

世界银行贷款师范教育发展项目改革课题

教育计量与评估

曾昭轰 郭 睿 编著

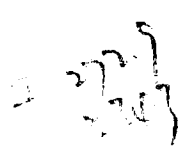
新华出版社

世界银行贷款师范教育发展项目改革课题

教育计量与评估

曾昭轰 郭睿 编著

新华出版社



责任编辑:李 琴

封面设计:陈 玮

图书在版编目(CIP)数据

教育计量与评估/曾昭轰 郭睿 编著. 北京:新华出版社
1997.12.

ISBN 7-5011-4375-7

I.教… II.①曾…②郭… III.教育-计量-评估-基础知识
IV.H.179.3

中国版本图书馆 CIP 数据核字(97)第 84217 号

教育计量与评估

曾昭轰 郭睿 编著

※

新华出版社出版发行

中共江西省委党校印刷厂印刷

※

850×1168mm 32K 10.625 印张 247 千字

1997 年 12 月第 1 版 1997 年 12 月第 1 次印刷

※

印数 1-1500 册

ISBN 7-5011-4375-7/Z·126

定价:11.80 元

目 录

前言	(1)
----	-----

教育计量篇

第一章 教育测量的历史、现状与局限	(4)
§ 1-1 教育测量的历史与现状	(4)
§ 1-2 教育测量的局限	(7)
第二章 教育计量理论产生的历史必然性	(12)
§ 2-1 教育计量理论产生的现实基础	(12)
§ 2-2 教育计量理论产生的内部动力	(14)
§ 2-3 教育计量理论产生的科技条件	(16)
第三章 教育计量理论	(19)
§ 3-1 教育计量的概念	(19)
§ 3-2 教育计量学与教育测量学的区别	(20)
§ 3-3 教育计量学的意义	(21)
§ 3-4 教育计量的特点	(23)
第四章 教育计量的基本要素	(26)
§ 4-1 教育计量的基本要素	(26)
§ 4-2 教育计量的步骤与方法	(28)
第五章 教育规划的计量模型	(31)

§ 5-1 教育规划的意义	(31)
§ 5-2 教育事业发展规划的内容	(32)
§ 5-3 编制教育规划的原则	(33)
§ 5-4 编制教育规划的步骤	(34)
第六章 学校德育工作的计量	(50)
§ 6-1 计量学校德育工作的意义	(50)
§ 6-2 学校德育工作计量的原则	(52)
§ 6-3 学校德育工作计量的内容	(53)
§ 6-4 计量学校德育工作的方法	(55)
第七章 学生个体思想品质计量	(64)
§ 7-1 学生个体思想品质计量的意义	(64)
§ 7-2 学生个体思想品质计量中的几个认识问题	(65)
§ 7-3 学生个体品德计量方法	(67)
第八章 学校体育卫生工作计量	(75)
§ 8-1 学校体育卫生工作计量的意义	(75)
§ 8-2 学校体育卫生计量的范围和方法	(76)
§ 8-3 学校体育工作的计量	(82)
第九章 教育经济效益计量	(98)
§ 9-1 计量教育经济效益的意义	(98)
§ 9-2 教育经济效益的计量范围和特点	(100)
§ 9-3 计量教育经济效益的方法	(102)
§ 9-4 学校内部办学效率的计量	(112)

第十章	教育评估概述	(116)
§ 10-1	教育评估概念	(116)
§ 10-2	教育评估的特点	(121)
§ 10-3	教育评估的类型	(123)
§ 10-4	教育评估的功能和作用	(132)
第十一章	教育评估的历史和现状	(136)
§ 11-1	教育评估的产生	(136)
§ 11-2	教育评估的发展历程	(139)
§ 11-3	教育评估的现状和发展趋势	(144)
§ 11-4	中国的教育评估	(150)
第十二章	教育评估的主要模式	(156)
§ 12-1	系统分析模式、行为鉴赏模式和鉴定模式	(156)
§ 12-2	行为目标模式	(157)
§ 12-3	CIPP 模式	(159)
§ 12-4	目标游离模式	(163)
§ 12-5	表象模式	(164)
§ 12-6	对手模式	(166)
§ 12-7	应答模式	(167)
第十三章	教育评估的依据、原则与方法	(169)

§ 13-1	教育评估的主要依据	(169)
§ 13-2	教育评估的原则	(172)
§ 13-3	教育评估的一般方法	(179)
第十四章	指标体系的设计	(190)
§ 14-1	目标、指标和指标体系	(190)
§ 14-2	设计指标体系的原则	(195)
§ 14-3	指标体系的设计程序与方法	(200)
第十五章	权系数的确定	(205)
§ 15-1	权系数的概念与作用	(205)
§ 15-2	权系数的确定方法	(208)
§ 15-3	权系数品质的检验	(219)
第十六章	评估标准的制定	(223)
§ 16-1	教育评估标准及其类型	(223)
§ 16-2	制定评估标准应注意的问题	(226)
§ 16-3	制定教育评估标准的方法	(228)
§ 16-4	评估标准的量表	(230)
第十七章	教育评估的实施	(236)
§ 17-1	准备阶段	(236)
§ 17-2	实施阶段	(241)
§ 17-3	有关实测的几个具体问题	(246)
§ 17-4	评估误差心理的调控	(252)
第十八章	评估信息的获取	(259)
§ 18-1	获取评估信息应注意的问题	(259)

§ 18-2 获取评估信息的主要技术	(260)
§ 18-3 关于教育测量中两个问题的讨论	(273)
§ 18-4 获取评估信息的方法和途径	(276)
第十九章 评估信息的统计分析	(279)
§ 19-1 教育统计的基本条件	(279)
§ 19-2 单组数据的统计方法	(280)
§ 19-3 多组数据的统计方法	(288)
§ 19-4 评估信息的综合分析	(299)
第二十章 评估结果的处理	(308)
§ 20-1 造成评估结果误差的分析	(308)
§ 20-2 评估结果的一致性检验	(309)
§ 20-3 评估指标的检验	(317)
§ 20-4 评估结果的再评估	(324)
§ 20-5 评估结果的反馈	(327)
参考文献	(329)

前 言

《教育计量与评估》一书是世界银行贷款“师范教育发展”项目改革课题——师范教育系统评估体系研究课题的研究成果之一。因为我们在完成师范教育系统评估方案的研制过程中，曾经碰到了大量的理论与实际问题。这些问题主要来自于两大方面：一方面是教育活动中的相关因素如何科学、准确计量；另一方面是教育活动的社会价值如何客观地进行评估。如果这些问题得不到较好的解决，课题研究就有搁浅的危险。为此，课题组组织力量，对在研制和实施师范教育系统评估方案过程中所碰到的教育计量与评估的有关理论与实际问题进行了专门、深入的研究，并将研究结果编成了《教育计量与评估》。

本书分教育计量篇、教育评估篇两大部分，共 20 章。教育计量篇的 9 章由曾昭轰编写，教育评估篇的 11 章由郭睿编写。全书由曾昭轰负责统稿。在成书的过程中，曾经得到江西师范大学高教研究所廖德文先生、厦门师范学校易化同志的热情指导和帮助，在此，向他们表示诚挚的谢意。

尽管国内外对教育计量与教育评估的研究已经多时，但至今仍不能说这方面的理论与实际问题就完全解决了，对它的研究过程仍在继续、深入之中。如果本书中我们的认识能有一点

点有助于推动教育计量与评估研究过程的话，我们将感到十分欣慰，但更使我们感到不安的是：教育计量与评估毕竟是一门综合技术，它涉及到教育学、心理学、数学、教育经济学、行为科学等多学科的知识，囿于编者知识面所限，书中可能有许多肤浅的认识，甚至是谬误的结论，在此，恳切希望同仁们批评指正。

编 者

1997 年 10 月 18 日

教育计量篇

第一章 教育测量的历史 现状与局限

§ 1-1 教育测量的历史与现状

从实践意义上说的教育测量,已经过了一个十分长远的历史过程。可以说,从教育活动诞生之日起,施教主体就已开始采用一些简单、直观的形式,去检测施教客体对施教主体所传知识与技能的掌握程度。以后,随着社会经济的发展以及教育自身发展的需要,教育测量的内容不断充实,教育测量的手段不断更新,教育测量的理论不断完善,教育测量的精确度不断提高。

在文字出现以前相当长时间内,知识载体是显现的语言和动作,那时教学活动是一种极其原始的教育活动,其目标是维系低水平的生存和繁衍,其内容是原始的生产、生活经验,其形式是口耳相传。在这种原始教育状态下,考察受教育者掌握知识和技能的程度,只需要也只能从受教育者的语言复述和动作再现状况与传授者传授时的语言、动作的吻合程度。显然,这种原始的教育测量,其形式极其简单,其内容极其狭窄,其标准又因人而异差别极大,因此,其测量结果因缺乏可比性,除对受教育者本人的生活和生产有其一定的作用外,对离开传授者以外的教育作用不大。

文字,尤其是中国的纸张、印刷术出现后,极大地推动了教育的发展,也极大地推动了教育测量的发展。由于文字、纸张、印刷术的产生,已使知识的载体形式,开始脱离原始状态,人们可以利用文字、纸张将前人的知识、经验记录下来,并通过分析、研究、整理,形成教科书,这样就使知识可以脱离教育者个人的局限,而把知识的时空范围变得空前的广泛;也使受教育者接受知识和再现知识的形式发生了深刻的变化,由简单的模仿变成记忆、理解、应用等一系列的复杂的内化过程,从而极大地促进了受教育者思维能力、创造能力的提高。正因为教育内容和形式的深刻变化,使考查、考核受教育者知识和能力的教育测量的形式也需要有新的变化。于是,除采用语言复述和动作再现形式来考查受教育者知识技能以外的新教育测量形式出现了,这主要是笔试。在这方面,我国具有悠久的历史。据《礼记·学记》记载:早在我国古代西周时的“国学”中不仅已有笔试形式,而且已经形成了定期的学业考查制度,依据不同年级学生确定不同的考试目标、考查内容和考查标准:第一年考查他们能否分析经义、辨别志愿而决定学习趋向;第三年考查他们能否专心学习,与同学互相切磋研讨;第五年考查他们的应用知识能力;第九年考查他们推理论事,触类旁通,解决实际问题的能力等等,这种考查与今天的考试已相当接近了。

后来经过秦、汉,我国考试制度更为严格,只有通过考试才可以毕业。考试形式已有“口试”、“策试”、“射策”三种,是现在常用的“口试”、“笔试”、“演示操作”考试形式雏形。自隋朝科举制度建立以后,考试不仅在人才培养中起着重要作用,而且在人才选拔中也起着十分重要的作用,后经唐、宋、元、明、清等数代,

科考制度在中国延续了 1300 年之久,虽然科举制度在学校教育中存在许多诸如只重视记诵,不求义理;束缚学生思想发展,障碍学生眼界,甚至摧残人才等弊端,但是,历经多年的科举制度,对于以学生学业成就考试的教育测量理论与方法,在中国乃至世界都功不可没。它所创立的考试形式、组织方式、考场规则乃至试卷编制办法,至今仍在影响着世界的教育测量。中国古代的考试方法传到国外以后,曾经得到较大的发展,尤其是以蒸气动力技术为先导的第一次技术革命促进了资本主义的发展,也促进了教育的发展,特别是对学校教育的人才培养提出了新的要求,这就促使原有考查学生学业成就的办法也应有新的改进。教育测量最初的奠基人——美国的菲奢在 1864 年首次编制了学生成绩量表,1897 年莱斯进行了拼字测验,1904 年桑代克发表了《心理与社会测验导论》一书,标志着教育测量学的真正诞生。从此之后,教育测量作为一种可行性工具在鉴别智愚、选拔人才、指导就业、预测、评价学生学业成就等方面得到了广泛的应用。1905 年法国心理学家比纳和西蒙合作编制的《比纳—西蒙量表》首先在学校使用,用于确定儿童的智力水平;英国人高尔顿、美国人武德沃斯先后编制了人格测验量表,用于系统的人格测验等等。不仅把教育测量的使用范围由学生学业成就扩展到人的智力水平和人格特征,而且也把教育测量的功能由评价学生学业成就延伸到人的心理、人格等品质水平进行鉴定、预测。同时,由于数理统计理论的日臻成熟,促进了教育统计学的发展,进而又加快了教育测量理论和方法的完善过程,特别是电子计算机的出现与普及和现代通讯技术的发展,既极大的改善了教育测量的手段,也推进了教育测量理论体系的建构,对于测

验题目编制,常模制定,评分体系,分数解释,测验结果质量检验指标与方法等一系列问题均有比较科学、可行的理论阐述和操作设计,从而使教育测量可以脱离狭小的地域限制,而使不同地域,甚至不同国别的可以进行联考、统测,考试结果可以比较。

二次大战后,由于电子计算机的出现与普及以及现代交通、通讯技术的发达,又使教育测量在原有基础上向前跨进了一大步,这就是标准化考试。所谓标准化考试,即考试工作要求、工作程序的规范化、统一化。具体说是,考试工具亦即试卷的标准化,试卷的信度、效度、难度、区分度均符合考试要求;考试过程的标准化,严格按照考试规则,尽量排除对考试的干扰;评分的标准化,一般都采用阅卷机评分;分数解释和使用的标准化,将原始分数转换成百分等级或标准分,并提供常模,用于对照。由于这种标准化考试对考试过程中的每个环节都有严格的要求,对于可能造成误差的因素进行严格控制,从而保证考试结果准确可靠;标准化考试题量较大,内容覆盖面大,考核全面;评分客观、准确,可比性强;阅卷自动化,可减少大量的评卷组织工作和经费支出等。正因为如此,近几十年来世界各国都在广泛推行标准化考试。近些年来,我国的出国人员外语考试、高等学校招生考试以及各级学校校内的大考均全部或部分采用了标准化考试。

§ 1-2 教育测量的局限

尽管教育测量已有几千年的历史,现代意义的教育测量也有近百年的历史,且其理论体系和方法论体系应当说已经发展

到了相当完善的程度,但是,应当承认,从真正意义上说的教育测量至今仍有许多局限性,主要是:

1、教育测量范围的狭窄性

教育测量从它诞生的第一天起,其目标就是考查受教育者的知识和技能水平,虽然经过长时期的发展,已经延伸到考查受教育者的心理、智力等方面来,但其最终目标仍然是为预测学生的知识、技能发展倾向服务的。所以,至今见到的有关书刊中对教育测量的界定大多是:“教育测量学是一门以学业成绩的考查和评定为研究对象的科学,也就是专门研究考试问题的;”^①“教育测量仅以学生的认知领域中的成就为对象,专门研究如何测量学生在文化科学知识方面的学习成就”。^②“教育测量是对学生的学力,包括学习能力、学习成绩及其相关的心理状态进行了解的一种方法。”^③等等。然而,教育是一种培养人的社会活动,广义说,泛指一切增进人们的知识、技能、影响人们的思想、意识的活动。狭义(即学校教育)说,教育是教育者根据一定社会(或一定阶级)的要求和年轻一代身心发展的规律,对受教育者所进行的一种有目的、有计划、有组织的传授知识技能,培养思想品德,发展智力和体力的活动。显然,目前这种仅以学生学业成绩测定的教育测量,是难以对教育活动的全部内容进行测量的。而且,随着社会的进展,教育的功能将愈加宽泛,教育测量范围的这种狭窄性与教育活动内涵的扩充矛盾愈加突出。致使学生的思想品德、基本技能以及教育的社会、经济效益等本属

① 漆书青等译著:《教育测量纲要》,1984年内部发行。

② 肖鸣政:《试卷编制的方法与技巧》,江西教育出版社,1989年版。

③ 朱作仁:《教育辞典》,江西教育出版社,1987年版。

教育活动的直接结果,现在的教育测量就很难测量。

2、教育测量的浅表性

教育测量的主要手段——考试,是按照一种恰好与教学过程逆向的运行机制进行的(如图 1-1):

知识体系 $\xrightarrow[\text{考试,通过学生外化}]{\text{教学,通过学生内化}}$ 学生的智能体系

教学过程的机制是,教师依据对教学大纲和教材的理解与把握,形成一个较为完整的知识体系而体现在教案之中,然后教师通过语言、文字、图表、动作、教具等工具把知识体系变成学生可以接受的多种信息,学生则通过对这些教学信息的接收、加工、处理,最后内化成自己知识和技能后储存在头脑之中。考试的目的则在于检测学生头脑中的智能体系与原先规定的教学要求是否吻合,进而达到改进教学工作的目的。但是,贮存在学生头脑中的智能是隐含的,不容易计量,需要通过一种方式或刺激,使之外化而变成可以测量、评价,这种方式或刺激就是通常所说的考试。考试就是以测量学生的学业水平为目标的测量,其机制是通过标准反应模式(即试卷)为中介,学生在对标准反应模式(试卷)应答(答题)时,将其隐含在头脑中的知识和能力外化在标准反应模式(试卷)上。教师就可以通过学生答题情况,来判断学生对所教授内容的掌握情况。如果这种判断要认可,则考试至少应当具备如下二个前提条件:其一是用于测量的标准反应模式(试卷)应能全面反映教授内容的广度和深度;其二,考试的监考、评分等各个环节应能有利于学生潜在智能的外化,有利于学生知识和能力的发挥。然而,在目前情况下,这二