MOONS, Filip

Investigating the use of recommender systems (AI) to make the reuse of feedback faster when assessing handwritten mathematics tasks

1. Introduction

Written feedback is a powerful formative assessment practice in math teaching yet highly labour-intensive. Ways to provide it include digital assessments with fully automated feedback. Unfortunately, learners solve higher-order thinking questions more naturally using paper and pencil (Hoogland & Tout, 2018). Making the process of giving feedback on handwritten tasks less labour-intensive is thus crucial. Therefore, in this research project, we investigate semi-automated assessment (SA): teachers write feedback on a computer on handwritten mathematics tasks by students, and the computer saves the given feedback so that it can easily be reused when a following student makes a similar mistake. To write 'reusable' feedback, we invented atomic feedback: feedback is written as a hierarchical list of feedback items (see below), making each item a separate option to reuse. In a previous study, the SA software tool was developed to provide atomic feedback and suggested items in a non-intelligent way: it tried to match verbatim what a teacher was writing with previously used items. However, one of our study results is that teachers tend to forget the exact formulation of their feedback items, preventing them from retrieving them in the current system (Moons et al., 2022). That is why we are integrating recommender systems in our SA software tool, making the suggestion system smarter; removing the need to remember (words of) previously written feedback by heart. In the following sections, we (1) introduce the idea of atomic feedback, (2) give an overview of the previous research results (including a description of the SA tool) and conclude with a (3) description of our current research on integrating recommender systems.

2. Atomic feedback

How to write feedback that can easily be reused for other students? Long pieces of classic feedback are often too targeted to a specific student. Hence, we suggest atomic feedback (see Figure 1): a collection of form requirements for written feedback that have been shown to make feedback significantly more reusable (Moons et al., 2022). To write an atomic feedback item (= single bullet point), teachers must:

- identify independent errors,
- write small feedback items for each error separately, or
- if an error is a structural mistake (misconception), create 2 feedback items:

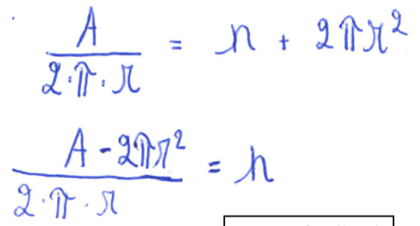
In: P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht & P. Scherer (Hrsg.),
Beiträge zum Mathematikunterricht 2024.
57. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.
https://doi.org/10.37626/GA9783959872782.0

- one item containing feedback on the misconception in general,
- one or more sub-items addressing specific mistakes.

Atomic feedback items ultimately form a point-by-point list covering only feedback items relevant to a student's solution. The list can be hierarchical in order to cluster feedback items that belong together. A comparison of classic (PP condition) and atomic feedback (SA condition) is presented in Figure 1. This comprehensive example demonstrates that classic feedback reports can be rephrased as atomic.

Student's solution

Manipulate the formula: $A = 2\pi rh + 2\pi r^2$ to h



Classic feedback

Mind the fact that the dominant operation on the right-hand side of the equation is an addition! The division of the left-hand side by $2\pi r$ is, therefore, not helpful. Moreover, $2\pi r$ is a common factor of the right-hand side, but the sum wasn't completely divided by it (second addend not divided). Although your final answer is correct, the way it is written makes it look like a coincidence. Going from the first to the second step, you would normally subtract $2\pi r^2$ from both sides, meaning that it shouldn't be placed directly in the numerator, as you should make the denominators the same.

Atomic feedback

- First step
  - Dominant operation on the right side is an addition!
    - ⋆ Division of left-hand side is not helpful
    - ⋆ $2\pi r$ is a common factor of the right side, but:
      - · sum wasn't completely divided by it
      - · the second addend was not divided
- Second-step
  - Your final answer is correct, but:
    - ⋆ It looks like a coincidence.
    - ⋆ You should subtract $2\pi r^2$ from both sides.
    - ⋆ Mistake with making the denominators the same!
      - · $2\pi r^2$ shouldn't be directly in the numerator.

Fig. 1: A comparison between classic (PP) and atomic (SA) feedback

3. Previous research results

For a crossover study with 45 math teachers in Flanders (Belgium), we developed an SA assessment tool for the e-learning platform Moodle (see Figure 2). When formulating feedback with the SA tool, they could use keyboard shortcuts (e.g., to indent, insert, or remove items) to create a hierarchical list of atomic feedback items (see Figure 2a). When a teacher typed something, the system searched the feedback items that had already been entered to detect possible matches. Thus, the suggestion system was non-intelligent. In order to compare the SA approach, we had a traditional, pen-and-paper-based written feedback condition (PP). Teachers received 60 student tasks on linear equations: on half of them, they wrote feedback using SA, the other half using PP.

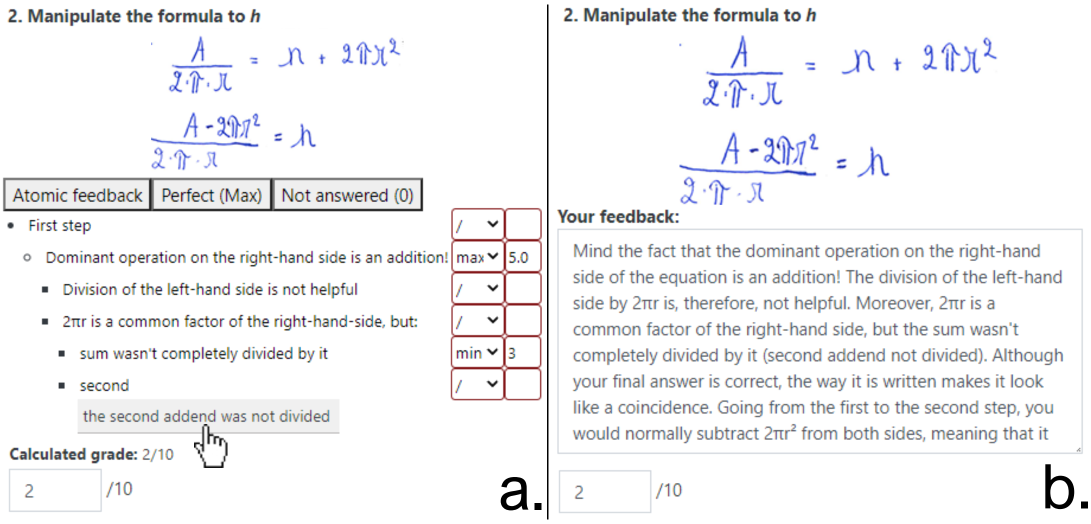


Figure 2: Screens of the tool in the SA condition (a) & PP condition (b)

Our study (Moons et al., 2022) showed that atomic feedback items were significantly more reused than items that did not adhere to the definition of atomic feedback (odds ratio: 2.6). Meanwhile, results showed no significant time differences between conditions (see Figure 3): teachers tend to become faster and faster in writing feedback in both conditions, as they get used to the task under assessment. Nevertheless, the teachers used the gained time to write significantly more feedback with a medium effect size ($d = 0.41$).

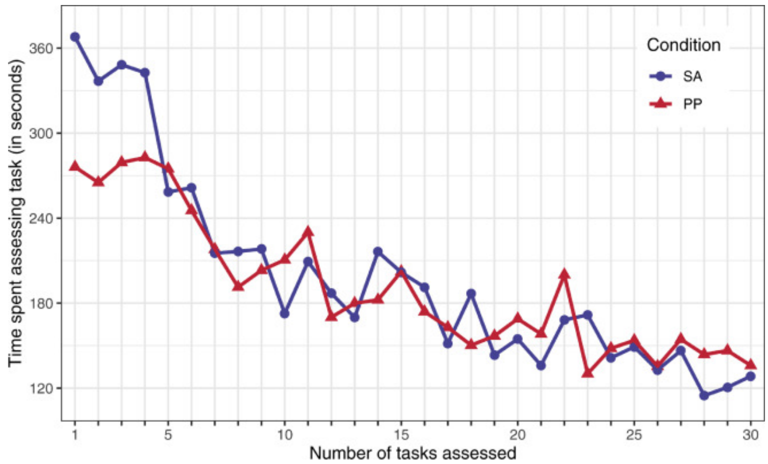


Figure 3: Evolution of teachers' assessment time in both conditions

Even though the atomic feedback items were reused significantly more than the non-atomic feedback items, most of the atomic items were not reused. One possible explanation is the non-intelligent suggestion system of feedback items during the experiment, which led us to integrating recommander systems.

4. Integrating recommender systems

A recommender system (Mohanty et al., 2020) is a type of software application or algorithm designed to analyse user preferences, behaviours, and patterns to provide personalised suggestions or recommendations. These systems are commonly used in various online platforms, such as e-commerce websites, streaming services, and social media. Recommender systems aim to enhance user experience by offering relevant content, products, or services tailored to individual tastes and preferences. There are different approaches to building recommender systems, including collaborative filtering, content-based filtering, and hybrid methods, each with its unique way of predicting and delivering recommendations. We experience some limiations with our integration: first, we experience a so-called *cold start*: no items are available for re-use for a new task. Moreover, we can only rely on content-based filtering as we can not suggest feedback items from other teachers.

Concretely, we experiment with feeding our recommending system with the following information:

- *Item popularity:* Suggest more popular items first, by just counting the number of times a feedback item was re-used.
- *Item distance:* Calculate the average distance of a feedback item to already selected items. The distance is measured in how many times items co-occured previously. Closer items are suggested first.
- *Error location:* We now allow teachers to indicate where an error has occurred in a handwritten solution, which provides helpful information about which feedback items are appropriate. However, this is not perfect as some students have a different way of writing down their solutions.

The first results indicate a recommender system can make the reuse of feedback items much more straightforward, making the highly labour-intensive - yet critical - job of giving feedback to students easier for teachers. Of course, ensuring feedback quality remains vital.

5. References

Hoogland, K., & Tout, D. (2018). Computer-based assessment of mathematics into the twenty-first century: Pressures and tensions. *ZDM, 50*(4), 675–686. https://doi.org/10.1007/s11858-018-0944-2

Mohanty, S. N., Chatterjee, J. M., Jain, S., Elngar, A. A., & Gupta, P. (2020). *Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries*. Wiley.

Moons, F., Vandervieren, E., & Colpaert, J. (2022). Atomic, reusable feedback: A semi-automated solution for assessing handwritten tasks? A crossover experiment with mathematics teachers. *Computers and Education Open, 3*, 100086. https://doi.org/10.1016/j.caeo.2022.100086