The Best Measure of The Accuracy of Metacognitive Monitoring: Measures Determining the Consistency and Generality of Metacognition

En-Ping Liang   Ju-Whei Lee

Department of Psychology, Chung Yuan University
Tau-Yuan, Taiwan. 320

The Goodman-Kruskal gamma correlation coefficient (denoted as G) was evaluated by Nelson (1984) as the best measure for assessing the accuracy of metacognitive monitoring. Consequently, it was widely adopted by many researchers in the field of metacognition. Recently, some researchers found that the value of the G measure could not accurately reflect the individual's metacogntion (Schwartz & Metcalfe, 1994), it varied with item difficulty (Weaver & Bryant, 1995), and it was unstable within a single domain (Thompson & Mason, 1996). Meanwhile, some scholars used the mean probability score, bias, the calibration index, the discrimination index, and the adjusted normalized discrimination index (denoted separately as $\overline{PS}$, Bias, CI, DI, and ANDI) to evaluate a subject's accuracy of metacognitive monitoring. (e.g., Koriat & Goldsmith, 1996; Maki, 1998; Schraw, Dunkle, Bendixen, & Roedel , 1995). With these vaiours measures, one enigmatic issue concerns which measure best reflects the accuracy of metacognitive monitoring. In constructing a test, one needs to establish the validity and the reliability of that test. Same conceptions apply to the measures of metacognitive monitoring. From the viewpoint of definitions and mathematical formulas, it is apparent that each of thest existing measures possesses construct validity. Nevertheless, few studies examined thoroughly the reliability of these meausures. The present study, thus, empirically compared the stability of these six measures in terms of three criteria: The stability across item difficul-

ty, the stability within a single domain, and the stability across domains.

Three experiments, each with a single factor design (item difficulty: easy/medium/difficult), were conducted to assess the stability of these six measures. Fifty-nine college students repeatedly participated in three experiments. Although these experiments belonged to different domains (the word recognition test, the face recognition test, and the general knowledge test), they all adopted the confidence-judgment accuracy paradigm to measure the subject's accuracy of metacognitive monitoring. The experiments of word recognition and face recognition were conducted on IBM-compatible PCs. Each of these two experiments began by asking participants to memorize a set of items, then followed by a two-alternative recognition test. As to the general knowledge experiment, it was a one-stage recognition test. For each recognition item, regardless of the experiment, the subject had to choose the correct answer from the two alternatives, then gave a confidence rating (in the range of 50% to 100%) for the chosen answer to be regarded as correct.

For each experiment, the values of $\overline{PS}$, Bias, CI, DI, ANDI, and G were computed. A Kruskal-Wallis test was then conducted to examine the effect of item difficulty on each of those measures. The Spearman correlation of each index was computed from split-halves of each test to evaluate the stability of each index in a single domain. Spearman correlations were also computed among experiments to reflect sta-

bility across domains. Results from the experiments showed that the values of the ANDI and G did not change with item difficulty. Those of $\overline{PS}$, Bias, CI, and DI were stable within a domain. The values of $\overline{PS}$, Bias, CI, and DI showed stability across domains. In conclusion, none of the examined measures was an entirely stable measure for the accuracy of metacognitive monitoring. Consequently, it is necessary to develop a new stable measure to assess the accuracy of metacognitive monitoring. In addition, results from various measures ($\overline{PS}$, Bias, CI, and DI) indicated stability over time and among tasks, implying the existence of a general and consistent metacognitive ability. The present study, thus, suggests that previous controversy about the nature of the metacognitive ability is partly due to different measures.

Keywords: Metacognitive monitoring, Stability, Domain-specific ability, Domain-general ability, Confidence-judgment accuracy paradigm.