Learning and Instruction 44 (2016) 1–10

Contents lists available at ScienceDirect

Learning and Instruction

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

A two-process model of metacognitive monitoring: Evidence for general accuracy and error factors

Antonio P. Gutierrez [a, *], Gregory Schraw [b], Fred Kuch [b], Aaron S. Richmond [c]

[a] Department of Curriculum, Foundations, & Reading, Georgia Southern University, P.O. Box 8144, Statesboro, GA 30460-8144, USA
[b] Department of Educational Psychology, University of Nevada, Las Vegas, 4505 Maryland Parkway, Box 453003, Las Vegas, NV 89154-3003, USA
[c] Department of Psychology, Metropolitan State University of Denver, 890 Auraria Parkway, Denver, CO 80204, USA

ARTICLE INFO

Article history:
Received 16 October 2014
Received in revised form
16 February 2016
Accepted 20 February 2016
Available online 1 March 2016

Keywords:
Metacognition
Monitoring
Self-regulation
Judgment accuracy and error

ABSTRACT

We examined the latent structure of metacognitive monitoring judgments using hierarchical confirmatory factor analysis to compare five competing theoretical models with respect to domain-specific versus domain-general monitoring processes. We expected our results to support a domain-general monitoring model. Of the five models, the domain general monitoring model provided the best fit. In this model, level-1 domain-specific accuracy and error factors for each of the three tests loaded on second-order domain-general accuracy and error factors, which then loaded on a third-order general monitoring factor. This model suggest that metacognitive monitoring consists of two different types of cognitive processes, one that is associated with accurate monitoring judgments, and one that is associated with error in monitoring judgments. We discuss the theoretical and practical instructional implications of our findings.

1. Introduction

Metacognitive monitoring refers to the relationship between task performance and a judgment about that performance (Boekaerts & Rozendaal, 2010; Efklides, 2008; Winne & Nesbit, 2009). Monitoring may be applied in a variety of activities such as judgments of learning, understanding, and performance either before or after a task to promote self-regulation (Efklides, 2011). The prototypical format in performance monitoring studies is to answer a test item and judge whether one's answer is correct or not (i.e., make a performance accuracy judgment) as shown in Table 1. This framework yields a 2 × 2 data matrix with four distinct outcomes. Cell *a* corresponds to correct performance that is judged to be correct. This outcome represents prototypically accurate monitoring. Cell *d* represents incorrect performance that is judged to be incorrect, which also represents accurate monitoring. In contrast, cell *b* reflects incorrect performance that is judged to be correct. These outcomes have been characterized as *overconfidence* (Koriat, 2012; Pieschl, 2009; Stankov, 2000; Stankov & Crawford, 1996) or as an *illusion of knowing* (Serra & Metcalfe, 2009). Research shows that males often are more overconfident than females (Lundeberg & Mohan, 2009; Stankov & Lee, 2008), but that overconfidence for both genders decreases over time and with practice (Hadwin & Webster, 2013). Cell *c* reflects correct performance that is judged as incorrect. These outcomes have been characterized as *underconfidence* (Dinsmore & Parkinson, 2013) or as an *illusion of not knowing* (Serra & Metcalfe. 2009). We refer to the contents of cell *b* and cell *c* throughout the remainder of this study as overconfidence and underconfidence respectively. Previous research reported a general confidence factor based on confidence ratings across different tasks that was statistically independent of ability, metacognitive knowledge, and beliefs (Stankov & Lee, 2008; Stankov, Lee, Luo, & Hogan, 2012). However, these studies did not attempt to partition variance attributable to accurate judgments versus inaccurate judgments. Although polytomous performance judgments may be made leading to an *m* x *n* matrix, we focus on dichotomous judgments in the present research to better compare the relationships between the four mutually exclusive types of performance-judgment outcomes in Table 1.

* Corresponding author.
E-mail addresses: agutierrez@georgiasouthern.edu (A.P. Gutierrez), gschraw@unlv.nevada.edu (G. Schraw), kuch@unlv.nevada.edu (F. Kuch), arichmo3@msudenver.edu (A.S. Richmond).

http://dx.doi.org/10.1016/j.learninstruc.2016.02.006
0959-4752/

Table 1
A 2 × 2 performance-judgment data array for monitoring Accuracy.

	Performance		
Monitoring Judgment	Correct	Incorrect	Row Marginals
Correct	*a*	*b*	a + b
Incorrect	*c*	*d*	c + d
Column Marginals	a + c	b + d	a + b + c + d

2. Domain-specific versus domain-general views of monitoring

Previous research suggests that metacognitive monitoring can be explained in two different ways, which we refer to as the *domain-specific* and *domain-general monitoring hypotheses*. The domain-specific hypothesis states that monitoring accuracy is a function of domain expertise that is situated within a specific content domain (e.g., mathematics), sub-domain (e.g., algebra), or task (e.g., proofreading). This perspective is consistent with theories of expertise and deliberate practice in which monitoring improves as knowledge and performance skills increase because knowledge and skills can be used as benchmarks to gage judgments about one's performance (Duncan, 2007; Kelemen, Frost, & Weaver, 2000; Tricot & Sweller, 2014). In addition, experts may be more automated in different aspects of the task, which enables them to allocate more cognitive resources to assessing the accuracy of their performance judgments (De Bruin, Rikers, & Schmidt, 2005). Domain-specific theories predict a positive correlation between level of expertise and monitoring accuracy within a content domain, but not between domains.

Kelemen et al. (2000) supported the domain-specific hypothesis by comparing four different types of metacognitive judgment tasks across college students, including ease-of-learning judgments, judgments of learning, feeling-of-knowing judgments, and metacomprehension predictions. Results showed low test–retest reliability of accuracy for all the predictions, which was interpreted as evidence in support of domain-specific monitoring. Similarly, Ozuru, Kurby, and McNamara (2012) compared two different types of metacomprehension judgments referred to as judgments of difficulty and predictions of performance and found that each type of judgment was related to different information sources within the to-be-learned materials. When taken together, these studies suggest that different types of metacognitive tasks draw on domain-specific processes.

In contrast to the domain-specific view, the domain-general monitoring hypothesis states that skilled adult learners construct a repertoire of general regulatory skills that enables them to make accurate judgments of performance even in low-knowledge domains. This repertoire may include a variety of metacognitive skills such as goal setting, strategy management, self-explanation and self-testing (Azevedo & Witherspoon, 2009), time (McNamara & Magliano, 2009), and comprehension monitoring (Huff & Nietfeld, 2009; Nietfeld, Cao, & Osborne, 2005). This perspective is consistent with the claim that general skills emerge later in development, are preceded by domain-specific skills, and emerge only after considerable effort has been devoted, either implicitly or explicitly, to integrating monitoring competencies across diverse domains (Chiappe & MacDonald, 2005; van der Stel & Veenman, 2010; Winne & Nesbit, 2009; Zohar & David, 2009). In particular, van der Stel and Veenman (2010) found in a 2-year study of middle school students that both domain-specific and domain-general skills were present. However, they concluded that domain-general skills emerge from and become dominant to domain-specific skills as children grow older. For this reason, the domain-general monitoring hypothesis predicts positive correlations across content domains, as well as a significant correlation between monitoring accuracy within a domain and future performance (Mengelkamp & Bannert, 2010).

Schraw and colleagues (Schraw, Dunkle, Bendixen, & Roedel, 1995; Schraw & Nietfeld, 1998) provided support for the domain-general monitoring skill by comparing accuracy across a variety of different content domains such as vocabulary knowledge, analogies, and working memory tasks. They found that confidence and