

教育考试国家题库 理论与实践

李光明 等著

高等教育出版社

JIAOYU KAOSHITI GUOJIA TIKU LIJUN YU SHIJIAN

教育考试国家题库 理论与实践

李光明 等著

高等教育出版社·北京

图书在版编目(CIP)数据

教育考试国家题库理论与实践/李光明等著. --北京:高等教育出版社,2014. 11
ISBN 978-7-04-041063-1

I. ①教… II. ①李… III. ①考试-题库-研究-中国 IV. ①G424.79

中国版本图书馆 CIP 数据核字(2014)第 199040 号

策划编辑	朱丽娜	责任编辑	张耀明	封面设计	张楠	版式设计	马敬茹
插图绘制	邓超于博	责任校对	刁丽丽	责任印制	刘思涵		

出版发行 高等教育出版社
社 址 北京市西城区德外大街 4 号
邮政编码 100120
印 刷 北京人卫印刷厂
开 本 787mm×960mm 1/16
印 张 23.75
字 数 400 千字
购书热线 010-58581118

咨询电话 400-810-0598
网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landaco.com>
<http://www.landaco.com.cn>
版 次 2014 年 11 月第 1 版
印 次 2014 年 11 月第 1 次印刷
定 价 49.00 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

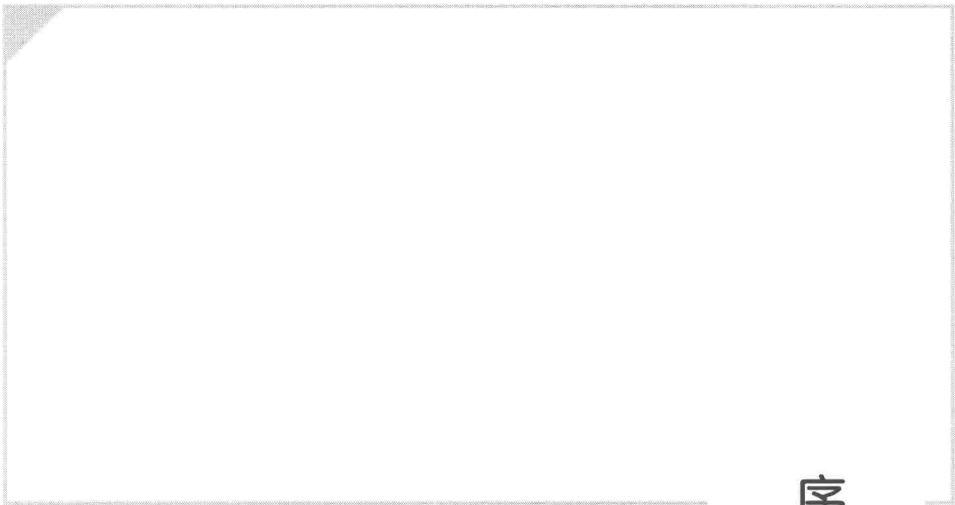
物 料 号 41063-00

审 图 号 GS(2014)1406 号

本书为全国教育科学“十一五”规划 2009 年度教育部重点课题

“教育考试国家题库的研究与应用”

[课题号:GFA097013]的研究成果



序

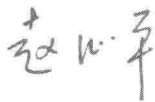
题库是专业化考试机构的重要基础设施。数十年前,不少发达国家的专业考试机构已建立了适应大规模考试的题库系统,为考试技术和服务水平不断提高发挥了重要作用。而我国的考试机构长期以来一直没有将题库建设作为自身的建设内容,影响了考试技术的现代化进程。

教育部考试中心于2006年初决定开展题库建设工作,我对此非常赞同,也争取到国家财政部门的理解和支持。建设教育考试国家题库,是教育考试领域的重要技术进步,是考试机构专业化的必然选择,同时,也是教育部考试中心在考试改革中的重要举措。由于题库建设的重要性和迫切性,我国2010年颁布的教育规划纲要明确提出了“完善国家考试科目试题库”的要求。

题库的建立使考试技术发生了根本性的变革。教育部考试中心经过8年的建设,题库已初具规模,不仅积累了大量试题,提高了试题质量,而且大幅度提升了我国教育考试的安全性、科学性和突发事件处置应变能力。题库的投入与使用,帮助政府顺利地应对了2008年汶川地震、2010年玉树地震等自然灾害给当年教育考试带来的巨大困难,维护了学生的利益和社会的稳定;在题库的支撑下,逐步实现了高考的“多题多卷”、部分考试项目的一年多次考试,以及计算机辅助考试等形式的转变,提升了为考生服务的能力,提高了考试的科学性,进一步保证了考试的公平公正性。

科研先行,技术推动。教育部考试中心在建设题库的过程中,将科研意识和科研方法引入题库工作机制,形成研究性工作氛围,以科研带动工作,以工作推进科研。因此,教育考试国家题库建设既是一项以应用为目标的系统研究,也是将研究成果体现在题库建设实践当中的复杂工程。工作与科研相结合的原则使题库研究成果丰富,从《教育考试国家题库理论与实践》一书即可管中窥豹。该书系统地总结了题库建设的经验,全面反映了题库建设中的所思、所得,不仅吸收了西方考试机构的先进技术和方法,还提出了一系列具有我国特色的题库建设理论,较好地解决了题库建设中的许多难题。可以肯定地说,没有题库研究人员的不懈追求与努力,就没有教育考试国家题库建设今天所取得的成绩。

题库建设不能一蹴而就,也不能一劳永逸,它是一项长期的、复杂的系统工程。希望教育部考试中心和题库研究人员对前期成果不断巩固和深化,继续做好题库建设的理论研究与应用实践,为党的十八大提出的深化国家教育考试的内容与形式改革做出更大的贡献。



2014年3月

前 言

为了确保考试安全、维护社会稳定,使教育考试主动服务社会,根据教育部领导的指示,我中心承担的教育考试国家题库建设项目于2006年初正式启动,迄今已走过了八年历程。根据部领导要求,围绕题库建设,教育部考试中心“举全中心之力”建题库,我们“精心论证、精心设计、精心实施”,为“建成高水平实用化题库”而努力。

题库建设是一项庞大的系统性工程,为了更好地指导国家题库建设、服务国家题库建设和推动国家题库建设,我们从科学实施题库建设的战略高度出发,专门成立“教育考试国家题库的研究与应用”课题组,以课题研究形式全面解决题库软件开发、试题命制、试卷设计、标定与等值、试题和试卷评价等题库建设环节的理论问题和实践问题,并探讨了题库工作的长效机制。在课题研究中,我们采取“没有经验,探索经验;没有办法,创新办法”的题库建设推进策略,不断研究和总结不同考试项目不同学科的题库建设办法及经验,形成了一套具有我国特色的大规模考试题库建设的操作“范式”,确保了题库建设朝着预期的目标迈进,促进题库建设的科学化、规范化、合理化和考试安全。

今天我们编辑出版《教育考试国家题库理论与实践》,既是对过去工作的梳理和总结,更是为了开拓更加广阔的未来,因为我们知道,题库既是迎接考试领域高新技术挑战的需要,也是今后考试内容与形式改革的重要利器。本书反映了我们在基于题库的试题命制、试卷设计、标定、等值与难度控制、统计分析与评价、题库软件的设计与实现、题库的维护与管理、题库在计算机辅助考试中的运用,以及题库与考试制度改革等多个方面较为系统的研究与思考。这些题库建设中的重点问题和难点问题的解决为科学合理地建设教育考试国家题库提供了理论依据和技术设计方案,一定程度上解决了题库长久、安全、有效运行的机制问题和技术问题。

本书是教育部考试中心很多工作人员和科研人员智慧的结晶。第一章由刘芃、关丹丹编写,刘芃负责统稿;第二章由张亚南、程蒙蒙、单旭峰、程力、陈昂编写,张亚南负责统稿;第三章由柳博、程力、程蒙蒙、邢博特、姚刚编写,柳博负责统稿;第四章由高升、焦丽亚、东晓华编写,高升负责统稿;第五章由任子朝、关丹丹、陈昂、王蕾、章建石编写,任子朝负责统稿;第六章由李光明、燕紫、刘晓华编写,李光明负责统稿;第七章由李勇、胡传勇、高升、钱林编写,李勇负责统稿;第八章由刘庆思、关丹丹、李骐编写,刘庆思负责统稿;第九章由李光明、高升、柳博、王蕾编写,李光明负责统稿。全书由李光明统稿;高升、关丹丹校对。本书的出版得到了教育部考试中心和全国教育科学规划领导小组办公室各级领导的全力支持,梅迪亚和高等教育出版社对本书出版给予了大力协助。在此,一并表示感谢。

《教育考试国家题库理论与实践》是我们对题库的思考与探索,希望我们的努力能够对考试同行们有所启发。我们也希望借此与更多的考试工作者们开展更广泛的交流,听到同行们对我们的指正和教诲。

编 者

2014年3月

目 录

第一章 教育考试国家题库建设	1
第一节 题库的回溯与思考	2
第二节 教育考试国家题库设计	19
第二章 基于题库的试题命制	28
第一节 命题理念与原则	29
第二节 学科评价目标与试题属性	34
第三节 命题流程与质量控制标准	43
第四节 题库试题设计的创新与实践	50
第三章 基于题库的试卷设计	68
第一节 试卷设计的理论基础	69
第二节 基于题库的试卷设计及应用	81
第三节 多种形式试卷的设计及应用	97
第四章 题库建设中的标定、等值与难度控制	108
第一节 标定	109
第二节 等值	128
第三节 试题的预估难度与实测难度	143
第四节 试题的难度控制	148
第五章 统计分析与评价	166
第一节 验证性因素分析在考试评价中的应用	168
第二节 多元概化理论在考试评价中的应用	175
第三节 潜在类别分析在考试评价中的应用	187
第四节 标准参照考试理论在考试评价中的应用	200
第六章 题库软件的设计与实现	210
第一节 题库系统需求分析	211
第二节 技术要点简介及系统技术选择	215
第三节 题库系统设计	228
第四节 题库系统实现	236
第七章 题库的维护与管理	262
第一节 题库的监测与评估	263
第二节 题库软件的维护与升级	271
第三节 题库试题沉淀问题及应对策略	279
第四节 题库安全与保密	290
第八章 题库与计算机辅助考试	300
第一节 计算机辅助考试	301
第二节 计算机自适应考试	321
第三节 PETS 计算机自适应序列考试的设计与实现	337
第九章 题库与考试制度改革	353
第一节 题库是考试机构专业化建设的需要	354
第二节 题库是深化考试内容和形式改革的基础	356
第三节 题库是发展考试评价功能的必然	362

第一章

教育考试国家题库建设

20 世纪中叶,迅猛发展的世界经济迫使人们寻找编制测验的新技术,以保证客观、准确、及时地选拔人才,题库正是顺应这一时代潮流需要的产物。近十几年来,我国也在题库建设方面进行了许多尝试,并取得了一些经验。2006 年,教育部考试中心开始建设教育考试国家题库,迄今为止,国家题库已初具规模,在很多考试项目上也初见成效。例如,成人高考已实现了日常化命题方式,全国硕士研究生招生考试中部分学科实现了不入围命题。入围命题封闭时间长、任务紧、压力大、管理严,是一种非常不人性化的工作方式,而题库带来的工作方式的变革深受学科秘书和命题教师的欢迎。高考项目由于其高利害、社会的高关注,短时间尚无法取消入围命题,但题库建设正在减少入围的人数、缩短入围的时间。更为重要的是,题库大大提高了统一考试的应急能力,如在 2008 年四川汶川特大地震和 2010 年青海玉树地震发生后,题库在试卷供应方面发挥了关键作用,最大限度地保障了考试的顺利举行,从而维护了灾后社会的稳定。

第一节

题库的回溯与思考

20 世纪 40 年代中期以后,由于战后经济的发展,职业需要对专业人员和直接劳动者的知识技能提出了更高的要求,考试获得了前所未有的发展。考试次数的增多,命题工作量的增大,对考试的要求日益严格,迫使人们寻找编制测验的新技术,以保证客观、准确、及时地选拔人才,考试的命题方式亟待革新,题库正是顺应这一时代潮流需要的产物。

一、题库的起源

(一) 什么是题库?

题库(Item Bank)一词,源于英国 20 世纪 60 年代的一个全国教育研究课题(Wood et al., 1969),本意是指测验试题的有序集合。更为准确的说法是,题库是适合于一定考核目标的、具有必要参数的大量优质试题的有机集合。计算机题库系统(Computerized Item Banking)的思想早在三十多年前就有了(Choppin, 1976),只是功能非常简单。由于人们对计算机

化考试一直持怀疑态度(Hiscox,1985),计算机题库系统主要是用于管理试题、组卷后打印试卷,用于纸笔考试。如今,题库系统是收集、管理和输出试题集合的软件系统,既可用于存储试题,又可以作为命题的工作平台和计算机辅助考试的后台系统。按建设题库的主管单位划分,有国家级、省级、校级等题库;按构成题库的学科(或通用性)可分为单一学科题库和通用性题库;按接受测验的对象来分,则有适宜于各级、各类在校学生的学绩测验题库与大规模考试题库。

题库,毋庸置疑,储存的是试题。什么是试题? 试题(Item)是题库中最小的可更换的对象,是包含一个或多个问题和应答的基础单元(IMS,2002)。质量审核过关的试题及其各项参数将被输入到题库系统中,按照一定的规则、分门别类地存储起来。题库管理人员可根据工作需要对题库进行维护,如了解题库中的试题情况,有针对性地布置征题任务,以手工或自动的形式构建考试所需的纸质或电子版试卷。

笼统来讲,建设题库是为了有序管理所命制的试题,以提高命题工作效率。具体来讲,题库建设的目的主要为:其一,储存处于各命题阶段的试题。命题工作的各个环节都会为所征集试题增加若干新的信息,如不进行系统化的管理,极易出现技术性问题,且难以弥补,而题库系统则可以用来规范和记录各个环节的工作成果。其二,储存试题的各项参数。试测完成后,试题即获得了考查内容、具体考点等定性信息和难度、区分度等定量信息,这些信息是组配试卷、等值处理时的必须内容,应与相应试题绑定在一起,但又不能显示在试卷中,应分别存放。其三,随时检索和使用所命制的试题。试题积累到一定数量时,如不进行系统化、电子化的管理,极可能会使检索和调取变得异常困难。其四,搭建计算机自适应考试的试题调用平台。作为提供自适应考试所必须的题池的母库,题库系统的任务是为题池储存大量的试题,并根据考试需要向其导入所需试题。

题库首要的基本特性是它的存储性,因此,题库中大量的试题都满足某一考试项目的《考试大纲》,并具有统一量表上的必要参数,还被有序地组织起来。这样,按照指定的试卷设计要求,可灵活而便捷地从题库中抽取试题,组成试卷,并生成彼此等值的多份试卷,以适应社会的需要。所以题库的生成性(即组卷功能)是其价值和功用的直接体现。

(二) 建立题库的基本条件

建立题库一般应具备如下基本条件:

1. 有一个成熟、稳定、明确的《考试大纲》。大纲应对不同性质、层次和目的的考试的内容、能力要求、考试方式和对象做出明确说明,以便明

确题库的内容、题型、权重和完成时量。

2. 有一个命题和审题的专家队伍。命题的专家队伍由学科秘书和学科教师组成,其任务是根据《考试大纲》命制题目。而审题专家,则是由经验丰富的命题教师和学科专家组成,从政策性、科学性、公平性和规范性等多个角度鉴定题目能否入库。

3. 有科学有效的测量学指标。对于一个题目,必须判明其适宜性、难度、区分度、及格要求度、信息函数值等才能决断该题能否入库。

4. 有建立题库的分类系统。为便于管理,对题库试题必须进行分类存储。整个分类系统由试题代码、试题属性、试题内容、测量学指标、试题答案等组成。

5. 有大量的题目。一般实际考题数与题库总题数之比不低于 1 : 10,尤其是国家级题库,题库基准容量应该更高。

6. 有合理完善的保存试题的方法。无论是用题卡还是用计算机保存试题,都必须完整地记载分类细目,而且要利于检索和管理。目前,计算机管理的题库已成大势所趋。

(三) 关于题库的几个问题

建题库涉及一系列问题,这些问题包括:作为题库基本成分的题目的产生、选择、分类及管理;由题目组合成试卷的流程、实施、评分及评价;题库系统的开发、硬件与软件的特征及管理;题库使用的原则等诸多问题。在每一个问题中又存在着许多具体的小问题。这里仅尝试探讨四个最常见的问题。

1. 题库应该包含多少试题?这是一个比较难明确回答的问题。不同的测量对象领域有着不同的自身特点,形成的题目也有着不同的难度,而且题库中的题量与题库的使用范围有着直接的联系。基本上而言,题库内的试题当然是愈多愈好,但是应该考虑所加入题库的试题,是否具有内容效度和统计质量应达成的标准,以及考虑测验的目的何在。Prosser (1974) 建议每个概念至少要包含 10 个试题,每一单元课程内容至少要包含 50 题。Reckase (1981) 则建议 100—200 个难度均匀分布,且具有合理的鉴别度的试题,便可适用在计算机自适应考试中。尽管说法不一,也没有一个绝对化的数量指标来作为题量的标准,但每一个题库的构建都必须选择其一定的题量标准。

2. 题库的题目是否必须要试测与校准?试测与校准问题实质涉及两个问题,一个是不是否需要试测和校准的问题,另一个是如何试测和校准的问题。尽管建题库一般都强调只有经过试测校准的题目才能纳入题库,然而在题库构建理论中也存在认为不需要试测和校准的论点。反对

试测和校准的人认为,通常试测群体的动机很难与正式考试状态一致,更为重要的是,很多时候用于校准的统计手段的假设不成立,校准的正确性就值得怀疑。如题目反应理论的单维性假设、局部独立性假设等在实践中很难满足,因而校准的手段首先就值得怀疑。总的来说,如果题目的统计学参数在组成试卷时非常重要,题库中的题目显然需要经过试测和校准。如果题目的统计量信息在组卷过程中只是一个参考,则试测与校准的意义不大。另外,考试模式也决定了试测与校准的需求,如支持计算机自适应考试的题库,则必须要经过试测和校准。

3. 题库是否安全? 题库的建立,能最大程度稀释泄密的风险,能应对突发事件,及时提供需要的试题或试卷。从这层意义上来说,题库使考试更安全。但是,就题库的安全性来说,至少有三种威胁题库安全的因素:一是偷窃,以出售盈利或自己备考为目的盗取题库中的试题;二是丢失,失误操作,或是难以预料的灾害都有可能造成试题、试卷甚至是整个题库的丢失;三是曝光,考试过程中,试题的重复使用使得题库中有些试题被潜在考生群体记住和窃取。鉴于题库的建立使得安全保密工作由阶段性向长期性转变,一是要对题库服务器的管理和数据备份必须做到万无一失,否则一失万无;二是要经常更新题库的内容,提高试题的数量和质量。

4. 题库是否省钱? 对于没有太多要求的认证或资格考试的题库,试题的平均成本是 300 美元。需要进行试测和校准以获得统计学指标的题库,可能试题的开发成本要达到 1000 美元甚至更多(Vale, 2004)。以美国研究生管理科学入学考试(GMAT)为例,试题的平均成本就达到 2000 美元。再加上题库中不可避免会有一定比率的试题沉淀。因此,就建题库而言,决不省钱。

(四) 题库的优势与劣势

题库主要具有以下几个优点:

(1) 灵活性:可随时根据既定的测验目标构造各种测验,可根据测验的要求选定知识范围、控制试卷难度,适合不同的测验要求。

(2) 保密性:专家提出测量要求以后,只要工作人员操作计算机就可以生成试卷,可以最大限度地限制涉密人员。

(3) 一致性:可以构造平行测验,生成对不同考生,在不同的时间,施测内容不同但水平相同的试卷,保证两次考试的等值。

目前,各个国家的大型题库主要显示出三个方面的优越性。第一,管理学上的优越性。即高效、灵活、高度保密。第二,测量学上的优越性。即高质量、可预控、等值可比、稳定一致。第三,计算机实现技术的先进

性。即试题的动态管理和试卷的自动生成和分发。

题库在实践中也会暴露它的一些弱点,例如题库容量不够,试题出现的重复率就会偏高,试题的保密性反而下降。而且,采用题库也不能完全阻止试题由命题教师或参与题库建设的其他技术人员有意无意之间的泄露。

另外,在题库建设的过程中,要克服的困难和需解决的问题也很多。首先,就是对课程大纲的要求,不仅要明确、具体,更要稳定,所以有一个成熟、稳定的教材及其知识、能力点体系是题库建设的前提。其次,征题入库的过程是题库管理的基础,但也是题库建设的核心和重点,命题质量的好坏、试测方法的设计与试测过程的组织、参数估计和等值处理的准确性和稳定性等都直接决定了题库的质量,这对命题人员和测量技术人员的理论和经验水平要求相当高。第三,题库软件的设计人员由于在教育测量理论方面存在欠缺,在题库的测量功能上很难实现参数估计和等值处理等繁琐的统计功能。如果要健全题库的测量功能,对软件编写人员的要求很高,编程技术既要跟得上计算机的发展,又要有教育测量背景,这样才能够根据题库的需求和技术的进步不断更新和升级软件。因此,题库建设需要教育测量学专家、相关学科专家和计算机专业人员的共同努力。

二、国内外题库现状

题库的建立使编制测验技术发生了根本性的变革,在发达国家建立的不同级别和不同类型题库的实践中,已显示出题库的优越性。随着国内教育测量理论的提高和统计软件的发展,我国在题库建设方面也进行了多次尝试,如前几年开发的通用题库系统,还有汉语水平考试(HSK)和部分省开发的一些题库软件。其中,到目前为止,投入使用且取得了显著成果的是全国英语等级考试(PETS)。该题库建立于1999年,用于存储和管理PETS五个级别的试题和生成各级别逐次考试的试卷,有效地保证了PETS考试实施的机动性、命题的灵活性、科学性和考试标准的稳定性,并为相关考试研究(如CBT、iBT等)奠定了基础。另外,我国台湾也已建立了适应大规模考试的题库。

(一) 国外题库研究与应用现状

1. 美国ETS、ACT与PEM的题库

这三个考试机构都自行开发了拥有自主知识产权的题库管理软件。

三家软件的开发思路基本上都源自 1994 年 Ward 和 Murray-Ward 提出的题库开发理论,大体分为录入和储存试题、修改试题、构建试卷三个环节。其中,录入和储存试题环节,要输入试题的所有属性,包括试题代码、作者、分类计划、题型、案例、图表、试题来源、统计数据、注释和使用范围等;修改试题环节,包括审查、修改试题;构建试卷环节则用于生成试卷。通过这三个环节完成从录入试题、修改编辑,到生成试卷,排版等整个命题工作。生成试卷可以采用人工组卷,也可以自动组卷,但实际操作中以人工组卷为主。

美国教育考试服务中心(ETS)使用的命题和组卷管理系统称为 TCS (Test Creation System),它不仅是一个试题库管理系统,而且已经发展为一个命题考试的工作平台。从录入试题、修改编辑,到生成试卷,排版印刷,都在这个平台上完成。系统能够跟踪每一道试题的运动和所有的修改变化,除了存储试卷,还能够帮助和管理试题的编写和审定、试题分析、编辑印刷、评估有关人员等。

这三家的题库试卷构建理论有的使用经典测量理论,有的使用项目反应理论;在项目反应理论中,有的使用三参数模型,有的使用单参数 Rasch 模型。在题库中采用哪种理论模型构建试卷,主要根据具体考试项目的要求。例如,ETS 题库主要使用的是项目反应理论的三参数模型,试题在入库之前要经过试测、等值、分析适宜性、偏异性等大量的工作。美国 ACT 公司的题库根据不同项目采用了不同的理论模型,美国大学考试(ACT 考试)使用的是经典测量理论,工商管理硕士考试使用的是单参数 Rasch 模型;美国 PEM 从各州客户的不同要求出发,有的项目使用三参数模型,有的项目使用单参数模型。

为使参加不同考次的考生成绩具有可比性,美国各考试机构对不同考次间的考生成绩均进行了等值并进行了分数转换。为了确保考试科学、公平且得到客户的认可,有两道工作防线。第一道防线是用严格的试题开发规范,保证历次测验的内容范围、试卷结构、统计和心理测量指标、行为目标等高度相似,对试题进行试测并允许客户对试卷的信度、效度等统计指标进行评价。第二道防线,即“最后一道防线”则是:通过考后分数处理,剔除质量有问题的试题,并对试卷进行等值,调整不同考次间的试卷难度差别,确保考生成绩的可比性。其中,使用的等值方案有经典理论等值法、项目反应理论等值法和专家定性等值法等多种,但多数采用的是经典理论等值法。

美国各考试机构题库的试题储量与单次考试所需试题数量的倍数并不相同。有的考试为了使试题更加安全,要求题库试题储量至少达到 15