

心理与教育的 测量和评价

(上)

[美] R. L. 桑代克 E. P. 哈根 合著
叶佩华 邹有华 刘蔚成 主译



人
民
教
育



外国教育丛书

心理与教育的测量和评价

上册

[美] R. L. 桑代克 E. P. 哈根 合著
叶佩华 邹有华 刘蔚成 主译

人民教育出版社

**MEASUREMENT AND EVALUATION
IN PSYCHOLOGY AND EDUCATION**

FOURTH EDITION

ROBERT L. THORNDIKE

Richard March Hoe professor
of psychology and education
Teachers College
Columbia University

ELIZABETH P. HAGEN

professor of psychology and education
Teachers College
Columbia University

copyright © 1955, 1961, 1969, 1977, by John Wiley & Sons, inc

外国教育丛书

心理与教育的测量和评价

上册

[美] R. L. 桑代克 合著
E. P. 哈根

叶佩华 邹有华 刘蔚成 主译

*

人民教育出版社出版

新华书店北京发行所发行

北京顺义永利印刷厂印装

*

开本 850×1168 1/32 印张 11 字数 263,000

1985年10月第1版 1987年12月第1次印刷

印数 1—6 200

ISBN 7-107-10169-2/G·1049

书号 7012·01000 定价 2.25 元

译者的话

教育测量在教学、教育过程中的地位和作用是不容抹杀的。因为实施教学、教育当然要有目的、内容和方法，但又必须考查其效果。只顾耕耘不问收获，决不能提高产量。现代教育理论证明教学效果的反馈和强化，使教学从“开放系统”中构成一个“闭合回路”，正是提高教学质量的关键所在。思想品德教育也要根据学生的思想品德在其语言行动上的表现，采取有针对性的措施，反复进行教育，才能实现目的。学校中考查学生的学习成绩，社会上通过考试来选拔人才，可以说是古今中外普遍存在的现象，问题只在于人们是否把教育测量作为一门科学来看待。

教育测量学在本世纪初，即以新兴科学的姿态出现于欧美，传播于全球，教育发达国家的师范院校都把教育测量与统计定为必修课程。我国早在本世纪十年代就有教育测验法的输入，二十年代又有美国专家来华讲学；国内的教授专家也投入教育测量学的研究，并编造(或修订)出智力测验和教育测验几十种；中高等师范学校普遍开设了教育测验与统计课程，培养了一批人才；测验的理论和实践对改革旧的考试和评分制度也起了良好的作用，中小学师生也很感兴趣。可是，近三十多年教育测验与统计这个学科的教学和研究都停顿了。1982年，国务院和教育部把教育测量、教育统计等，归入教育科学研究法，列为培养教育科学研究生的专业之一。近来各师范院校也积极创造条件，恢复和发展这些学科，教育界也有愈来愈多的同志重整旗鼓，研究起教育测量和统计问题。

我们认为要恢复、发展我国的教育测量学，非先了解国际上的

最新成就和发展方向,则不能跻入时代的行列。鉴于美国教育测量学的著述丰富,我们选定哥伦比亚大学桑代克和哈根两教授合著的《心理与教育的测量和评价》一书1977年的第四版,组织力量,利用业余时间赶译,历尽艰辛,终于脱稿。

此书著者在序言中指出:他们的著述是以教育测量过去70年奠定的基础与70年代以来的最新成就相融合。我们认为著者作出这样的处理是正确的,因为科学发展都有它自身的脉络,掌握这个脉络对寻求它继续发展的突破点或转折点就大有好处。因此,我们也采取温故知新的态度,就教育测量学的基本理论、最新成就和发展方向等方面作精简扼要的论述,而以评介此书为主,以供理论和实际工作者参考。

(一) 教育测量学,简单地说,就是研究 心理与教育的测量和评价的科学。

教育测量学在本世纪初的兴起,既有实际的需要,如低能与犯罪的关系,心理诊断与特殊教育的关系等;又有理论的兴趣,如研究实验心理学和优生学等。著名的先驱者有法国研究智力测验的比纳和西蒙,美国研究教育测验的莱斯和E.L.桑代克等。

心理测量与教育测量的关系,可以说是两者并列,也可以说是两者部分重叠。在高等师范学校中,可作两门课来开设,但教育测量学的书籍常把心理测量包括进去,而教育测量所测的也不外是学生在教学教育的影响下的心理活动,所以心理测量与教育测量是不能绝对分开的。

人们较易理解的测量多是物理测量。如生活上的度、量、衡都是司空见惯的;自然科学上的各种测试,由于测量技术的进步和测量工具的改善,可以进行直接的测量,并获得比较准确的结果。把心理测量与物理测量相比,可以明显地看到心理测量的对象不是

容易控制的物,而是变动不居的人,而且要测的还是人脑、感官及神经系统的活动,生理心理学者也很难用某些测量工具放在人脑上进行直接的测量。至今,心理测量都是用一群标准刺激去激发有机体的反应。从反应的质量去窥测人脑活动的机密。这样间接的心理测量的准确性自不及直接的物理测量。但美国心理与教育测量学家桑代克和麦柯尔早就提出一个假说:“凡是存在必有数量,既有数量即可测量。”事物的质也可以转化为量来计算。心理测量必须在技术和工具上狠下功夫,将来的发展就可以与物理测量比美。《自然杂志》1980年第8期报导:最近美国发明了脑电机,可以纪录下人们说出或想到每个词的波型,不同的波型就成为相应的词的译码,因而可以揭开人脑的机密。《人工智能学》也在研究人脑与电脑的关系,这些都能增进人们研究心理测量的信心和勇气。

心理与教育的测量和评价,顾名思义,显然需要测量的工具——测验,和评价的工具——进行评价时应当参照的标准。测验就是一靠标准刺激去激发有机体的反应,对反应的质量要用数量来表示。但教育工作总是有目的和要求的,因此必须把测验的结果放在依据教育目的要求而建立的参照标准(教学大纲、集体的成绩水平和个人的学习潜力等)上来比较,才能评定其高低或优劣,并作出教育的处理。所以这两种工具都是必不可少的。所谓量表一般即指测验和参照标准的总称。

(二) 测量工具的理论 and 类型的发展

测量工具决定着测量结果的精确度,测量技术的改革也要物化于测量工具之中。心理与教育测量的工具都叫测验(卷子、册子)。测验理论上的研究,几十年来的主题称《优良测验的必备条件》或《测量方法的质量要求》,基本内容如下:

1. 效度： 它是指测验要能测出所要测的东西；或者说，测验能提供与决策有关的信息的程度。新近研究测验的题目能否与教学的目的和内容相适应，称“内容效度”；研究测验所得分数与所代表的人的品质的有效程度，称“结构效度”；求这次测验所得分数与标准测验所得分数的相关，并用来判定这次测验的效度，称“标准相关效度”。标准测验在这里就称为“效标”。

2. 信度： 它主要指测验施行的准确度和评分的准确度。信度的指标是表示特定测验先后结果的一致性或可以再现的程度。因此求测验信度的基本方法是：求一个测验先后举行两次所得分数的相关；求两个质量相等的测验（即复份）所得分数的相关；把一个测验分为等质量的两半，求这对分的两半测验所得分数的相关。加长测验可以提高信度。信度与效度是交迭的，信度低，效度不可能高，效度低，信度不一定低。

3. 鉴别力（区分度）： 它指测验能鉴别被试者高低、优劣的能力。这可体现在三种测验上：难度测验是用来鉴别被试者在试题按由易到难编排的测验上能够答对多难的试题。试题的难度是在编造测验的预测过程中，统计答对每题的人数百分比来决定的。广度测验是用来鉴别被试者在同等难度的平面上，能够答对多广的试题。速度测验是用来鉴别被试者在规定的时间内能够完成多少工作，或完成规定工作需要多少时间。

4. 复份（或等效测验、交替测验）： 它是指一种测验宜有质量相等的若干份，借以避免练习和泄密的影响，来提高测验的信度，编造复份的方法是用大量试题进行预测，求每题答对人数的百分比，选出难度距离相等的题目，用往复配搭法，编成质量相等的若干份。现代化的新办法是建立题库，储存大量经过筛选符合需要的题目。以后就可以随时按照需要选用一定的题目。

5. 常模： 它原是优良测验的必备条件之一，用作评价测验

结果的参照标准。但因现代教育测量，用作参照标准的，除常模外，还有其他，故把它归入评价工具的标题下并论。

6. 测验施行与校阅的便利和经济：包括详尽地规定测验施行与校阅的手续，节省测验的时间和费用等。

上述关于测验的理论是用来指导各种测验的编造的。现代测验的类型愈来愈多，那是与心理和教育测量在教育过程中地位的提高和作用的扩大互为因果的。

心理测量包括历史悠久的智力测验，近来发展迅速的学习能力倾向测验，两者性质相似，都起预测作用。此外尚有抽象的认知能力测验、特殊能力测验、行为测验与品格测验。这些都很重要，但也很困难。近来运用投射技术、情境技术以及行为观测，别人评价和自我陈述的方法进行，有了新的发展。

教育测量，新近的发展有两个特征：一是从局部的分散的测量发展到全面的系统的测量，如由单科测验发展到成套测验；还有用调查性测验先了解基本情况，然后对成绩优秀者进行预测测验和发展测验；对学习困难者进行诊断测验，并采取补救措施。二是由小学、幼儿园测验发展到中学、大学的各种重要测验。如小学的阅读和数学的基本能力测验。中学的调查性成套测验，连贯性测验（如中学数学各分支的连贯），学习进步测验。大学的重点学科和课程重点的单科测验，文化推理智力测验，成年人智力量表，能力倾向分类测验，教育技能测验（大学新生的安排测验），大中学教育程度测验，职业兴趣调查等。此外还有用来承认通过军队等途径而取得大学同等学力的“大学水平考试大纲。”

上述心理与教育的两大类测验都是从测验的目的来分类的。若以测验的编造和适用的范围来分，有全国标准化的测验，地方编造的测验，学校或教师编造的测验；用于全州和用于少数民族的测验。若以测验的方法分，有团体测验和个人测验。以测验的材料

分,有文字测验和非文字测验。

(三) 评价工具的理论 and 类型的发展

评价是把测验的结果放在一定的参照标准上来评定其高低、优劣。评价工具就是按照参照标准而编造出来的。这些工具与物理测量的量具同理,也必须具备下列要素:

1. 参照点(或组、系): 这是计算的起点。物理测量的量具常用绝对零点(如各种度、量、衡器具上的零点)或人定零点(如经纬度以格林威治天文台、赤道为零点)。心理与教育测量的评价工具的参照点随其所参照的标准而变化。

2. 单位: 这是计算和读数所必需的。物理测量的单位要求距离相等、大小适宜。心理与教育测量必需有相应的单位。

心理与教育测量的评价工具所参照的标准,经历史发展,有如下三种:

1. 目标参照标准: 这是以教学的目标和内容(教学大纲、教材)为标准,来衡量测验结果达成教学目标的“完满程度”,实质上是衡量测验所测出的学生掌握知识的质量。用这个标准是很自然的,因而也是早已有之,因近来批评长期以来多以常模为标准的情况下,又强调了它的重要性。这个标准的参照点应当是最低年级的教材;但就单位说,教材的章节或单元的质量距离并不相等;教材质量是否确指知识内容和思维品质,各科也不相同。所以讲究了效度的测验结果,放在评价标准上来衡量就不一定有高的准确性,值得研究的问题还不少。

2. 常模参照标准: 这是以同类集体在某种测验的预测上所得的成绩分布为标准,来衡量以后个人在这测验上所得的成绩在这集体成绩分布中所占的地位。同类集体可以是全国的、地方的或一个学校的。心理与教育测量常用的常模有四种,这里只作简

单评介:

(1) 年龄常模 这是许多被试者的年龄与他们在某种测验中所得的平均成绩的对照。以后个人在测验上所得的成绩便可以从这“对照”中查得其相应的成绩,评定其相当的年龄。这种常模最先用于智力测验,可以评定被试者的智力年龄。用实足年龄去除智力年龄,再乘以 100,可得智力商数。这种常模的参照点是人的出生日,单位是岁和月。因为人的身心发展不是按年龄等量匀速地前进,所以它的单位是不相等的。

(2) 年级常模 这是许多被试者的年级与他们在某种测验中所得的平均成绩的对照。以后的个人成绩也可以从这“对照”中评定其相当的年级。这种常模多用于教育测验,其参照点是入学日,单位是年级和学月。由于学生在各年级中学习的进展不尽相同,它的单位也不是相等的。

(3) 百分位数常模 这是不受年龄或年级的限制,而把一个集体在某种测验中所得的分数,划分为若干组距(区间),计算各组距的频率及其累计数,并求出这集体总人数百分之一的数值。这样,就可以作出各个分数和它们在总人数中各占的百分地位的对照。以后,个人分数就可以从这“对照”中评定它在集体中应占的百分地位,称百分位数或百分等级。这种常模用处较广,其参照点是零,单位是百分之一。因为它的单位和所包含的人数百分比,还不符合常态分配(或正态分布)原理,所以它的单位还是不相等的。

(4) 标准分数常模 这是把一个集体在某种测验中所得的分数,划分为若干组距(区间),各计其频率,求出算术平均数和标准差,然后以平均数为零点,以标准差为单位,向左右各展开 5 个标准差,算出各单位在常态分配曲线中各占的面积(人数)。这个标准差单位与常态曲线面积的对照表,在教育统计学书籍中都已列出。实际使用这个常模来评定个人分数在集体中的地位时,可用

下列公式计算：

$$Z = \left(\frac{X - \bar{X}}{\sigma} \right)$$

Z = 标准分数, X = 个人分数, $\bar{X}$ = 集体分数的平均数, σ = 集体分数的标准差。教育统计学上所说的 T 分数, 就是 Z 分数变形和改进:

$T = 10 \left(\frac{X - \bar{X}}{\sigma} \right) + 50$ 。公式中, “乘10”表示以一个标准分数的 $\frac{1}{10}$ 为 T 分数的1分; 公式中“+50”, 为的是避免用负数表示 T 分数。如上所述, 标准分数常模的参照点和单位, 都是出自常态分配原理, 所以它在四种常模中是比较科学的、用途最广的常模。

3. 潜力参照标准: 这是晚近提出的参照标准, 是要以被试者的全部潜力为标准, 来衡量被试者在这次测验中发挥了其潜力的程度。在开发智力的世界浪潮中, 这个标准引起了广泛的重视。现有的智力测验, 学习能力倾向测验, 特殊才能测验, 都是朝这个方向用力, 现在也可以求出各种成绩测验与智力测验这两种结果的关系。但是应怎样测出人的潜力的总量? 能够做什么是否已经尽了他的最大潜力? 为什么常见的聪明伶俐的人却不一定是成绩优秀者? 所以这个参照标准意义重大, 很有前途, 但亟需研究解决的问题也不少。

上述三种参照标准是从不同的角度来建立的, 应用起来还可以发挥相辅相成的作用, 所以不应把它们看成是矛盾对立的或相互取代的。

(四) 心理与教育测量和教学与教育决策

心理与教育测量的地位和作用在历史上是不断提高和扩大的。起初的意图在于甄别儿童的智愚, 考查学生学习的成績。随后又感到不应满足于获得测验的结果, 而应进而分析学生学习成败的原因, 作为改进教学的依据, 信息论所说: 信息的传递与反馈,

程序教学所说:教学效果的反馈和强化,都是很有道理的。

当今的世界是充满着决策的世界,而教育测量学又有迅速的发展,如果学校能对学生的心理和教育进行全面系统的测量,那所得的结果必可成为作出教学、教育决策的依据。这一方面表现为,根据客观情况,采取教学教育措施。例如,全面掌握关于学生的信息,即可对学生日常的学习、工作、生活以及以后的升学就业作出适当的指导。信息愈全面愈准确愈好,但起决定作用的还是决策者的智慧,决策还要随客观情况的改变而改变。另一方面又表现为:根据客观需要,作出适当决策。包括:1. 根据学生的智能,作出安排的决策,如编班、分组、开特殊课,课后作A、B处理等。2. 根据学生已有的知识技能,作出日常的教学决策,如要不要再教“质数”、“借位减法”,怎样编排学生的学习程序的进度? 3. 诊断学生学习的困难,作出补救的决策。如分析阅读能力的基本因素是:阅读的速度;把视觉符号转换为声音的技巧;从原文获得直接的书面意义;作超出直接说明事物的范围之外的推论;对作者的倾向性或观点的了解。根据这些因素,编造诊断测验,通过测验即可找到学生阅读的障碍在哪里,然后进行有针对性的教学,必能见效。4. 根据单元考试,期中考试和期末的综合考试,作出评价学生的决策。5. 根据教学效果,评价教材教法,作出改革课程的决策。6. 根据学生中学毕业成绩、高等学校入学考试成绩和招工的测试成绩,分别作出录取新生、招收和安排职工的决策。

按照惯例,全国性的标准化测验通常是作出重大决策的根据。如教育方针、课程改革、升学就业、辅导指导和选拔人才等。地方编造的测验通常也是作出日常决策的根据,如诊断补救、学生鉴定等。

(五) 在心理与教育测量上,要正确处理的社会和政治问题

教育测量学在其发展过程中,时常遭到某些人的批评和非议。至今存在比较突出的社会和政治问题有二,学者专家们提出了正确处理意见:

1. 对私事侵犯问题: 测验都在吸取有关人们的信息,而受试者还不知道信息正在被吸取,也不知这些信息将作何用;如果测验测及有损受试者的私人秘密时,他必不愿提供信息;这些都会发生受试者与主试者的矛盾。教育测量学家认为关键在于测验的目的,如果目的是为了社会性的事业,那受试者就应该提供帮助,测验的主持人也应该在社会利益和个人利益之间,作出公平合理的调整。

2. 对少数民族的测验问题: 社会上各种社会群体的文化背景、教育水平、语言差别和学习需要等,与测验应有的内容和方法都有密切的关系。只用一种测验去测不同的社会群体,从而评价学校和学生,不免脱离实际,失去公平;如果用来预测他们的发展前途则争论更大。所以必须弄清对非典型背景的人,作什么样的测量和预测,才能作出不偏不倚的公平合理的决策和推论。

(六) 教育测量学的发展方向

作为《心理与教育的测量和评价》一书的著者之一的桑代克还和其他 22 位专家合著《教育测量》一书,他写了该书的第一章《七十年代的教育测量》。我们从这两书的概要,可以估计出教育测量学今后继续发展的方向主要的是:

1. 面对整个教学系统,进行大量表的配套测验,从而作出教学决策。

2. 大量表所需的大量试题,将在大量试题的预测中,由电子计算机予以选择,并建立题库予以保存。

3. 测量工具和评价工具的继续研究,要加强知识、技能、行为品格等方面的测量和评价的研究。

4. 提高测量与评价的社会、政治效果,防止可能产生的危害。

总之,我们深感此书的编著是实现了固有基础与最新成就相结合的计划,是当代教育测量学的一部代表作。现已译成,出版之后,可以填补我国三十多年来教育测量学方面著译的空白;可以作为教育学界研究教育测量问题,编写我国教育测量专著的借鉴;可以作为中、高等师范学校恢复教育测量课程时师生的基本参考书,可以作为各级学校教师和教育行政干部的教育科学读物。

至于要建立我国的教育测量学,还有大量工作要做:1. 根据我国的具体情况和实际需要,编著我国的教育测量学。2. 迅速引进外国心理与教育的测量工具和评价工具,修订或自编我国急需的各种量表。3. 师范院校要迅速恢复教育测量与统计课程,师生都要学会使用电子计算机和程序编制;同时要尽可能开办教育测量与统计讲习班,向广大教师和干部推广这些知识技能。4. 教育行政部门与教育科学研究机构结合,组织力量去研究各级学校学生知识智力、思想品德的测量和评价问题,改革现行的高中和大学的入学考试制度,以及评定职工、提升干部制度。5. 应当建立全国性的教育测量、教育统计学会,引起各界重视,并依靠群众力量,推动学术发展。6. 当前国内此项人才十分缺乏,所以选派优秀中青年出国留学也是必要的。

《心理与教育的测量和评价》一书,涉及的学科很广,用到数理和统计的知识也很多,译它,要靠普通英语与专业英语相结合。我们为了保证翻译质量,分译、校、统三个步骤进行。承我校外语系主任麻贵宾、总支书记左钜生的支持,请了五位教授、副教授、讲师、助教、

同我们校阅、修正,最后由我们三人总校、统编。为了尊重劳动,表明职责,这里把参加译、校、统工作的同志表列于下:

章次	译 者	校阅修正者	总校和统稿者
序	叶佩华		
1	邹有华	叶佩华	刘蔚成
2	吴炯平	梁锦祥	邹有华
3	曾桂兴、张敏强	罗万贤、杨立勇	叶佩华
4	曾桂兴	梁锦祥	叶佩华
5	周新泉	曾美倩	刘蔚成
6	宾雨珠	郭应阳	邹有华
7	万晋卿	刘蔚成	叶佩华
8	万晋卿	刘蔚成	叶佩华
9	刘志棠	梁锦祥	邹有华
10	周新泉	曾美倩	叶佩华
11	高凌飏	刘蔚成	邹有华
12	高凌飏	郭应阳、黎柱辉	叶佩华
13	黎柱辉	邹有华	刘蔚成
14	黎柱辉	郭应阳	刘蔚成
15	张敏强	刘蔚成	邹有华
16	邢最智	郭应阳	刘蔚成

本书专业词汇英汉对照表,由各章译者提出,经黎柱辉同志校订,再由三位主译者审定。

此外,又承我校地理系绘图室潘国璠、马蓉、罗敏三同志绘制了插图 6b 幅,编制图目一份,在译校过程中,曾桂兴同志始终担任了组织联络工作,他还同李志好同志编制全书的章节目录、表目,并整理全稿。对他们的辛勤劳动,我们均表深切的感谢!

最后,由我们三人协商,由邹有华执笔写成《译者的话》,借以概括全书,引导阅读,刊在卷首,献给读者。

此书承人民教育出版社怀着发展我国教育科学和教育事业的宏旨,热情支持,接受出版,还校正了我们译述中的部分疏忽和错漏,我们更表由衷的感谢!

此书质量较重,依靠集体力量赶译,在翻译过程中,虽已尽力克服遇到的困难,但荒废已久的学科,不免仍有错漏不周之处,深望读者、行家,惠予指正,以利修改!

叶佩华 邹有华 刘蔚成

1984年7月于华南师范大学

序 言

22年前，本书第一版发行时，美国公众和美国教育界因俄国空间开拓的影响而掀起了一个争取优势的热潮。从此，大批人造地球卫星遨游太空，阵阵改革之风吹遍了教育界及教育心理测量界。在社会内部，要求平等待遇、平等权利及平等就业的压力，使人们对许可加入会社、就业及许可开业等测验的适宜性产生了怀疑。在学校中，恢复个别教学的建议使人们认识到，用相同的组来比较学生并不一定能提供人们作出教育决策所需要的信息。然而，对于以客观的、可靠的信息为依据来作出决策的需要仍旧没有改变。测验与测量程序就是这种信息的一个重要来源。

在本版(第四版)中，我们估价了人们最近强调指出的关于“公平测验”、“个人与标准参照测验的真相”的重要性，从新的资料中吸取了其中有用的东西而不是整个地吞下去，并使之与过去七十年所形成的教育与心理测量的坚实基础相融合。我们的目标在于提出一种既能适应七十年代事物发展变化，又不致丧失过去的坚实基础的“公平处理办法”。

在这一版里，我们把那种有助于制定决策的测验与测量作为中心题目来看待。我们指出了各式各样的教育决策和个人决策，并且提出了测验与测量怎样为每一项决策作出贡献的问题。我们认为这是事情的实质。测验与测量之所以有用，是由于它允许而且被用来作出较好的决策。若经常记住这一目的，那末，用以达到这一目的的手段——对测验的合理评价看来是可能的。