

附件 2:

浙江大学优秀博士学位论文申请表

论文题目		测量学与涉考群体视阈下的英语水平测试公平性研究：构建测试公平性评估模型			
论文英文题目		Evaluating the fairness of an English proficiency test from the psychometric and stakeholders' perspectives: Toward a model of test fairness evaluation			
作者学号	作者姓名	获得博士学位日期	论文涉及的研究方向		
12005002	张娟	2025 年 6 月 20 日	语言测试		
作者电话	作者邮箱	现工作或学习单位			
17816862122	jzhang2277@outlook.com	伊利诺伊大学厄巴纳-香槟分校			
一级学科或专业学位类别代码	一级学科或专业学位类别名称	二级学科或专业学位领域代码	二级学科名称或专业学位领域名称		
0502	外国语言文学				
主导师姓名	何莲珍	主导师研究方向	语言测试、外语教学		
导师组其他成员姓名 <sup>①</sup>		导师组其他成员研究方向			
申请学位时论文的总体评价		优秀（4）个	良好（1）个	其他（0）个	
作者攻博期间及获得博士学位后一年内与博士学位论文密切相关的浙大为第一单位的代表性成果 <sup>②</sup>	序号	成果名称 <sup>③</sup>	成果出处 <sup>④</sup>	获得年月 <sup>⑤</sup>	查询信息 <sup>⑥</sup>
	1	学术论文《语言测试的公平性：内涵、公平观及研究启示》	《外语教学与研究》	2022 年 1 月	2022 年第 1 期
	2	学术论文《“公平”的理论向度——兼论对语言测试公平性研究的启示》	《浙江大学学报（人文社会科学版）》	2024 年 5 月	2024 年第 5 期
	3				
	4				
	5				
论文主要创新点	<p>本研究从测量学与涉考群体两个视角着手评估某高风险语言考试项目的公平性。本研究在理论、实践及方法层面均具有重要启示。</p> <p>理论层面，本研究提出了一个测试公平性评估模型。该模型不仅凸显了测试公平性的双重属性（即测量学属性和价值判断属性），还通过呈现测试公平性的四个重要维度（即可比性、可及性、一致性和问责制）、核心涉考群体以及影响涉考群体对测试公平性认知的四大因素（即社会文化、教育机会、测试机构和考生个体因素），彰显了测试公平性的多维性、社会建构性以及情境依赖性等本质特征。该模型为高风险语言考试项目公平性评估工作的顺利开展提供了不可或缺的理论参考。</p>				

	<p>实践层面，针对本研究聚焦的某高风险英语水平考试，实证研究发现能够为该考试的质量提升和相关测评实践提供重要参考。此外，研究结果还将助力校本语言测试公平性标准或指南的起草工作。</p> <p>方法层面，本研究通过实践探索，初步验证了将混合研究方法应用于高风险语言考试项目公平性评估工作的可行性。混合研究方法有助于开展兼具全面性与系统性的语言测试公平性评估工作。</p>
学位论文作者承诺	<p>本人的学位论文申请参评浙江大学优秀博士学位论文，本人承诺：以上所填全部信息及代表性成果证明材料准确无误、真实可靠；本学位论文不涉密，可在互联网上公开评审并全文公示；<b>本人无学术失范或学术不端问题</b>。如信息不实，愿承担由此带来的一切后果和法律责任。</p> <p>承诺人签名：</p> <p>年 月 日</p>
指导教师意见	<p>本人确认该论文无涉密内容，可以在互联网上公开评审、公示。该论文“代表性成果”等相关材料和数据准确无误、真实可靠，同意推荐参加学校优秀博士学位论文评选。</p> <p>导师签名：</p> <p>年 月 日</p>

所在学院（系） 形式审查意见	<p>经学院（系）审查，该论文符合申请条件，以上所填信息及相关材料属实。</p> <p>学院（系）负责人签字：</p> <p>单位公章</p> <p>年 月 日</p>
学科级学位评定委员会 推荐意见	<p>学科级学位评定委员会主任签字：</p> <p>年 月 日</p>
学部级学位评定委员会 推荐意见	<p>学部级学位评定委员会主任签字：</p> <p>年 月 日</p>
学校学位评定委员会 评定意见	<p>(盖章)</p> <p>年 月 日</p>

填表要求:

- ① 其他导师应与研究生管理系统一致,如有多位,可添加行,如没有,可填无。
- ② “代表性成果”限填作者攻博期间及获得博士学位后一年内与博士学位论文密切相关、并能反映学位论文水平的成果。可填学术论文、专著、专利、奖励等,但总数不得超过 5 项,且必须是在规定时间内公开发表(含网络在线发表)或审批的。学术论文请标注影响因子。在规定时间内已录用而未发表的学术论文、已受理而未审批的专利和已公示而无批文的奖励等成果,以及在规定时间内外获得的成果一律不计入。请准确填写各项成果的查询信息,确保按此查询信息能查询到该成果,以便于专家评议时核查。
- ③ “成果名称”栏,可填写论文题目、专著名称、专利名称、奖励名称等。
- ④ “成果出处”栏,可填写刊物名称、出版机构、奖励发放单位等。
- ⑤ “获得年月”栏,可填写论文公开发表、专著公开出版、专利授予、奖励获批的具体年月。
- ⑥ “查询信息”栏,应填写论文检索号、国际标准书号 (ISBN)、专利号、获奖证书号等。填写“检索号”时,若论文被 SCI、SSCI、EI、A&HCI 等检索,则填写论文检索号;否则填写刊物的出版年期。



附件 3:

代表性成果证明材料

序号	论文题目	刊物名称	发表时间	本人排序/ 总人数/ 导师排序	浙大是否 第一单位	备注(SSCI/一级、SCI/EI 收录、5 年平均影响因子等)	他引 次数	学位论文 对应章节
1	语言测试的公平性：内涵、公平观及研究启示	《外语教学与研究》	2022 年 1 月	2/1	是	权威、1.45	15	第二章 （文献综述）
2	“公平”的理论向度——兼论对语言测试公平性研究的启示	《浙江大学学报（人文社会科版）》	2024 年 5 月	2/1	是	一级、1.13	1	第二章 （文献综述）
3								
4								
5								

注：1. 备注(SSCI/一级、SCI/EI 收录、影响因子等)，影响因子是五年影响因子；建议影响因子、他引次数均来自 web of science。  
2. 无特别必要，请不要加页。  
3. 证明材料请按以上顺序装订在推荐表后面。

国家社科基金资助期刊·教育部高校学报名刊·国家期刊奖百种重点期刊·美国LLBA收录期刊  
中国人文社会科学A刊权威期刊·CSSCI来源期刊·中国期刊方阵双效期刊

# 外语教学与研究

FOREIGN LANGUAGE TEACHING AND RESEARCH

外国语文双月刊

Vol.54 No.1

2022 / 1

外语  
教学  
与  
研究

第  
54  
卷  
  
第  
1  
期

二  
〇  
二  
二  
年  
一  
月

# 外语教学与研究

WAIYU JIAOXUE YU YANJIU

北京外国语大学主办 中国外语与教育研究中心承办

## FOREIGN LANGUAGE TEACHING AND RESEARCH

Directed by Beijing Foreign Studies University

Managed by the National Research Centre for Foreign Language Education

编审委员会 (Editorial Board)

主 任 (Chair): 杨 丹 (Yang Dan)

副主任 (Associate Chairs): 文秋芳 (Wen Qiufang), 王克非 (Wang Kefei)

委 员 (Members) (按姓氏音序):

陈 平 (Chen Ping), 昆士兰大学 (University of Queensland, AUS)

顾曰国 (Gu Yueguo), 北京外国语大学 (Beijing Foreign Studies University)

胡文仲 (Hu Wenzhong), 北京外国语大学 (Beijing Foreign Studies University)

胡壮麟 (Hu Zhuanglin), 北京大学 (Peking University)

黄国文 (Huang Guowen), 华南农业大学 (South China Agricultural University)

黄 衍 (Huang Yan), 奥克兰大学 (The University of Auckland, New Zealand)

蒋 楠 (Jiang Nan), 马里兰大学 (The University of Maryland, USA)

李行德 (Thomas Hun-tak Lee), 香港中文大学 (The Chinese University of Hong Kong)

刘润清 (Liu Runqing), 北京外国语大学 (Beijing Foreign Studies University)

申 丹 (Shen Dan), 北京大学 (Peking University)

沈家煊 (Shen Jiaxuan), 中国社会科学院语言研究所 (Chinese Academy of Social Sciences)

石定栩 (Shi Dingxu), 广东外语外贸大学 (Guangdong University of Foreign Studies)

石毓智 (Shi Yuzhi), 新加坡国立大学 (National University of Singapore, SG)

王初明 (Wang Chuming), 广东外语外贸大学 (Guangdong University of Foreign Studies)

吴一安 (Wu Yi'an), 北京外国语大学 (Beijing Foreign Studies University)

许 钧 (Xu Jun), 浙江大学 (Zhejiang University)

徐盛桓 (Xu Shenghuan), 河南大学 (Henan University)

许余龙 (Xu Yulong), 上海外国语大学 (Shanghai International Studies University)

袁博平 (Yuan Boping), 剑桥大学 (Cambridge University, UK)

周流溪 (Zhou Liuxi), 北京师范大学 (Beijing Normal University)

主 编 (Editor-in-Chief): 王克非 (Wang Kefei, Beijing Foreign Studies University)

副主编 (Associate Editor-in-Chief): 王文斌 (Wang Wenbin, Beijing Foreign Studies University)

本刊自 2000 年 1 月起实行同行专家匿名审稿制度

(本刊不以任何形式收取版面费; 全国哲学社会科学工作办公室举报电话: 010-55604027)

# 《外语教学与研究》第 54 卷第 1 期

(2022 年 1 月)

## 目 录

### · 语言研究 ·

- 英语多重递归关系从句的分布和限制 ..... 丁彧藻 陈保亚 ( 3 )  
英汉时间状语结构的句法分析: 基于介词短语制图理论的研究  
..... 张庆文 刘 惟 ( 16 )  
跨语言视角下的两类局域性条件 ..... 胡笑适 ( 29 )  
汉英对比视角下全量表达的“非全量”用法 ..... 冯予力 ( 42 )  
焦点类型和语言水平对“花园路径式”误译中读者反应的影响  
..... 蔡 妍 林 璋 ( 53 )

### · 外语教育 · (教育部哲学社会科学学报名栏)

- “云连接论”的构想和应用 ..... 文秋芳 ( 66 )  
语言测试的公平性: 内涵、公平观及研究启示 ..... 何莲珍 张 娟 ( 79 )  
中国英语学习者非宾格结构的在线加工研究 ..... 曾 涛 高雅蓉 代学群 ( 90 )  
中国英语学习者语用充实启动研究 ..... 贾 莉 杨连瑞 ( 103 )

### · 翻译研究 ·

- 隐喻视译的认知加工模式——眼动追踪和译语分析的证据  
..... 卢 植 郑有耀 ( 115 )  
翻译方向对信息加工过程及质量的影响  
——基于眼动和屏幕记录等数据的多元互证 ... 王湘玲 王 律 郑冰寒 ( 128 )

### · 书刊评介 ·

- 《动词语义的词根》(2020) 评介 ..... 张梦杰 林允清 ( 140 )  
《跨语言视野下的粤语句末助词研究》(2019) 述介 ..... 罗虞欢 林华勇 ( 145 )  
《扩展的概念隐喻理论》(2020) 述评 ..... 马俊杰 ( 150 )  
致 谢 ..... ( 127 )  
本刊 2021 年总目录 ..... ( 155 )  
本刊稿约 ..... ( 封四 )  
英文摘要 ..... ( 158 )

编辑部副主任: 郭 洁

本期执行编辑: 刘 华

[ 期刊基本参数 ] CN11—1251/G4 \* 1957 \* b \* 16 \* 160 \* zh—en \* P \* ¥25.00 \* 9000 \* 14 \* 2022—01

# FOREIGN LANGUAGE TEACHING AND RESEARCH

---

Volume 54, Number 1

January 2022

---

## Articles

- Multiply-recursive relative clauses in English: Distributions and constraints  
..... DING Yuzao & CHEN Baoya (3)
- The syntactic analysis of English and Chinese temporal adverbial constructions:  
A study based on the cartography of prepositional phrases  
..... ZHANG Qingwen & LIU Wei (16)
- Two types of locality conditions from a cross-linguistic perspective..... HU Xiaoshi (29)
- On the “non-universal” use of universal quantificational expressions in Chinese  
and English..... FENG Yuli (42)
- The influence of focus types and mother tongue proficiency on readers’ response  
to “garden path mistranslation” ..... CAI Yan & LIN Zhang (53)
- Cloud Interconnectedness Theory for developing an online teaching-research community of  
foreign language teachers ..... WEN Qiufang (66)
- Test fairness: Conceptualization, perspectives, and implications  
..... HE Lianzhen & ZHANG Juan (79)
- Online processing of unaccusative constructions by Chinese English learners  
..... ZENG Tao et al. (90)
- A study of pragmatic enrichment priming among Chinese English learners  
..... JIA Li & YANG Lianrui (103)
- A cognitive processing model for English-Chinese metaphor in sight translation  
— Evidence from eye-tracking and product analysis ..... LU Zhi & ZHENG Youyao (115)
- Impact of translation directions on information processing and translation quality  
— A triangulated study using eye-tracking and screen-recording  
..... WANG Xiangling et al. (128)

## Book Reviews

- J. Beavers & A. Koontz-Garbonden: *The Roots of Verbal Meaning* (2020)  
..... ZHANG Mengjie & LIN Yunqing (140)
- Iida Maki: *Cantonese Sentence-final Particles from a Cross-linguistic  
Perspective* (2019) ..... LUO Yuhuan & LIN Huayong (145)
- Z. Kövecses: *Extended Conceptual Metaphor Theory* (2020) ..... MA Junjie (150)

---

## Information about the Journal of FLTR

**Foreign Language Teaching and Research** is a bimonthly journal for academics in the field of foreign/second language teaching and research. All articles to be published in the journal are refereed by members of the Editorial Board or other peer reviewers.

**Major Focus:** General and applied linguistics, and learning and teaching of foreign/second languages.

**Manuscript Form:** Mainly in accordance with the *Linguistic Inquiry* style sheet.

**Manuscript Evaluation:** Manuscripts are refereed by at least three reviewers.

**Time from Submission to Decision:** Normally within 90 days; Time from acceptance to publication usually no longer than 180 days.

**Subscription:** China International Book Company (Guoji Shudian), P.O. Box 399, Beijing, P. R. China (Tel: 0086-10-68413063)

**Correspondence:** Editorial Office of *Foreign Language Teaching and Research*, P. O. Box 8936-45, Beijing Foreign Studies University, Beijing 100089, P. R. China  
[bwyys@126.com](mailto:bwyys@126.com)

# 语言测试的公平性：内涵、公平观及研究启示<sup>\*</sup>

浙江大学 何莲珍 张 娟

**提要：**考试是迄今为止所实施的相对公平的人才选拔制度，对守护教育及社会公平起到了至关重要的作用。本文在回顾语言测试公平性内涵的基础上，梳理了学界主流的公平观，即“标准”观、论证观和伦理道德观，以及相应的研究进展。本文指出，为保证语言测试的公平性和科学性，学界需继续探究如何处理测试公平性与效度的关系，如何对本土语言考试项目的公平性进行研究，以及如何在测试公平性研究中引入涉考者视角并对其予以关切。

**关键词：**语言测试、公平性、内涵、公平观、研究进展

[ 中图分类号 ] H310.4 [ 文献标识码 ] A [ 文章编号 ] 1000-0429 ( 2022 ) 01-0079-11

DOI:10.19923/j.cnki.fltr.2022.01.007

## 1. 引言

测试的公平、公正、公开一直以来就承载了民众的深刻关切和强烈诉求，大规模和高风险测试尤甚。然而，测试反映了测试开发者、使用者及政策制定者的价值观，其间充斥的隐性权力极易滋生教育和社会不公 ( Madaus 1990 )。20 世纪 60 年代末，教育测量学界开始关注题项及测试的公平性 ( Cole & Zieky 2001 )。20 世纪 80 年代末，测试公平性进入语言测试研究者的视野 ( Alderson 1988 )，并于 1996 年第十八届国际语言测试研讨会 ( Language Testing Research Colloquium, LTRC ) 被正式提上研究议程 ( Kunnan 2010 )。

教育评价与教育公平密切相关，我国政府对此给予了极大关注。2020 年，中共中央、国务院印发了《深化新时代教育评价改革总体方案》，从国家层面明

<sup>\*</sup> 本文系国家社科基金项目“教育公平背景下的语言测试公平性研究”(20BYY107)的阶段性成果。



确提出要“扭转不科学的教育评价导向”(中共中央、国务院 2020: 1)。2021 年 3 月 11 日,国务院总理李克强在人民大会堂出席记者会时明确指出:“机会公平中,教育公平是最大的公平”。同年发布的《中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要》提出要“深化新时代教育评价改革,建立健全教育评价制度和机制”(国家发展和改革委员会 2021: 131-132)。在当下中国,保证测试公平无疑是提高教育评价科学性的关键,对守护教育公平乃至社会公平大有裨益。本文聚焦语言测试公平性这一研究议题,梳理语言测试公平性的内涵、学界主流的公平观及研究进展,以期为开展本土语言考试项目的公平性研究提供更多思路。

## 2. 语言测试公平性的内涵

总体而言,测试公平性的内涵经历了从狭义到广义的嬗变。20 世纪 80 年代,心理和教育测量学领域对公平性的定义局限于考试的技术维度,把测试公平性视作统计学问题(如 Shepard *et al.* 1981; Angoff 1982)。狭义视角下,测试公平性指在不同考生群体中不存在考试偏颇(test bias)(AERA *et al.* 1985)。若考试存在偏颇,则意味着测量过程中的系统性误差会对某一考生群体的表现造成不利影响。这一时期的实证研究尝试运用统计手段对考试偏颇进行检测,其中的项目功能差异(differential item functioning, DIF)检验是最常见的技术手段之一(参见 Zumbo 2007)。然而,将“测试公平性”技术化存在一定局限,因为这一界定方式与涉考者对测试公平性的理解严重脱节(Crocker 2003)。此外,考试结果所涉及的决策过程还包含价值判断,测试公平性的内涵也会随之延伸至社会维度。广义视角下,除技术维度外,考生的学习机会(Kunnan 2004)、涉考者对考试内容和评分标准的看法(Crocker 2003)、考试结果的使用(Zwick 2019)、考试对社会的影响(Shohamy 1998; 杨惠中、桂诗春 2007)和针对特殊考生群体的考试环境调适(test accommodation)(Li & Suen 2012)等问题均需予以关切。

迄今为止,“测试公平性”在语言测试界还没有统一的定义,部分学者参考效度这一概念对其进行定义。在这一视角下,定义测试公平性的关键在于厘清公平性与效度的关系(Xi 2010),主要有以下几种观点。第一种观点将测试公平性视作效度的一部分。如 Messick (1988) 和 Weir (2005) 认为考试偏颇造成的不良后果归根结底是考试的效度问题,公平性与考试结果的使用息息相关。Willingham & Cole (1997) 将测试公平性定义为考生个体间和群体间的可比效度(comparable validity)。具体而言,效度可比不仅要求考试分数解读及使用具有可比性,还要求考生展现自己语言水平的机会、评估的程序及评分结果也具有可比性。第二种观点将效度视作公平性的一个方面。如 Kunnan (2000b) 认为

效度是实现测试公平性的必要但非充分条件,除效度外,公平性还应关注不同背景的考生群体在考试过程中是否受到平等对待和应有尊重。这一视角下的公平性强调考生个体的权利和义务。第三种观点认为公平性与效度相互联系,研究范畴或有交叉,但二者研究范围孰大孰小不在讨论之列。Davies *et al.* (1999)指出,测试公平性关乎考试结果对考生个人、考生群体乃至整个社会产生的影响,它不仅关涉考试效度,也贯穿整个考试过程,在一定程度上反映或影响社会公平。Kane (2010)认为测试公平性和效度涉及测试的方方面面,研究议题或有重叠,但各有侧重,两者无法互相替代。另外,《教育测试实践公平性准则》(JCTP 2004)和《ETS 质量和公平标准》(ETS 2014)均主张减少与考试构念无关的因素以保证测试公平,这一主张与维护考试构念效度的观点不谋而合。

另有一些学者尝试借鉴其他学科视角对测试公平性进行定义。在法律视角下, Kane (2010)将测试公平性分为程序性公平 (procedural fairness) 和实质性公平 (substantive fairness) 两类。前者强调从考试开发到评分各环节,考生应被同等对待;后者强调考试分数解释和用于决策的依据在不同考生群体间的合理性和适切性。在哲学视角下, Davies (2010)认为对公平性的追求徒劳且无必要,解决一个不公平事件会引发另一不公平事件,因此对公平性的追求纯属空想。在社会学视角下, Spaan (2000)将测试公平性视作考试开发者、使用者和考生之间的社会契约,三方有义务共同维护考试公平。考试开发者在确保考试本身科学有效的同时,需指导考试使用者正确解读及使用考试分数;考试使用者需了解考试目的和考试分数的含义,确保考试结果不被滥用;考生则有义务主动了解考试内容及题型,防止题型对自己的考试表现产生不利影响。从政治学视角出发, Cronbach (1988)认为政治因素在保证考试公平方面有着无可替代的作用,其影响力远超法律、哲学等层面的考量。结合社会学和政治学视角, McNamara & Roever (2006)认为测试研究者在提高测量精度的同时,需充分考量语言测试所处的社会及政治环境,关注测试实践中蕴含的价值判断。Kunnan (2000b)认为审视测试公平性时需考量测试语境所涉及的各个维度,因此他主张将社会、伦理、法律、哲学等视角纳入测试公平性研究。多学科视角的引入极大丰富了测试公平性的内涵,为后续的理论 and 实证研究持续赋能。

### 3. 语言测试的公平观

#### 3.1 标准观

20 世纪 50 年代以来,教育测量界的相关机构和学者开始通过制订标准性文件来保证测试质量并规范测试从业人员的行为,测试公平性的“标准”观应运而生。测试公平性是诸多标准性文件的重要组成部分,常作为独立的章节而存



在。“标准”是“测试从业人员在开发考试、管理并实施考试项目、解读考试内容及其结果时需遵循的特定程序”(Stansfield 1993: 190)。20 世纪 90 年代,测量标准的制定、颁布和实施成为热点议题。因篇幅有限,本节仅简要介绍最具影响力的标准性文件,即《教育与心理测试标准》(*Standards for Educational and Psychological Testing*, 简称《标准》)。

《标准》由美国教育研究协会(AERA)、美国心理学会(APA)和美国国家教育测量委员会(NCME)联合发布,旨在指导考试开发,评价考试质量,为评价考试分数解读及考试结果使用的效度提供参考。1985 版的《标准》对公平性的讨论较为零散,主张一个公平的考试不应系统性地高估或低估某一考生群体的能力(AERA *et al.* 1985)。后于 1999 和 2014 年进行了两次修订。1999 版《标准》包括三个部分,第二部分聚焦测试公平性,其中“测试与测试使用公平”这一章提出了 12 条具体标准,涵盖四个方面:考试无偏颇、考试过程中平等对待所有考生、考试结果不受群体特征影响、考生学习机会均等。在 2014 版《标准》中,“测试公平性”与“效度”“信度/测量精度与测量误差”是并列的章节,强调测试公平性需贯穿从测试开发到使用的全过程。“测试公平性”这一章共包含 20 条标准,与 1999 版相比,除保留“考试无偏颇”“考试过程中平等对待所有考生”这两个特征以外,还要求“与考试构念无关的因素不影响考试结果”“考试结果的解释需有效”。该版标准进一步指出,与考试构念无关的因素极易威胁测试公平,考试内容、考试环境、考生作答及其学习机会的差异都可能引入与考试构念无关的因素,需要予以特别关注。作为指导测试开发和使用的重要参考性文件,《标准》为提升测试质量、促进测试公平做出了重要贡献,影响深远。

### 3.2 论证观

随着语言测试效度论证观的出现和效度验证体系的日臻完善,有学者尝试吸纳效度验证的理论研究成果,主张通过收集证据的方式论证语言测试的公平性。Bachman & Palmer (2010/2016) 提出的评估使用论据(Assessment Use Argument, AUA)为测试公平性研究提供了思路。他们认为公平性并非测试的独立特征,而是与测试过程和测试使用密切相关。测试过程的公平性主张: 1) 考生受到平等对待; 2) 测试过程、测试记录和基于测试记录所做的解读无偏颇(同上: 132)。测试使用的公平性主张: 1) 决策对不同群体的考生同样适切,即决策者在不同考生群体中需参考相同标准做出决策; 2) 不同考生群体应通过相同渠道充分了解决策依据,且有权知晓决策过程中相关主体是否严格贯彻该依据(同上: 134-135)。AUA 围绕测试过程和测试使用的四条主张收集证据,并根据以下依据对测试使用进行了论证: 1) 测试的使用及决策对所有涉考

者有益；2) 基于测试结果所做的决策需考虑教育现状、社会价值观及法律要求，且对所有涉考者公平；3) 对考生能力的解读有意义、无偏颇、可概推、与决策相关，且能为决策过程提供充足信息；4) 不同考生群体的测试记录在考查相同能力构念的不同考试中需具有一致性。由此观之，测试公平性在 AUA 的四条主张中均有体现。需要注意的是，该框架并非专为论证测试公平性而生，在学界多用于指导测试研发及效度证据的收集。

为提升测试公平性研究的系统性和可操作性，Xi (2010) 将公平性视作效度的一个方面，巧妙地将测试公平性纳入语言测试效度验证框架进行研究。她认为测试公平性指各评估环节（设计、开发、施测、使用）在不同群体间的可比效度，任何影响测试公平性的因素（如与构念无关的因素、构念代表性不足、不一致的施测行为、不恰当的决策程序等）均会削弱考试效度。具体而言，Xi 的测试公平性论证框架借鉴了 Toulmin (1958/2003) 的论证模型和 Chapelle *et al.* (2008) 的效度验证推理链。前者由主张 (claim)、依据 (data/grounds)、理据 (warrant)、支撑 (backing)、反驳 (rebuttal)、限定 (qualifiers) 等六个部分构成；后者包含目标域定义 (domain definition)、评估推理 (evaluation)、概化推理 (generalization)、解释推理 (explanation)、类推推理 (extrapolation)、使用推理 (utilization) 等六个环环相扣的推理过程。Xi (2010) 的公平性论证框架是一个逻辑严密的有机整体，前一个推理环节的“主张”经论证成立后，可以作为下一环节的“依据”，公平性论证伴随着效度论证贯穿测试始终。Xi 认为这一处理方式有利于研究人员精准定位公平性推理链中的薄弱环节，从而确定公平性研究的优先级，有针对性地解决威胁测试公平性的诸多问题。

### 3.3 伦理道德观

测试公平性作为社会、政治等因素驱动下的研究议题 (McNamara & Roever 2006)，若仅从“标准”和“论证”视角着手探讨，可能无法兼顾其背后的伦理及道德哲思。因此，有学者提出用基于伦理道德的理论框架来检验测试公平性。Kunnan (2000b, 2004) 将测试公平性视作多维的复杂概念，认为测试公平性包含效度、机会、正义三个维度，并提出了测试公平性框架 (Test Fairness Framework, TFF)。该框架由正义原则 (the Principle of Justice) 和惠益原则 (the Principle of Beneficence) 构成，即考试需对考生公平且于社会有益。该框架充分体现了测试实践对考生的人文关怀，能够触及效度验证框架的疏漏之处（如考生学习机会、考试注册费用等）。

基于道德原则的测试评价模式是伦理道德视角下的另一理论成果。从公平、正义的哲学思想出发，Kunnan (2018) 再次提出测试评价应包含两大原则：

公平原则 (the Principle of Fairness) 和正义原则。前者以考生为中心,主张考试应对所有考生公平,考试过程中考生应受到同等对待;后者指评估机构应主持公正,弘扬积极的价值观(同上:80-81)。这两大原则分别由若干个次原则构成。考虑到两大原则的可操作性,Kunnan(同上)将2014版《标准》中提及的测试公平性各要素纳入原则和次原则的考量范围,并采用Toulmin(1958/2003)的论证模型构建了公平与正义论证框架。参照两大原则和若干个次原则,提出相应的主张和次主张。收集公平性和正义方面的证据后,对次主张的成立逐一进行论证。若次主张均成立,主张的有效性也将得到支撑。

#### 4. 语言测试公平性的研究进展

标准观、论证观和伦理道德观推进了测试公平性的研究进程,在指导语言测试公平性的理论研究和实践过程中发挥了积极作用。

一些学者参照2014版《标准》,对语言测试实践的公平性进行了较为全面的审视。如Fan(2018)参考《标准》调查中国高校师生对国内现行语言测试实践的看法,所用问卷涵盖测试信息透明度、测试开发、施测、测试的公平性及社会影响等五个维度。研究发现师生普遍认为现行的语言测试实践较为公平。Fan & Jin(2020)对照《标准》,从考试说明与目的、考试内容与题型、题项开发与试测、施考与考试管理、评分与分数报告等六个方面全面审视了国内英语分级测试实践。该研究指出,考试的质量控制环节缺失和涉考者评估素养缺乏等因素对英语分级测试实践的规范化十分不利,致使测试公平性缺乏保障。除实证研究以外,Jonson *et al.*(2019)结合《标准》的若干要求,拟定了测试公平性评价量表,并附上测试实践案例指导该量表的使用。这一量表将《标准》与测试实践紧密结合,有助于进一步落实《标准》对测试公平性的要求。

另有学者在研究语言测试公平性的诸多议题时持论证观。如Deygers(2017)就两个荷兰语入学考试及基于考试所做的决策开展了历时研究,围绕六个涉及语言测试公平性和决策正义性的高风险主张收集证据,揭示了该地区入学考试语言政策的不合理之处,为该地区高校入学政策的修订提供了重要参考。此外,Xi(2010)的公平性论证框架也被用于Williamson *et al.*(2012)的评分研究和Yan *et al.*(2019)的语言测试题型研究之中。

还有部分学者沿袭伦理道德观的理念审视语言测试的公平性。如Ypsilandis & Mouti(2019)将测试实践按照测试开发、施测和测试影响三个大类进行了细分,并将具体的测试实践纳入道义论(deontologism)或结果论(consequentialism)的伦理道德范畴加以评判。此外,Kunnan(2004)提出的测试公平性框架能够有效指导语言考试项目的质量评估(Loh & Shih 2016;



Moghadam & Nasirzadeh 2020)。受 Kunnan (2004, 2018) 影响,除考试本身的质量以外,考生是否享有同等的学习机会和考试机会也常被用作评估考试项目公平性的重要依据 (Isbell & Kremmel 2020)。

## 5. 启示

基于以上对测试公平性定义、公平观及研究进展的梳理,笔者认为测试的公平性研究需进一步思考以下三个问题: 1) 如何看待测试公平性与效度之间的关系? 2) 如何对本土语言考试项目开展公平性研究? 3) 如何引入涉考者视角以促进考试公平?

### 5.1 测试公平性与效度的关系

测试公平性和效度这两个概念本身的定义及其各自的研究范畴不完全固定,因此两者之间的关系值得商榷。

笔者认为,在现阶段,测试公平性和效度均无法完全替代对方而独立存在。首先,测试公平性一直以来都是社会的共同追求。效度虽兼具技术性和社会性,但较为抽象,不易理解。仅从效度这一层面对测试公平性加以审视无法很好地回应社会对测试公平性的关切。其次,测试公平性和效度是提升测试科学性的不同路径。测试公平性的侧重点在于考试结果在不同考生群体中“无偏颇”,效度则更加关注考试结果的“准确性”和“一致性”。前者运用反证法,聚焦威胁测试公平性的主张,通过收集证据将其证伪,从而论证考试的科学性;后者运用枚举法,旨在提出测试效度的正面主张,收集证据并加以论证,从而为考试结果的使用提供理据。最后,测试公平性和效度的关注点不尽相同,具体采用何种定义、运用何种理论框架,不可一概而论,需视研究目的而定。

### 5.2 本土语言考试项目的公平性研究

本土考试项目旨在满足当地的测评需求,反映该语境下教育体系的价值观 (Dimova *et al.* 2020)。然而,相较于大规模、标准化的国际语言测试项目,学界较少关注本土语言考试项目的公平性。就我国而言,一方面,保证测试质量、规范测试从业人员行为的标准性文件较少,在很大程度上阻碍了测试行业的专业化和规范化;另一方面,多数本土语言考试项目缺乏公平性审查机制,对考试项目研发、实施及考试结果的使用等诸多关键环节缺乏有效监督。测试公平性总是不可避免地受其所处的社会环境影响 (Bachman 1990),语境不同,考试结果的解释和使用也会随着考试用户新需求的出现而不断变化 (Kane 2013)。因此,有必要为本土语言考试项目建立常态化的公平性审查机制,持续关注考试结果的使用情况。为此,本文提出以下两点建议:

第一,学界及相关部门应尽快制订基于我国国情的语言测试标准性文件,保障我国测试行业的健康发展,维护教育公平。早在世纪之交, Bachman (2000) 就强调了语言测试领域专业化建设的重要性。其中,值得努力的方向之一就是建立测试实践标准,并健全相关机制保障其实施。在中国,杨惠中、桂诗春 (2007) 曾提出过类似的倡议,呼吁政府职能部门为语言测试这一重要的社会实践活动提供质量标准。标准的缺失将直接导致考试质量的评价无据可依,考试结果的使用缺乏监督。尽管语言测试标准的制订道阻且长,但作为一项基础性工程,于千千万万的考生而言、于教育公平而言,意义重大。

第二,学界应逐步建立我国语言考试项目的公平性审查机制。Kunnan (2005) 的测试语境框架 (Test Context Framework) 为这一目标的实现提供了思路。该框架主张从宏观层面对测试公平性展开分析,关注施考地区的政治、经济语境,教育、社会、文化语境,技术、设施语境以及法律、伦理语境。语境类别可结合具体情况进行调整。本土语言考试项目的公平性审查需立足当地语境,综合考量施测当地的政策、法律法规、教育与社会背景以及涉考者的价值观。

### 5.3 涉考者视角的引入

涉考者在测试实践中扮演着重要角色,在测试公平性研究中是否引入涉考者视角这一问题值得深思。Weiss (1983) 认为引入涉考者视角有助于提升语言测试的公平性,使测试评价的过程更加民主,但不能保证涉考者提供的信息相关且合理。因此,在审视测试公平性的过程中,如有必要引入涉考者视角,测试专家不仅要为涉考群体提供一个阐明疑虑、表达想法的平台,更要顾及其评估素养,主动为他们提供必要、易解读的信息,帮助其做出合理评价和判断。

笔者认为,在测试公平性研究中引入涉考者视角有其必要性。一方面,涉考者作为制定决策或受决策影响的主体,有权利对测试不公平现象提出质疑,且有义务监督测试从业者的言行。另一方面,涉考者对测试公平性的看法能帮助测试研究者和从业者明确不同涉考群体的需求及其对测试公平性的期待,并对涉考者评估素养的发展模式予以关注 (参见 Yan & Fan 2021)。基于涉考者视角,测试研究人员可以面向不同涉考群体有针对性地组织语言测试工作坊并设计相应的培训材料,逐步加深涉考群体对语言测试专业知识和技能的了解,帮助他们明确自己在测试实践中所扮演的角色,并强化其责任意识。随着涉考群体评估素养的提升,考试将在整个教育体系及社会中持续发挥积极的导向作用。

## 6. 结语

本文聚焦 20 世纪 80 年代以来语言测试公平性的内涵、公平观及研究进展,

在回顾和反思的基础上,提出了三个有待深入探讨的议题,呼吁学界在厘清测试公平性与效度关系的同时,关注本土语言考试项目的公平性,并尝试在公平性研究中引入涉考者视角,为科学命题和公平决策提供依据,促进教育及社会公平。测试公平于考生个人命运、教育公平和社会公平而言意义重大,需要测试从业者与涉考者携手并进,共同维护测试实践的科学性与专业性。

#### 参考文献

- Alderson, C. 1988. This test is unfair: I'm not an economist! [A]. In P. Carrell, J. Devine & D. Eskey (eds.). *Interactive Approaches to Second Language Reading* [C]. Cambridge: CUP. 168-182.
- American Educational Research Association (AERA), American Psychological Association (APA) & National Council on Measurement in Education (NCME). 1985. *Standards for Educational and Psychological Testing* [M]. Washington: American Psychological Association.
- Angoff, W. 1982. Use of difficulty and discrimination indices for detecting item bias [A]. In R. Berk (ed.). *Handbook of Methods for Detecting Test Bias* [C]. Baltimore: Johns Hopkins University Press. 96-116.
- Bachman, L. 1990. *Fundamental Considerations in Language Testing* [M]. Oxford: OUP.
- Bachman, L. 2000. Modern language testing at the turn of the century: Assuring that what we count counts [J]. *Language Testing* 17(1): 1-42.
- Bachman, L. & A. Palmer. 2010/2016. *Language Assessment in Practice: Developing Language Assessments and Justifying Their Use in the Real World* [M]. Beijing: Foreign Language Teaching and Research Press.
- Chapelle, C., M. Enright & J. Jamieson. 2008. *Building a Validity Argument for the Test of English as a Foreign Language* [M]. London: Routledge.
- Cole, N. & M. Zieky. 2001. The new faces of fairness [J]. *Journal of Educational Measurement* 38(4): 369-382.
- Crocker, L. 2003. Teaching for the test: Validity, fairness, and moral action [J]. *Educational Measurement: Issues and Practice* 22(3): 5-11.
- Cronbach, L. 1988. Five perspectives on validity argument [A]. In H. Wainer & H. Braun (eds.). 1988. 3-17.
- Davies, A. 2010. Test fairness: A response [J]. *Language Testing* 27(2): 171-176.
- Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamara (eds.). 1999. *Dictionary of Language Testing* [Z]. Cambridge: UCLES & CUP.
- Deygers, B. 2017. Assessing High-stakes Assumptions: A Longitudinal Mixed-methods Study of University Entrance Language Tests, and of the Policy that Relies on Them [D]. Ph.D. Dissertation. KU Leuven.
- Dimova, S., Xun Yan & A. Ginther. 2020. *Local Language Testing: Design, Implementation, and Development* [M]. New York: Routledge.
- Educational Testing Service (ETS). 2014. *ETS Standards for Quality and Fairness* [S]. Princeton: Educational Testing Service.
- Fan, Jinsong Jason. 2018. A survey of English language testing practices in China: Students' and

- teachers' perspectives [A]. In D. Xerri & P. Briffa (eds.). *Teacher Involvement in High-stakes Language Testing* [C]. Cham: Springer. 283-300.
- Fan, Jinsong Jason & Yan Jin. 2020. Standards for language assessment: Demystifying university-level English placement testing in China [J]. *Asia Pacific Journal of Education* 40(3): 386-400.
- Isbell, D. & B. Kremmel. 2020. Test review: Current options in at-home language proficiency tests for making high-stakes decisions [J]. *Language Testing* 37(4): 600-619.
- Joint Committee on Testing Practices (JCTP). 2004. *Code of Fair Testing Practices in Education* [M]. Washington, D.C.: Joint Committee on Testing Practices.
- Jonson, J., P. Trantham & B. Usher-Tate. 2019. An evaluative framework for reviewing fairness standards and practices in educational tests [J]. *Educational Measurement: Issues and Practice* 38(3): 6-19.
- Kane, M. 2010. Validity and fairness [J]. *Language Testing* 27(2): 177-182.
- Kane, M. 2013. Validating the interpretations and uses of test scores [J]. *Journal of Educational Measurement* 50(1): 1-73.
- Kunnan, A. (ed.). 2000a. *Fairness and Validation in Language Assessment* [C]. Cambridge: CUP.
- Kunnan, A. 2000b. Fairness and justice for all [A]. In A. Kunnan (ed.). 2000a. 1-13.
- Kunnan, A. 2004. Test fairness [A]. In M. Milanovic & C. Weir (eds.). *European Language Testing in a Global Context* [C]. Cambridge: CUP. 27-48.
- Kunnan, A. 2005. Language assessment from a wider context [A]. In E. Hinkel (ed.). *Handbook of Research in Second Language Learning* [C]. Mahwah, N.J.: Lawrence Erlbaum. 779-794.
- Kunnan, A. 2010. Test fairness and Toulmin's argument structure [J]. *Language Testing* 27(2): 183-189.
- Kunnan, A. 2018. *Evaluating Language Assessments* [M]. New York: Routledge.
- Li, Hongli & Hoi K. Suen. 2012. Are test accommodations for English language learners fair? [J]. *Language Assessment Quarterly* 9(3): 293-309.
- Li, Keqiang [ 李克强 ]. 2021. Premier stresses education equality [OL]. [http://www.gov.cn/premier/2021-03/23/content\\_5595090.htm](http://www.gov.cn/premier/2021-03/23/content_5595090.htm) (accessed 02/11/2020). [ 机会公平中教育公平是最大的公平 ]
- Loh, Tai Lee & Chih-Min Shih. 2016. The English language test of the Singapore Primary School Leaving Examination [J]. *Language Assessment Quarterly* 13(2): 156-166.
- Madaus, G. 1990. *Testing as a Social Technology: The Inaugural Annual Boise Lecture on Education and Public Policy* [M]. Boston: Boston College.
- McNamara, T. & C. Roever. 2006. *Language Testing: The Social Dimension* [M]. Oxford: Blackwell.
- Messick, S. 1988. The once and future issues of validity: Assessing the meaning and consequences of measurement [A]. In H. Wainer & H. Braun (eds.). 1988. 33-45.
- Moghadam, M. & F. Nasirzadeh. 2020. The application of Kunnan's test fairness framework (TFF) on a reading comprehension test [J]. *Language Testing in Asia* 10: Article No. 7.
- National Development and Reform Commission [ 国家发展和改革委员会 ]. 2021. *The Outline of the 14th Five-Year Plan (2021-2025) for National Economic and Social Development and the Long-Range Objectives through the Year 2035 of the People's Republic of China* [M]. Beijing: People's Publishing House. [ 《中华人民共和国国民经济和社会发展第十四个五年规划和



- 2035 年远景目标纲要》。北京：人民出版社 ]
- Shepard, L., G. Camilli & M. Averill. 1981. Comparison of procedures for detecting test-item bias with both internal and external ability criteria [J]. *Journal of Educational Statistics* 6(4): 317-375.
- Shohamy, E. 1998. Critical language testing and beyond [J]. *Studies in Educational Evaluation* 24(4): 331-345.
- Spaan, M. 2000. Enhancing fairness through a social contract [A]. In A. Kunnan (ed.). 2000a. 35-38.
- Stansfield, C. 1993. Ethics, standards and professionalism in language testing [J]. *Issues in Applied Linguistics* 4(2): 189-206.
- The Central Committee of the Communist Party of China & The State Council [ 中共中央、国务院 ]. 2020. *Overall Plan for Deepening Educational Evaluation Reform in the New Era* [M]. Beijing: People's Publishing House. [ 《深化新时代教育评价改革总体方案》。北京：人民出版社 ]
- Toulmin, S. 1958/2003. *The Uses of Argument* [M]. Cambridge: CUP.
- Wainer, H. & H. Braun (eds.). 1988. *Test Validity* [C]. New York: Routledge.
- Weir, C. 2005. *Language Testing and Validation* [M]. New York: Palgrave Macmillan.
- Weiss, C. 1983. The stakeholder approach to evaluation [A]. In A. Bryk (ed.). *Stakeholder-based Evaluation: New Directions in Educational Evaluation* [C]. San Francisco: Jossey-Bass. 3-14.
- Williamson, D., Xiaoming Xi & F. Breyer. 2012. A framework for evaluation and use of automated scoring [J]. *Educational Measurement: Issues and Practice* 31(1): 2-13.
- Willingham, W. & N. Cole. 1997. *Gender and Fair Assessment* [M]. Mahwah, N. J.: Lawrence Erlbaum.
- Xi, Xiaoming. 2010. How do we go about investigating test fairness? [J]. *Language Testing* 27(2): 147-170.
- Yan, Xun & Jason Jinsong Fan. 2021. “Am I qualified to be a language tester?”: Understanding the development of language assessment literacy across three stakeholder groups [J]. *Language Testing* 38(2): 219-246.
- Yan, Xun, Lixia Cheng & A. Ginther. 2019. Factor analysis for fairness: Examining the impact of task type and examinee L1 background on scores of an ITA speaking test [J]. *Language Testing* 36(2): 207-234.
- Yang, Huizhong & Shichun Gui [ 杨惠中、桂诗春 ]. 2007. The sociology of language testing [J]. *Modern Foreign Languages* (4): 368-374. [ 语言测试的社会学思考, 《现代外语》4 ]
- Ypsilandis, G. & A. Mouti. 2019. Fairness and ethics in multiple choice (MC) scoring: An empirical study [J]. *Journal of Language and Education* 5(1): 85-97.
- Zumbo, B. 2007. Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going [J]. *Language Assessment Quarterly* 4(2): 223-233.
- Zwick, R. 2019. Fairness in measurement and selection: Statistical, philosophical, and public perspectives [J]. *Educational Measurement: Issues and Practice* 38(4): 34-41.

收稿日期：2021-03-15；修改稿 2021-08-09；本刊修订 2021-11-10

通讯地址：310058 浙江省杭州市 浙江大学外国语言文化与国际交流学院



inner circle includes three overlapping hypotheses, that is, cognitive interconnectedness, social interconnectedness, and emotional interconnectedness, with ‘problems’ placed on the top of the hypotheses as a node of their links; and the outer circle demonstrates the working process of the online activities, consisting of three phases: group preparation, CTTC gathering, and follow-up activities. The proposed theory assumes that the closer the interconnectedness is, the better the learning outcomes are; the degree of interconnectedness depends on the quality of the designed ‘problems’ based on specific tasks. The three hypotheses must function during the entire working process to achieve the desired outcome.

**Test fairness: Conceptualization, perspectives, and implications (p. 79)**

*HE Lianzhen & ZHANG Juan* (School of International Studies, Zhejiang University, Hangzhou 310058, China)

Tests, as talent selection and screening instruments, are relatively fair mechanisms implemented widely in the contemporary world and play a pivotal role in ensuring educational fairness and social justice. Following the introduction of the conceptualization of test fairness, this article reviews the approaches to exploring test fairness (i.e. standards-based, argument-based, and ethics-based approaches) as well as research along those lines. The article concludes with recommendations of avenues for further research. Specifically, researchers need to examine the relationship between test fairness and validity when conducting test fairness research. In addition, research on the fairness of local language tests is highly encouraged. Further, the perspectives of various stakeholders could be incorporated into the investigation of test fairness.

**Online processing of unaccusative constructions by Chinese English learners (p. 90)**

*ZENG Tao & GAO Yarong* (School of Foreign Languages, Hunan University, Changsha 410082, China)

*DAI Xuequn* (School of Foreign Languages and Literature, Shandong University, Ji'nan 250100, China)

The present study investigates Chinese English learners’ online processing of unaccusative/unergative constructions and subtypes (less core/peripheral) of unaccusative constructions through self-paced reading tasks. The results show that: 1) Chinese English learners spend more time processing unaccusative constructions, which supports the Unaccusative Hypothesis on the one hand but reflects the limitations of the internal interface theory in the Interface Hypothesis of language acquisition on the other. 2) Chinese English learners have a better acquisition of the peripheral unaccusative constructions than the less core unaccusative constructions, reflecting the larger problems of the acquisition of verbs with a stronger meaning of external causation, hence casting doubt on the Unaccusative Hierarchy Hypothesis; Chinese English learners can more effectively process peripheral unaccusative constructions which do not involve syntactic movement, lending support to the Interface Hypothesis that single language modules are easier to be acquired.

**A study of pragmatic enrichment priming among Chinese English learners (p. 103)**

*JIA Li & YANG Lianrui* (College of Foreign Languages, Ocean University of China, Qingdao 266100, China)

The present study aims to examine the pragmatic enrichment priming effect (PERE) and its factors by control experiments using EFL learners among English and non-English majors as subjects. The results show that: 1) PERE is found in interactions where pragmatic enrichment appears, which is not affected by the number of primes. 2) There are significant differences in pragmatic enrichment production across contexts in the four prime-target types, and the PERE in the same prime-target context is significantly higher than in different prime-target contexts. 3) The PERE of high proficiency was significantly higher than that of low proficiency; language

## 本刊稿约

### 一、本刊欢迎以下各类来稿:

1. 对语言学各领域、各交叉语言学科、语言学史及语言哲学的原创性研究和评述;
2. 对各国语言文字所做的具有一定普遍意义和理论价值的单独研究, 以及与中国语言文字的比较研究;
3. 关于第二语言习得与外语教学理论和方法的研究、实验和调查报告, 以及对中外外语教育政策、历史的考察和分析等;
4. 对翻译理论、实践和方法以及翻译史等的考察和分析;
5. 对域外文化摄入和中华文化走出的双向探讨, 以及对中外思想文化的比较研究;
6. 与上述内容有关的书刊评介、新书简介、札记、访谈、报道等。

二、来稿请力求精炼, 凡论文以 12000 字以内为宜, 书刊评介、札记、访谈 6000 字以内为宜, 简介、报道等不超过 3000 字。论文请附中、英文摘要和题目。

三、投稿及查询稿件格式要求和稿件处理情况请登陆本刊网站: <http://www.fltr.ac.cn>。

四、本刊按国际学术界通行做法实行同行专家匿名审稿, 故来稿正文请勿写作者姓名和单位。

五、本刊力求高效处理稿件, 一般稿件三个月内给出处理结果。

六、来稿一经发表即付稿酬, 优稿优酬, 并赠当期刊物 1—2 册。

本刊地址: 北京市海淀区西三环北路 19 号 北京外国语大学西院国内大厦《外语教学与研究》编辑部  
电 话: (010) 88816466 Email: [bwyys@126.com](mailto:bwyys@126.com)

外语教学与研究 (北京外国语大学学报) 1957 年创刊

2022 年第 1 期

主管单位 中华人民共和国教育部

主办单位 北京外国语大学

编辑单位 《外语教学与研究》编辑部

100089 北京西三环北路 19 号

电话: (010) 88816466

电子邮件: [bwyys@126.com](mailto:bwyys@126.com)

本刊网址: <http://www.fltr.ac.cn>

主 编 王 克 非

出版单位 外语教学与研究出版社

印刷单位 北京盛通印刷股份有限公司

订 阅 处 全国各地邮局

国内总发行 北京报刊发行局

国外总发行 中国国际图书贸易集团有限公司

国内统一刊号: CN11-1251/G4

国内邮发代号: 2-130

国外代号: BM203

2022 年 1 月 20 日出版

国内定价: 25.00 元

ISSN 1000-0429





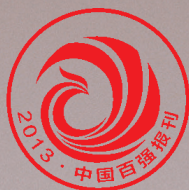


JOURNAL OF  
ZHEJIANG UNIVERSITY  
HUMANITIES AND SOCIAL SCIENCES

# 浙江大学学报

人文社会科学版

●首批国家社科基金资助期刊



**5/2024**

Vol. 54, No. 5, May 2024

Zhejiang Daxue Xuebao (Renwen Shehui Kexueban)

JOURNAL OF ZHEJIANG UNIVERSITY

# 浙江大学学报

(人文社会科学版)

## 编辑委员会

主 任 黄先海

顾 问 (以下按姓氏音序排列)

束景南 田正平 王重鸣

编 委 白谦慎 褚超孚 何莲珍 胡 铭 黄 旦 黄先海 李 实

刘 斌 刘海峰 刘海涛 刘同舫 罗卫东 毛 丹 倪梁康

桑 兵 史晋川 眭依凡 王 杰 韦 路 魏 江 吴晓波

徐 枫 徐小洲 许 钧 姚先国 余逊达 郁建兴 袁亚春

张俊森 张涌泉 赵鼎新 赵志荣 周江洪

特约编委 卜宪群 孙笑侠 袁振国 张 军 张永清

国际编委 安守廉(William P. Alford):美国哈佛大学教授

伯纳黛特·贝克(Bernadette Baker):美国威斯康星大学麦迪逊分校教授

杰瑞米·克莱格(Jeremy Clegg):英国利兹大学教授

主 编 黄先海

执行主编 朱希伟

执行副主编 陈 洁 钱文荣 赵 骏



# 目次

# CONTENTS

● 2024 年 第 5 期 ● 第 54 卷

浙江大学学报

(人文社会科学版)

## 主题栏目:马克思主义研究

中国式现代化的中华文明意蕴

庞 虎 金姚炫(5)

资本如何驯服劳动力?

——马克思《资本论》的福柯式权力主义解读

吴奕乐 陈嘉明(18)

本栏目特约主持人:刘同舫教授

## 主题栏目:数字传播研究

数字“记忆之场”:社交媒体中英烈纪念空间的重建

邵 鹏 赵若邑 潘仲菁(29)

AIGC时代虚假信息的制造传播与国家安全及公民权益的维护

陆建平 党自强(42)

本栏目特约主持人:邵培仁教授

## 法学研究

“法规引用标准”解释论

——以关键词为中心

柳经纬(59)

## 管理学研究

包容性视角下农业产业集群提升农民工居留意愿的路径研究

阮建青 张雨薇(71)

期刊基本参数: CN 33-1237/C\*1955\*B\*A4\*160\*ZH\*P\*¥ 70.00\*1500\*12\*2024-05



本刊鼓励学术创新,严守学术规范,秉承浙江大学求是创新的学风,积极传播先进文化,严格按国际惯例实行同行专家双向匿名审稿制,努力为具有原创性的人文社会科学学者提供与世界学术对话的机会。

## 匿名审稿 欢迎订阅

本刊不以任何形式收取版面费、审稿费  
全国社科工作办举报电话:010-63098272

被美国《剑桥科学文摘》等14个著名国际索引收录

### 主题栏目:新出文献与中古文学研究

以“燕许”为宗的唐文变革考论

张兴武(87)

唐诗人崔备家族及任官考

沈文凡 谢梦莹(97)

本栏目特约主持人:胡可先教授

### 主题栏目:艺术与考古研究

16世纪中西跨艺术诗学互文比较

应宜文(115)

本栏目特约主持人:陈野教授

### □ 语言学研究

“公平”的理论向度

——兼论对语言测试公平性研究的启示

何莲珍 张娟(122)

### □ 文学研究

南宋湖州《思溪藏》的刊刻历程

李周渊(131)

初唐文士与文学仕进的演进

徐焕(147)

校史专栏:科学精神:竺可桢谈科学与人文的共性

卢军霞(17)

档案所见杭州大学出版社成立始末

宋旭华(114)

学术补白:晚清民国时期西泠印社篆刻书籍的出版

李寒晴(121)

征订启事·征稿简则

(28)(160)

出版日期:2024年5月15日

本期执行编辑:黄山杉 刘双庆 周梦烨 郭琳琳

# Journal of Zhejiang University

Humanities and Social Sciences

(Monthly)

Volume 54: Number 5, May 2024

## Contents

The Implication of Chinese Civilization in Chinese Path to Modernization .....	Pang Hu Jin Yaokai ( 5 )
How Capital Tames Labor Power: A Foucaultian Powerism Interpretation of Marx's <i>Capital</i> .....	Wu Yile Chen Jiaming ( 18 )
Digital "Fields of Memory": The Reconstruction of Heroic Commemorative Space in Social Media .....	Shao Peng Zhao Ruoyi Pan Zhongjing ( 29 )
Manufacture and Dissemination of Fake Information vs. Protection of National Security and People's Rights and Interests in the AIGC Era .....	Lu Jianping Dang Ziqiang ( 42 )
An Explanation of the "Reference of Regulations to Standards": Centered on Keywords .....	Liu Jingwei ( 59 )
Research on the Realization Path of Agro-industrial Clusters Enhancing Migrant Workers' Settlement Willingness from the Perspective of Inclusiveness .....	Ruan Jianqing Zhang Yuwei ( 71 )
A Study of the Reform of Tang Literature Based on "Yan-Xu" .....	Zhang Xingwu ( 87 )
Clan and Official Career Life of Tang Poet Cui Bei .....	Shen Wenfan Xie Mengying ( 97 )
An Intertextual Comparison of Chinese and Western Poetics of Ekphrasis in the Sixteenth Century .....	Ying Yiwen (115)
Theoretical Perspectives on Fairness: Implications for Fairness Research in Language Testing .....	He Lianzhen Zhang Juan (122)
An Analysis of the Engraving Process of <i>Sixi Canon</i> in Huzhou during the Southern Song Dynasty .....	Li Zhouyuan (131)
Early Tang Dynasty Literati and the Evolution of Political Career Advance via Literature .....	Xu Huan (147)

Translators and revisers of this issue : Huang Tianhai

# “公平”的理论向度

## ——兼论对语言测试公平性研究的启示

何莲珍 张 娟

(浙江大学 外国语学院, 浙江 杭州 310058)

**[摘 要]** 语言测试公平性研究近年来受到测试学界的持续关注。然而,“测试公平性”的含义、研究范畴和研究方法等重要议题至今仍未有定论。厘清“公平”的基本含义对理解“测试公平性”的内涵至关重要,这是语言测试学界开展理论争鸣与实证研究的必经之路。“公平”作为一个跨学科概念,具有多维性。从教育测量、规程标准和哲学三个视域观之,有关“公平”的学理主张各不相同、各有侧重。这些主张对“测试公平性”相关研究具有指导意义。具体而言,研究的开展需要考虑具体的测试语境、诉诸公开透明的证据、倾听来自不同涉考群体的声音。若将测试公平性视作一个连续统,则每一个不公平因素的消弭都能让测试实践趋向公平。

**[关键词]** 公平;测试公平性;理论视角

### 一、引 言

20世纪90年代,伴随着1996年和1997年两届国际语言测试研讨会的召开,公平性作为一个研究议题,正式在语言测试的研究议程中占有一席之地<sup>[1]</sup>。时至今日,“测试公平性”这一概念的重要程度已然与“效度”“信度”等现代语言测试理论中的核心概念比肩。在某种意义上,甚至可以说测试公平是我国教育系统乃至整个社会共同的价值诉求,没有任何一项测试特征比测试公平性更能牵动民众和媒体的神经。

作为一个多维概念,“公平”的含义在不同学科体系和不同社会文化语境中不断经历着演化和发展。与之类似,“测试公平性”的含义在测试学界仍未形成共识<sup>[2]</sup>。鉴于“公平”这一概念在多个领域已有一定的学理积淀,有必要了解其他视域下有关“公平”的理论探讨,掇菁撷华,反哺语言测试公平性研究。囿于篇幅,剖析“公平”的基本含义之余,本文仅梳理教育测量、规程标准和哲学三个视域下有关“公平”的理论观点,探讨其可资借鉴之处,以期为进一步开展语言测试公平性研究提供启示。

**[收稿日期]** 2023-03-13

**[在线优先出版日期]** 2024-05-06

**[基金项目]** 国家社会科学基金一般项目(20BYY107)

**[作者简介]** 1. 何莲珍(<https://orcid.org/0000-0002-8480-7017>),女,浙江大学外国语学院语言与翻译系教授,博士生导师,文学博士,主要从事应用语言学、语言测试与评估研究;2. 张娟(<https://orcid.org/0000-0002-1253-9287>),女,浙江大学外国语学院语言与翻译系博士研究生,主要从事语言测试与评估研究。

**[本刊网址·在线杂志]** <http://www.zjujournals.com/soc>

**[网络连续型出版物号]** CN 33-6000/C



## 二、“公平”的基本含义

“公平”是一个内涵极为丰富的概念,身处不同地区、不同历史阶段,人们对“公平”的理解迥然不同<sup>[3]</sup>。可见,穷尽其含义的难度着实不小。本节拟结合我国和少数几个西方国家的语境,归纳“公平”的若干基本含义。选取这些含义加以呈现,原因有二:其一,这些含义在不同文化语境中具有共性;其二,这些含义与语言测试界学者对测试公平性的界定有契合之处。

若把“公平”视作一个关系范畴,按照关系主体加以区分,“公平”有三层基本含义。就第一对关系主体(即个体与其所得之间的关系)而言,“公平”的两层基本含义分别是:(1)个体的条件和努力与所得或回报相匹配<sup>[4]</sup>; (2)个体的合理期待能够得到满足<sup>[5]</sup><sup>3</sup>。这两层含义表明,包括命题、施考、评分、基于考试结果所作决策在内的测试实践如果违背考生、教师、家长等涉考群体的合理预期,那么考试很有可能被认为是不公平的。针对第二对关系主体(即个体与他人之间的关系),“公平”意味着“同等情况同等对待”<sup>[5]</sup><sup>3</sup>。这一基本含义要求我们待人接物时需要遵循既定的规则 and 标准,面对相同或相似的情况要做到一视同仁,不偏不倚。也就是说,如果语言考试结果仅取决于考生的语言水平,不受考生国籍、种族、性别等与考试构念无关因素的影响,那么考试在这个意义上就是公平的。这是因为语言水平相同或相似的考生,无论其背景如何,如果最终取得的成绩相同或相似,那么考生之间作横向比较时,内心不会产生不平衡之感。

若将“公平”视作基本的价值追求,“公平”意指“对人的权利和尊严予以同等的尊重”<sup>[6]</sup><sup>80</sup>。国际语言测试学会制定的《道德规范》要求语言测试工作者“尊重每一位考生的人性及尊严”,其中包括“不歧视或利用不同年龄、性别、种族、民族、性取向、语言背景、信仰、政治团体或宗教的考生”<sup>[7]</sup><sup>2</sup>。比如,语言测试开发人员需要确保考试内容适切、语言中立,考试材料中的观点不会因为违背考生的价值观而导致考生情感上的不适。Kunnan提出的测试公平原则同样强调了尊重考生的重要性,这一原则要求测试从业人员尊重考生平等的学习机会以及在考试过程中被平等对待的基本权利<sup>[6]</sup><sup>80</sup>。

若将“公平”置于时间维度观之,有“事后弥补由于过往原因所造成的不公平”之意<sup>[5]</sup><sup>5</sup>。以美国为例,历史上的种族隔离政策曾导致非裔美国学生的学业成绩显著低于美国白人学生的成绩。如Lawlor等人采集了美国维克森林大学非裔美国人毕业生和美国白人毕业生入学时的SAT成绩及其本科绩点,发现非裔美国学生的SAT成绩比白人学生的平均成绩低80分左右,且SAT总分不能显著预测学生在校期间的学业表现<sup>[8]</sup>。因此,作者有理由认为SAT充斥着种族歧视。如果美国高校仅将申请者的SAT成绩作为录取依据,对于非裔美国学生来说,录取结果显然有失公平。为确保美国少数族裔和妇女等弱势群体在接受教育和就业等方面享有平等的权利和机会,20世纪60年代美国平权运动兴起。《1964年民权法案》颁布后,美国诸多高校招生时会采取平权措施。除了评估申请者的学业成绩以外,这些高校会将申请者的种族背景作为重要的考虑因素,录取时实行种族平衡政策<sup>[9-10]</sup>。平权措施旨在弥补由于过往政策、经济、社会、文化等因素造成的考生能力起点不平等的问题。

此外,“公平”还意味着“同一竞争或活动的参与者需要受相同规则制约”<sup>[5]</sup><sup>3</sup>。规则统一能够确保参与者拥有同样的机会展现个人能力、拥有同样的机会获得成功。任何破坏规则的行为均违背了考试的公平竞争原则。其中,舞弊就是典型的违规违纪行为。

## 三、“公平”的理论向度

上一小节抽丝剥茧,足见“公平”含义之丰富。但是,不同学科视域下,“公平”的定义各有侧重,

对“公平”的理论考量也各不相同。本节将依次阐述并简要评价教育测量、规程标准和哲学视域下有关测试公平性的理论观点和主张。

### (一) 教育测量向度

教育测量学界最初对“公平”的定义较为技术化,认为公平的考试应该具备以下两个特征:(1)测量精准;(2)不存在偏颇<sup>[11-12]</sup>。测量精准能够确保考生得到的考试成绩与其能力相匹配,符合其合理预期;而考试不存在偏颇则能确保考试对所有考生一视同仁,考生的种族、性别等与考试构念无关的因素不影响考试结果。通过以上定义,我们可以看到,教育测量学界将“公平”视作一个关系范畴。

随后,教育测量学界逐渐意识到测试公平性的定义不应被狭隘地定义在技术层面,转而采用一种较为广义的定义方式<sup>[13]81</sup>。如Cronbach认为,考试作为一种客观的评价方式,可以发挥一定的政治功能。比如,考试结果可以用于决定谁可以得到什么<sup>[14]5</sup>。这里的“所得之物”并不是直接的物质利益,而是个体在社会中所能获得的机会、服务和社会地位<sup>[15]</sup>。由此,除测评工具本身的质量以外,测试公平性的研究范畴进一步外延至考试之前和考试之后两个阶段。前者主要关注考生受教育的机会是否公平<sup>[16]</sup>,后者主要关注考试结果的使用是否公平<sup>[17-18]</sup>。可见,测试公平性与考试赖以生存的政治、经济和社会文化环境有着千丝万缕的联系<sup>[19]</sup>。

狭义和广义的定义方式各有利弊。狭义的定义方式使得测试公平性研究更具可操作性。收集考生的作答数据之后,题项、题组或测试功能差异(differential item/bundle/test functioning,以下简称DIF、DBF和DTF)等检测技术能够甄别试卷中的潜在偏颇,辅助考试开发人员判断与考试构念无关的因素是否影响考试结果。但是,狭义的测试公平往往是相对于考生群体而言的,并未虑及考试对考生个体而言是否公平。这可能是源于DIF、DBF等统计方法自身的局限性,因为这些方法几乎无一例外地将能力相同、显性背景特征不同的考生群体视作分析对象,进而检验不同考生群体正确作答题目或题组的概率是否相同。同时,由于DIF、DBF等偏颇检测技术只能在考试结束之后用于分析考生的作答数据,狭义视角下的公平性研究存在滞后性。此外,从狭义的角度来看,“测试公平性”只能涵盖“个体的条件和努力与所得或回报相匹配”“同等情况同等对待”等“公平”的基本含义,而无法兼顾其他基本含义。广义视角下的测试公平则能比较全面地涵盖“公平”的诸多基本含义,同时能更好地关切考试对考生个体而言是否公平。但是,广义视角下,“测试公平性”与“效度”两个概念之间有所重叠,各自的研究范畴也会有所交叉<sup>[20-22]</sup>。

### (二) 规程标准向度

鉴于测试公平性已逐渐成为考试质量评价的重要维度,司法机构、国际组织和学界相继颁布了与测试公平性相关的规程标准,用以指导规范的测试实践,提升考试服务质量。按照严厉程度从高到低排序,规程标准文件大致可以分为四类:法律法规、法律授权性标准、权威性的行业标准和组织机构自行颁布的规程标准<sup>[5]53</sup>。由于诸多规程标准在制订目的、制订主体、所针对的涉考群体、所关涉的考试环节等方面各不相同,针对每一类别中较为典型或影响力较大的文件,本节将择其重点加以介绍。

法律法规通常对考试过程中明令禁止的行为及惩罚措施作出相应规定,这些规定多为强制性规范,用以维护考试程序公平。我国作为考试制度的发源地,历史上有着丰富的考试立法经验<sup>[23]</sup>,此处不作赘述。本节仅对近年来国内外颁布的法律法规作简要回溯。《中华人民共和国教育法》第七十九条至八十一条明令禁止包括作弊、替考在内的舞弊行为,并对举办国家教育考试的行政部门和考试机构提出了明确的法律责任要求。此外,《中华人民共和国刑法修正案(九)》第二百八十四条明确规定了组织考试作弊罪,非法出售、提供试题、答案罪和代替考试罪。这些法律法规对考试

违纪行为起到了极大的震慑作用,有助于营造公平竞争的良好氛围。另有一些法律法规旨在保障残障人士的合法权利,主张消除不利于残障人士平等获得学习、工作机会的障碍。比如《美国残疾人法案》(Americans with Disabilities Act)、英国的《残疾人歧视法案》(Disability Discrimination Act)和《平等法案》(Equality Act)等法律文件均提出要保证残障人士拥有平等的考试机会,考试机构必要时应根据残障考生的实际情况为其提供便利,如提供盲文试卷、适当延长考试时间等。

法律授权性标准的颁布主体一般是法定的考试监管机构。较为典型的是英格兰地区的资格与考试监管局,该机构颁布的《国家测评监管框架》将减少考试偏颇视作保证考试效度的重要举措,规定“考试应最大限度地减少偏颇,考试结果应仅取决于考生的能力水平”<sup>[24]</sup>。适用于英国威尔士地区的文件《通过考试设计实现机会公平》旨在通过优化考试设计,为每一位考生提供公平的考试机会,使得考生能够最大限度地展示自己的能力水平<sup>[25]</sup>。该文件提供了一份较为全面的测试实践规范性审查清单,涵盖诸如试题措辞、内容选材、卷面设计等方面需要注意的细节,值得测试从业人员细细翻阅。但是,这两份文件目前仅适用于英国的部分地区,其影响范围较为有限。

权威性的行业标准虽然不具法律效力,但这些标准的颁布极大地提高了教育测评实践的专业化程度,为考试内容的敏感性审查提供了重要的参考依据,因而在业内备受尊崇。其中,最具影响力的行业标准是美国的《教育与心理测试标准》(Standards for Educational and Psychological Testing,以下简称《标准》)。截至目前,该文件共修订六次,其中对测试公平性的定义和章节编排能够反映学界在不同历史阶段对这一概念所达成的共识。鉴于已有文献对《标准》中测试公平性的相关内容进行了详细介绍<sup>[26]</sup>,本节不作赘述。除《标准》之外,面向考试开发和考试结果的使用,《教育测试实践公平性准则》(Code of Fair Testing Practices in Education,以下简称《准则》)从考试的开发和选择、施考和评分、考试结果的报告和解读、考试信息的公布四个方面规定了考试开发者和使用者各自应尽的义务。最新版《准则》指出,测试公平性并非凭借考试开发者一己之力就可以实现,需要考生和考试使用者共同维护<sup>[27]</sup>。关于考试结果的使用,国际测试委员会在其颁布的《国际测试使用指南》中明确了考试使用者应具备的素养和能力。该文件指出,合格的使用者在基于考试结果作出决策时,应当遵循良好的测试实践准则,并站在伦理道德高度,对所作决策负责<sup>[28]</sup>。

最后一类规程标准是由考试开发机构自行制订的文件。其中,较为典型的是美国教育考试服务中心颁布的《ETS质量和公平标准》(ETS Standards for Quality and Fairness,以下简称《ETS标准》)。该标准是ETS所有考试质量与服务规范化程度的审查依据。2014版《ETS标准》认为,公平的考试应对不同的考生群体均有效<sup>[29]</sup>。该文件包括若干具体标准,如确保考生特征不影响考试结果、考试内容无偏颇、考试能够为考生提供平等的学习机会、测试方法适切、考试能够为特殊考生群体提供便利等。

概而论之,多数规程标准均指出,考试开发人员应尽最大努力减少与考试构念无关的因素对不同考生群体所造成的不利影响。同时,这些文件也让我们意识到,除考试质量以外,测试实践的规范化程度也是涉考群体判断考试是否公平的重要考量因素。但是,面对不同的考试语境,目前已经颁布的规程标准是否适用,还有待进一步探究。

### (三) 哲学向度

上一小节提及的诸多规程标准对公平的测试实践提出了相应要求。但是,我们不禁会思考这样一些问题:制订上述规程标准的依据是什么?测试从业人员为什么需要遵从其中的准则?公平何以成为测试实践过程中重要的价值诉求?这些问题的解答需要诉诸公平背后的哲学思想。

测试公平性这一概念的萌芽与发展受到哲学思想的影响。前文提到“公平”的基本含义之一是“对人的权利和尊严予以同等尊重”<sup>[6]</sup><sup>[80]</sup>,这一含义与康德的哲学思想有着极深的渊源。康德主张尊



重人的尊严,并且认为判断某一行为道德与否的关键在于其动机,而非该行为产生的结果<sup>[30]21</sup>。若将康德的思想用于指导测试实践,这就意味着,测试实践过程中因考生处于弱势地位而不尊重考生权利、将考生视作牟利工具等行为都是不道德的。康德的思想在国际语言测试学会制订的《道德规范》<sup>[7]</sup>中得到了充分体现。

受康德影响,罗尔斯同样主张尊重人的基本权利和自由<sup>[31]13</sup>。与功利主义追求“大多数人的最大幸福”原则不同,罗尔斯认为弱势群体的权利同样值得捍卫。通过“无知之幕”(the veil of ignorance)的思想实验,罗尔斯阐释了正义的社会制度应遵循怎样的分配原则。具体而言,罗尔斯假定一群理性人需要签订一份社会契约,依此决定社会资源的分配。但是,站在“无知之幕”背后,他们对自己的背景特征(如种族、性别、天赋、健康状况、职业、财富状况等)一无所知。因此,罗尔斯认为这群人能够不偏不倚地拟订契约内容,以便落幕之时,自己能够置身于一个拥有良好秩序的正义社会,充分享受自由和平等。由此,罗尔斯提出了实现社会公平与正义的两个基本原则。第一个原则,即平等自由原则(the principle of equal liberty),主张“每个个体应该拥有平等的、不可剥夺的权利,享有最广泛的基本自由权,其所享有的基本自由权与其他个体所享有的自由权相同”<sup>[31]42</sup>。罗尔斯认为,默认情况下,社会、经济资源应被平等分配,不平等亦可被接受,但需要满足两个限制条件:第一个限制条件可以用公平的机会平等原则(the principle of fair equality of opportunity)来概括,该原则要求社会地位和职位必须在公平的机会平等的条件下向所有人开放;第二个限制条件可以归结为差别原则(the difference principle),其核心思想是使社会中处于最不利地位的成员获得最大的利益。这两个限制条件共同构成了第二个原则<sup>[31]42</sup>。可以看出,罗尔斯将公平视作实现正义的先决条件,类比“无知之幕”这一隐喻,只有当决策机制公平、客观、不存在偏颇时,才能实现决策正义<sup>[15]</sup>。

与罗尔斯的正义观不同,森的正义观带有明显的实用主义色彩。他认为“无知之幕”的思想实验有不切实际之嫌,与其探讨建构正义的社会机构所应遵从的原则,不如甄别并祛除社会中的非正义现象<sup>[32]262</sup>。森主张通过收集所有利益相关者主观看法的方式实现决策机制的公平性和客观性。此外,社会政策的制订和选择需要诉诸公共推理(public reasoning),确保该政策最终能够得到所有社会成员的拥护<sup>[32]352</sup>。哲学向度中有关公平和正义的探讨对测试公平性的界定和实证研究的开展具有重要启示意义。

## 四、启 示

### (一) 关于测试公平性的定义方式

毋庸置疑,测试公平性是一个多维概念,这主要是因为“公平”一词本身蕴含着多种基本含义和解读方式。因此,在测试公平性研究伊始,有必要明确测试公平性的具体含义、涵盖的具体测试实践以及具体的涉考人员。

“测试公平性”是一个跨学科概念。狭义的定义方式从教育测量视角着手,将其细化为教育测量学意义上可以量化的指标。广义的定义方式则以“公平”为出发点,从哲学、法学、社会学等多学科视角入手,勾勒测试公平性的内涵和研究范畴。狭义和广义相结合的定义方式能够加强测试学界与社会各界的沟通交流。但是,社会对测试公平性的关切于测试从业人员而言,意味着更高的职业道德要求。测试从业人员在保证考试服务质量之余,应深入思考考试的社会功能以及自身在测试实践中应该承担的责任。由于公平性在不同话语体系中的含义不同<sup>[33]</sup>,测试从业人员还应加强与不同涉考群体的交流<sup>[34]</sup>,在充分了解他们对测试公平性价值诉求的基础上,用易于理解的方式回应不同涉考群体对测试公平性的期待,从而增强考试的公信力。

## (二) 关于测试公平性的研究理念

为增强语言测试公平性研究的可操作性,笔者提出两点建议。第一,研究人员可以在具体语境中研究测试公平性。国际上现行的教育考试质量与公平性标准有一个共同特点,即“不具有明显的本土化特征”<sup>[35]45</sup>。鉴于测试公平性往往会受到某一语境下政治、文化、经济等因素的制约<sup>[36-37]</sup>,测试公平性研究需要明确考试所处的具体语境<sup>[38]</sup>。与孤立的、仅关注考试本身测量学特征的研究视角不同,该视角将考试视作某一时间、某一地点发生的具体事件,基于对考试历史沿革、考生特点、考试结果使用等方面的了解,对某考试的公平性展开研究<sup>[5]155</sup>。

第二,研究人员可以从“测试不公平”现象着手开展研究。虽然前文提及的规程标准对每个测试环节的公平性都提出了具体要求,但是,测试公平性的必要条件是什么?对这一问题的求索往往无疾而终。尽管“公平”难以定义,人们对现实生活中的不公平现象却极为敏感。诸多学者秉承实用主义精神,通过消除“非正义”的方式维护正义<sup>[39][40]57</sup>。借鉴“正义”这一概念的研究视角,保证语言测试的公平性同样可以通过剔除潜在的“测试不公平”(如测试偏颇)加以实现。我们可以将测试公平性视作一个连续统,完美的测试公平虽难以企及,但每一个不公平因素的消弭都能让测试实践趋向公平。

## (三) 公平、正义和效度之间的关系

目前,语言测试领域学者对“公平”“正义”与“效度”三个概念的内涵与外延仍然存有较大分歧,三者之间的关系未有定论,三者讨论的场域也有所不同。在哲学视域下,公平是实现正义的先决条件<sup>[15]</sup>。而在语言测试视域下,有学者认为公平与正义都是效度的一部分。如 McNamara 和 Ryan 认为,公平关涉考试的测量学特征,正义则关涉测试实践过程中蕴含的价值判断及测试实践带来的影响<sup>[12]</sup>。就测试公平性与正义的研究范畴而言,尽管测试公平性的定义方式仍无定论,测试偏颇在测试公平性的研究中始终占有一席之地;正义的研究范畴则相对较广,其研究议题主要涵盖考试机构的社会责任<sup>[6]80</sup>、考试政策<sup>[41]</sup>和考试的社会影响<sup>[12]</sup>等多个方面。

可以肯定的是,对于测试公平性、正义和效度的证明,需要依托充足、公开、透明的实证证据<sup>[42][43]110</sup>。同时,在证据收集过程中,需要倾听来自不同涉考群体的声音<sup>[32]392</sup>。

## 五、结 语

“公平”的基本含义非常丰富,“测试公平性”作为一个跨学科概念,具有多维性。需要注意的是,测试公平性同时涉及测试实践和涉考群体的价值判断,而学界对测试公平性的含义、研究范畴和研究方法等重要议题至今尚未有定论,对不同社会文化语境下涉考群体对测试公平性的看法也知之甚少。通过描绘测试公平性在教育测量、规程标准、哲学三个视域下的定义方式和理论主张,本文指出,可以将测试公平性视作一个连续统。绝对公平难以实现,测试公平性的证明也绝非易事。但是,伴随着语言测试领域专业化程度的不断提升,考试开发人员和考试使用者在测试实践的各个环节应将维护考试公平的责任铭记于心。此外,测试公平性研究需立足于具体的考试语境,必要时可从考试的不公平之处着手收集证据,回应不同涉考群体、教育体系乃至整个社会对考试公平的深刻关切。

## [参 考 文 献]

- [1] Kunnan A., "Test fairness and Toulmin's argument structure," *Language Testing*, Vol. 27, No. 2 (2010), pp. 183-189.
- [2] Zieky M. J., "Developing fair tests," in Lane S., Raymond M. R. & Haladyna T. M. (eds.), *Handbook of Test Development*, New York and London: Routledge, 2015, pp. 81-99.
- [3] Nisbet I. & Shaw S., "Fair high-stakes assessment in the long shadow of COVID-19," *Assessment in Education: Principles, Policy & Practice*, Vol. 29, No. 5 (2022), pp. 518-533.
- [4] 万光侠:《公平范畴的社会哲学审视》,《探索》2001年第1期,第66-69页。
- [5] Nisbet I. & Shaw S., *Is Assessment Fair?* London: Sage, 2020.
- [6] Kunnan A., *Evaluating Language Assessments*, New York and London: Routledge, 2018.
- [7] International Language Testing Association, "ILTA Code of Ethics," 2000-03-01, [https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/ILTA\\_2018\\_CodeOfEthics\\_Engli.pdf](https://cdn.ymaws.com/www.iltaonline.com/resource/resmgr/docs/ILTA_2018_CodeOfEthics_Engli.pdf), 2023-03-13.
- [8] Lawlor S., Richman S. & Richman C. L., "The validity of using the SAT as a criterion for black and white students' admission to college," *College Student Journal*, Vol. 31, No. 4 (1997), pp. 507-515.
- [9] Alon S. & Tienda M., "Diversity, opportunity, and the shifting meritocracy in higher education," *American Sociological Review*, Vol. 72, No. 4 (2007), pp. 487-511.
- [10] Dworkin R., "Affirmative action: is it fair?" *The Journal of Blacks in Higher Education*, No. 28 (2000), pp. 79-88.
- [11] Dorans N. J. & Cook L. L., "Introduction," in Dorans N. J. & Cook L. L. (eds.), *Fairness in Educational Assessment and Measurement*, New York: Routledge, 2016, pp. 1-6.
- [12] McNamara T. & Ryan K., "Fairness versus justice in language testing: the place of English literacy in the Australian citizenship test," *Language Assessment Quarterly*, Vol. 8, No. 2 (2011), pp. 161-178.
- [13] McNamara T. & Roever C., *Language Testing: The Social Dimension*, Oxford: Blackwell, 2006.
- [14] Cronbach L. J., *Essentials of Psychological Testing* (4th ed.), New York: Harper & Row, 1984.
- [15] Deygers B., "Fairness and social justice in English language assessment," in Gao X. (ed.), *Second Handbook of English Language Teaching*, Cham: Springer Nature, 2019, pp. 541-569.
- [16] Kunnan A., "Test fairness," in Milanovic M. & Weir C. (eds.), *European Language Testing in a Global Context*, Cambridge: Cambridge University Press, 2004, pp. 27-48.
- [17] Lynch B. K., "In search of the ethical test," *Language Testing*, Vol. 14, No. 3 (1997), pp. 315-327.
- [18] Zwick R., "Fairness in measurement and selection: Statistical, philosophical, and public perspectives," *Educational Measurement: Issues and Practice*, Vol. 38, No. 4 (2019), pp. 34-41.
- [19] Kunnan A., "Fairness and justice for all," in Kunnan A. (ed.), *Fairness and Validation in Language Assessment*, Cambridge: Cambridge University Press, 2000, pp. 1-13.
- [20] Kane M., "Validity and fairness," *Language Testing*, Vol. 27, No. 2 (2010), pp. 177-182.
- [21] Min S. & He L., "Test fairness: examining differential functioning of the reading comprehension section of the GSEEE in China," *Studies in Educational Evaluation*, Vol. 64 (2020), <https://doi.org/10.1016/j.stueduc.2019.100811>.
- [22] Xi X., "How do we go about investigating test fairness?" *Language Testing*, Vol. 27, No. 2 (2010), pp. 147-170.
- [23] 刘海峰:《考试立法应以史为鉴》,《湖北招生考试》2008年第6期,第1页。
- [24] Ofqual, "Regulatory Framework for National Assessments: National Curriculum and Early Years Foundation Stage Assessments," 2018-03-01, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/685610/National\\_assessment\\_regulatory\\_framework\\_-\\_March\\_2018.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685610/National_assessment_regulatory_framework_-_March_2018.pdf), 2023-03-13.
- [25] Qualifications Wales & CCEA Regulation, "Fair Access by Design: Guidance for Awarding Organisations on Designing High-quality and Inclusive Qualifications," <https://qw-website-prod-master.azurewebsites.net/media/avudal1dk/fair-access-by-design.pdf>, 2023-03-13.
- [26] 何莲珍、张娟:《语言测试的公平性:内涵、公平观及研究启示》,《外语教学与研究》2022年第1期,第79-89页。
- [27] Joint Committee on Testing Practices, "Code of Fair Testing Practices in Education," <https://www.apa.org/science/>

- programs/testing/fair-testing.pdf, 2023-03-13.
- [28] International Test Commission, “International Guidelines for Test Use,” *International Journal of Testing*, Vol. 1, No. 2 (2001), pp. 93-114.
- [29] Educational Testing Service, “ETS Standards for Quality and Fairness,” <https://www.ets.org/pdfs/about/standards-quality-fairness.pdf>, 2023-03-13.
- [30] [德]康德:《道德形而上学奠基》,杨云飞译,邓晓芒校,北京:人民出版社,2013年。
- [31] Rawls J., *Justice as Fairness: A Restatement* (2nd ed.), Cambridge: Belknap Press of Harvard University Press, 2001.
- [32] Sen A., *The Idea of Justice*, Cambridge: Belknap Press of Harvard University Press, 2009.
- [33] AERA, APA & NCME, “Standards for Educational and Psychological Testing,” <https://www.apa.org/science/programs/testing/standards>, 2023-03-13.
- [34] Taylor L., “Reframing the discourse and rhetoric of language testing and assessment for the public square,” *Language Testing*, Vol. 40, No. 1 (2023), pp. 47-53.
- [35] 王后雄、杨季冬:《国际教育考试质量与公平标准的特点与趋势——基于NVivo 11.0的编码分析》,《现代教育管理》2020年第3期,第45-51页。
- [36] Stobart G., “Fairness in multicultural assessment systems,” *Assessment in Education*, Vol. 12, No. 3 (2005), pp. 275-287.
- [37] Kunnan A., “Language assessment from a wider context,” in Hinkel E. (ed.), *Handbook of Research in Second Language Learning*, Mahwah: Lawrence Erlbaum, 2005, pp. 779-794.
- [38] McArthur J., *Assessment for Social Justice: Perspectives and Practices Within Higher Education*, London: Bloomsbury Academic, 2018.
- [39] Sen A., “Values and justice,” *Journal of Economic Methodology*, Vol. 19, No. 2 (2012), pp. 101-108.
- [40] Deygers B., *Assessing High-stakes Assumptions: A Longitudinal Mixed-methods Study of University Entrance Language Tests, and of the Policy That Relies on Them*, <https://lirias.kuleuven.be/retrieve/456108>.
- [41] Deygers B., van den Branden K. & van Gorp K., “University entrance language tests: a matter of justice,” *Language Testing*, Vol. 35, No. 4 (2018), pp. 449-476.
- [42] Sireci S. & Randall J., “Evolving notions of fairness in testing in the United States,” in Clauser B. E. & Bunch M. B. (eds.), *The History of Educational Measurement: Key Advancements in Theory, Policy, and Practice*, New York: Routledge, 2021, pp. 111-135.
- [43] Rawls J., *Political Liberalism*, New York: Columbia University Press, 2005.

## Theoretical Perspectives on Fairness: Implications for Fairness Research in Language Testing

He Lianzhen Zhang Juan

(School of International Studies, Zhejiang University, Hangzhou 310058, China)

**Abstract:** Test fairness has emerged as a prominent topic in the field of language testing since the 1990s. The 18th and 19th Language Testing Research Colloquiums, held in 1996 and 1997 respectively, generated a surge of research interest in this research topic. Until now, fairness has been regarded as a central value of assessment, alongside validity and reliability, in the field of language testing. Despite its importance, an accepted definition of test fairness remains elusive within the language testing community. Clarifying the definition of “fairness” is crucial for understanding the concept of “test fairness”—a foundational step for the subsequent theoretical debates and empirical research in language testing. Given

the well-developed theoretical underpinnings of “fairness” in various domains, it is imperative to examine different perspectives adopted to define and operationalize fairness. Meanwhile, valuable insights could be extracted to advance the research on test fairness in the realm of language testing. Following a delineation of various senses of fairness, this article reviews conceptualizations of test fairness through the lenses of educational measurement, professional standards and guidelines, and philosophy. This review article endeavors to provide insights into the theoretical and empirical inquiry of fairness in language testing.

Fairness is a complex and multifaceted concept that embraces several senses. The first sense of “fair” can be characterized as “relational”, denoting: (1) receiving what an individual merits or demerits; (2) meeting an individual’s legitimate expectation; and (3) treating like cases alike. The second sense is linked to the value of “respect”, denoting “the respect of each individual’s rights and esteem”. The third sense of “fair” can be labelled as “retrospective”, which refers to “rectifying past unfairness”. The fourth sense of “fair” is “formal”, indicating “adherence to procedural rules by players of a leveling field”. The abovementioned senses of fairness align well with the conceptualization of test fairness as proposed by scholars in the field of language testing.

Conceptualizations of fairness differ across different lenses. From the lens of educational measurement, two approaches have been used to conceptualize test fairness: a narrow approach and a broad approach. The former views test fairness as “absence of measurement inaccuracy and bias”. Whereas the latter perceives test fairness as essentially a social concern that extends beyond the psychometric properties of an assessment instrument. In a broad sense, the research scope of test fairness encompasses not only test-takers’ learning opportunities but also test use consequences. Seen through the lens of professional standards and guidelines, a fair test is characterized by the absence of construct irrelevance, as supported by legal documents, rules or guidance authorized by law, authoritative professional guidance, and requirements set by international and regional organizations. Additionally, philosophical thinking has long been concerned with the concept of fairness. Central to Kantian philosophical thought is the notion of “respect for human dignity”. Rawls contends that fairness in decision-making mechanisms is a prerequisite for achieving justice. Sen takes a pragmatic stance, advocating that the recognition and rectification of injustices should be prioritized instead of striving for perfect justice.

By delineating the definitions and theoretical propositions of fairness from the perspectives of educational measurement, professional standards and guidelines, and philosophy, this article suggests that research on test fairness should be contextualized, be supported by transparent empirical evidence, and incorporate the voices of different stakeholder groups. Test fairness is not a black-and-white concept. Instead, it exists along a continuum, and by addressing unfairness, fairer assessment instruments and practices can be expected.

**Key words:** fairness; test fairness; theoretical perspectives





- 首批国家社科基金资助期刊
- 2012—2019 年国家社科基金优秀期刊
- 2013 年、2015 年连续两届新闻出版广电总局全国“百强社科期刊”
- 新闻出版总署“中国期刊方阵双效期刊”
- 教育部高校哲学社会科学“名刊工程”入选期刊
- 第四届、第五届、第六届华东地区优秀期刊
- 浙江省期刊方阵“精优型期刊”“浙江省一级期刊”
- 第 26 届、第 30 届浙江省政府出版奖“树人奖”获奖期刊
- 中国人文社会科学各评价机构、各数据库历年核心期刊及来源期刊
- 被 CSA、UPD、EBSCO、Gale、OCLC、IC、ICIJ 等著名国际机构的 14 个重要索引收录

浙江大学学报 (人文社会科学版)

Zhejiang Daxue Xuebao (Renwen Shehui Kexueban)

2024 年 第 54 卷 第 5 期

1955 年创刊 (月刊)

2024 年 5 月 15 日出版

Journal of Zhejiang University

(Humanities and Social Sciences)

Vol. 54, No. 5, May 2024

Started in 1955 (Monthly)

Published on May 15, 2024

主管部门	中华人民共和国教育部	Administrated by	Ministry of Education of the People's Republic of China
主办单位	浙江大学	Sponsored by	Zhejiang University
主 编	黄先海	Editor-in-Chief	Huang Xianhai
编辑单位	浙江大学学报编辑部 (杭州市天目山路 148 号 邮编: 310028)	Edited by	Editorial Board of Journal of Zhejiang University (148 Tianmushan Road, Hangzhou 310028, China)
网 址	http://www.zjujournals.com/soc	Web Site	http://www.zjujournals.com/soc
电子信箱	zdxw_w@zju.edu.cn	Email	zdxw_w@zju.edu.cn
电 话	0571-88273210 0571-88925616	Tel	86-571-88273210 86-571-88925616
出版单位	浙江大学出版社有限责任公司	Published by	Zhejiang University Press Co., Ltd.
印刷单位	浙江省邮电印刷股份有限公司	Printed by	Zhejiang Post and Telecommunications Printing Co., Ltd.
国内发行	浙江省报刊发行局	Distributed Abroad by	China International Book Trading Corporation (P. O. Box 399, Beijing 100044, China)
订 阅 处	全国各地邮政局 (所)		
国外发行	中国国际图书贸易集团有限公司 (北京 399 信箱 邮编: 100044)		

中国·杭州 HANGZHOU, CHINA (国内外公开发行)

乌利希国际期刊分类代码 (Dewey #): 300

连续出版物号: ISSN 1008-942X 邮发代号: 国内 32-35 国内定价: 70.00 元 / 期  
CN 33-1237/C 国外 BM 372 840.00 元 / 年

ISSN 1008-942X

