how much detail students inspected their student models. Student model views that lasted shorter than 2 s were omitted from analysis: these views were considered too short for students to be able to interpret the contents of the student model.[1] This concerned 173 student model views, out of a total of 2710. Regarding the detailedness of student model inspection, a Chi-Square proportion test served to examine whether students tended to select the categories with the lowest scores for further inspection, if they opened any categories at all. For all statistical tests, a significance level of .05 was used. To enable an interpretation of the frequency of student model views, working sessions were defined. Following Chen, Breslow, and DeBoer (2018) working sessions were defined as series of student actions in the DME in which the time period between two actions was never longer than 1 h. Working sessions were mapped over time to determine in what proportion of working sessions students viewed the inspectable student models and to investigate whether students kept inspecting the student models during the course period. To enable further analysis at the level of individual students rather than at the level of student model views, students were assigned to groups based on their feedback-seeking behavior, as will be explained in the results section.

To answer RQ2 on how consulting a student model affects students' decision making on subsequent actions, only student model views after which the student continued working in the DME were included. Three

[1] Assuming a reading speed of approximately 250 msec per word (H. van Oostendorp, personal communication, April 9, 2019), reading the concepts listed in Fig. 2 (left) would take 2 s, which makes 2 s a reasonable lower bound.

general decisions were possible for students who continued working after viewing a student model, namely work on (1) the homework set for which the student model was viewed ("Homework"), (2) extra practice related to the homework set for which the student model was viewed ("Practice"), or (3) a homework set or extra practice on a different topic than addressed in the student model just viewed ("Other topic"). Students were grouped based on which of the three decisions they made at least once. This resulted in seven groups, namely:

- HPO: Homework-Practice-Other topic, students who made all three decisions at least once;
- HP: Homework-Practice, students who made the decisions Homework and Practice at least once and never made the decision Other topic;
- HO: Homework-Other, students who made the decisions Homework and Other topic at least once and never made the decision Practice;
- PO: Practice-Other topic, students who made the decisions Practice and Other topic at least once and never made the decision Homework;
- H: Homework, students who always continued to work on the homework set after viewing the student model;
- P: Practice, students who always worked on an extra practice set after viewing the student model;
- O: Other, students who always worked on another topic after viewing the student model.

To compare student model scores between different decisions within each group, for each student model view a mean student model score was calculated: the mean of all topic scores currently in the student model. Next, for each decision within each group, the median of the mean student model scores preceding that decision was calculated. Medians and non-parametric tests were used, since the distribution of mean student model scores was negatively skewed. For group HPO, a Friedman's ANOVA and follow-up pairwise Wilcoxon signed rank tests were used to compare median scores for the three decisions. A Bonferroni correction was used to control for the inflated chance of a type I error in multiple comparisons (Shaffer, 1995). For groups HP, HO and PO, the median scores were compared using Wilcoxon signed rank tests.

To examine the relations between feedback-seeking behavior, decision-making behavior and exam results (RQ3), a Chi-square test was used to assess whether feedback-seeking and decision-making behavior were independent. The seven groups for seven possible combinations of decisions were supplemented with an eighth group, "Nothing", for students who never viewed a student model or who never continued working in the DME after viewing a student model. A possible confounding variable in relations between feedback-seeking behavior, decision-making behavior and exam results was the students' activity in the learning environment. More active students may be more likely to view and use the student models and may also be more likely to perform well on the exam. To assess the influence of this confounding variable, the total time students worked on the tasks in the DME was calculated (including breaks of up to 5 min). Two one-way ANOVAs were carried out to examine the relations between feedback-seeking and decision-making behavior on the one hand, and time on task on the other hand. When the ANOVAs yielded significant results, they were followed up with pairwise comparisons with Bonferroni correction. Finally, a hierarchical multiple linear regression model was set up to assess the relations between students' exam grade as outcome variable and feedback-seeking behavior, decision-making behavior and time on task as predictor variables. Because of possible interaction effects between time on task and feedback-seeking and decision-making behavior, a regression model was deemed more suitable than an ANCOVA, in which interaction effects between grouping variables and covariates are not included.

## 5. Results

### 5.1. Students' feedback-seeking behavior

To gain insight into students' feedback-seeking behavior (RQ1), we summarized all students' views of the inspectable student models. Furthermore, the distribution of student model views over the course period was examined and students were grouped according to their feedback-seeking behavior.

Table 1 gives an overview of the students' working sessions in the DME, the number of sessions in which they viewed a student model and the number of times they inspected student models more closely by opening one or more categories. The table reveals that students viewed a student model in 25% of all working sessions in the DME (1874 out of 7410 sessions). There were more student model views (2522) than sessions in which a student model was viewed (1874), which implies that students viewed the same student model more than once or viewed the student models for more than one homework set, in some sessions. In most sessions, however, students consulted only the student model concerning the homework set they were working on and consulted it just once. Students inspected the student model more closely in 40% of all student model views (997 out of 2522 views). Closer inspection in most cases entailed opening all categories, namely in 713 (72%) of the 997 views. When students did select categories, they tended to select the one or two categories with the lowest score(s): 246 (87%) out of 284 views, $\chi^2(1, N = 284) = 150.88, p < .001$.

Regarding duration of student model views, the logs revealed that most student model views were rather short; the median viewing duration was 6 s. Longer views also occurred: 155 views lasted longer than 30 s. Logs of student work in the DME also revealed that students mostly viewed student models after they had finished all tasks in the corresponding homework set: this was the case in 2030 out of 2522 views (80%).

The upper part of Fig. 3 displays the distribution of the students' working sessions over time. Each bar represents the number of working sessions for one day in the course period and the dashed lines indicate the dates of the intermediate and final exams. The figure shows that students kept inspecting the student models throughout the course and that on days before exams both the number of working sessions and the number of student model views increased rapidly. The lower part of Fig. 3 shows the percentage of sessions in which students inspected a student model, as percentage of the total number of sessions that day, together with a fitted linear regression line. It reveals that the percentage of sessions in which students inspected a student model decreased slightly but significantly over the course period. Taking the values from the regression line, the percentage dropped by 0.26 percentage points per day, from 34% to 21%.

On average, the 531 students who viewed a student model at least once viewed student models for 2.9 out of five homework sets. Student models for all five homework sets were viewed by 103 students.

**Table 1**
Summary of working sessions and student model views.

| | Total number | Number of unique students | Mean[a] per unique student (*SD*) | Median[a] |
|---|---|---|---|---|
| Working sessions in the DME | 7410 | 599 | 12.4 (6.2) | 12 |
| Working sessions with student model view | 1874 | 531 | 3.5 (2.3) | 3 |
| Student model views | 2522 | 531 | 4.7 (4.1) | 4 |
| Views with inspection of categories | 997 | 337 | 3.0 (2.5) | 2 |

[a] Means and medians were calculated over the students involved.

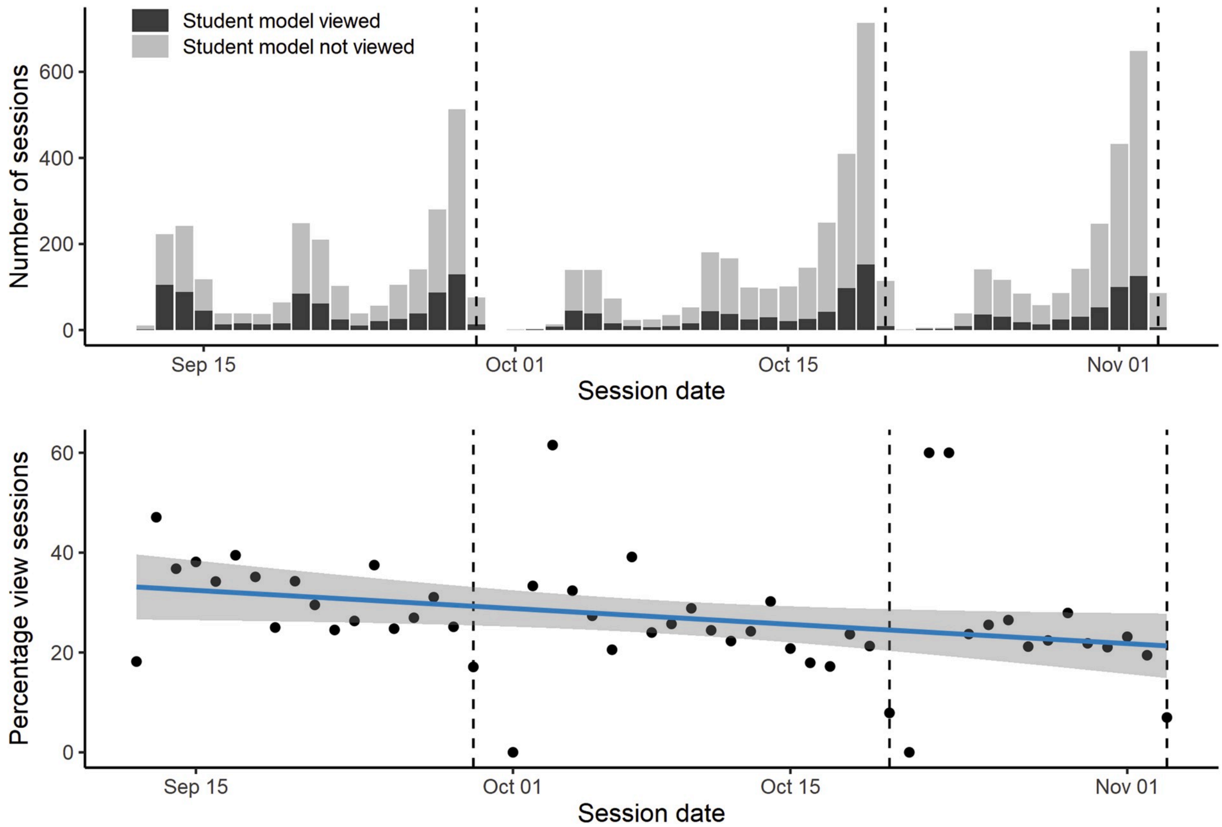


**Fig. 3.** Number (top) and percentage (bottom) of working sessions in which students viewed a student model per day over the course period.

Based on the number of homework sets for which students viewed their student model over the course period and the number of times they opened categories for closer inspection, students were assigned to one of three groups: limited, moderate and extensive feedback seekers. The following definitions led to approximately equal group sizes:

- Limited: student viewed student model of at most one homework set;
- Moderate: student viewed student models of at most four homework sets. If four student models were viewed, the student never inspected categories further;
- Extensive: student viewed student models for four or five homework sets. If four student models were viewed, the student inspected categories at least once.

This resulted in a group of 190 limited feedback seekers, a group of 222 moderate feedback seekers and a group of 187 extensive feedback seekers.

**Table 2**
Students' decisions after viewing student model and median student model scores preceding the different decisions (H = Homework, P = Practice, O = Other topic).

| Group | $N$ | Median score preceding decision Homework | Median score preceding decision Practice | Median score preceding decision Other topic | Value test statistic | $p$ |
|---|---|---|---|---|---|---|
| HPO | 62 | 79.1 | 83.0 | 84.8 | $\chi^2$ = 17.61 | <.001 |
| HP | 48 | 80.8 | 83.6 | – | $T$ = 399 | .082 |
| HO | 62 | 78.2 | – | 82.0 | $T$ = 305 | <.001 |
| PO | 36 | – | 79.9 | 81.0 | $T$ = 216.5 | .108 |
| H | 98 | 78.8 | – | – | – | |
| P | 35 | – | 81.5 | – | – | |
| O | 76 | – | – | 79.5 | – | |

### 5.2. *Students' decision-making behavior*

With respect to the students' decision-making behavior (RQ2), the following results were found. From the 1244 student model views after which the student continued to work in the DME (49% of all 2522 student model views), 587 (47%) were followed by the decision to work on the homework set for which the student model was viewed (Homework). For 281 views (23%), the student's decision was to work on extra practice tasks related to the just viewed student model (Practice), and for the remaining 376 views (30%), the student decided to work on a different topic (Other). Table 2 summarizes the allocation of students to the seven decision-combination groups, as well as the medians of the mean student model scores for each of the decisions in each group at the moment of student model consultation.

From the table it can be inferred that median scores when students continued with Homework were generally lower than those for the decisions Practice and Other topic. Furthermore, median scores for Practice and Other topic seemed fairly similar. These impressions were confirmed by the tests comparing the median scores in the four groups in which students had made multiple decisions. For students who made all three decisions at least once, group HPO, the Friedman's ANOVA yielded that median scores differed significantly between possible decisions, $\chi^2(2) = 17.61, p < .001$. Follow-up pairwise Wilcoxon signed rank tests yielded that scores when students decided to work on Homework were significantly lower than scores when students (1) decided to work on Practice ($T = 385, p < .001, r = -.35$) and (2) decided to work on Other topic ($T = 402, p < .001, r = -.30$). The scores did not differ significantly between the decisions Practice and Other topic ($T = 873.5, p = .608, r = -.05$). For group HO, the Wilcoxon signed rank test yielded a significant difference ($p < .001$, $r = .37$), indicating that students in this group chose to work on the homework set for lower scores and chose to work on another topic for higher scores. For groups HP and PO, scores did not differ significantly for the two decisions. Altogether, students tended to continue to work on the homework set for lower student model scores and started working on extra practice or another topic for higher student model scores.

### 5.3. Relation between feedback seeking, decision making and exam results

Before looking at exam results (RQ3), we first examined the relations between feedback-seeking behavior, decision-making behavior and time on task. Table 3 characterizes students by their feedback-seeking and by their decision-making behavior. A Chi-square test yielded that the characterizations were strongly related: $\chi^2(14, N = 599) = 323.86$, $p < .001$, Cramer's $V = 0.52$. Table 3 reveals that most limited feedback seekers, if they viewed a student model at all, indeed made just one single decision after viewing, while many extensive feedback seekers made different decisions on different occasions of viewing the student model. Hence, both characterizations seem to describe how intensively students used the student models. Although a strong relationship was found, Table 3 also reveals that students' decision-making behavior varied considerably among students exhibiting similar feedback-seeking behavior, especially for moderate feedback seekers.

The final row and column of Table 3 summarize student activity, as measured by time (in hours) that students worked on the tasks in the DME. Time on task was found to be significantly different for students with different feedback-seeking behaviors ($F(2, 596) = 74.88$, $p < .001$, $\eta^2 = .20$). Post-hoc comparisons revealed that all differences between the three groups were significant (all $p$-values smaller than .001). As expected, extensive feedback seekers spent most time on the tasks and limited feedback seekers the least. Time on task also differed significantly between groups of decision-making combinations, ($F(7, 591) = 16.26$, $p < .001$, $\eta^2 = .16$). Post-hoc comparisons revealed that students in the Nothing group worked significantly shorter than students in all other groups, that students in group O worked significantly shorter than students in groups HP and HPO, and that students in groups H and HO worked significantly shorter than students in group HPO.

Table 4 summarizes the parameter estimates and model fits of the hierarchical regression model predicting exam grade from time on task, feedback-seeking behavior, decision-making behavior and interactions. The base model shows that time on task was a significant predictor and explained 13% of the variability in exam grade. Adding feedback-seeking behavior resulted in a significantly better model ($F(2, 561) = 17.82$, $p < .001$), explaining an additional 5% of the variance. Time on task was a confounding variable in this relationship between feedback-seeking behavior and exam grade, as indicated by the relations found between time on task and both feedback-seeking behavior and exam grades. Still, the positive model parameters for moderate and, especially, extensive feedback seekers (compared to the reference group limited feedback seekers) suggest that, regardless of time spent on tasks, more extensive feedback seeking resulted in higher exam grades. Adding decision-making behavior to the model added another 1.5% to the amount of explained variance and significantly improved the model ($F(7, 554) = 2.56$, $p = .013$), but to a lesser extent than adding feedback-seeking behavior did. The parameter values for the different decision-making groups (compared to the reference group Nothing) are difficult to interpret, given the interplay between feedback-seeking behavior and decision-making behavior that is illustrated by Table 3. Finally, interactions between the predictor variables were added to the model. Only the interaction between feedback-seeking behavior and time on task significantly improved the model ($F(2, 552) = 5.17$, $p = .006$) and explained an additional 1.2% of the variance in exam grade. Fig. 4 illustrates this interaction effect. It reveals that for moderate and limited feedback seekers the time worked in the DME was strongly related with exam grade. For extensive feedback seekers, however, there seemed to be no relation between time on task and exam grade.

**Table 4**
Parameter estimates and model fits for the linear regression model predicting exam grade from time on task (in hours), feedback-seeking behavior, decision-making behavior, and the interaction between time on task and feedback-seeking behavior.

| | Base model with time on task | + Feedback seeking | + Decision making | + Time on task × Feedback seeking |
|---|---|---|---|---|
| Intercept | 8.90*** | 8.65*** | 8.77*** | 8.22*** |
| Time on task | 0.25*** | 0.17*** | 0.18*** | 0.28*** |
| Feedback seeking: Extensive | | 1.65*** | 2.23*** | 4.00*** |
| Feedback seeking: Moderate | | 0.79** | 1.14*** | 1.90*** |
| Decision making: HPO | | | −1.14* | −1.00* |
| Decision making: HP | | | −0.60 | −0.58 |
| Decision making: HO | | | −0.66 | −0.70 |
| Decision making: PO | | | −1.33** | −1.32** |
| Decision making: H | | | −0.21 | −0.29 |
| Decision making: P | | | 0.20 | 0.20 |
| Decision making: O | | | −0.96** | −0.95** |
| Time on task × Extensive | | | | −0.23** |
| Time on task × Moderate | | | | −0.13 |
| Adjusted $R^2$ | 0.129 | .176 | .192 | .204 |
| Adjusted $R^2$ change | | .047 | .015 | .012 |
| $F$ change | | 17.82*** | 2.56* | 5.17** |

*$p < .05$; **$p < .01$; ***$p < .001$.

## 6. Discussion

In this study, we investigated whether and how first-year university students used inspectable student models in a statistics course, and whether students seemed to benefit from these student models. We examined the students' feedback-seeking behavior (RQ1), decision-making behavior (RQ2), and the interplay between student behavior and exam grades (RQ3).

**Table 3**
Student characterization by feedback-seeking and decision-making behavior and time on task (in hours) for all groups.

| Decision making | HPO | HP | HO | PO | H | P | O | Nothing | Total | Time on task ($SD$) |
|---|---|---|---|---|---|---|---|---|---|---|
| Feedback seeking | | | | | | | | | | |
| Limited | 0 | 0 | 3 | 0 | 23 | 12 | 22 | 130 | 190 | 5.3 (3.9) |
| Moderate | 9 | 19 | 26 | 18 | 44 | 14 | 45 | 47 | 222 | 8.0 (3.4) |
| Extensive | 53 | 29 | 33 | 18 | 31 | 9 | 9 | 5 | 187 | 9.6 (3.1) |
| Total | 62 | 48 | 62 | 36 | 98 | 35 | 76 | 182 | 599 | 7.6 (3.9) |
| Time on task ($SD$) | 10.7 (3.8) | 9.1 (3.3) | 8.2 (3.2) | 8.9 (3.2) | 7.9 (3.4) | 8.3 (3.6) | 6.7 (3.5) | 5.8 (3.9) | 7.6 (3.9) | |

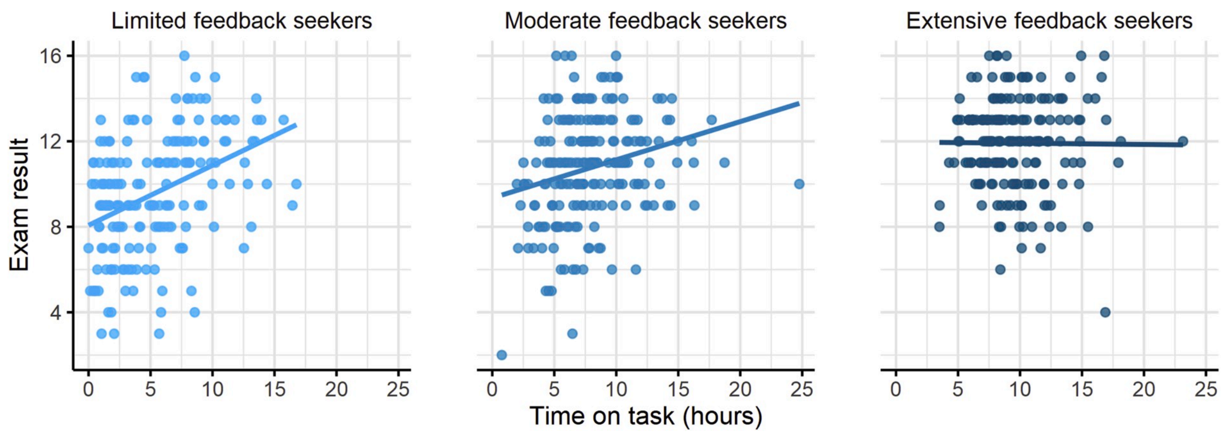


**Fig. 4.** Relation between exam result and time on task for limited, moderate and extensive feedback seekers.

Concerning RQ1, a wide variety was found in students' feedback-seeking behavior, or, more specifically, in frequency, timing, duration and amount of detail of student model views. This diversity seems to reflect a variety in self-motives underlying feedback-seeking behavior (Anseel et al., 2007), both among students as well as within students over time. For example, a student model view of a few seconds that takes place before the student has fully completed a homework set may be driven by a self-verification motive (i.e., quickly verify what one's weaker and stronger topics currently are). A long view that takes place after completing the homework set is more likely to be driven by a self-improvement motive (i.e., consider what to do next to improve one's mastery). Regardless of their exact motives, for most students the perceived values of the inspectable student models outweigh the costs, and hence, in line with earlier research, students seem to appreciate the availability of inspectable student models (Bull, 2004; Mitrovic & Martin, 2007).

Concerning RQ2, student appreciation is an important factor for inspectable student models to affect learning, but appreciation alone is not enough. Students also need to actively process the provided feedback and use it to decide on subsequent study steps (Timmers et al., 2013). Students in this study made a wide variety of decisions, which is in line with earlier findings (Bull et al., 2008). It suggests that inspectable student models fit into many different learning paths and, thus, allow students to take responsibility for their own learning. Across this variety of learning paths, students seemed inclined to improve their homework sets when student model scores were low, and to work on extra practice tasks or other topics when student model scores were higher. Hence, lower scores may encourage students to devote more effort to the homework sets than they would do without student models available, which is a valuable – hidden – effect of feedback (Hattie & Yates, 2014). This effect implies that inspectable student models may indeed support students in reducing the academic procrastination that is common in many introductory statistics courses (Onwuegbuzie, 2004).

The aim of RQ3 was to evaluate how students' use of the inspectable student models, i.e., their feedback-seeking and decision-making behavior, related to performance on the final exam, as an indication of how this operationalization of formative assessment can contribute to student performance. While student activity, as measured by time on task, was found to be an important predictor of exam result, the frequency of student model viewing explained significant additional variance in students' exam grades. So did, to a lesser extent, the amount of variety in decisions students made after viewing the student model. These findings suggest that frequently inspecting student models and using them to inform subsequent study steps seems a fruitful learning strategy. Furthermore, these findings support the assumption that feedback-seeking behavior and decision-making behavior are influenced by students' individual self-motives (Anseel et al., 2007), and not only by the amount of time they spend in the learning environment. Especially in the group of frequent feedback-seekers, no relation was found between time on task and exam grade, suggesting that other factors than activity determined how efficiently and effectively these students could use inspectable student models in their learning strategies.

While these findings imply that inspectable student models can be a valuable enrichment for electronic learning environments, especially in university statistics courses, this study has some limitations. First, due to the explorative nature of this study no causal inferences could be made about the influence of students' feedback-seeking and decision-making behavior on exam grades. It is, for example, likely that students differ in self-regulated learning abilities and that stronger self-regulated learners have strong self-improvement motives, which results in a high frequency of viewing student models. At the same time, these stronger self-regulated learners are also likely to perform well on an exam. Future research, with a randomized control design, is needed to establish whether there is a causal relation between availability of inspectable student models and exam results.

A second limitation relates to the main source of data for this study: logs from student work. While they provide valuable information and have the large advantage that collecting them is minimally invasive for students, for this study they also have a drawback: it is difficult, if not impossible, to infer students' intentions or self-motives from log files. We do not know whether long student model views indicate intensive engagement with a student model, or off-task behavior. Likewise, we assumed, but cannot prove, that students' decisions were influenced by the contents of the student models. Meanwhile, the ways in which students could benefit from inspectable student models might vary along with their varying self-motives. For example, students with weaker self-improvement motives might be expected to benefit relatively much from inspectable student models, because of the low costs of seeking feedback from them (Timmers et al., 2013) and the support they can give for selecting appropriate subsequent tasks (Corbalan, Kester, & Van Merriënboer, 2006). Future research that more directly addresses the students' self-motives and self-regulated learning capabilities, for example through questionnaires or interviews with focus groups, could provide more insight into how feedback from inspectable student models can best be tailored to the students' individual needs and preferences.

A final limitation is that the studied decision-making behavior concerned quite general decisions: continue working on the same homework set, go to extra practice or move to another topic. Inspectable student models have the potential to inform more specific decisions about topics for which students need to exert their effort and thinking (Hattie & Yates, 2014). In this study, however, due to the design of homework and extra practice sets, only a few such topic-specific decisions could be identified. In the homework sets, connections between topics and tasks were not made explicit for students and in the extra practice sets, topic descriptions for the tasks did not align completely with the terminology used in the inspectable student models. This may have hindered students in responding to the feedback according to their intentions (Kinicki et al., 2004). Consistent and explicit connections between tasks and topics could better support students in making deliberate decisions on topics to work on (Brusilovsky et al., 2009; Kicken, Brand-Gruwel, & van Merriënboer, 2008). This should receive careful attention in both further research and implementation of inspectable student models in practice, to realize their full potential for supporting more efficient and effective learning processes.

As concluding remarks, we note that the current research has revealed that students exhibit a wide variety in feedback-seeking and decision-making behavior when inspectable student models are available. Hence, this operationalization of the tailored feedback element that is essential for a cyclical formative assessment approach (Gikandi et al., 2011) seems to fit well within many learning paths. This allows students to take the responsibility for their learning that is required in university education (Krause & Coates, 2008). Furthermore, students' decision making appeared to be, at least partly, informed by the provided feedback, suggesting that inspectable student models also facilitate a second essential element of the formative assessment cycle: deciding on subsequent actions to enhance learning. Regarding performance, this study supports the claim that feedback-seeking behavior positively relates to performance, as well as the hypothesis that performance is enhanced by a high variety in decisions based on inspectable student models. For practice, this suggests that inspectable student models can indeed be a valuable enrichment of electronic learning environments, even in cases where student models do not inform task selection directly. While our implementation required students to actively seek for feedback by clicking a button, many other systems automatically show students their student models, which reduces the cost of seeking feedback. Whether this would result in more students engaging with the feedback and using it to decide on subsequent study steps than in our implementation is uncertain, though, because the students' self-motives also play a crucial role in engaging with feedback (Anseel et al., 2007). This could be an interesting venue for further research.

Finally, to answer the title question, a note on implementation effort is in place. Once the infrastructure within the electronic learning environment is set up, a simple inspectable student model implementation – like the one used in this study – only requires a list of concepts in the domain and connections between the tasks and these topics. Given that students value the inspectable student models, that students seem to practice more when such models are available, and that performance seems positively related to both feedback-seeking behavior and variety of decisions, our answer to the title question would be positive: implementing inspectable student models does seem to be worth the relatively small effort, and, while this study was conducted in the frame of a statistics course, we expect this to be worthwhile for other domains as well.

## Declarations of interest statement

None.

## CRediT authorship contribution statement

**Sietske Tacoma:** Conceptualization, Visualization, Formal analysis, Software, Writing - original draft. **Corine Geurts:** Conceptualization, Methodology, Writing - original draft. **Bert Slof:** Conceptualization, Methodology, Writing - review & editing. **Johan Jeuring:** Writing - review & editing, Supervision. **Paul Drijvers:** Writing - review & editing, Supervision.

## Acknowledgements

Design and implementation of the educational materials used in this study was supported by Utrecht University's Education Incentive Fund. We are grateful to Jeltje Wassenberg-Severijnen, teacher of the course, and Peter Boon, developer of the Digital Mathematics Environment, for the fruitful collaboration.

## References

Abraham, J. D., Burnett, D. D., & Morrison, J. D. J. (2006). Feedback seeking among developmental assessment center participants. *Journal of Business and Psychology, 20* (3), 383–394. https://doi.org/10.1007/s10869-005-9008-z.

Ang, S., Cummings, L. L., Straub, D. W., & Earley, P. C. (1993). The effects of information technology and the perceived mood of the feedback giver on feedback seeking. *Information Systems Research, 4*(3), 240–261. https://doi.org/10.1287/isre.4.3.240.

Anseel, F., Beatty, A. S., Shen, W., Lievens, F., & Sackett, P. R. (2015). How are we doing after 30 years? A meta-analytic review of the antecedents and outcomes of feedback-seeking behavior. *Journal of Management, 41*(1), 318–348. https://doi.org/10.1177/0149206313484521.

Anseel, F., Lievens, F., & Levy, P. E. (2007). A self-motives perspective on feedback-seeking behavior: Linking organizational behavior and social psychology research. *International Journal of Management Reviews, 9*(3), 211–236. https://doi.org/10.1111/j.1468-2370.2007.00210.x.

Antoniou, P., & James, M. (2014). Exploring formative assessment in primary school classrooms: Developing a framework of actions and strategies. *Educational Assessment, Evaluation and Accountability, 26*(2), 153–176. https://doi.org/10.1007/s11092-013-9188-4.

Ashford, S. J., & Black, J. S. (1996). Proactivity during organizational entry: The role of desire for control. *Journal of Applied Psychology, 81*(2), 199–214. https://doi.org/10.1037/0021-9010.81.2.199.

Ashford, S. J., & Cummings, L. L. (1983). Feedback as an individual resource: Personal strategies of creating information. *Organizational Behavior & Human Performance, 32* (3), 370–398. https://doi.org/10.1016/0030-5073(83)90156-3.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678.

Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., … Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education, 13*(1), 117–140. https://doi.org/10.1177/1478210314566733.

Black, P., & Wiliam, D. (2012). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 206–229). London, UK: SAGE.

Brusilovsky, P., Sosnovsky, S., & Yudelson, M. (2009). Addictive links: The motivational value of adaptive link annotation. *New Review in Hypermedia and Multimedia, 15*(1), 97–118. https://doi.org/10.1080/13614560902803570.

Bull, S. (2004). Supporting learning with open learner models. In *Proceedings of 4th hellenic Conference in Information and communication Technologies in education* (pp. 47–61) (Athens, Greece).

Bull, S., & Kay, J. (2007). Student models that invite the learner in: The SMILI:( ) Open learner modelling framework. *International Journal of Artificial Intelligence in Education, 17*(2), 89–120.

Bull, S., Mabbott, A., Gardner, P., Jackson, T., Lancaster, M., Quigley, S., et al. (2008). *Supporting interaction preferences and recognition of misconceptions with independent open learner models. Adaptive hypermedia and adaptive web-based systems*. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70987-9_9.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., & Wood, B. (2016). *Guidelines for assessment and instruction in statistics education (GAISE) college report 2016*. American Statistical Association.

Castro Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98–113. https://doi.org/10.1016/j.edurev.2007.04.001.

Chen, X., Breslow, L., & DeBoer, J. (2018). Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment. *Computers & Education, 117*, 59–74. https://doi.org/10.1016/j.compedu.2017.09.013.

Chew, P. K. H., & Dillon, D. B. (2014). Statistics anxiety update. *Perspectives on Psychological Science, 9*(2), 196–208. https://doi.org/10.1177/1745691613518077.

Corbalan, G., Kester, L., & Van Merriënboer, J. J. (2006). Towards a personalized task selection model with shared instructional control. *Instructional Science, 34*(5), 399–422. https://doi.org/10.1007/s11251-005-5774-2.

Crommelinck, M., & Anseel, F. (2013). Understanding and encouraging feedback-seeking behavior: A literature review. *Medical Education, 47*(3), 232–241. https://doi.org/10.1111/medu.12075.

Dirkx, K. J., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research, 107*(5), 357–364. https://doi.org/10.1080/00220671.2013.823370.

Drijvers, P., Boon, P., Doorman, M., Bokhove, C., & Tacoma, S. (2013). Digital design: RME principles for designing online tasks. In C. Margolinas (Ed.), *Proceedings of ICMI Study 22 Task Design in Mathematics Education* (pp. 55–62). Clermont-Ferrand, France: ICMI.

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351. https://doi.org/10.1016/j.compedu.2011.06.004.

Harks, B., Rakoczy, K., Hattie, J., Besser, M., & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology, 34*(3), 269–290. https://doi.org/10.1080/01443410.2013.785384.

Hattie, J., & Yates, G. (2014). Using feedback to promote learning. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (5th ed., pp. 45–58). Washington, DC: Society for the Teaching of Psychology.

Heitink, M. C., Van der Kleij, F., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review, 17*, 50–62. https://doi.org/10.1016/j.edurev.2015.12.002.

Hendriks, M. A. (2014). The influence on school size, leadership, evaluation, and time on student outcomes. *Four reviews and meta-analyses. (Doctoral dissertation, University of Twente, Enschede, the Netherlands)*. https://doi.org/10.3990/1.9789036538008. Retrieved from https://research.utwente.nl/en/publications/the-influence-of-school-size-leadership-evaluation-and-time-on-st.

Herder, E., Sosnovsky, S. A., & Dimitrova, V. (2017). Adaptive intelligent learning environments. In E. Duval, M. Sharples, & R. Sutherland (Eds.), *Technology enhanced learning* (pp. 109–114). Cham, Switzerland: Springer.

Kicken, W., Brand-Gruwel, S., & van Merriënboer, J. J. (2008). Scaffolding advice on task selection: A safe path toward self-directed learning in on-demand education. *Journal of Vocational Education and Training, 60*(3), 223–239. https://doi.org/10.1080/13636820802305561.

Kinicki, A. J., Prussia, G. E., Wu, B. J., & McKee-Ryan, F. M. (2004). A covariance structure analysis of employees' response to performance feedback. *Journal of Applied Psychology, 89*(6), 1057–1069. https://doi.org/10.1037/0021-9010.89.6.1057.

Krause, K., & Coates, H. (2008). Students' engagement in first-year university. *Assessment & Evaluation in Higher Education, 33*(5), 493–505. https://doi.org/10.1080/02602930701698892.

Long, Y., & Aleven, V. (2011). Students' understanding of their student model. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Artificial Intelligence in education, 15th international conference* (pp. 179–186). Berlin Heidelberg: Springer-Verlag. https://doi.org/10.1007/978-3-642-21869-9_25.

Mitrovic, A., & Martin, B. (2007). Evaluating the effect of open student models on self-assessment. *International Journal of Artificial Intelligence in Education, 17*, 121–144.

Morrison, E. W., & Cummings, L. L. (1992). The impact of feedback diagnosticity and performance expectations on feedback seeking behavior. *Human Performance, 5*(4), 251–264. https://doi.org/10.1207/s15327043hup0504_1.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education, 29*(1), 3–19. https://doi.org/10.1080/0260293042000160384.

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Frontiers in Psychology, 8*, 1196. https://doi.org/10.3389/fpsyg.2017.01196.

Robinson, J., Myran, S., Strauss, R., & Reed, W. (2014). The impact of an alternative professional development model on teacher practices in formative assessment and student learning. *Teacher Development, 18*(2), 141–162. https://doi.org/10.1080/13664530.2014.900516.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology, 46*(1), 561–584.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Tacoma, S., Sosnovsky, S., Boon, P., Jeuring, J., & Drijvers, P. (2018). The interplay between inspectable student models and didactics of statistics. *Digital Experiences in Mathematics Education, 4*(2), 139–162. https://doi.org/10.1007/s40751-018-0040-9.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157–167. https://doi.org/10.1016/j.chb.2014.05.038.

Timmers, C. F., Braber-van den Broek, J., & Van den Berg, S. (2013). Motivational beliefs, student effort, and feedback behavior in computer-based formative assessment. *Computers & Education, 60*(1), 25–31. https://doi.org/10.1016/j.compedu.2012.07.007.

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education, 20*(2), 1–55. https://doi.org/10.1080/10691898.2012.11889641.

Torenbeek, M., Jansen, E., & Suhre, C. (2013). Predicting undergraduates' academic achievement: The role of the curriculum, time investment and self-regulated learning. *Studies in Higher Education, 38*(9), 1393–1406. https://doi.org/10.1080/03075079.2011.640996.

Tuckey, M., Brewer, N., & Williamson, P. (2002). The influence of motives and goal orientation on feedback seeking. *Journal of Occupational and Organizational Psychology, 75*(2), 195–216. https://doi.org/10.1348/09631790260098677.

VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback-seeking process. *Journal of Applied Psychology, 82*(3), 390–400. https://doi.org/10.1037/0021-9010.82.3.390.

Van der Kleij, F., Timmers, C. F., & Eggen, T. (2011). The effectiveness of methods for providing written feedback through a computer-based assessment for learning: A systematic review. *Cadmo, 1*, 21–38. https://doi.org/10.3280/CAD2011-001004.