

Standards for Educational and
Psychological Testing

燕妮琴 谢小庆 译

教育 与 心理测试标准

美国教育研究协会
美国心理学协会
全美教育测量学会
主编

沈阳出版社

美国教育研究协会
美国心理学协会
全美教育测量学会

主编

教育

心理测试标准

燕妮琴 谢小庆 译

沈阳出版社

图书在版编目(CIP)数据

教育与心理测试标准 / 美国教育研究协会, 美国心理学协会, 全美教育测量学会主编; 燕妮琴, 谢小庆译.

沈阳: 沈阳出版社, 2003. 3

ISBN 7-5441-2208-5

I. 教… II. ①美… ②美… ③全… ④燕… ⑤谢… III. ①考试—标准—美国 ②心理测验—标准—美国 IV. G424. 7

中国版本图书馆 CIP 数据核字(2003)第 047033 号

本书 1999 年版的版权 © 属美国教育研究协会、美国心理学协会、全美教育测量学会拥有。除美国 1976 年《版权法》允许条款外, 未经出版单位的事先书面许可, 任何人不得将本书任何部分以任何形式重新制版、发行、储存于文件库或检索系统中。

© 1999 by the American Educational Research Association

美国国会图书馆书号: 99-066845

国际书号: 0-935302-25-5

版权人授权沈阳出版社在全球出版、印刷、销售本书中文图书版

辽宁省版权局著作权合同登记 06-2003 年第 106 号

出 版 者: 沈阳出版社 (地址: 沈阳市沈河区南翰林路 10 号 邮编: 110011)

印 刷 者: 中共沈阳市委机关印刷厂

发 行 者: 沈阳出版社

开 本: 880mm × 1230mm 1/32

印 张: 11.375

字 数: 292 千字

印 数: 1—3100

出版时间: 2003 年 3 月第 1 版

印刷时间: 2003 年 3 月第 1 次印刷

责任编辑: 华玉洪 刘 杨

封面设计: 任晓岚

版式设计: 王珏菲

责任校对: 赵 琳

责任监印: 张建荣

定 价: 28.00 元

联系电话: 024-24112447

邮购热线: 024-24124936

E-mail: sysfax_cn@sina.com

译者前言

历经两年多的孕育,《教育和心理测试标准》第六版(1999年)(以下简称《标准》)的中文版终于在2003年元月初的纷纷瑞雪中诞生了!欣喜和慰藉之余,忍不住要对这个“新生儿”的出世回顾一番。

在2000年4月的全美教育研协研究会年会上,两译者初次相识,并意外发现在此之前我们已分别翻译了《标准》的“概述”一章。通过交谈,我们一致认为随着教育和心理测试在中国的应用日益广泛,测验结果的影响对人们的生活和工作日显重要,建立一套适合中国的测试标准已迫在眉睫。以译者之一在美从业十数年的经验而知,《标准》是几乎所有在美从事教育与心理测量工作的有关人员必备的重要工具书之一。同时,在新版《标准》和上一版《标准》之间相隔的15年里,教育和心理测量领域在测试理论、实践和技术上都有许多重大突破。由美国在教育 and 心理学研究方面的三大专业协会共同合作,历时六年编纂而成的新版《标准》不仅详细论述了这些进展和基于这些进展而制订的一系列新标准,还对旧版的内容结构作了彻底的调整。因此,将刚刚出台的《标准》翻译成中文不仅可为中国的测试实践解燃眉之急,同时也可中文世界的相关从业人员了解美国在这方面的最新进展助一臂之力,使之成为今后制作自己的测试标准时的参考。

经过一番讨论,我们决定合作翻译新版《标准》,并对申请中文版权、寻找出版社等事宜作了具体分工。

从此便开始了漫长、艰难的行程。虽然中国的考试制度已有近两千年的历史，但教育和心理测量作为一个交叉学科被引入中国则是近二十年的事。在翻译的过程中，我们本着忠实于原文的原则，追求“信、达、雅”的目标，特别是对名词术语的翻译做了以下尝试。

(1) 沿袭现行译法。有些术语在中文世界通用已久，且得体传神，如“施测(test administration)”、“机测(computerized administration, computer-based administration)”、“信度(reliability)”、“效度(validity)”等。对此类术语，我们采用了拿来主义。

(2) 修正现行译法。有些名词虽然已在中文世界里通用，但容易和其他名词混淆。例如，**criterion** 一词目前在中国通译为“标准”或“效度标准”，但它在本专业里很多情形下被用来表达参照或评判准则的意思。而 **standard** 一般也被译为“标准”。很明显，两者的定义是有差异的。为此，对 **criterion**，我们根据上下文有时译作“准则”，有时译作“测量准则”或“判别准则”，而 **standard** 则统一译作“标准”。

(3) 沿用其他学科的译法。有些术语在数学、统计等学科里已有既定译法，如“回归分析(regression analysis)”“因素分析(factor analysis)”、“收敛(convergence)”等。对此，我们照搬沿用。

(4) 一词多译。有些名词在英文里只有一种说法，如 **item**，**performance**。但在中文里，前者可译为考题、试题或题目；而后者根据不同上下文则可分别译为成绩、表现或能力。但要注意的是，前者虽然译法不同，但意思是相同的，可交换使用；而后者可能因上下文的内容而有不同的意思。对此，我们具体情况具体对待。

(5) 增加词义。有些词虽然在教育和心理测量领域里已有通用的译法，但它并没有完全表达出该名词或术语的所有词义。最典型的名词之一是“量表(scale)”。该词在测量专业里表示某个

检测工具时译作“量表”很合适(如 intelligence scale), 因这里的 scale 具体指一种测验。但是, 在将该量表的原始考试分数通过数学公式转换为导出分数, 然后将分数放在同一个测量标度上或数轴上, 使得分数之间可以互相比较而不会改变其意义的形式时, 译成“量表”就表达不出这层意思(如 test score scale, scale score)。同时, 这个标度或数轴也不是现成就有的, 需要根据具体测验的目的和用途, 通过数学方法得出。鉴于后者使用的情形居多, 我们便在原义的基础上扩充词义引入了“标度”作为新名词。标度的制作也就顺理成章地成为“制标(scaling)”。

(6) 造词。由于计量心理学既是社会科学里的一门交叉学科, 又是一门新兴学科, 中国文字虽然丰富多彩, 但有时仍然会对一些专业用语没有现成的词句应对。对此我们不得不造些新词。例如, 将 differential item functioning 译作“试题差异作用”, 将 test equating 译作“试卷等值”, 将 construct 译作“构念”和将 anchor test 译作“定位考试”等。

(7) 规范译法。有些常用词意思相近(如“测验”、“评价”、“评估”), 且常被人们在未明确其词义细微差别的情况下交换使用。对此, 我们有所规范, 例如, 将 test 统一译为“测验(或考试)”, 将 evaluation 译为“评价(或评鉴)”, 将 assessment 译为“评估”等(详见书末词汇表)。

本《标准》共有十五章。目录、前言、概述、第一至九章及第十五章由燕妮琴译。第十至十四章及索引由谢小庆组织初译。具体分工如下: 第十章和第十二章由柴省三译, 第十一章由武晓宇译, 第十三章、第十四章和词汇表由谢小庆译, 索引由马新芳译。为了将不同译者的用词和语气统一, 并尽量使译文语句通顺, 意思正确, 燕妮琴前后多遍对译文初稿逐字逐句做了实质性的修改并对全书(除索引外)最后定稿。尽管如此, 由于这是我们第一次尝试, 疏忽和错误在所难免, 有些名词术语的翻译可能会显得生硬或词不达意。为此, 恳请有心人士在阅读的过程中发现

错译或不妥的地方，或是有更贴切的译法以及其他任何意见和建议，通过以下途径向我们指正，不胜感谢。

在翻译此书的过程中，得到一些同事和朋友的关心和帮助。特别要提到的是，菲利浦·弗莱切 (Philip Fletcher) 对译者之一提出的问题总是有问必答。吴军华在译稿成形的最后阶段不仅帮忙编辑审阅部分章节，还对译文的句子结构和某些词句的译法提出了很好的修正意见。同时，没有全美教研协会出版部门及时授予我们中文版版权和中国沈阳出版社积极承接出版，这本书也不可能现在就和大家见面。在此一并致谢。

燕妮琴

美国，华盛顿特区

电邮: yanj@westat.com

谢小庆

中国，北京

电邮: hsk3@blcu.edu.cn

2003 年 1 月 20 日

主要译者简介:

燕妮琴: 佛罗里达大学(University of Florida)教育测量专业硕士及密西根州立大学(Michigan State University) 教育测量专业博士。1989 年至 1999 年先后服务于美国密西根州最高法院和密西根州教育厅，1999 年至今任 Westat, Inc.

谢小庆: 1998 年获北京师范大学心理学系博士。1999 至 2000 年在美国教育测试服务中心(ETS)进行博士后研究。现任北京语言大学教授，汉语水平考试(HSK)中心副主任。

前 言

主持《标准》修订的三家单位在指导测验的研制和使用方面迄今已出了五种文献。第一种,《心理测试和分析方法的技术建议》,由美国心理学协会的一个委员会主持并经该协会于1954年出版发行。第二种,《智力测试的技术建议》,由美国教研协会和全美教育测量学会的一个联合委员会主持,美国教育协会于1955年出版发行。第三种由美国心理学协会、美国教育协会和全美教育测量学会的一个联合委员会主持,由美国心理学协会于1966年出版。它取代了前两种文献,称为《教育和心理测验的标准和说明》。第四种,《教育和心理测验标准》,由上述三方再度合作修订并于1974年出版。第五种《教育和心理测试标准》发表于1985年,仍是由三方共同主持。

1991年美国心理学协会下属的心理测验和评价委员会提出修订1985年版《标准》的建议。美国教研协会、美国心理学协会和全美教育测量学会之三方代表遂开会讨论了有关《标准》修订之举、修订的指导原则和联合修订委员会的预定成员事宜。到1993年时,三家组织的会长各自指定了本协会出任联合修订委员会的成员名单。该委员会便在1993年11月举行了第一次会议。

联合修订委员会成员如下:

Eva Baker (联合主席)

Paul Sackett (联合主席)

Llody Bond

Leonard Feldt

David Goh

Bert Green

Edward Haertel

Jo-Ida Hansen

Sharon Johnson-Lewis

Suzanne Lane

Joseph Matarazzo

Manfrey Meier

Pamela Moss

Esteban Olmedo

Diana Pullin

Charles Spielberger 在 1993—1996 年间担任该委员会的联合主席。主持单位各方委派至多两名联络员参与委员会的事务，担任所在协会和联合委员会之间的传声筒。来自美国心理学协会心理测验与评估委员会的联络员因为该委员会成员改组而几度更换。

联合委员会联络员如下：

美国教研协会：William Mehrens

美国心理学协会：Bruce Bracken, Andrew Czopek, Rodney Lowman, Thomas Oakland

美国教育测量学会：Daniel Eignor

美国心理学协会和教育测量学会各另设委员会跟踪修订过程并及时知会有关方面。

美国心理学协会代表理事会特别委员会成员有：

Melba Vaseuez

Donald Bersoff

Stephen DeMers

James Farr

Bertram Karon

Nadine Lambert

Charles Spielberger

全美教育测量学会标准及测验用途委员会成员有:

Gregory Cizek

Allen Doolittle

Le Ann Gamache

Donald Ross Green

Ellen Julian

Nambury Raju

项目执行初期成立了一个管理委员会, 监管本项目的行政和财务安排, 并向各主持单位提出有关建议。

管理委员会成员有:

美国心理学协会: Frank Farley

美国教研协会: George Madaus

全美教育测量学会: Wendy Yen

《标准》修订项目的其他工作人员包括:

项目主管: Dianne Brown Maranto

工作人员联络员: Dianne L. Schneider

1993—1994 年项目主管: Wayne J. Camara

心理学协会的法律顾问对《标准》作了合法性方面的审查。William C. Howell 和 William Mehrens 对全书各章节的连贯性作了审查。Linda Murphy 为全书作索引。

联合委员会在技术性方面和公平性方面的一些章节的草稿请有关知名专家作初审。名单如下:

Marvin Alkin

Philip Bashook

Bruce Bloxom

Jeffrey P. Braden

Robert L. Brennan
John Callender
Ronald Cannella
Lee J. Cronbach
James Cummings
John Fremer
Kurt F. Geisinger
Robert M. Guion
Walter Haney
Patti L. Harrison
Gerald P. Koocher
Richard Jeanneret
Frank Landy
Ellen Lent
Robert Linn
Theresa C. Liu
Stanford von Mayrhauser
Milbrey W. McLaughlin
Samuel Messick
Craig N. Mills
Robert J. Mislevy
Kevin R. Murphy
Mary Anne Nester
Maria Pennock-Roman
Carole Perlman
Michael Rosenfeld
Jonathan Sandoval
Cynthia B. Schmeiser
Kara Schmitt

Neal Schmitt

Richard J. Shavelson

Lorrie A. Shepard

Mark E. Swerdlik

Janet Wall

Anthony R. Zara

《标准》修订的未定稿曾三次大面积分发，广泛征求意见。委员会共收到近 8000 页的评论意见。递交评论的各组织名单列举如下。许多个人对该组织的评论作了贡献。尽管我们希望向每位做出贡献者致谢，但我们无法做到，因为我们不知具体是谁向其所属组织提出的这些评论。没有这些专业人士的认真审阅，联合委员会不可能完成这个工作。

主持单位名单如下：

美国教研协会 (AERA)

美国心理学协会 (APA)

全美教育测量学会 (NCME)

相关科学、行业、贸易和协助团体名单如下：

美国高等教育协会 (AAHE)

美国医疗专科理事会 (ABMS)

美国心理咨询协会 (ACA)

美国评鉴协会 (AEA)

美国职业病疗法协会 (AOTA)

美国心理学学会 (APS)

美国心理学协会心理咨询分会(十七分会)

美国心理学协会发育心理分会(七分会)

美国心理学协会评估、测量及统计分会(五分会)

美国心理学协会智障及发育残障分会(三十三分会)

美国心理学协会药理学及滥用毒品分会(二十八分会)

美国心理学协会复健心理分会(二十二分会)

美国心理学协会学校心理分会(十六分会)
全美亚裔心理学协会(AAPA)
咨询评估协会(AAC)
测验出版商协会(ATP)
澳大利亚教育研究学会(ACER)
芝加哥工业心理 / 组织心理学会(CIOP)
执照管理及法规学会(CLEAR) 测验方法及顾问委员会(ERAC)
平等就业顾问委员会(EEAC)
复健证书、教育及研究基金会
南非人体科学研究学会
国际交叉文化心理学协会
国际电工兄弟会
国际语言测试协会
国际人事管理协会评估分会(IPMAAC)
测验实务联合委员会(JCTP)
全美有色人种进步协会法律辩护和教育基金组织
全美公平、公开测试中心(Fairtest)
全美确保竞争机会组织(NOCA)
华盛顿地区人事测试理事会(PTC/MW)
南加州人事测试理事会(PTC/SC)
人事管理协会(SHRM)
印度心理学家协会(SIP)
美国心理学协会工业心理及组织心理分会(十四分会)
美国心理学协会少数族裔事务心理研究分会(四十五分会)
操作测验的评估和学生标准技术指导州际联盟(TGPA)
电信业员工招聘论坛
西部地区政府部门间公务员评估委员会(WRIPAC)
资格审定委员会名单如下:
美国体能及医疗复健委员会

美国医疗技师执照测验委员会

复健咨询师执照测验委员会

全美注册心理咨询师委员会 (NBCC)

全美验光师执照测验委员会

全美医师执照测验委员会

全美州级护士执照委员会协会

各级政府部门名单如下:

陆军研究所 (ARI)

加州高速公路巡警、人员配备及培训局研究处

达拉斯市公务员管理局

弗吉尼亚州教育厅

国防人力数据中心 (DMDC) 人事测验处

国防部助理部长办公室 (DOD)

教育部提高教育质量办公室全国教育统计中心

司法部移民归化局 (INS)

劳工部就业培训管理局 (DOL/ETA)

美国平等就业机会委员会 (EEOC)

美国人事管理办公室 (OPM) 人力资源及开发中心

测验研制 / 出版单位名单如下:

全美大学测试公司 (ACT)

CTB/ 马格诺 - 希尔公司

大学联考委员会

教育测试服务公司 (ETS)

高地出版社

个性及能力测试研究所 (IPAT)

职业测验服务公司 (PES)

学术机构名单如下:

创意领导中心

加罗得大学全国患耳障专业人员的公平测试专案办公室

海法大学以色列小组

肯色斯州立大学

全国教育成果中心 (NCEO)

宾州州立大学

北卡大学夏洛特分校

南密西西比大学心理系

联合委员会完成《标准》修订工作后将文献分别送交给三家主持单位批准。各单位有其自己的审批机制和程序，以及他们“批准通过”的定义。具体诠释如下：

美国教研协会：批准本《标准》的修订完成表明，原则上我们相信《标准》代表了本领域公认的专家们就测量实践中应遵循的准则达到目前的共识。测验的研制、施行、出版发行及使用各方人士均需照此办理。

美国心理学协会：《标准》的通过表明本协会将采用该文献作为准则。

全美教育测量学会：本学会批准《教育与心理测试标准》修订完成，同时也认为这些标准将用于促进测试工作日臻完善。《标准》修订的完成和通过也促使本学会会员在工作中按《标准》行事。

虽然《标准》属规定性质，但它本身并非带有强制性。各项标准的制定力求和三家主持单位下的其它现行标准、指南和实施规则保持一致。希望读者参阅这些文献。其中有些文献是专门针对某些特殊情形下的测试与评估行为而制定的。

参考文献

美国教研协会。(1992年6月)。《美国教育研究协会的道德标准》。华盛顿特区：本协会。

美国教师联合会、全美教育测量学会及全美教育协会。(1990年)。《学生教育评估过程中教师应具备的能力标准》。华

盛顿特区:全美教育测量学会。

美国心理学协会。(1992年12月)。《心理学家的道德操守和实施规则》。《美国心理学家》期刊,第47卷,第12期,1597-1611页。

测试实践联合委员会。(1988年)。《教育系统中公平测试实践实施规则》。华盛顿特区:美国心理学协会。

全美教育测量学会。(1995年)。《教育测量中职业责任的实施规则》。华盛顿特区:本协会。

目 录

译者前言 前 言

概 论	1
测试过程中的参与者	1
制定《标准》的目的	3
《标准》的分类	3
本《标准》适用的测验及测验用途	5
使用《标准》的注意事项	6
《标准》的数目	7
测验作为测量构念的手段	8
本版《标准》的组织结构	8

第一部分 测验的制作、评估和文件存档

第一章 效度	12
背景介绍	12
标准 1.1-1.24	25
第二章 信度和测量误差	38