



中国浦东干部学院博士文库

资格考试的 理论与实践

赵世明·著

The Theory and Practice of Credentialing Examination



上海社会科学院出版社



中国浦东干部学院博士文库

资格考试的 理论与实践

赵世明·著

The Theory and Practice of Credentialing Examination



上海社会科学院出版社

图书在版编目(CIP)数据

资格考试的理论与实践/赵世明著. —上海:上海社会科学院出版社, 2007. 11

(中国浦东干部学院博士文库)

ISBN 978-7-80745-103-7

I. 资… II. 赵… III. 资格考核-研究-中国 IV. C975

中国版本图书馆 CIP 数据核字(2007)第 150803 号

资格考试的理论与实践

著 者: 赵世明

责任编辑: 徐祝洁

封面设计: 王斯佳

出版发行: 上海社会科学院出版社

上海淮海中路 622 弄 7 号 电话 63875741 邮编 200020

<http://www.sassp.com> E-mail: sassp@sass.org.cn

经 销: 新华书店

印 刷: 上海新文印刷厂

开 本: 890×1240 毫米 1/32 开

印 张: 7.5

字 数: 205 千字

版 次: 2007 年 11 月第 1 版 2007 年 11 月第 1 次印刷

ISBN 978-7-80745-103-7/C · 003

定价: 20.00 元

版权所有 翻印必究

总 序

P R E F A C E

创办中国浦东、井冈山、延安干部学院，是党中央从推进中国特色社会主义伟大事业和党的建设新的伟大工程全局出发作出的一项重大决策。

中国浦东干部学院自 2005 年 3 月正式开办以来，始终坚持胡锦涛同志提出的切实把学院建设成为进行革命传统教育和基本国情教育的基地、提高领导干部素质和本领的熔炉以及开展国际培训交流合作的窗口的办学目标，秉承“实事求是、与时俱进、艰苦奋斗、执政为民”的办学要求，努力体现“国际性、时代性、开放性”的办学特色，取得了较为突出的业绩，正在向“国内一流、国际知名”的新型领导学院的方向稳步迈进。

办好一所学院的关键是教师。人才是事业之本，人才兴则事业兴、事业旺。“所谓大学者，非大楼之谓也，有大师之谓也”，这句话对中国浦东干部学院来讲同样有指导意义。中国浦东干部学院汇集了一批优秀的教师，他们当中，既有国外学成归来的学子，也有来自国内著名

高校、科研机构的青年才俊。他们有火热的创业激情，有对干部教育培训事业的执著和热爱。他们大多拥有博士学位，在自己所属的学科领域已崭露头角。这支队伍是建设好中国浦东干部学院的人才支持和智力保证。为他们搭建平台，促进他们成长，引领他们发展，是学院义不容辞的职责。

支撑一所学院的基础是学术。学院之称，有学科、学养、学理之意蕴。没有了学术，学院也就失却了原动力和根基。中国浦东干部学院创办以来，坚决贯彻并创造性地执行中央的战略决策和一系列办学要求，明确了教育培训、科学研究、咨询服务、领导测评、网络教育五位一体的功能定位，突出强调了学术研究、学科建设在学院发展中的重要地位。整合学术资源、加强学科建设对学院发展至关重要。

体现一所学院办学水平的重要标志是品牌。品牌汇集了办学的智慧，凝聚了办学的精华，同时也提升了学院的美誉度。中国浦东干部学院以问题为核心，以能力为导向，以现场教学资源为依托，自创办以来培训了大批学员，培训成果显著，一批具有较高质量的课程品牌、教学品牌正在形成。从一所学院的发展来看，既要有教学的品牌，也要有科研的品牌；既要有品牌学员，也要有品牌教员；而这一切都需要长期积累。求木之长必先固其本。积累品牌素材，探寻品牌来源，滋养品牌发展，是学院发展的长远大计。

基于上述认识,我们组织出版了这套中国浦东干部学院博士文库。入选文库的书稿均为学院青年教研人员的博士学位论文,并经过了严格的“双盲”评审。作者根据评审意见和所论问题的发展以及研究的深化,都进行了认真修改,可以说基本反映了所论问题的学科前沿。我们希望,这套分辑出版的文库能开启和激励我们的后续研究,促进学院自身研究特色和学术传统的形成,促进相关学科领域的建设,促进学术交流与繁荣。

文库的出版得到了上海社会科学院出版社领导和编辑同志的鼎力支持和帮助,借此表示诚挚谢意。同时,对为文库的建设作出贡献的评审专家和付出辛劳的同志表示诚挚的感谢。

文库中存在的不足,敬恳广大读者批评指正。

中国浦东干部学院博士文库编委会

2007年元月

ABSTRACT

The credentialing examination system has been established in China since the beginning of 1990s. To the initial period of 21st century, the credentialing examination enterprise experienced a swift development; the population of examinees of whole country exceeds nearly ten million examinees. The majority of professions have implemented the professional credentialing tests or examinations which providing the powerful personnel support to our country's economic and social development.

The credentialing examination are mainly used to appraise the examinee's professional knowledge and skills of necessary to be competent the specific profession, in order to assure that they can safely and effectively carry out the occupational activity. From the views of its psychometric nature and characters, the credentialing examination belongs to the typical criterion-referenced test. The development and the implementation of the criterion-referenced test have many differences from the traditional norm-reference examination. Some theories and methods of norm-reference model are not suitable to the credentialing examination. But in our country, the research fundament of the criterion-referenced test is

still relative weak, applications of the criterion-referenced testing techniques are rarely seen reported, some theoretical and the technical problems of professional credentialing examination are urgently awaited to be solved. For instance, many professional credentialing examination did not specify the test domain content according to the practice analysis; standard-setting procedure are not based on the test content; the item analysis and the reliability estimation still continues to use the statistical analysis pattern of norm-reference test; and the majority of credentialing examination has not carried on the equating program. These problems in a certain extent affected the fairness of the credentialing examination and limited the further enhancement of the credentialing examination's scientific level.

This research took the National Licensing Examination of Nursing (LEN) and Medical Licensing Examination as the sample, explored to establish the criterion-referenced model which including test development and the statistical analysis procedure for the needs of professional credentialing examination practices in China, and at the same time, made some positive attempt in some aspects of theoretical research and technical application of criterion-referenced test.

The entire book introduced systematically the basic concepts, theories and techniques of criterion-referenced test regarding to the professional credentialing examination; conducted some experimental studies on key technical question faced by current credentialing examinations, and proposed some innovational techniques and methods to practice and implement the

credentialing examination; initially established a theoretical and technical system based on the criterion-referenced model for the credentialing examinations. Many research results have the very good promotion and application values for the ongoing professional credentialing examinations in China.

Key words: credentialing examination, criterion-referenced test

前言

P R E F A C E

我国从20世纪90年代初期开始推行行业资格考试制度。到了21世纪初期,资格考试事业得到迅猛发展,考试人口逾千万人,大部分行业已建立起专业人才资格考试制度,对于推动我国经济社会的不断发展提供了强有力的人才支撑。经过多年的研究与实践,我们已经初步建立起资格考试的标准参照模式和专业技术标准,其中许多研究成果对于我国目前开展的各项行业资格认证考试具有很好的推广和应用价值。

资格考试(credentialing examination)主要用来评价应试者所具有的专业知识与技能是否达到胜任特定行业的基本专业要求,以确证他们能够安全、有效地开展职业活动。在国内外的许多行业中,专业人才必须通过政府机构、专业协会或专业组织实施的资格考试,才能获得该行业的从业资格。因此,资格考试的科学性和权威性问题的始终是行业领域和应试者关注的焦点。

从心理测量学的性质和特点来看,资格考试属于典型的标准参照测验。标准参照测验的编制和实施与传统的常模参照测验(选拔性考试,如高考)存在很多区别。常模参照测验的一些理论与方法在资格考试中并不适用。但是目前在我国,标准参照测验的理论研究相对滞后,标准参照测验的技术应用尚不多见,行业资格考试的理论与技术问题亟待解决。比如,许多行业资格考试

没有基于行业的实际需要确定考试的内容,在确定考试的合格标准或及格分数时没有参照具体的考试内容,题目分析和信度估计仍沿用常模参照的统计分析模式,大部分资格考试没有进行等值。这些问题的存在,在一定程度上影响了资格考试的公平性,限制了资格考试科学化水平的进一步提高。

本研究以全国护士执业考试(LEN)和医师资格考试(MLE)为样本,重点研究国家级行业资格考试的标准参照模式,探索建立适合我国行业资格考试特点的测验编制和统计分析体系,同时在标准参照测验的理论研究与技术应用方面做出积极尝试。对行业资格考试进行系统的心理测量学研究,这在我国心理测量和资格考试发展史上还是第一次,对于建立和完善我国行业资格考试制度具有重要意义,对于标准参照测验技术的应用和发展也具有很好的实践和理论价值。

全书比较系统地介绍了与行业资格考试相关联的标准参照测验的基本概念、理论和技术,对我国资格考试目前面临的主要的、关键的技术问题进行了实验研究,提出了一些创新技术和方法,初步建立起一套标准参照模式的资格考试理论与技术体系。

本研究提倡基于工作分析或实践分析确定资格考试的内容领域,有助于提高资格考试的有效性,改变目前依据学校教学大纲确定资格考试内容的不恰当做法;提出以多重判区间判断法确定资格考试的合格标准,该方法集中了多种合格标准确定方法的合理成分,保证并兼顾了资格考试合格标准的可靠性、有效性和可操作性,比较符合我国行业资格考试的特点和需求;采用分半法估计资格考试的分类一致性信度,可以提高大规模资格考试区分信度估计的可操作程度,由此得到的分类一致性信度指标具有常规指标的基本特性;在系统介绍标准参照测验信度指标体系的基础上,利用基于方

差分析的概化理论,对资格考试领域分数估计与测验分数信度进行了试验研究,制定了资格考试信度估计的基本方法和程序;讨论了基于标准参照模式的题目分析模式,强调了基于题目内容的判断分析和统计方法对资格考试的适用性;深入研究了资格考试中的题目功能差异问题,建议将题目功能差异的检测纳入资格考试题目分析和效度检验的常规程序,进一步提高考试的公平性;利用协方差结构方程模型对资格考试的内容结构进行验证性因素分析,认为该方法是估计标准参照测验和资格考试结构效度的适合方法,比较好地解决了资格考试的结构效度估计问题。最后,本研究还探讨了资格考试的等值设计模型,建议采用标准卷法处理资格考试的题目预测、考试等值和报告分数问题,研究并提出了大规模资格考试中作弊雷同的判定标准。

目 录

CONTENTS

前 言	001
第一章 资格考试的发展现状	001
一、国内行业资格考试的发展现状	001
二、国外行业资格考试的发展现状	007
三、研究的意义与问题	016
第二章 资格考试的理论基础:标准参照测验	019
一、标准参照测验的产生与发展	020
二、标准参照测验概念的界定	025
三、标准参照测验的分类	027
四、标准参照与常模参照的区别	028
五、标准参照测验的编制程序	030
六、标准参照测验存在的理论问题	032
第三章 考试内容的界定	035
一、实践分析的方法	038
二、考试大纲的编制	046
三、内容权重的确定	049
四、试卷长度的确定	053

第四章 合格标准的确定	058
一、确定合格标准的基本方法	058
二、多重区间法的基本内容	066
三、多重区间法的实验研究	071
四、合格标准的等值问题	080
第五章 题目质量分析	082
一、题目分析的基本方法	082
二、题目分析方法的适用性	088
三、题目的技术特征	090
第六章 题目功能差异	104
一、题目功能差异的基本概念	105
二、题目功能差异常用检测方法	107
三、题目功能差异的实验研究	112
第七章 分类一致性信度估计	122
一、标准参照测验的信度体系	122
二、分类一致性信度的估计方法	125
三、分类一致性信度估计的实验研究	132
第八章 考试分数信度估计	139
一、领域分数估计信度	139
二、测验分数信度估计	142
三、考试分数信度估计的实证研究	145

第九章 考试的效度研究	152
一、效度估计的基本方法	153
二、资格考试结构效度研究方法	156
三、资格考试结构效度的实证研究	160
第十章 考试等值设计	168
一、测验等值的基本方法	171
二、资格考试的等值模型设计	176
三、题目参数等的试验研究	180
第十一章 作答雷同的判定标准研究	183
一、错同率的检验方法	184
二、四选项选择题的错同率研究	185
三、五选项选择题的错同率研究	187
四、错同率的常模研究	191
附录 1 资格考试的统计分析模式	197
附录 2 资格认证测验的心理测量学标准(节选)	200
附录 3 协方差结构方程分析 LISREL 程序	204
附录 4 方差分量计算方法	207
参考文献	209

CONTENTS

Preface	001
Chapter 1 The development condition of credentialing examinations	001
1.1 The development condition of Chinese credentialing examinations	001
1.2 The development condition of credentialing examinations abroad	007
1.3 The significance and objectives of this research	016
Chapter 2 The theoretical base of credentialing examination: Criterion-referenced test	019
2.1 The origin and development of criterion-referenced testing	020
2.2 The definitions of criterion-referenced test	025
2.3 The classification of criterion-referenced test	027
2.4 The differences between criterion-referenced and norm-referenced tests	028
2.5 The development procedure of a criterion-referenced test	030



2.6 The theoretical problems of criterion-referenced testing	032
--	-----

Chapter 3 Specifying the domain of content of credentialing examination	035
--	------------

3.1 The practice analysis	038
3.2 Developing the outlines of examination	046
3.3 Determining content weight	049
3.4 Determining test length	053

Chapter 4 Standards setting for qualification	058
--	------------

4.1 Basic methods of standards setting	058
4.2 The Multi-interval Judgment Method	066
4.3 The performance of the Multi-interval Judgment Method	071
4.4 Equating problem of the standards	080

Chapter 5 Item analysis	082
--------------------------------------	------------

5.1 Basic methods of Item analysis	082
5.2 Adaptability of the Item analysis methods	088
5.3 The item technical characters	090

Chapter 6 Differential item functions	104
--	------------

6.1 Concepts of differential item functions	105
6.2 Detecting methods of Differential item functions	107
6.3 Experimental researches on Differential item functions	112