

Using student models to generate feedback in a university course on statistical sampling

Sietske Tacoma, Paul Drijvers and Peter Boon

Utrecht University, Faculty of Science, Utrecht, The Netherlands; s.g.tacoma@uu.nl

Due to the complexity of the topic and a lack of individual guidance, introductory statistics courses at university are often challenging. Automated feedback might help to address this issue. In this study, we explore the use of student models to provide feedback. The research question is how student models can be used to generate feedback to university freshman in an online course on statistical sampling. An online activity was designed and delivered to 40 Biology freshmen. Instruments for generating student models were designed and student models were generated. Four students were interviewed about the generated models, and about the differences with their own estimation of their understanding. Results show that it is possible to generate individual feedback from student work in an online learning activity and suggest that discussing differences between own estimations and generated student models can be a fruitful teaching strategy.

Keywords: Statistics, feedback, educational technology, higher education, student model.

Introduction

Many bachelor programs offer introductory courses in statistics. Success rates for these courses are often low, which makes them an obstacle for students in obtaining a bachelor's degree (Murtonen & Lehtinen, 2003; Tishkovskaya & Lancaster, 2012). One challenge is that such courses are often taught to large groups of students, making it difficult for teachers to provide individual guidance. The main challenge concerns the complex topics in these courses. One of these topics is statistical sampling, which involves concepts such as sampling distributions and sampling variability. Understanding statistical sampling does not only require students to understand these concepts, but also the connections between them (Castro Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007).

To address both the problem of individual guidance and the complexity of the subject matter, an approach using automated individual feedback can be promising. If appropriately designed and timed, feedback has the potential to increase student learning (Hattie & Timperley, 2007; van der Kleij, Feskens, & Eggen, 2015). Especially for large groups of students, automated feedback can add individual support that would otherwise be unattainable. For the case of statistical sampling, such automated feedback should aim at student understanding of the complex concepts involved. This asks for feedback on a global level, aggregated per concept, rather than on a local, task level. To gather input for such an overview, a specific digital assessment activity might be set up, but this would take valuable instruction time. A less time-consuming option may be to use student work in an online learning activity. Although such an activity is primarily designed for learning rather than assessing, it might still be possible to track the students' evolving understanding (VanLehn, 2008).

The question now is: how can we use students' solutions to tasks in an online learning activity to generate feedback on their knowledge of statistical sampling? This paper makes a start in answering this question by describing a prototypical approach through the use of student models.

Theoretical background

The theoretical background of this study includes two main elements: an analysis of the difficulties in the domain of statistical sampling, and the notion of student model.

Difficulties in the domain of statistical sampling

Samples are the key instruments to make inferences about a population. An important idea in making these inferences is that samples provide useful, but not complete, information about the population. This idea relates to two concepts: sample representativeness and variability. Sample representativeness means that for properly selected samples, sample characteristics will likely resemble those of the population. Sample variability means that not all samples are equal, and that sample characteristics do not necessarily meet population characteristics, and may not even be close. Making inferences from a sample involves a trade-off between these two concepts; a balance that is influenced by factors such as sample size and population variability (Batanero, Godino, Vallecillos, Green, & Holmes, 1994).

Castro Sotos et al. (2007) identify three main misconceptions that students may have about samples and sampling distributions. The first concerns the effect of sample size on the variance of the sample mean: as sample size increases, sample characteristics are likely to approach the population characteristics more and more. Many students misinterpret this so-called law of large numbers and use the sample representativeness heuristic to conclude that any sample's characteristics should be very similar to those of the population. The second misconception concerns the different distributions involved. Students often confuse the distribution of one sample of data with the distribution of sample means for several samples (Chance, del Mas, & Garfield, 2004). This can, for example, result in confusion between the standard deviation in a sample, the mean of the standard deviation over many samples, and the standard error of the sample mean. The third misconception concerns the central limit theorem, which states that for sufficiently large sample sizes, the sampling distribution of the sample mean can be approximated by a normal distribution. Students tend to wrongly extrapolate this theorem and believe that the larger the sample size, the closer the distribution of any statistic in the population will approximate a normal distribution (Bower, 2003).

Automated feedback through student models

In many online learning environments, automated feedback is offered at a task level or even at a step level. However, as we are interested in the students' conceptual understanding of sampling, it seems more relevant to provide the students with an overview of their knowledge of the entire domain, which, of course, is still based on the scores on single tasks. Such an overview for an entire domain is often called a student model (Brusilovsky & Millán, 2007; Bull, 2004). Student models can be used to adapt the educational intervention (i.e. the series of tasks) to the specific needs of the individual learner (Bull, 2004) and are in this role often invisible to the student. However, opening up the student model can promote learner reflection on his knowledge and understanding, and may help learners to monitor and plan their learning (Bull & Kay, 2007; Sosnovsky & Brusilovsky, 2015).

A student model contains a domain model and an overlay. The domain model consists of knowledge components (KC's) that each describe a piece of knowledge in the domain. All tasks in the learning activity are connected to one or more KC's in the domain model. The overlay contains a score for each KC, based on the student's performance on connected tasks, which describes the student's current understanding. The KC's in a domain model can be more or less coarse grained. An advantage of a fine-grained domain model is that it enables a very sophisticated and precise diagnose of the student's current understanding. However, when using a course-grained domain model, connections between tasks and the domain model are much easier to manage, while still a reasonable diagnosis can be accomplished (Sosnovsky & Brusilovsky, 2015).

In the light of this theoretical framework, the research question addressed here is: How can student models be used to generate feedback to university freshman in an online course on statistical sampling?

Methods

To address the research question, a prototypical environment to generate feedback on the understanding of statistical sampling through the use of student models was set up. In this explorative design research, 40 freshmen Biology participated. The design included an online activity on statistical sampling, and a domain model and Q-matrix for generating student models. Data collection included digital student work, a questionnaire and interviews with students. Analysis aimed at choices in generating overlays and describing the students' reactions to their student model.

Design of an online activity on statistical sampling

Gemiddelde en standaardafwijking van de steekproef

Hiernaast zie je weer zes steekproeven uit de populatie van lentetemperaturen. Op deze pagina bekijk je of de gemiddeldes en standaardafwijkingen van die steekproeven goede schattingen geven voor de gemiddelde dagtemperatuur en de standaardafwijking van de populatie.

Vul de tabel met gemiddeldes en standaardafwijkingen verder in. Rond af op één decimaal.

steekproef	gemiddelde (°C)	standaardafwijking (°C)
1	8,7	4,4
2	10,1	4,3
3
4	10,0	...
5	...	4,2
6	9,1	5,3

Je hebt hiervoor gezien dat het populatiegemiddelde is $\mu = 9,4$ °C en dat de standaardafwijking van de populatie is $\sigma = 4,5$ °C. Bij welke steekproeven is het steekproefgemiddelde m groter dan het populatiegemiddelde μ ? En bij welke is de steekproefstandaardafwijking s groter dan de populatiestandaardafwijking σ ?

gemiddelde m							standaardafwijking s						
stp	1	2	3	4	5	6	stp	1	2	3	4	5	6
$m > \mu$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	$s > \sigma$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
$m < \mu$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	$s < \sigma$	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Selecteer groter of kleiner en klik op **Klaar**

Selecteer groter of kleiner en klik op **Klaar**

Instellingen

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 alles opnieuw

Figure 1: Example page of the DME-activity on statistical sampling

The online activity on statistical sampling was designed in the frame of the Utrecht University project “Innovative remedial digital learning modules for statistics”, by an educational designer and the researcher (first author), in close collaboration with the teacher of the statistics course for biology students. For the design, the Freudenthal Institute’s Digital Mathematics Environment (DME, see Drijvers, Boon, Doorman, Bokhove, & Tacoma, 2013) was used. Aim of the designed activities was to deepen the students’ understanding of statistical sampling and sample variability. The activities contained theory, a simulation on sampling and questions about the students’ intuitions, the simulation, and the theory. The difficulties described in the theory section were addressed extensively. Students were able to enter answers to all questions and receive immediate feedback on the correctness of their response. For many tasks, hints and feedback on incorrect responses were designed. For an example page of the activity, see Figure 1.

Development of a domain model and Q-matrix

Through studying theory on statistical sampling¹, a set of knowledge components (KC’s) for the domain of sampling was identified by the researcher. As the intended use of the domain model was to present it to students, a rather coarse-grained approach was chosen and too detailed KC’s were avoided. Moreover, for a clear presentation to students, complete descriptions of the KC’s were formulated, as opposed to one or two words per KC. Four main KC’s were identified: Taking samples (procedure), Estimations based on a sample, Distribution of the sample mean, and Standard error. For each main KC, four detailed KC’s were identified.

After the domain model was designed, all tasks in the module were linked to the corresponding KC’s by the researcher. This resulted in a Q-matrix, in which entry (i, j) is 1 if task j is related to KC i, and 0 otherwise. The module contained 45 (sub-)tasks in total. For twelve tasks, the researcher judged that no KC’s were relevant. In Figure 1, for example, the students are asked to read off values from a table. This activity helps students understand the table-tool they will be working with in this module, but how well students read off the values does not involve their knowledge of any of the KC’s. For four subtasks, more than one KC was judged to be relevant. All other subtasks were connected to one KC.

Data collection: student work, questionnaire and interviews

The participants in this study were 40 biology students, who participated in the first year introductory course Experiment & Statistics at Utrecht University. The students first attended a lecture on sampling and worked on the designed online module in the week following the lecture. Three sources of data were collected in this small-scale explorative study:

- Student work: the DME stores all student work, including all attempts that students do before reaching a final answer. Student work was collected for all 40 students;

¹ Sources used are two text books on introductory statistics: Whitlock, M., & Schluter, D. (2009). *The analysis of biological data*. Greenwood Village, Colo: Roberts and Co. Publishers, and Field, A. P. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll)*. Los Angeles [i.e. Thousand Oaks], Calif.: SAGE Publications.

- A questionnaire, in which the KC's from the domain model were presented to students. Students were asked to give their own estimation of their understanding of each KC. The questionnaire was completed by seven students.
- Interviews, in which students were questioned about the appropriateness of the domain model and generated overlay, and about differences between their estimated and generated overlay. Out of the seven students who completed the questionnaire, four were interviewed.

Data analysis

In the analysis, the students' attempts were extracted from the DME, exported to Excel and prepared for generating overlays. For each student, an overlay was generated. Next, the generated overlays for the four students who would be interviewed were studied and remarkably high and low scores were recorded. Explanations were sought by studying tasks connected to the remarkable KC's and by analysing student explanations in the interviews.

The interviews were transcribed and the students' answers were aggregated by topic. Next, summaries for each topic were written to create a general image of the students' reactions to the overview, and to identify issues in the current calculation of the overlays.

Results

In addition to the domain model and the Q-matrix, which are described in the Methods section, the results of this study include the generation of overlays, and the students' reactions to their student model.

Generating overlays

To generate overlays, the students' interactions with the DME had to be translated into scores for each KC. Student interactions with the DME are stored as attempts. Three attempt types are possible: correct attempts, half-correct attempts (for example when a student still needs to round off an answer) and incorrect attempts. To generate an overlay, correct attempts were counted as 1, half correct attempts as 0.5 and incorrect attempts as 0. For each task, the mean attempt score was calculated by dividing the sum of the attempts by the number of attempts. For each KC, the overlay score was calculated as the mean attempt score of all tasks that were connected to this KC in the Q-matrix. See Figure 2 for an example.

KC	P1	P2	P3	P4	E1	E2	E3	E4	D1	D2	D3	D4	S1	S2	S3	S4
Generated overlay	1,00	1,00	0,88	—	0,89	0,50	0,61	0,75	0,83	0,67	0,75	0,67	0,80	0,50	0,57	0,83
Student's estimation	0,80	0,80	0,70	0,70	0,70	0,80	0,80	0,70	0,70	0,80	0,70	0,60	0,70	0,70	0,80	0,80

Figure 2: One student's generated overlay and his own estimation

The method used is not the only possible calculation method. Another option that was considered is taking only the student's first attempt for each task into account. The student's subsequent attempts are guided by the DME's immediate feedback and therefore do not directly reflect the student's knowledge, but rather a combination of this knowledge and the student's reaction to the immediate feedback. However, this method neglects the fact that students are likely to learn from the immediate feedback, and therefore this approach is left out of the analysis.

The generated overlays are different for different students, which confirms that they provide individual feedback indeed. Moreover, the scores for each individual student were more or less spread out, so students did score different for different KC's. This shows that our way of calculating discriminates between KC's, and hence can inform students on their understanding of the different KC's.

Students' reactions to their student model

The domain model was first presented to students in the questionnaire, in which students were asked to estimate their own overlay. In the interviews, the domain model was presented again, this time with the generated overlay. All four students said that they understood the domain model well and thought it formed a useful summary of the domain of statistical sampling. One student explicitly mentioned that he would use the domain model in his exam preparation.

The comparison between the generated overlays and the students' own estimations resulted in fruitful exchanges. Most students seemed to adopt the generated overlay as a true representation of their knowledge. Three students seemed to adjust their own estimation to the generated overlay. For example, when seeing the generated overview, one student concluded: "Apparently I can give myself higher grades than I did." The fourth student, however, thought that the activities in the DME were easier than other activities in the course, and therefore thought her knowledge of the topic was not as good as her work on the DME-activity suggested.

When asked about differences between their generated and estimated overviews, students came up with some meaningful explanations:

- One student made some initial mistakes on a certain KC, because he did not understand it correctly yet. Therefore, his generated score was low. But with the help of the DME's immediate feedback, he realized what he had understood wrongly, and therefore learned from these mistakes and the feedback. In his own overlay, he rated his knowledge in this KC as high.
- Some students were tempted to guess answers to see what immediate feedback the DME would provide. As in the previous explanation, such trial-and-error-behavior results in lower generated scores, but it is likely that students learn from the immediate feedback they obtain.
- Initial confusion about what students were required to do in the DME also resulted in low scores for some tasks. Here, a low score indicates difficulties with DME-interaction, rather than little knowledge or skills.

These explanations for difficulties show two things. First, these are exactly the difficulties that arise when learning material as opposed to assessment material is used for generating feedback. The tasks are above all designed to make students learn, and it is sometimes difficult to determine whether that learning has taken place before or after the student has answered a task. Second, the discussions with the students seemed to help them get a clearer picture of which KC's they understood and which they did not. So using student work to generate an overlay, confronting that with the student's own estimation and discussing differences seems to be a fruitful teaching strategy.

Conclusion and discussion

In this explorative study, we have shown an example of the use of student models for providing individual feedback in a university statistics course. We developed an online activity on statistical sampling, a domain model for the domain of statistical sampling and a Q-matrix connecting the tasks from the online activity to the domain model. Next, we used the students' work to generate overlays and presented the generated student models to the students, to give them more insight in their understanding of the different concepts involved in statistical sampling.

The generated student models were different for different students, which indicates that they indeed provided individual feedback. Students regarded the domain model as a useful summary of the domain of statistical sampling. As such, the domain model seems a useful instrument to confront students once again with difficult aspects of the domain of statistical sampling. Moreover, students regarded the generated overlay as a more or less true representation of their knowledge of the domain.

Asking students to compare their generated student model with their own estimations resulted in fruitful exchanges and therefore seemed a promising teaching strategy. Students tended to adjust their own estimation according to the generated model, but were also able to explain remarkable differences between their estimation and the generated model. These explanations often involved the immediate feedback provided by the DME, or, more general, the fact that the feedback is based on student interaction with an activity that is designed to learn from. Calculation methods that account more for this fact are available (VanLehn, 2008) and can be taken into account in future experiments.

Another lesson we have learned is that some tasks are important for the learning activity, but are not useful for the generation of overlays. This concerns, for example, tasks that serve to explain a tool or simulation to be used. Therefore, careful considerations should be made whether or not to include specific tasks in the Q-matrix.

In this study, with its explorative nature, we have shown that it is possible to generate useful student models, based on student work in an online learning activity. A next step is to investigate how these models can best be embedded in education to help students monitor and plan their learning.

Acknowledgments

We would like to thank Rutger Hermsen, teacher of the course, and Henk Hietbrink, co-designer of the course, for their engagement and all fruitful discussions on statistical sampling. We would also like to thank Johan Jeuring for his comments on previous versions of this paper.

References

- Batanero, C., Godino, J. D., Vallecillos, A., Green, D., & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematical Education in Science and Technology*, 25(4), 527-547.
- Bower, K. M. (2003). Some misconceptions about the normal distribution. Paper presented at the *Six Sigma Forum*, Milwaukee, WI: American Society for Quality.

- Brusilovsky, P., & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The adaptive web* (pp. 3-53). Berlin Heidelberg: Springer-Verlag.
- 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).
- 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.
- 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.
- Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi, & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Dordrecht: Kluwer Academic Publishers.
- 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.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Murtonen, M., & Lehtinen, E. (2003). Difficulties experienced by education and sociology students in quantitative methods courses. *Studies in Higher Education*, 28(2), 171-185.
- Sosnovsky, S., & Brusilovsky, P. (2015). Evaluation of topic-based adaptation and student modeling in QuizGuide. *User Modeling and User-Adapted Interaction*, 25(4), 371-424.
- 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.
- van der Kleij, F., Feskens, R., & Eggen, T. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes. A meta-analysis. *Review of Educational Research*, 85(4), 475-511.
- VanLehn, K. (2008). Intelligent tutoring systems for continuous, embedded assessment. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 113-138). New York, NY: Erlbaum.