Learning and Instruction 33 (2014) 92–107

Contents lists available at ScienceDirect

Learning and Instruction

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

Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving

Martine Baars [a,*], Sigrid Vink [a], Tamara van Gog [a], Anique de Bruin [b], Fred Paas [a,c]

[a] Institute of Psychology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
[b] Department of Educational Development and Research, Maastricht University, The Netherlands
[c] Interdisciplinary Educational Research Institute, University of Wollongong, Australia

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form
14 February 2014
Accepted 18 April 2014
Available online 21 May 2014

Keywords:
Self-assessment
Judgments of learning
Monitoring accuracy
Worked examples
Problem solving

ABSTRACT

Both retrospective and prospective monitoring are considered important for self-regulated learning of problem-solving skills. Retrospective monitoring (or self-assessment; SA) refers to students' assessments of how well they performed on a problem just completed. Prospective monitoring (or Judgments of Learning; JOLs) refers to students' judgments about how well they will perform on a (similar) problem on a future test. We investigated whether secondary education students' SA accuracy could be improved by training (Experiment 1 and 2), or by providing assessment standards (Experiment 2), and whether this would also affect the accuracy of JOLs. Accurate assessment of past performance might provide a good cue for judging future performance. Both Experiment 1 and 2 showed no effect of training on SA or JOL accuracy, but SA and JOLs were positively correlated with each other and negatively with effort. Providing standards did improve SA and JOL accuracy on identical problems, and performance on all problems.

1. Introduction

Self-regulated learning can only be optimally effective for learning outcomes when students are able to accurately monitor their own performance and use this information to choose what to study again or what to study next (e.g., Metcalfe, 2009; Winne & Hadwin, 1998; see also recent special issues by Alexander, 2013; De Bruin & Van Gog, 2012). As such, accurate monitoring seems to be a pivotal aspect of self-regulated learning. Monitoring can be measured both *retrospectively*, by asking students to judge their performance on a task just completed, which is also known as self-assessment (Kostons, Van Gog, & Paas, 2012) or a self-score judgment in verbal tasks (Lipko et al., 2009; Rawson & Dunlosky, 2007) and *prospectively*, by asking students to predict their performance on that task on a future test, which is also known as a *Judgment of Learning* (JOL; e.g., Koriat, Ackerman, Lockl, & Schneider, 2009a, 2009b; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991). Monitoring accuracy can then be determined by comparing students' self-assessed or predicted performance with their actual performance on a task. The more accurate monitoring is, the better participants are assumed to be able to keep track of their learning process, and the better they might be able to regulate it. Research has shown, however, that accurately monitoring their own performance is hard for students, and accuracy of both self-assessments (Bjork, 1999; Dunlosky, Rawson, & McDonald, 2002) and JOLs (Dunlosky & Lipko, 2007; Maki, 1998; Serra & Metcalfe, 2009; Thiede, Griffin, Wiley, & Redford, 2009) is often low, but there are instructional techniques that seem to improve accuracy.

However, even though problem-solving tasks play an important role in education, only very few studies have investigated monitoring accuracy when learning to solve problems in educational contexts, and it is therefore a major question whether and how JOL accuracy when learning to solve problems can be *improved* (for an exception, see Baars, Visser, Van Gog, De Bruin, & Paas, 2013). In the next sections, we will discuss research showing how monitoring accuracy can be improved, in light of the aims of the present study, which were to investigate whether the accuracy of secondary education students' self-assessment of problem-solving tasks can be enhanced by training, providing standards, or both, and whether increased self-assessment accuracy (retrospective) would also enhance JOL accuracy (prospective).

* Corresponding author. Tel.: +31 10 408 9021; fax: +31 10 4089009.
E-mail addresses: m.a.baars@gmail.com, m.a.baars@hr.nl (M. Baars).

http://dx.doi.org/10.1016/j.learninstruc.2014.04.004
0959-4752/

1.1. Improving self-assessment accuracy

Two techniques that have been shown to improve self-assessment (SA) accuracy are training and using standards to assess performance. Kostons et al. (2012) showed that SA skills can be trained by means of modeling examples that show students how to assess performance on biology problem-solving tasks. They trained secondary school students to assess their own performance and, based on that assessment, to select an appropriate task for further studying. In this training students were shown four computer screen recordings of human models who verbally explained how they: 1) solved a heredity problem, 2) assessed their own performance on that problem by assigning 1 point for each step in the procedure they felt they had performed correctly (i.e., no standard was used), and 3) selected a new task to study next, at an appropriate level of difficulty given their assessed performance in combination with the amount of mental effort they had to invest to reach that performance. They also investigated the effects of training only SA skills, with modeling examples showing only step 1 and 2, or training only task selection (step 1 and 3). Kostons et al. found that SA and task-selection skills improved when these had been trained, and that it was necessary to train both; they found no effects of training SA skills on task selection accuracy or vice versa (Experiment 1). Moreover, when students engaged in self-regulated learning after an SA and task selection training, their learning gains were higher than for students who had not been trained (Experiment 2).

Another technique that was found to improve SA accuracy, at least when learning from text, is using standards. Rawson and Dunlosky (2007) showed that students were better able to assess the correctness of their own test performance when they were provided with an assessment standard, that is, a description of the correct answer to compare their own answer to. They asked college students to self-assess the quality of their recall of key concepts from textbooks, by assigning themselves no credit, full credit, or partial credit. Students overestimated their own recall performance, but when they were provided with a standard, consisting of the correct definition of the key concepts, their overestimation was smaller. Lipko et al. (2009) replicated these findings with middle school children. Thus, by having correct definitions available as standard for evaluation, students are better able to recognize incorrect responses, which reduces their overconfidence and leads to better calibration of their assessment and their actual performance. In a cyclical learning process, in which learners continue studying after the self-assessment, standards might also improve learning outcomes, because they also provide learners with feedback regarding their own performance and correct responses (Butler & Winne, 1995). Indeed, Rawson and Dunlosky (2007) found that performance on a criterion test improved when students had used standards to assess their performance on a practice test of definitions of the key concepts. To the best of our knowledge, the effects of standards on calibration of SA, JOLs, and learning outcomes, have not yet been tested with problem-solving tasks.

In sum, SA accuracy can be improved by training (Kostons et al., 2012) or by using standards (Lipko et al., 2009; Rawson & Dunlosky, 2007). It would be interesting to investigate whether the findings by Kostons et al. can be replicated using written worked examples to train SA (instead of video-based modeling examples) as these might be easier to create and implement, and whether *combining* training prior to making SAs of problem-solving tasks with standards provided while making those assessments, would be more effective than either method alone.

1.2. Improving accuracy of judgments of learning

Many studies have investigated the accuracy of JOLs when learning word pairs (for a review see Rhodes & Tauber, 2011) or when learning from expository texts (for a review see Thiede et al., 2009). They have shown that relative accuracy of JOLs when learning from more complex materials like expository text or problem-solving tasks is very low, but that adding so-called 'generation strategies', helps students to make more accurate JOLs. Such strategies make students actively generate (part of) the learning materials after studying them, focusing their attention on the gist of the material or the underlying structure of the material. For example, it was found that generating keywords (De Bruin, Thiede, Camp, & Redford, 2011; Thiede, Anderson, & Therriault, 2003), making summaries (Thiede & Anderson, 2003), making concept maps (Thiede, Griffin, Wiley, & Anderson, 2010), and self-explaining (Griffin, Wiley, & Thiede, 2008) improved the accuracy of JOLs when learning from text. In addition, generating keywords did not only improve JOL accuracy but also affected regulation and led to greater test performance (Thiede et al., 2003).

The cue utilization framework (Koriat, 1997) can explain the effect of generation strategies on JOL accuracy. According to this framework, JOL accuracy is the result of the cues that are used to make a JOL and the extent to which these cues are diagnostic for future test performance. Generating keywords or summaries, or self-explaining a text, are all activities that provide participants with insight into the quality of their representation of the text, and they can use this information when making JOLs. The cues provided by such generation strategies are more indicative for future test performance than cues learners would spontaneously use, and therefore lead to more accurate JOLs (Thiede, Dunlosky, Griffin, & Wiley, 2005; Thiede et al., 2009). It should be noted that accuracy in those studies was mostly defined as *relative* accuracy, measured by calculating a Goodman-Kruskal gamma correlation (Thiede et al., 2009). This shows whether participants are able to discriminate between different items, it does not give any information on how accurate they were in predicting their performance per item. However, *absolute* measures of accuracy that do show the precision of JOLs, like bias (i.e., the difference between self-assessed and actual performance, with positive values indicating overestimation and negative values underestimation) or absolute deviation scores (i.e., without direction), have also been used to analyze JOL accuracy (Baars, Gog, Bruin, & Paas, 2014; Maki, Shields, Wheeler, & Zacchilli, 2005; Mengelkamp & Bannert, 2010; Schraw, 2009).

With regard to problem solving tasks, only few studies have investigated JOL accuracy and how to improve it (e.g., De Bruin, Rikers, & Schmidt, 2005, 2007). Recent research showed that when acquiring problem-solving skills from worked examples, which is an effective and efficient way of learning to solve problems compared to engaging in problem-solving practice (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller, Van Merriënboer, & Paas, 1998; Van Gog & Rummel, 2010), the use of a generation strategy was effective for improving JOL accuracy in terms of bias (Baars et al., 2014). That is, when students tested their knowledge after studying a worked example, by means of solving a problem on their own, bias was reduced. Presumably, engaging in problem solving after worked example study provides a learner with relevant cues on which to base their JOLs.

This would mean, basically, that students use cues from past performance (i.e., the problem they just solved) to judge future performance, which is what Griffin, Jee, and Wiley (2009) refer to as the postdiction route to JOLs. The postdiction route is based on the Memory for Past Test (MPT) heuristic (Finn & Metcalfe, 2007). According to the MPT heuristic, when making a JOL about word pairs following a practice test, learners will use their feeling of