Intelligence 55 (2016) 44–56

Contents lists available at ScienceDirect

Intelligence

Do individual differences in test preparation compromise the measurement fairness of admission tests?

Martin E. Arendasy [a,*], Markus Sommer [a], Karin Gutiérrez-Lobos [b], Joachim Fritz Punter [b]

[a] University of Graz, Austria
[b] Medical University of Vienna, Austria

ARTICLE INFO

Article history:
Received 3 December 2015
Received in revised form 13 January 2016
Accepted 17 January 2016
Available online xxxx

Keywords:
Admission testing
Test preparation
Measurement invariance
Cognitive ability tests
Declarative knowledge tests

ABSTRACT

The existence of individual differences in the use of test preparation has raised concerns regarding the fair and valid use of admission tests. Measurement specialists argued that the extent to which individual differences in test preparation affects the measurement fairness and construct validity of admission tests depends on the processes that lead to an increase in admission tests score due to test preparation. Four theoretical models have been advanced in the literature to account for the effect of test preparation on admission test scores. The four theoretical models make competing predictions with regard to the processes that lead to an increase in admission test scores and the extent to which measurement fairness can be assumed across test-takers differing in test preparation at the time-point of the actual admission exam. In the present article, latent class analysis was used to model individual differences in the use of test preparation methods used by applicants to a medical university ($N = 1768$). Four latent classes of test-takers were identified that differed qualitatively and quantitatively in terms of test preparation. Item response theory analyses and multi-group means and covariance structure analyses indicated strict measurement invariance across the four latent classes at the level of the individual subtests. However, group differences in test-takers admission test scores were not related to differences in either psychometric g, or general natural science knowledge ($G_{k\text{-}ns}$), respectively. The findings were consistent with theoretical models that attribute the effect of test preparation on test-takers' admission test scores to an increase in test-specific cognitive abilities and/or domain-specific knowledge.

1. Introduction

In the last decade there has been renewed interest in the effect of test preparation on test-takers' admission test scores and the extent to which individual differences in test preparation compromise the measurement fairness of admission tests (cf. Anastasi, 1981; Messick, 1982; Powers, 2012). Several meta-analyses (cf. Bangert-Drowns, Kulik, & Kulik, 1983; Becker, 1990; Briggs, 2001, 2004, 2009; Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007; Kulik, Kulik, & Bangert, 1984; Messick & Jungeblut, 1981; Powers & Rock, 1999; Witt, 1993) indicated that test preparation increases test-takers' admission test scores. The effect size of the increase in admission test scores has been shown to depend on the method(s) used to prepare for an upcoming admission test. Furthermore, research indicated that test-takers differ in their use of various test preparation methods (e.g. Buchmann, Condron, & Roscigno, 2010; Kirchenkamp & Mispelkamp, 1988; Loken, Radlinski, Crespi, Millet, & Cushing, 2004; Powers, 1988; Powers & Rock, 1999; Ryan, Ployhart, Greguras, & Schmit, 1998). These findings raised the question on whether individual differences in test preparation call the fair and valid use of admission tests into question. Messick (1982) already pointed out that whether individual differences in test preparation compromise the measurement fairness depends on how test preparation leads to an increase in admission tests scores. He outlined four types of theoretical models that differ (1) in the processes assumed to lead to an increase in admission test scores and (2) their implications regarding the level of measurement fairness of admission tests at the time-point of the actual admission test. Despite the practical and theoretical relevance of this topic, the predictions deduced from these four theoretical models have not been empirically tested thus far. Therefore the present study has been conducted to evaluate the predictions of the four competing theoretical models with regard to the level of measurement fairness of a medical university admission test.

1.1. Classification of test preparation methods

Researchers (e.g. Anastasi, 1981; Briggs, 2009; Messick, 1982; Powers, 1988) have found it useful to distinguish between *test familiarization* and *test coaching*. Test familiarization methods aim to familiarize test-takers with the mechanics of admission test taking. They constitute

* Corresponding author at: Department of Psychology, University of Graz, Universitätsplatz 2, 8010 Graz, Austria.
E-mail address: martin.arendasy@uni-graz.at (M.E. Arendasy).

http://dx.doi.org/10.1016/j.intell.2016.01.004
0160-2896/

informal, test-taker-driven forms of test preparation (Briggs, 2009). The main aim of test familiarization methods is to reduce construct-irrelevant variance in test-takers' admission test scores attributable to individual differences in test familiarity (cf. Anastasi, 1981; Briggs, 2009; Burns, Siers, & Christiansen, 2008; Messick, 1982; Powers, 1988; Powers & Alderman, 1983; Ryan et al., 1998). By contrast, professional test coaching constitutes a formal instructor-driven test preparation method. They primarily aim to increase test-takers' admission test scores (cf. Allalouf & Ben-Shakhar, 1998; Messick, 1982; Powers, 2012). The next sections briefly summarizes research on the use of these two kinds of test preparation methods, and their effect on test-takers' admission test scores.

1.1.1. Test familiarization methods and their effect on test-takers' admission test scores

Test familiarization booklets constitute one of the most commonly used test familiarization methods (cf. Kirchenkamp & Mispelkamp, 1988; Powers, 1988, 2012). Usually, test familiarization booklets contain information on (1) the specific tests administered, (2) the number of items and the time-limits for each subtest, (3) effective time-management strategies, (4) how to use the answer sheets, and (5) a short practice test. Research indicated, that test familiarization booklets increase test-takers' test-wiseness and confidence (cf. Burns et al., 2008; Powers & Alderman, 1983). However, their effect on admission test scores has been shown to be small to negligible (cf. Burns et al., 2008; Powers & Alderman, 1983; te Nijenhuis, Voskuijl, & Schijve, 2001). In general, test familiarization booklets with a focus on practicing sample items have been shown to improve test-takers' admission test score to a larger extent than test familiarization booklets with a focus on basic information on the upcoming admission test (cf. Burns et al., 2008; Powers & Alderman, 1983; te Nijenhuis et al., 2001).

Sample tests constitute another form of test familiarization. Usually, these sample tests comprise larger sets of items that are no longer used operationally. They provide test-takers with means to familiarize themselves with the task demands and to practice the sample test items. Several meta-analytic studies (cf. Hausknecht et al., 2007; Kulik et al., 1984) indicated that practicing sample test items increases test-takers' admission test scores with effect size estimates ranging from small to moderate. The effect size estimates have been shown to depend on the number of sample test items a test-taker practiced before taking an admission test (cf. Hausknecht et al., 2007), the particular admission test used (cf. Burke, 1997; Kulik et al., 1984; te Nijenhuis, Vianen, & van der Flier, 2007), and the general mental ability of the test-taker (cf. Arendasy & Sommer, 2013a; Freund & Holling, 2011; Kulik et al., 1984). In general, effect sizes were larger for cognitively more able test-takers and for less g-saturated tests. Furthermore, the effect size has been shown to logarithmically increase with the time devoted to practicing the sample test items (Hausknecht et al., 2007).

1.1.2. Test coaching and its effect on test-takers' admission test scores

Commercial and non-commercial *test coaching courses* constitute formal, instructor-driven methods of test preparation (Briggs, 2009). Test coaching programs usually consist of the following components: (1) test familiarization, (2) extended practice on sample items with feedback, (3) formal instruction on topics covered by the admission test, and (4) instruction on specific test-taking strategies (cf. Allalouf & Ben-Shakhar, 1998; Briggs, 2009; Messick, 1982; Powers, 2012). Several meta-analyses and large-scale studies indicated that test coaching increases test-takers' admission test scores (cf. Allalouf & Ben-Shakhar, 1998; Bangert-Drowns et al., 1983; Becker, 1990; Briggs, 2001, 2004, 2009; Hausknecht et al., 2007; Kulik et al., 1984; Messick & Jungeblut, 1981; Powers & Rock, 1999; Witt, 1993). For instance, Hausknecht et al. (2007) reported a large (d = .70) meta-analytic effect size for a combination of test familiarization and test coaching. However, the incremental effect of test coaching over and above practicing sample tests and other test familiarization methods has been reported to be considerably lower, with mean effect sizes varying from d = .10 to d = .43 (cf. Bangert-Drowns et al., 1983; Becker, 1990; Briggs, 2001, 2004, 2009; Kulik et al., 1984; Messick & Jungeblut, 1981; Powers & Rock, 1999; Witt, 1993). Research also indicated that the magnitude of the effect size depends on the particular admission test used (Bangert-Drowns et al., 1983; Becker, 1990; Briggs, 2001, 2004, 2009; Kulik et al., 1984; Powers & Rock, 1999) and the time devoted to practicing sample test items in test coaching courses (Becker, 1990; Messick & Jungeblut, 1981). In general, effect size estimates have been reported to be higher for less g-saturated test and higher for test coaching courses focusing on practicing sample test items.

1.2. Individual differences in test preparation

Several studies indicated that test-takers differ in the kind of methods used to prepare for an upcoming admission test and in the time devoted to test preparation (e.g. Buchmann et al., 2010; Kirchenkamp & Mispelkamp, 1988; Loken et al., 2004; Powers, 1988; Ryan et al., 1998). Although the frequency of different kinds of test preparation methods differed across studies, practicing sample items and reading test familiarization booklets have been consistently reported to be among the most frequently used test preparation methods (Kirchenkamp & Mispelkamp, 1988; Powers, 1988; Ryan et al., 1998). By contrast, test coaching courses and commercial test preparation books were among the least commonly used test preparation methods (Kirchenkamp & Mispelkamp, 1988; Powers, 1988; Ryan et al., 1998). Several authors hypothesized that differences in the use of various test preparation methods may be attributable to self-selection effects (cf. Ryan et al., 1998). Although the self-selection effects are far from completely understood, prior studies indicated that test-takers' awareness of test preparation resources, their self-efficacy and test anxiety, and their ambition constitute important determinants in choosing test preparation methods (cf. Kirchenkamp & Mispelkamp, 1988; Powers, 1988; Ryan et al., 1998). These individual difference variables may also account for the observed differences in the time and effort devoted to test preparation. Research consistently indicated that a small number of test-takers either took the admission test without test preparation, or spent considerable time and effort in preparing for the admission test (Kirchenkamp & Mispelkamp, 1988; Powers, 1988). By contrast, most admission test-takers seem to devote a small to moderate amount of time to prepare for the admission test and hardly use more than four or five different test preparation methods (Kirchenkamp & Mispelkamp, 1988; Powers, 1988).

2. Theoretical models on the effect of test preparation

Four theoretical models have been advanced in the literature to explain how different test preparation methods increase test-takers' admission test scores cf. Anastasi, 1981; Lievens, Reeve, & Heggestad, 2007; Messick, 1982; Powers, 2012). The main differences between these models concern (1) the processes responsible for the increase in test-takers' admission test scores and (2) the effect of individual differences in test preparation on the measurement fairness of admission tests.

2.1. The concept of measurement fairness

In educational admission testing it is common practice to rank-order test-takers according to their admission test scores. However, this common procedure requires that the admission test measures the same latent trait(s) for all test-takers in a fair and valid manner (cf. Lubke, Dolan, Kelderman, & Mellenbergh, 2003; Millsap, 2007, 2011; Mislevy et al., 2013; Rajo, Laffine, & Byrne, 2002; Wu, Li, & Zumbo, 2007). If measurement fairness can be assumed, test-takers with the same standing on the latent trait(s) have equal expected item- and/or test scores. Thus, differences within- and between groups of test-