An Argument-Based Approach to Test Fairness:
The Case of Multiple-Form Equating in the College English Test

Yan Jin, School of Foreign Languages, Shanghai Jiaotong University, Shanghai, China
Eric Wu, University of California, Los Angeles (UCLA), Temple City, CA, USA

ABSTRACT
This article aims to demonstrate how innovative testing practices can effectively prevent high-tech mass cheating and improve fairness in language assessment. The article first introduces Xi’s (2010) view of validity and fairness and her proposal of an argument-based approach to empirically examining test fairness. The article then describes the threat to fair testing posed by high-tech cheating on the College English Test (CET). A study of multiple-form equating was conducted and reported in the article, which was aimed at achieving alternate form reliability when multiple versions and multiple forms were used in the CET. The article then concludes with a discussion on the usefulness of an argument-based approach to empirically examining test fairness.

KEYWORDS
Argument-Based Approach, College English Test (CET), High-Tech Cheating, Test Fairness

1. INTRODUCTION
In the field of language testing and assessment, technology is generally seen as a blessing: the latest developments in information and communications technology (ICT), for example, have improved not only the efficiency of assessment practices through delivering tests on computers or over the internet, but the validity of language tests through introducing new language constructs and bringing about changes in language teaching and learning (Chalhoub-Deville & Deville, 1999; Chapelle, 2008; Chapelle & Douglas, 2006; Chapelle & Voss, 2016). However, the blessing of technology may well become a threat if used improperly by ill-intentioned people. Technology-assisted cheating on the College English Test (CET), a national language testing system in China (see Jin, 2010, 2016; Yang, 2003; Zheng & Cheng, 2008), is a case in point. In recent years, with the rise in the stakes of the CET, high-tech cheating has become a major concern among the test developer and its stakeholders. By “high-tech cheating on tests”, we mean the use of communications devices such as mobile phone, needle camera, invisible watch, mini-earpiece, and multifunctional Bluetooth transmitter for sending and/or receiving messages in a testing situation. The purpose of this article is to discuss the misuse of technology for cheating and introduce an innovative, preventive measure to combat high-tech cheating.

DOI: 10.4018/IJCALLT.2017070104

Copyright © 2017, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
There is no denying that cheating on tests constitutes a serious threat to test fairness (AERA, APA, & NCME, 1999, 2014). Cheating also raises ethical concerns that require ongoing consideration by test developers and administrators (Cizek, 1999). To prevent high-tech cheating on the CET, the National Education Examinations Authority (NEEA), a government institute in charge of the operation of the CET, designed a measure called multiple versions and multiple forms (MVMF), which has been implemented by the National College English Testing Committee (NCETC), the developer of the CET, since 2013. One of the technical challenges facing the implementation of MVMF is score equating, that is, the differences in difficulty levels across alternate forms need to be adjusted so as to produce comparable score scales. In this article, we adopted the argument-based approach advocated in Xi (2010) to examine, from the CET developer’s perspective, the issue of high-tech cheating that threatens the fairness of the test.

2. CONCEPTUALIZATION OF TEST FAIRNESS

2.1. Validity and Fairness

Validity has been a central theme and the most important concept in the field of language testing and assessment since the publication of *Language Testing* by Robert Lado in the early 1960s (Lado, 1961), and especially so after Samuel Messick proposed “unified validity” in the 1980s (Messick, 1989, 1995). In the mid-1990s, Bachman and Palmer (1996) proposed a framework of test usefulness, which included six essential qualities for evaluating the usefulness of a language test for intended purposes: reliability, construct validity, authenticity, interactiveness, washback, and practicality. Test fairness, though recognized as a basic requirement of assessment practices, was not explicitly articulated as an essential aspect of validity or an independent quality of test usefulness till the late 1990s, when the field of language testing and assessment began to take a social turn (see Kunnan, 2000, 2004; McNamara & Roever, 2006; Yang & Gui, 2007).

The social dimension highlights the role of fairness in language assessment and calls for a clearer conceptualization of test fairness. Given the centrality of validity to testing and assessment, there is a legitimate question as to whether fairness is considered as an indispensable aspect of test validity. Xi (2010) summarizes three views of the relationship between validity and fairness and argues for an argument-based approach to examining test fairness.

The first view sees fairness as an independent test quality that is separate from validity. Similar to validity, fairness is treated as a test quality that permeates the whole assessment process. The *Standards for Fairness and Quality* by Educational Testing Service (2002) and the *Code of Fair Testing Practices in Education* by Joint Committee on Testing Practices (1998, 2004) are representative of this view. The second view gives primacy to test fairness and defines it as an all-encompassing test quality, which includes validity, absence of bias, access to the test, administration conditions, and test consequences. The most enthusiastic and influential advocate of this view is Antony Kunnan (2000, 2004), whose test fairness framework broadens the span of fairness in language testing by drawing on social justice theories. In the third view, validity is treated as the fundamental test quality and fairness as a test quality directly linked to validity. This view is built upon the conceptualization of fairness proposed in Willingham and Cole (1997) and Willingham (1999), which conceives fairness as comparable validity for all individuals and groups. In this view, whatever weakens fairness also compromises the validity of a test.
Social Presence Functions in Task-Based Language Activities in a Virtual Classroom in Second Life
www.igi-global.com/chapter/social-presence-functions-in-task-based-language-activities-in-a-virtual-classroom-in-second-life/172870?camid=4v1a

A Blended Chinese-as-a-Foreign-Language Short Course: Design and Perceptions
www.igi-global.com/article/a-blended-chinese-as-a-foreign-language-short-course/153894?camid=4v1a