Computers & Education 86 (2015) 260–267

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems

Yigal Attali

Educational Testing Service, Rosedale Rd., MS-10-R, Princeton, NJ 08541, USA

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form 19 August 2015
Accepted 21 August 2015
Available online 25 August 2015

Keywords:
Computer-assisted instruction
Feedback
Formative assessment
Mathematics
Transfer

ABSTRACT

In a study of mathematics problem solving, the effect of providing multiple-try feedback on later success in solving similar problems was examined. Participants solved mathematics problems that were presented as either multiple-choice or open-ended questions, and were provided with one of four types of feedback: no feedback (NF), immediate knowledge of the correct response (KCR), multiple-try feedback with knowledge of the correct response (MTC), or multiple-try feedback with hints after an initial incorrect response (MTH). Results showed that gains in performance were larger in the open-ended than multiple-choice condition. Furthermore, gains under NF and KCR were similar, gains were larger under MTC than KCR, and gains were larger under MTH than MTC. The implications of these results for the design of assessments for learning are discussed.

1. Introduction

Providing feedback regarding task performance is one of the most frequently applied psychological interventions. Simple feedback helps learners verify performance expectations, judge their level of understanding, and become aware of misconceptions. Instructional feedback may also provide clues about the best approaches for correcting mistakes and improving performance. In one historical review on feedback, Kulhavy and Stock (1989) summarized that effective feedback provides the learner with verification, a judgment of whether an answer is correct, and elaboration, additional information to help the learner. However, despite a huge body of literature, the specific mechanisms relating feedback to learning are still not well understood. Historical reviews and meta-analyses on the subject describe the findings as “inconsistent,” “contradictory,” and “highly variable” (Azevedo & Bernard, 1995; Kluger & DeNisi, 1996; Shute, 2008).

Even one simple feedback mechanism, asking the student who did not answer a question correctly to try and correct the answer, is not widely researched and understood. Whereas human tutors rarely provide the correct answer immediately after an incorrect response (Lepper & Woolverton, 2002), multiple-try feedback is rarely applied in assessment, although it has a long history in educational technology. The first attempt to automate this approach was Pressey (1926, 1950). His “teaching machine” presented a multiple-choice question and provided immediate feedback on the correctness of a response (selected by pressing the appropriate key). The student repeatedly selected answers until the correct answer was chosen, hence the term “answer-until-correct.” Pressey (1950) reviewed several studies that showed positive long-term learning effects of answer-until-correct. However, later research showed mixed results (some of these results in programmed instruction are

E-mail address: yattali@ets.org.

http://dx.doi.org/10.1016/j.compedu.2015.08.011
0360-1315/© 2015 Published by Elsevier Ltd.

reviewed by Jaehnig & Miller, 2007). An explanation of these mixed results was proposed by Clariana and Koul (2005). Their review of the literature suggests that multiple-try feedback is less effective than other forms of feedback when criterion test items are taken verbatim from lesson materials or when students are asked to remember facts, but more effective when criterion test items require higher-order generalizations and transfer of learning. This distinction is explained from a generative learning perspective (Wittrock, 1992). When test items require comprehension and understanding of material, such as in mathematical problem solving, the request to try again provides an opportunity for elaboration and reorganization of information that may be beneficial for learning. However, when the situation requires a simpler stimulus–response association, such as in fact retrieval, additional trials (necessary when these associations are not established yet) may in fact interfere with memory processes.

The purpose of this study was to further explore the effect of multiple-try feedback for a higher-order assessment for learning in mathematics problem solving. To do that, participants solved problems in different feedback conditions. They were then assessed in their ability to transfer their understanding from the initial problem solving and feedback phase to different problems that had the same underlying structure. Because of the shared underlying structure of the initial and similar problems, we expect performance on the similar problems to be dependent on problem solving experiences during initial problem solving. In particular, because we experimentally manipulate these experiences (in ways that will be described below), differences in performance gains (from initial to similar problems) across conditions are indicative of experiences that foster more (or less) access to and understanding of this underlying structure.

Two other issues were explored in this study. The first concerns the possible difference between selected-response and constructed-response assessments. Selected-response tests are widely used because of the speed and ease of grading them. However, they have long been criticized for not encouraging productive or creative thinking (Martinez, 1999).

One possible concern with selected-response items in the context of assessments for learning is that the distractors (incorrect options) expose students to erroneous information. Even if students select the correct option, they may acquire incorrect knowledge just by reading and processing the distractors (Roediger & Marsh, 2005). Butler and Roediger (2008) found that in comparison with a no-feedback condition, both immediate and delayed feedback increased the proportion of correct responses and reduced the proportion of intrusions (i.e., distractor responses from the initial test) on a delayed cued recall test.

Another concern with the selected-response item format is that it allows students to produce answers effortlessly and mindlessly. The concept of mindfulness (Salomon & Globerson, 1987) has been used to explain the successful implementation of assessments for learning and feedback in particular (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991). Feedback can promote learning if it is received mindfully, but can inhibit learning if it encourages mindlessness. Therefore, the lower effort required in answering selected-response items may reduce their effectiveness in promoting learning.

In a multiple-try test in particular, it is possible that students would exert less effort in correcting initial errors in selected-response items than constructed-response items, because of the temptation to simply click on a different option without much thought. Indirect evidence for this was recently reported by Attali, Laitusis, and Stone (2013), who found that in a multiple-try mathematics problem solving test, students tended to answer much more quickly in correcting selected-response errors than constructed-response errors. In addition, partial credit scores based on revised answers showed significantly higher measurement precision for the constructed-response format but not for the selected-response format, suggesting that revised answers were not useful in better measurement of partial knowledge.

Despite these concerns, comparisons of selected-response and constructed-response formats in the context of assessments for learning have produced mixed results. For example, Clariana and Lee (2001) compared learning of definitions used in a university course using either multiple-choice or constructed-response questions with immediate feedback. In a post-test with the same terms no significant differences in performance were found across study conditions.

Similarly, in Smith and Karpicke (2014) participants read texts, then answered questions about the texts (either multiple-choice or short answer), and finally answered the same questions 1 week later. They found no advantage to practicing with short answer questions.

On the other hand, in Kang, McDermott, and Roediger (2007) participants similarly read texts, answered questions in different formats, and finally answered the same questions 3 days later. They found an advantage for intermediate practice with short answer questions, but only when feedback was provided for these answers (an initial study with no feedback found an advantage for multiple-choice questions).

These studies are similar in their focus on retention of studied verbal material: questions are answered a first time under different feedback conditions and question types and a second time after some time elapsed. In contrast, the focus of the current study is on mathematical problem solving and participants are not presented with the same problem twice. The use of different problems with the same underlying structure provides an opportunity to explore the learning support of different item types in a different cognitive setting. In this context it is not enough for participants to remember that a similar problem was previously answered. They also need to activate the appropriate procedures for solving the problem and apply these procedures correctly to arrive at the right answer. The increased effort required to answer constructed-response items may have more beneficial consequences in this procedural context than in the declarative context of previous research.

Another issue explored in this study concerns the possible beneficial effect of providing the student additional information following incorrect answers in the form of hints. In general, the answer-until-correct procedure has been applied with minimum correctness feedback ("your answer is incorrect, try again"). However, tutors will help and challenge students who are unable to answer a question by providing them with a hint – a partial solution that either suggests to the student the