现代教育测量

XIANDAI JIAOYU CELIANG

宋鸿表
刘兆美
张世华
张才增
张国奖
彭颂著
编成

现代教育测量

宋兆鸿　刘世表　张才美
张国华　张颂增　彭成奖
编　著

教育科学出版社

现 代 教 育 测 量

宋兆鸿　刘世表　张才美
张国华　张颂增　彭成奖　编著

教育科学出版社出版
（北京北环西路 10 号）

责任编辑　金宏瑛

新华书店北京发行所发行
中国科学院印刷厂印装

开本 850×1168 毫米 1/32　印张 7.5　字数 198,000
1986 年 11 月第 1 版　1986 年 11 月第 1 次印刷
印数 0,001—5,000 册

书号：7232·253　定价：1.40 元

内 容 提 要

本书阐述了把考试建立在科学基础上的理论与方法，内容新颖，反映了国际上教育测量学研究的新成果，从理论与实际的结合上探讨考试改革，是使考试更好地为选才、育才和用才服务的教育基础理论和应用技术读物。

本书注意从教育测量的起步知识讲起，对考试命题，考试结果的统计检验和预测，以及对编制分析考试结果的计算机程序等，都作了系统的阐述。广大教师可以依据本书提供的方法，科学地测量学生的成绩；拥有微电脑的学校使用本书更为方便。

前　言

教育测量学是教育科学的重要组成部分。它的任务是把考试建立在科学的基础上，即以客观的定量分析代替传统考试方法的主观随意性，使考试在选才、育才和用才上更有成效。教育测量学的原理和方法在国外已经沿用了七、八十年，目前仍在蓬勃发展着。相比之下，在这方面，我国却有不小差距。为了赶上国际上教育测量的发展，促进我国教育测量的现代化，1983 年冬广东省高教局召开了全省"教育测量"讨论会。我们受高教局之托，在暨南大学校领导和教务处的支持下，为会议准备了几个专题报告。我们现在把它们整理成册，奉献给读者。本书在编写过程中，遵循了下列原则，一是力求使本书具有广泛的适应性，编入的预备知识较多，便于缺乏数理统计知识的读者循序阅读、理解，二是加强实用性。教育测量学本身应用性就很强，为使它能够直接应用于考试实践，我们比较重视电脑在考试中的应用并附有软件，只要备有微型电脑，即可使用。

本书的编写工作是在我国著名的教育测量专家，广州师院院长陈一百教授和北京师院郝德元教授热心指导和帮助下进行的。他们严格细致地审校了本书的初稿。没有他们的帮助，本书是不能问世的。编写工作还得到暨南大学顾问王越教授的鼓励和支持，在此一并致谢！

编者

一九八四年春于暨南园

此为试读，需要完整PDF请访问：www.ertongbook.com

序

从历史发展趋向看学业成绩考评方法的现代化

陈 一 百

一、成绩考评必须实现现代化

学业成绩的检查与评定是整个教学过程的有机组成部分。通过成绩的考核，不仅可使学生及时了解自己学习的优缺点以及对各个方面学习目标的差距，自动调节努力方向，充分发挥学习的主观能动作用，而且对教师来说，也是自己了解教学效果，及时调整教学要求，改进教学方法，保证完成教学任务的必要依据。

成绩的检查与评定是教育测量中的一个门类，它同一切物理的测量一样，必须做到客观、准确。飞机着陆，飞行员如果看错仪表，俯仰角有几度的误差，便会造成严重事故。成绩考核如果组织不当，评分不够准确，对学生就会产生严重消极后果，不利于智能与德体的健康成长。教师和教育部门也会由此作出错误的教育决策，造成智力开发和培育人才的巨大浪费。

当前我国的教育测量方法(包括课堂考试以及升学考试等等)不够科学和准确，许多教师和学生都有切身感受，社会上也有不少事实可资说明。据《中国青年报》报道，有的考生通不过高考，却考取了研究生，有的人考不上研究生，却考取了研究所，当助理研究员。考生考入大学后，高分低能的事例亦时有所闻。

但当前问题的严重之处，还不在于考评办法不够健全，而在于许多人因循守旧，对传统的成绩考核办法的严重不合理之处，或是缺乏认识，或是讳莫如深，以致成绩考评始终成为教育工作中最薄弱的一个环节。

由于“极左”余毒未消，对借鉴国外先进测验理论和经验余悸犹存，安于现状，不敢大胆进行改革，这是我国现行成绩考核制度长期得不到改进的原因之一。我国现行的成绩考核办法，一般说来，无非是主讲教师根据教材任意出三、五道试题，然后根据学生书面答案的质量，评给一个分数，这实质上就是我国在二十年代以前所采用的那一套方法——也就是所谓旧法考试。旧法考试及其评分带有极大的主观性，早已为教育学者所熟知。远在几十年以前，著名教育学家斯太奇等就曾经做过试验，他把同一本语文试卷分给142位本学科中学教师评分，结果同一本试卷所得分数有35种，从50分到98分，高低不等。实验者又把同一本几何试卷分给116位本科教师评阅，结果所得分数有60多种，最低为28分，最高达92分。旧法考试评分不够客观，误差之大，一至于此！外国教育界还有一件引为笑谈之事：某年夏季许多大学教授在评阅历史考卷，有一位教授为评阅方便起见，自己写了一份答案作为典范。不料这份范卷和其它待评考卷混在一起，给另一教授评阅，竟得一个不及格的分数。为慎重起见，其他教授对这本不及格的试卷重复评定，结果所得分数差别之大，竟是从40分起直至90分为止！

类如这些早为科学准则所否定的过时的考试办法，我们今天仍在普遍使用，并且视之为天经地义丝毫不容置疑的可靠量具，岂非怪事。在这种思想僵化，盲目迷信分数的思想支配下，误用分数的事例，乃层出不穷，流毒社会。有的部门仅凭学生考分一两分之差，便对学生发展前程作出生死悠关的最后判决，大有“一试定终身”的意味。有的教师根据本班某科成绩平均提高零点若干分，或者学生不及格率有所下降，便毅然断言教学质量有了提高。曾有不少学校以高考多出“状元”为荣，报上也大肆宣扬。有些教育行政部门对学校进行排队，亦以高考分数或升学率的高低作为学校办得好坏的唯一标准，使学校偏离了正确的办学方向。

我们并不反对严格执行现有考试制度，不反对以学科考分作为甄别学习质量的依据之一。现行的考评方法虽尚有严重缺陷，毕竟比之取消考试，否定教师组织成绩考评而代之以行政推荐，民

主评议等等的方法要优胜得多。我们所必须反对的乃是把现行制度下分数之价值绝对化，而应当实事求是地对分数采取科学分析的态度。只有这样，才能有助于提高有关成绩考评的科学认识水平，才有助于减少误用分数的弊害！

在现行成绩考评方法中，存在着严重的思想混乱。这可从有关"分数贬值"的问题谈起。

教育部办公厅编的第48期《教育简报》，载有上海工大物理教研室进行统一测验，改变"分数贬值"现象的报导。据称，该校近年普遍出现成绩偏高、"分数贬值"的现象，许多课程的平均成绩在八十分以上；有的课程成绩高达92分（其中优秀占62.2%，良好占32.4%，最低的72分）。文章指出，在通常情况下，学生成绩的分布总是"两头小，中间大"的，因此断言该校所曾普遍出现的考试成绩偏高，实际上是"分数贬值"。

本人分析我院（广州师院）一九八〇年度下学期各门课程考试成绩的分布，象上海工大一样，绝大部分课程考试的平均成绩亦都在80或90分以上，优生（即90分以上的）在全班所占比例，绝大多数课程都大大超过7%以至20%。以文科一个系来说，优生比例占25%以上的有一半课程，比例最大的一门竟高达93%。理科一个系，优生比例占20%以上的有一半课程，优生比例最大的一门课占48.9%。

我曾就这两份资料征询院内外教师对这些考试是否"分数贬值"的看法，结果出现两种不同意见：

部分教师同意贬值之说，但仅凭直觉作出判断，讲不出什么客观依据。有的也提到不符合"两头小，中间大"一般的分布情况，但对"两头小，中间大"的确切意义，怎样才算违背了这个准则，各人的理解很不一致。大家更无法说明为什么不符合这个准则就是分数贬值。

另一部分教师对贬值之说，持保留态度，认为不能从表面上仅仅根据全班平均成绩较高或优生的比例较多，便断言是分数贬值。他们指出，如果试题对教学内容要求确有代表性，打分又确实客观

可靠，则绝大多数学生取得优秀成绩，正是体现了教学面向全班学生取得了实际效果，没有理由认为就是分数偏高。过去学习苏联，采用五级制记分法，不是要求学生人人争取 5 分吗？

这两种意见不能认为没有道理，谁也说服不了谁。那么，问题究竟出在什么地方呢？

依我看来，问题出在我们立论往往把两种性质截然不同的评分体制纠缠在一起，概念混淆不清。我们知道，当前世界上评分方法有两大类，一种称为"绝对评分"，一种称为"相对评分"。

绝对评分是以考生对测验所要求的全部知识内容究竟掌握得多少作为依据的，一般与百分制记分法相关联。答对了全部试题，便评给 100 分，对测验内容毫无所知，便给零分，对测验内容掌握了百分之六十，便评为 60 分，惯常以 60 分作为及格线（及格线的规定是任意的，国外亦有学校以 75 分为及格线的）。

相对评分是以考生测验成绩相比较按其好坏在全班学生中居于什么地位作为评分依据的，一般与 A、B、C、D、F 或优、良、中、差、劣（或不及格）这类文字式记分法（亦有转化为数字式记分的）相关联的。为使这种分数的意义明确，各级分数的价值等距，故常依据正态曲线下对应于横轴一定区间内五等分间距所占有的面积，规定各级分数所应评给的人数比例，例如规定成绩评 A 者应占 7%，评 B 者占 24%，评 C 者占 38%，评 D 者占 24%，评 F 者占 7%。此外尚有分级较多的类似记分法。

我国现行评分制度通用百分制，属于绝对评分法的范畴，但在衡量分数是否贬值的时候，人们又往往依据相对评分法的准则（"两头小，中间大"）来要求教师打分。这就构成了一种矛盾，使教师无所适从。

究竟是采取相对评分还是绝对评分？如果是绝对评分又应如何看待或解释分数的"当值"，防止出现分数"贬值"和"升值"？这些都是学校学业成绩考评办法所必须作出明确规定而还没有做到的。其他问题例如，以正态分布的规律来规范经过严格筛选的学生成绩是否合理？成绩分布出现了显著的正、负偏态是否就意味

着教师提高或降低了分数标准？依靠调整试题难度来取得考分分布的正态化，是属于绝对评分还是属于相对评分？如其他条件相同，对于提高学习质量、两种评分体制的效果孰较显著？两种体制互有优缺点，能否兼取其长，创立一种较为完备的评分方法？如此等等，都是我们必须进一步进行探索力求取得解决的理论和实际问题。

二、从历史发展看成绩考评的发展趋向

自有学校教育，即有成绩考评，历史已甚悠久。就近八十年的演变过程看来，大致可划分为两个阶段，先是（1）由旧法考试发展到客观测验，然后是（2）由常模参考性测验发展到目标参考性测验，由相对评分到绝对评分或绝对——相对评分相结合。

客观测验的兴起　二十年代以前，世界各国包括旧中国在内，采用的都是我们所已熟知的所谓"旧法考试"，直至第一次世界大战前后，在西方特别是美国，出自对科学地甄别人才、选拔人才的迫切需要，作为旧法考试否定物的种种标准化智力测验和教育测验，乃风起云涌，大量编制出来，并大大促进了教育测量这门科学的建立和发展。这些标准化的教育测验，其编制大都出自测验专家之手，是经过大量试测，其有效性和客观性业已证明达到了一定数量指标，然后公开发行供人使用的。这种测验的突出之点是：

（1）试题的取样亦即智能的复盖面广，效度高。

（2）题式方面不要求考生作长篇大论的书面回答，要求明确，答法单纯，定分客观、准确。

（3）从命题，测验实施到评分的一切方面，努力排除一切无关因素（如被试者对测验指导语的理解，精神紧张，试场噪音等等主客观因素）对得分的影响，并实施测验条件的规范化等，以确保得分的准确性，可靠性和可比性。

（4）每个测验制备有试题内容不同而价值相等的"复份"，对同一学生可以一再测试，而得分在不可避免的一定的误差范围内，大致相同。

（5）每个测验都根据对全国范围或极大规模的同年龄、同年

级或同性质的学生集体进行实测，据以制成集体常模，借助于这些常模，任何考生在测验中所得的原始分数（例如答对题数或答对要点的百分数）就可以转换成为对照学生集体水平的常模性分数（例如标准分），一看便可了解某一考生的作业亦即原始分数，与同一集体的其他学生相比，在具有价值等距的成绩量表中，居于什么一个位置。例如某一考生甲的标准分为 0.00，即表明其考试成绩相当于同一集体学生的平均或中位水平，乙生的标准分为－0.50，即表明该生成绩落后于集体平均水平半个标准分单位。假设成绩分布是正态的，便可推知同一集体约有 70% 的学生成绩超过他，水平不能算好；丙生的标准分为 1.00，即表明该生成绩超过集体平均水平一个标准分单位，可以推知同一集体约只有 16% 的学生成绩超过他，水平自然属于比较良好。

由此可见，标准分的意义十分明确，毫无含糊不清的毛病。尤其具有重要意义的是，标准分的单位价值是等距的，可以进行加减运算，可以比较不同学生在同一学科测验中彼此之间的成绩差别有若干等值等距的单位（譬如说，可以求知丙甲二生的成绩差距为丙乙二生的成绩差距的三分之一），也可以比较同一学生在不同学科之间的成绩差别有若干等值等距的单位（例如可以求知物理与化学的成绩差别远较物理与汉语的成绩差别为小，前者的差别仅为后者差别的二分之一，等等）。诸如此类有关个别差异和智能差异的有用信息，都是旧法考评所无法提供的。

标准化测验的蓬勃发展，为教育工作者提供了其客观性和正确性在某些方面堪与物质量具相比拟的教育量具，大大促进了教育研究的科学水平，也大大丰富了教育测量学的理论。但是经过标准化处理的各科教育测验，其品种毕竟是有限的。为了满足各科不同阶段，不同单元教学上的需要，教师自编的课堂测验仍然一贯居于不可或缺的首要地位。经过彻底改革的课堂考试，除了无须象标准测验那样要经过试测，计出常模，制备复份以及采用一系列的标准化手续之外，都是要求教师按照标准测验编制的主导思想与原则方法进行编制的，其性质，试题取样，试题形式，实施方法以至

此为试读，需要完整PDF请访问：www.ertongbook.com

评分所依据的原则跟标准测验都大体相同。应当说，自二十年代以来长达四十多年的悠长历史阶段中，所有一切心理与教育测验，都是以鉴别儿童或学生的个别差异为指导思想的，成绩评定亦是以个别差异的实际分布作为依据的，因而这一历史时期的各种测验往往被称为"常模参考性"测验，其评分方法则属于"相对评分法"的范畴。

目标参考性测验与绝对评分法的提倡　常模参考性测验的发展，使成绩考评实现了高度的数量化，定分的客观性与可比性达到了前所未有的水平，这是其最大的贡献所在。但是由于这种测验的立足点是个别差异，是比较不同学生在学业上"总"的成就，具有一般性的调查性质，而不是着眼于首先明确规定一个测验所要测验的各项目标，不是着眼于鉴别各类教学目标是否完成或完成得一样好。根据测验目标与测验内容的这一个特点，因此常模参考性测验又常被称为"调查性测验"或"概观性测验"，它对学生学习所起的作用主要是考核或监督的功能，而不能充分起到诊断学习缺点、难点，主动调节努力方向，确保完成各项学习目标的作用。迨及六十年代，在多元智能结构新理论和其他新教学论的影响以及课程改革迫切要求的推动下，一种与常模参考性测验相对立而被称为"目标参考性测验"的新型测验乃应运而生，骎骎然大有取代常模参考性测验之势。

目标参考性测验具有哪一些特点呢？只要同常模参考性测验相对照，便不难窥知其要。首先，二者差别在于所要提供的信息种类有所不同。前者的用处在于确知有哪一些规定的教学目标某一学生已经完成，而后者的用处则在于确知某一学生对于某科"总"的知识量掌握了多少。由实施目标参考性测验而制作的初步成绩记录往往是标明一系列业已完成或尚未完成的学习目标，而实施常模参考性测验后的初步成绩记录，则是全部测验试题中已被答对的总计题数。其次，二者用以解释所获得信息的依据有所不同。目标参考性测验的"目标"就是完成所有的教学目标。学生学习的好坏是以该生对预定的各项教学目标业已完成的数量或百分数来

判断的。而常模参考性测验的"常模"是指某一规定的学生集体在该测验的成就。某一特定学生的学习成就的好坏，乃是以该生成绩在这一规定集体所居地位如何作为判断的。

由于目标参考性测验具有上述两大鲜明特点，在欧美现代教学体制的许多革新中已获得广泛的应用，尤其是需要严格贯彻循序渐进，依靠自学或自动调节进度的各种教学体制包括计算机辅助、计算机管理的教学体制中，其应用的效果更为显著。在所有这些教学体制中，测验总是与教学结合为一的，在单元教学之前，中间和结束，都必须通过测验来核对必须具备的基础知识技能，诊断可能出现的学习困难，并预订后继的教学程序。教师自编的课堂测验，如果遵循目标参考性测验的原则方法进行改革，自将大大有助于提高学生学习的目的性，充分发挥测验对教学的反馈作用，从而使成绩考评成为保证实现教学目标的强大动力和武器。这在欧美已有大量的事实足以说明。

我国的情况　解放前的旧中国，学校一贯采用旧法考试。自二十年代开始，亦吹来了测验客观化的新风，数十种由专家或学术团体编制的各种心理和教育测验和标准化测验陆续问世，师范院校亦开设有教育测验与统计的课程。虽则国内绝大多数学校的课堂考试仍沿用旧法，原封未动，但就大势而论，实已开创了由旧法考试向客观测验过渡的新路。迨及新中国诞生，由于照搬苏联凯洛夫的那一套，客观测验被贴上了资产阶级的标签而受到全盘否定，以五级分制为中心的苏式成绩考评方法便一跃而成为举国一致奉行的准则。

苏联成绩考评法的特点　五十年代苏联成绩考评方法对旧法考试来说，是一个划时代的革新。与欧美的客观测验体制相对比，亦具有许多可取的独到之处。苏联成绩考评，出自某种偏见，对欧美教育测验的偏重常模与追求考评的数量化全盘否定，从而走向另一极端，影响到考评结果的客观性与准确性，是其主要缺点所在，但是就其涉及教学论的要求方面来说，却有许多地方是符合我国社会主义教育体制的实际要求，并且是与现代教育考评的最新发

展趋势相一致的。要而言之，苏联考评体制有如下几个特点：

1. 强调成绩考评要紧扣统一规定的教学大纲的具体要求，加强学生学习的目的性。

2. 强调成绩考评与教学过程的有机统一，保持成绩考评的经常性，系统性，使自我考核成为学生的自觉要求，强调考评要对学生掌握知识、技能、技巧、发展能力，养成良好习惯，理想、态度、情操以至世界观，起到积极的促进作用，而不是消极的监督作用。

3. 反对相对评分，而采用绝对评分的五级制记分法，制订了各级分数标准的统一的原则性规定，评分只分五级，易于鉴别，适用于课堂提问，以及其他一切经常性的，系统性的考评。

4. 强调教师要及时把分数告知学生，并说明给定这一分数的依据，使学生觉得公平合理。

在苏联及我国的实践证明，苏联考评方法对保证学习质量是富有成效的。苏联卓有成效地培养了大批堪与欧美争锋的高级科技人才，以及一九五七年的卫星上天，激发了美国对苏联教育包括成绩考评制度的重视。美国六十年代开创大规模的课程改革实验以及目标参考性测验的勃兴，可能也有借鉴苏联经验的因素在内。在此同时，苏联教育界亦已开始开放某些学术"禁区"，注意吸取欧美行之有效的某些测验方式方法，两种成绩考评体制正在出现取长弃短互为补充的新局面。

返观我国六十年代以来，凯洛夫教育学受到了批判。欧美式教育测验既以"资"字号被否定于前，苏联考评方法又以"修"字号被否定于后，而教育领导部门又长期未能为建立新的成绩考评制度打开局面，指明方向，制订办法，遂使各级学校考评陷于无所适从的困境，甚至自发倒退，承袭了封建、半封建时代的考试旧法而不自知。这种落后状态，已到了非改不可的时候了！

怎样才能促进考评方法的现代化呢？除了教育学专业队伍必须充分发扬勇于革新的精神，积极开展有关教育测量理论的和实验的研究外，我们广大教育工作者如能在辩证唯物主义的思想指导下，借鉴苏美各国学业成绩考评方法的最近发展趋势，吸取其中

某些已被验证的理论、方法，结合我国社会主义教育实际，在各种教学实践中不断摸索成绩考评的新经验，其重大意义也是显而易见的。

目　　录

第五章　考试的标准化

第六章　考试结果的统计检验与预测

第七章　分析考试结果的计算机程序

此为试读，需要完整PDF请访问：www.ertongbook.com