The effects of computer self-efficacy, training satisfaction and test anxiety on attitude and performance in computerized adaptive testing

Hong Lu [a, *], Yi-ping Hu [a], Jia-jia Gao [a], Kinshuk [b]

[a] College of Communication, Shandong Normal University, China
[b] Faculty of Science and Technology, Athabasca University, Canada

ARTICLE INFO

ABSTRACT

This study focused on test-takers' psychological effects on computerized adaptive testing (CAT). The development and implementation of CAT were based on item response theory (IRT), and two-parameter logistic model was chosen for the items. The total of 268 students from a high school in Jinan took part in the English adaptive test. A structural equation model was used to examine the potential connections among a series of individual variables (computer self-efficacy, training satisfaction, test anxiety, CAT attitude and CAT performance). The findings revealed significant positive paths from computer self-efficacy and training satisfaction to CAT attitude, as well as a negative path from test anxiety to CAT performance. Furthermore, there was significant correlation between the residual variances of CAT attitude and CAT performance. Thus, it could be seen CAT might produce an unfair disadvantage for test-takers with higher test anxiety. The relevant research and implications were further discussed.

1. Introduction

As information technology has become increasingly more prevalent and accessible for use in student assessment, innovative test delivery models are adopted to collect, analyze, and report student-level data. Among these models, computerized adaptive testing (CAT) based on item response theory (IRT) has been attracting more and more attention. The basic idea of CAT is that test items are selected by the computer to individually match the ability level of each student (Wainer, 2000). In this manner, the test is tailored to each student. There are some benefits associated with CAT, and it is logical to see why testing experts are making a push toward this testing modality. For example, by using more precise and efficient assessments that take less time to complete, teachers and students will get test results that are either just as accurate as traditional tests or more accurate. In addition to this, the tests tailor each question to the knowledge and abilities of the students, thereby theoretically keeping them appropriately challenged and more likely to stay engaged (Wainer, 2000). Based on the above mentioned advantages, CAT is becoming more and more common in high-stake assessment. For instance, the Graduate Record Examination (GRE), the nursing licensing exam, and the Graduate Management Admission Test (GMAT), are all now primarily offered in CAT. Additionally, in the U.S., many states were moving to put in place online testing tied to the common

---

core state standards in 2014–15, at least 20 states among them indicated they would plan to use new computer-adaptive versions of the tests (Davis, 2012). Moreover, the Smarter Balanced Assessment Consortium has received federal funding to develop English/language arts and mathematics adaptive tests for the common standards. He said his assessment would feature high-tech, interactive questions that incorporated video and graphics, and were designed both to identify what students knew and to be more engaging (Davis, 2012).

1.1. Review of literature

The trend that CAT was an evolutionary step toward future testing methodologies resulted in a growing number of studies dedicating to it. Most of them investigated the technical aspects of CAT, such as comparison of different item-selection methods (Finkelman, Kim, Weissman, & Cook, 2014; He, Diao, & Hauser, 2014; Wang, 2013a; Yao, 2012), item pool construction (He & Reckase, 2014; Lee & Dodd, 2012), test stopping rules (Choi, Grady, & Dodd, 2011; Wang, Chang, & Boughton, 2013; Yao, 2013). However, only a few studies dealt with CAT's psychological effects on test-takers. In early studies, chief among them was that it might increase the student's interest and motivation for taking the test. For instance, Weiss and Betz (1973) indicated that adaptive testing was suggested to avoid boredom for test-takers with high ability and prevent test-takers with low ability from experiencing anxiety. Johnson and Mihal (1973) found that blacks performed better on adaptive testing. Weiss (1975) found similar motivational effects when feedback on the correctness of a response was provided. These results seemed to suggest that in some cases CAT might be more motivating or less anxiety-producing than conventional testing.

In recent years, undesirable psychological reactions to CAT were discussed as following: Tonidandel and Quinones (2000) explored how specific aspects of adaptive testing influence test-takers' reactions. Fifty-three undergraduates were presented with descriptions of hypothetical selection tests manipulated to reflect characteristics of adaptive tests that differed from traditional paper-and-pencil tests (P&P). The results demonstrated that certain features of adaptive tests, such as the inability to skip questions, review items, or go back and change answers, might adversely impact test-takers' psychological reactions. Ortner and Caspers (2011) investigated the effects of test anxiety on test performance using computerized adaptive testing versus conventional fixed item testing. A total of 110 students from a German secondary modern school were tested. Findings showed that, when confronted with an adaptive matrices test, test-takers with high test anxiety had lower test scores compared to persons with low test anxiety. That was to say, adaptive testing might lead to a bias that produced a disadvantage for test-takers with higher test anxiety. In another study, Ortner, Weißkopf, and Koch (2014) examined the effects of computerized adaptive testing versus computerized fixed item testing of reasoning ability on current motivation. A group of 174 students from two German secondary schools was presented either an adaptive or a fixed version of a matrices test. Less motivation was reported using adaptive testing compared to fixed item testing.

1.2. The present study

The researches published on CAT and psychological effects have yielded mixed results, raising the question of whether or not CAT does support fair assessment procedures for test-takers. For example, some previous researches showed that CAT might be revealed to be unfair with reference to its potential to evoke success-related estimations in high performers, and then the perceived unfairness of CAT had a negative impact on their CAT performance (Ortner et al., 2014). The result contradicted early assumptions generally supported higher fairness for CAT that every test-taker would solve about 50% of the given items correctly independent of ability. The mixed results also made researchers turn their attention to whether CAT was fair or not (Fritts & Marszalek, 2010; Ortner & Caspers, 2011; Ortner et al., 2014; Tonidandel & Quinones, 2000). So the goal of the present study was to further investigate the influence of some individual characteristics on CAT, and provided empirical evidence for fairness or unfairness of CAT.

In order to achieve the goal of this study, three individual characteristics which might have relationship with individual CAT performance or CAT attitude were chosen. The computer was the essential tool during the process of CAT. Therefore, computer self-efficacy should play an important part in applying CAT. In addition, due to the significant difference between CAT and P&P, CAT training was particularly necessary for test-takers. Certainly, training satisfaction was considered to have significant influence on the implementation of CAT. What's more, test anxiety was an essential variable widely studied in the context of various academic achievements (Chapell, et al., 2005; Farooqi, Ghani, & Spielberger, 2012). Thus, computer self-efficacy, training satisfaction, test anxiety, CAT attitude and CAT performance were included in this study to set up a causal model of CAT. The relationships between these latent variables were analyzed by using high-level analysis software as well.

1.3. Research model

Based on previous studies, hypotheses developed to test the effect of the variables of computer self-efficacy, training satisfaction, test anxiety, CAT attitude and CAT performance on each other and their relation to each other were presented below.