Intelligence 67 (2018) 44–66

Contents lists available at ScienceDirect

Intelligence

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

Retest effects in cognitive ability tests: A meta-analysis

Jana Scharfen*, Judith Marie Peters, Heinz Holling

Institute of Psychology, Westfälische Wilhelms-Universität Münster, Fliednerstr. 21, 48149 Münster, Germany

ARTICLE INFO

Keywords:
Meta-analysis
Cognitive ability
Intelligence
Retest effect

ABSTRACT

Retest effects are referred to as the increase in test scores due to the repeated administration of cognitive ability tests. This meta-analysis attempts to update and extend previous meta-analyses by examining the size of retest effects and its determinants in a high number of cognitive ability tests for up to four test administrations. Strict inclusion and exclusion criteria were applied regarding study design, participant age and health status, and cognitive ability tests. An extensive literature search detected 174 samples from 122 studies, which resulted in 786 test outcomes and an overall sample size of 153,185. A comprehensive longitudinal multilevel meta-analysis revealed significant retest effects and no further score gains after the third test administration. Moderator analyses for multiple retests indicated that cognitive ability operation and content, equivalence of test forms, retest interval and participant age have a significant influence on the size of the retest effect. Implications for future research and retesting practice are discussed.

1. Retest effects in cognitive ability tests

Repeated administrations of cognitive ability tests occur frequently in selection settings, educational, and neuropsychological contexts. In fact, especially in personnel selection contexts, where cognitive ability tests find high acceptance and are often utilized as personnel decision-making tools (Ones, Viswesvaran, & Dilchert, 2005; Hülsheger, Maier, & Stumpp, 2007), retest effects have the potential to impede valid measurements. As long as causes, determinants and consequences of retest effects are not comprehended in detail, false decisions based on test results can be easily made (Randall & Villado, 2017).

A lot of research has been focusing on retest effects in cognitive ability tests (e.g., Arendasy & Sommer, 2017; Bartels, Wegryzyn, Wiedl, Ackermann, Ehrenreich, 2010; Freund & Holling, 2011; Lievens, Reeve, & Heggestad, 2007; Reeve & Lam, 2005; Villado, Randall, and Zimmer, 2016). To date, we have an approximate impression of the overall size of the effect thanks to important prior meta-analyses from Kulik, Kulik, and Bangert (1984), Hausknecht, Halpert, Di Paolo, and Moriarty Gerrard (2007), and Calamia, Markon and Tranel (2012). Hausknecht et al. (2007) made a contribution to the field by evaluating both coaching and retest effects. For 75 samples, the mere repetition of a test resulted in an improvement of almost a quarter of a standard deviation. From a clinical perspective, Calamia et al. (2012) analyzed retest effects in neuropsychological instruments for two test administrations and came to similar conclusions by giving estimates of practice effects for specific neuropsychological tests, however, effects were smaller in clinical than in healthy samples. Kulik et al. (1984) found a slightly larger effect of approximately one third of a standard deviation by analyzing 40 studies.

Most practical settings in which retesting takes place allow for more than two test administrations, or multiple retests. Findings in this field are relatively scarce, as only few studies explicitly focusing on retesting have administered more than three tests (Bartels et al., 2010). In the last years, retesting multiple times has gained more attention (e.g., Bartels et al., 2010; Puddey, Mercer, Andrich, & Styles, 2014). For three test administrations, Kulik et al. (1984) and Hausknecht et al. (2007) found increasing effects, which will be elaborated in more detail below. However, these prior findings have to be interpreted carefully as both analyses were based on a low number of samples and possible moderators of retest effects for more than two test administrations have not been examined meta-analytically. Updating and expanding results for multiple retest effects is of special interest, as due to theoretical deliberations a plateau effect would be expected (Donner & Hardy, 2015; Jaber & Glock, 2013; Newell & Rosenbloom, 1981). This is why it is important to investigate retest effects for more than two administrations meta-analytically.

The goals of the present meta-analysis are, on the one hand, to give an update of prior findings and expand results to multiple retests, and, on the other hand, to integrate methodological and theoretical developments from different perspectives into a more basic view on retest effects, including clinical and applied contexts of retesting. The three groups of causes of retest effects introduced by Lievens et al. (2007) will be elaborated on and hypotheses will be derived from a comprehensive theoretical framework with reference to this and to other important

* Corresponding author.
E-mail addresses: jana.scharfen@uni-muenster.de (J. Scharfen), holling@uni-muenster.de (H. Holling).

https://doi.org/10.1016/j.intell.2018.01.003
Received 18 July 2017; Received in revised form 11 January 2018; Accepted 16 January 2018

theoretical deliberations and prior research (e.g., Arendasy & Sommer, 2017; Freund & Holling, 2011; Randall & Villado, 2017; te Nijenhuis, van Vianen, & van der Flier, 2007). This approach will base this meta-analysis on a profound theoretical basis, possibly leading to new insights on underlying mechanisms that cause retest effects and pointing out relevant questions for future research. One major contribution of the current meta-analysis will be the differentiation between cognitive abilities according to the *Berlin Model of Intelligence Structure* (BIS; Jäger, 1982) and suggesting explanations for the disparity or similarity of retest effects between cognitive abilities.

Due to a forward and backward search of numerous important publications and an additional extensive literature search, it was possible to gather the potentially highest number of healthy and cognitively fully developed samples evaluated on this topic to date (122 studies and 174 samples), to extend findings to four test administrations and to investigate the influence of theoretically relevant determinants. Also, methodological shortcomings of prior works were addressed by, e.g., applying very strict inclusion and exclusion criteria, considering publication bias and not aggregating effect sizes for different cognitive ability domains if they were reported for the same sample, which was achieved by a comprehensive multilevel meta-analysis (e.g., Musekiwa, Manda, Mwambi, & Chen, 2016; Salanti, Higgins, Ades, & Ioannidis, 2008; Viechtbauer, 2010). All in all, this approach allowed us to expect new insights on the topic. Accordingly, this meta-analysis will update and expand prior findings, that is to say it will reliably summarize the current status of knowledge about the size of retest effects in cognitive ability tests and its determinants in healthy and cognitively fully developed samples for multiple retests on the basis of a profound theoretical framework.

1.1. The retest effect

The retest effect is defined as the change in test scores as a result of retaking the same or alternate cognitive ability test under comparable conditions (Lievens et al., 2007). It is also referred to as testing effect (Roediger & Butler, 2011), retest bias (Villado et al., 2016) or practice effect (Hausknecht et al., 2007). Although there is a broad acknowledgement of the existence of retest effects, not all of its determinants have been examined extensively, nor have the reasons for its occurrence and its impact on the psychometric quality of cognitive ability tests been fully understood (Lievens et al., 2007; Randall & Villado, 2017).

Three categories of causes of retest effects have been summarized by Lievens et al. (2007). Firstly, it is argued that the latent construct that is measured by the test could be enhanced by retesting, which leads to higher scores in repeated measurements. This explanation is seized in the research field of the so called testing effect. It is assumed that learning is enforced by test-taking, because retrieval practice might activate mnemonic enhancement (e.g., Roediger & Butler, 2011; Roediger & Karpicke, 2006). Indeed, if the latent construct was enhanced by retesting, taking a test several times would not have an effect on the validity of the test. However, several studies contradict this view and imply that validity changes as a consequence of retesting (e.g., Hausknecht et al., 2002; Lievens et al., 2005; te Nijenhuis et al., 2007), which speaks against this first cause that might lead to retest effects. Generally, as cognitive ability is defined as a stable construct, an improvement due to retesting is seen critically.

Secondly, retest effects could be explained by the reduction of distorting and construct-irrelevant factors (Lievens et al., 2007; Matton, Vautier, & Raufaste, 2009; Freund & Holling, 2011). Participants' test anxiety, lack of understanding, or lack of familiarity are assumed to decrease when retested, which in turn leads to an increase in test scores. A variation in motivation might also affect the size of retest effects (Randall & Villado, 2017). It is argued that a person who is, e.g., less anxious about a test, or who understands the test fully, can probably show better results compared to when they were firstly confronted with the test. For example, it can be assumed that cognitive capacity is restricted when anxiety is high (Eysenck, Derakshan, Santos, & Calvo, 2007; Ng & Lee, 2015). When taking into account decreasing test anxiety due to repeated stimulus presentation, e.g., following the concept of habituation (Lader & Wing, 1964; Grissom & Bhatnagar, 2009), a higher amount of cognitive capacity might be available when retested. Studies investigating the causal relationship between these construct-irrelevant variables and retest effects, such as Anastasi (1981), Matton et al. (2009), Reeve and Lam (2005), and Reeve, Heggestad, & Lievens, (2009) find evidence that they contribute to causing the effect.

Lastly, the development and application of test-taking strategies or test-specific skills could also lead to an improvement of test scores (Lievens et al., 2007; te Nijehuis et al., 2007). Strategies are likely to be developed due to test taking, which might facilitate a better test performance when retested. This idea generally serves as a basis for several test coaching programs, as elaborated by, e.g., Allalouf and Ben-Shakar (1998) and Messick and Jungeblut (1981), and which are often based on strategies of test-wiseness (Millman, Bishop, & Ebel, 1965). Empirical evidence suggests strategies are indeed developed when retesting takes place and that the use of strategies increases test scores (Allalouf & Ben-Shakar, 1998; Arendasy & Sommer, 2017; Hayes, Petrov & Sederberg, 2015; Messick & Jungeblut, 1981).

These three causes of retest effects form a theoretical basis from which hypotheses of retest effects will be developed in the following. Generally, there is a high approval of retest effects in cognitive ability tests, as most of the causes elaborated above find theoretical and empirical support (e.g., Arendasy & Sommer, 2017; Matton et al., 2009; Lievens et al., 2007; te Nijehuis et al., 2007; Reeve and Lam, 2005, Reeve et al., 2009), and retest effects are a stable finding from previous primary and meta-analytic studies (Calamia et al., 2012; Hausknecht et al., 2007; Kulik et al., 1984).

Hypothesis 1. a. Retaking a cognitive ability test leads to higher test scores.

b. Retest effects between consecutive tests decrease with the number of test administrations.

1.2. Number of test administrations

Retest effects for three administrations have been summarized by Hausknecht et al. (2007) who found an effect of *Cohen's d* = 0.51 from the first to third test for 15 samples, without evaluating retest effects for further repetitions. Kulik et al. (1984) found retest effects of *Cohen's d* = 0.53 from the first to third test and of *Cohen's d* = 0.69 from the first to fourth test, assuming a linear improvement. Following theoretical assumptions above, a linear improvement seems implausible. Since 1984, several new studies investigating retest effects administering more than two tests appeared (e.g., Albers & Höft, 2009; Bartels et al., 2010; Dunlop, Morrison & Cordery, 2011; Puddey et al., 2014), whose results mostly describe a large score gain from first to second test and retest effects becoming smaller with the number of tests, rather suggesting a non-linear progression.

From a theoretical view, the widely acknowledged *power law of practice* describes the assumption that learning curves show diminishing gains over time (Donner & Hardy, 2015; Jaber & Glock, 2013; Newell & Rosenbloom, 1981). After a first phase of improvement, no further gains are observed. As Newell and Rosenbloom (1981) state, this theory should hold for “all types of mental ability” (p. 33), and thus it can be applied to retest effects in cognitive ability tests when multiple tests are administered. According to the *power law of practice*, retest effects will decrease with the number of test administrations.

It can be assumed that for all of the three causes of retest effects outlined above, their influence decreases when the participant is retested multiple times. Their role is expected to be greater in the first repetitions of the test, with mechanisms described above leveling off after a first or second test experience. For example, after test-specific strategies have been developed and applied within first test session