



0011984

BIAO ZHUN HUA

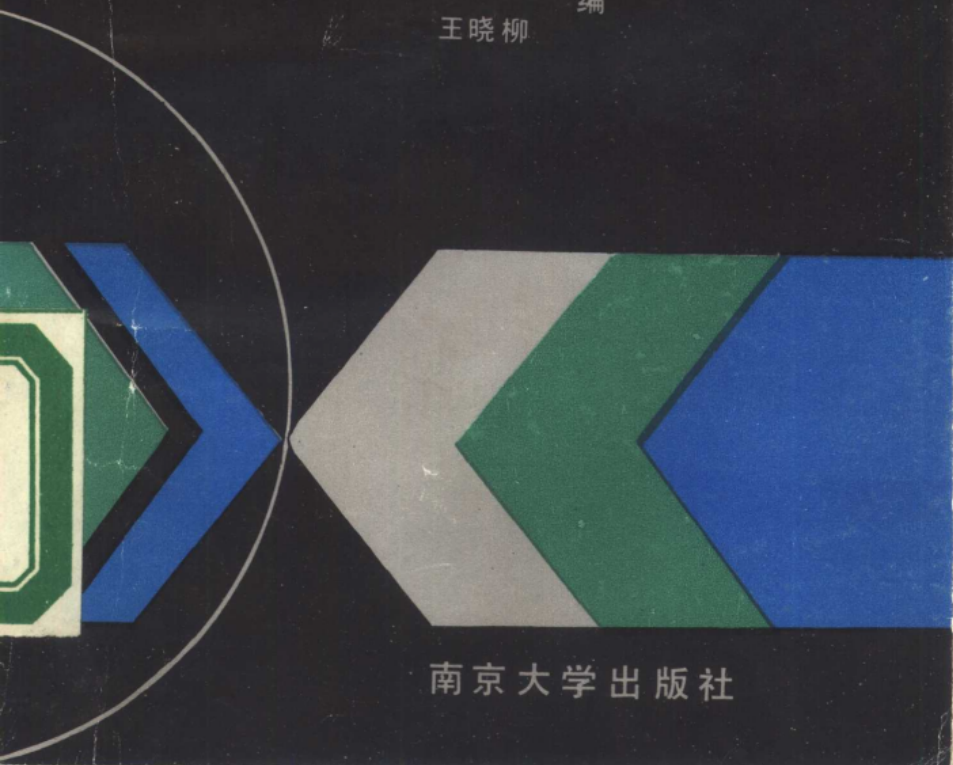
标准化 考试

KAO SHI

李绍珠

编

王晓柳



南京大学出版社



2 033 1906 0

(mk-28/6)

标 准 化 考 试

李绍珠 王晓柳 编



南 京 大 学 出 版 社

· 1988 · 南 京

责任编辑：李曾沛

标 准 化 考 试

李绍珠 王晓梅 编

南京大学出版社出版

(南京大学校内)

江苏省新华书店发行 栏江印刷厂印刷

1988年5月第1版 1988年5月第1次印刷

开本：787×1092 1/32 字数：114千

印张：5.375 印数：15000

ISBN 7-305-00241-0

G·27 定价：1.50元

前 言

近年来，标准化考试愈益受到人们普遍重视，这是因为标准化考试根据近代教育测量的理论，将知识和能力的测量建立在更加科学的基础上。国家教委已决定在我国高等学校招生考试中逐步实行标准化考试。

本书着重介绍考试为什么要标准化、什么是标准化考试和怎样实现考试标准化，力求通俗易懂、深入浅出，希望对广大教师、教育行政干部、各类考试的命题人员和组织者以及广大考生了解标准化考试有所帮助。

本书的问世，自始至终得到江苏省高等学校招生工作委员会办公室的支持和帮助。其中有关内容曾在省招办1988年元月举办的两期《标准化考试讲习班》上讲授。省招办万耀祖副主任和刘炳贵同志审阅了书稿，提出了不少宝贵意见，谨表示衷心的感谢。

本书第一至第五章由李绍珠编写，第六至第八章由王晓柳编写，全书由李绍珠定稿。

由于编者理论水平和实践经验所限，本书谬误之处，在所难免，敬希广大读者批评指正。

编者 1988年2月

目 录

前 言

第一章 什么是标准化考试.....	1
§ 1-1 考试为什么要标准化.....	2
§ 1-2 标准化考试的特点.....	9
§ 1-3 标准化考试的种类.....	12
§ 1-4 消除对标准化考试的疑虑.....	16
第二章 教育统计学基本知识.....	18
§ 2-1 平均数和标准差.....	18
§ 2-2 正态分布.....	21
§ 2-3 相关系数.....	26
§ 2-4 线性回归.....	33
§ 2-5 测量量表.....	37
第三章 试卷编制的标准化 (I).....	40
§ 3-1 确定考试目的和测量目标.....	40
§ 3-2 设计试卷蓝图.....	44
§ 3-3 制定考试大纲.....	47
§ 3-4 试题编写的原则.....	48
§ 3-5 主观性试题.....	49
§ 3-6 客观性试题.....	52
第四章 试卷编制的标准化 (II).....	58
§ 4-1 试题的试测.....	58
§ 4-2 试题分析.....	59
§ 4-3 试卷的拼配.....	76
§ 4-4 试卷的检验与修订.....	80
§ 4-5 试卷的版面和印刷.....	81

第五章 考试实施的标准化	84
§ 5-1 考试实施标准化的意义	84
§ 5-2 考场的组织	84
§ 5-3 实施的指导语和时限	86
§ 5-4 考生答题的注意事项	87
第六章 评分的标准化	89
§ 6-1 评分标准化的意	89
§ 6-2 主观性试题的评分	90
§ 6-3 主观性试题评分的监测与调整	94
§ 6-4 客观性试题的评分	100
§ 6-5 对客观性试题预测的校正	103
第七章 考分解释的标准化	107
§ 7-1 考分解释标准化的意义	107
§ 7-2 目标与常模	109
§ 7-3 标准化样本	110
§ 7-4 百分位量表	112
§ 7-5 标准分量表	118
§ 7-6 考分的合成	124
§ 7-7 考试之间分数的转换与调整	126
第八章 考试的评鉴	135
§ 8-1 考试评鉴依据的假设	135
§ 8-2 效度	137
§ 8-3 信度	148
附表1 正态分布的密度函数表	157
附表2 正态分布表	158
附表3 检验相关系数显著性的临界值	160
附表4 检验等级相关系数显著性的临界值	161
附录 1988年高等学校入学英语标准化考试说明	162

第一章 什么是标准化考试

标准化考试在国际上已广为流行。目前，我国高等学校招生考试也开始逐步采用这种考试方法，这是高考制度改革的一项重要措施。有关部门已在部分地区对个别学科的考试开展试验工作，并取得了成功的经验。

众所周知，高考是目前我国选拔人才的一项重要制度。无疑，有关高考改革的任何一项措施必将受到人们的关注。

然而，对于什么是标准化考试，人们却有不同的理解，有些甚至是误解。例如有人以为标准化考试就是采用选择题进行考试，这显然是很片面的。有的还存在各种疑虑，怀疑是否有必要实行标准化考试，或者担心是否会影响考生水平的发挥，或者担心自己能否适应这一考试等。因此，考试为什么要标准化，什么是标准化考试，怎样才能使考试标准化以及如何评价标准化考试等这样一些与考生利益密切相关的问题，自然受到主考机关、管理人员、命题人员、广大教师、考生乃至考生家长的普遍关心。加强有关宣传是十分必要的。

本章将着重讨论前两个问题，即考试为什么要标准化和什么是标准化考试。后两个问题将在以后几章中加以讨论。

§ 1-1 考试为什么要标准化

我们相信，任何阅读本书的读者都曾参加过各种各样的考试。比如，当您在校学习期间，每读一门课程都要参加一次或多次考试；在您毕业时，要参加毕业考试；在您升学时，要参加升学考试；甚至在您参加工作或晋升时，还要参加某种资格考试。也许您还组织实施过对他人的考试。尽管“考试”在现代社会业已司空见惯，然而，怎样使考试特别像高考这样的重大考试能够更加科学，更加合理，是要做很多细致工作的。

在教育上，考试属于“教育测量”的范畴。为了说明考试为什么要标准化，我们先从教育测量谈起。

一、教育测量

教育测量与物理测量既有相似之处，又有很大质的差别。先举一个简单的物理测量的例子。假如你想知道一段布的长度，可用米尺进行丈量。步骤是将布的一端对准尺的零刻度处，读出布的另一端与尺相对应的那一点的数字，便可知道布的长度是多少厘米。若布的末端在尺的85厘米处，那么数字85和单位厘米合在一起便是对这段布的长度的一种量度。从这一测量的步骤可以看出，测量就是按照一定的法则对事物的属性加以定量的描述。按照史蒂芬斯的说法，测量就是“按照法则给事物指派数字”。因此，任何一种测量都包含三个要素：

一是事物的属性，例如，布的长度，

二是法则，包括工具和按一定的程序进行的施测过程，

例如，尺子和丈量；

三是数字和单位，即测量的结果，测得的数字必须与某种单位联系起来才有意义。

“教育测量”是对受教育者接受某种教育后，在知识的掌握、能力的提高、性格的培养等方面进行的测量。简单地说，就是对作为教育效果的学生行为进行测量。

在作物理测量时，所用的手段是工具或仪器（如米尺、天平等）；进行教育测量时，所用的手段是测验或是考试。尽管教育测量和物理测量的对象不同，手段也不相同，但两种测量所遵循的原则和程序是一样的。

例如，若想测量学生的阅读能力，可用下述测验来进行：要求学生在一定的时间内阅读一个片段，然后数出他阅读的字数，再除以阅读时间，就可以得出评价学生阅读能力——阅读速度的一种量度——每秒阅读多少字（字/秒）。可以看出，这种测量的步骤与长度的测量是很相似的。它也包含三个要素，阅读速度（事物的属性），按照一定的要求和程序实施测验（法则）和测验结果——每秒多少字，就是数字和单位。这是与物理测量一致之处。

二、测量的标准化

实践证明，任何测量都不可能绝对的准确，即或多或少地存在不同程度的误差。例如在丈量布的时候，所使用的尺是否准确，布的松紧是否一致，起点是否对准，终点读数是否准确，在丈量较长的布时，第1米的终点是否恰好是第2米的起点，甚至在读数时有没有读错（多读或少读）等。所有这些都会引起误差。

为了使测量的结果有意义，并将测量误差尽可能减少到

最低限度，必须解决好下面四个问题：

(1) 要有一个统一的单位。也就是说，测量没有单位不行，单位不统一也不行。现在为了便于国际交流，国家已颁布了以国际单位制为依据的计量法。例如，长度单位以米代替(市)尺，重量以千克代替(市)斤等。这些就是计量单位标准化的问题。

(2) 要有一个参照点，即零点。没有一个参照点，任何测量都是无法进行的，有了参照点，但不统一，测量结果也是无法进行比较的。例如测量布以布的一端为零点(绝对零点)，测量地面高度以海平面为零点(相对零点)，测量身高则以站立地面为零点(也是相对零点)，这就是参照点标准化的问题。

(3) 量具要标准。就是要求按一定的规程进行生产，并且在使用过程中按期进行检定，这就是量具标准化的问题。

(4) 测量的各个环节都要按照一定的法则进行，这就是测量方法标准化的问题。

此外，操作要细心，必要时进行复核等也是减少误差的重要方面。

总之，测量的标准化问题，对测量的结果来说，意义是十分重大的。为此，各行各业的许多测量问题，常以“标准”的形式，如“国家标准”或“部颁标准”等在法律上予以确认。

三、教育测量的复杂性

1. 教育测量是一种间接测量

一般地说，在长度、时间、重量这样一些物理量的测量

中，对测量结果的描述是比较明确的。这是因为在测量时可用性能良好的测量工具，按照人们共同接受的方法和法则进行直接测量，测量的结果可用一定的数字和单位表示，测量结果的准确度较高。即使对于需要作间接测量的量，也是通过直接测量与该量有精确关系的其他量，再通过计算求得的。如欲测量电压，可通过测量出电阻和电流，通过电压 = 电阻 × 电流的公式进行计算。归根到底，这是直接测量。

然而在教育测量中，迄今人们既未找到用来量度知识、能力或性格的绝对的单位，也未能研制出一种可以直接测量它们的仪器。所幸的是这些欲测量的属性必定会在人们的具体活动或行为中有所表现。人们可以用下面的方法对这些行为进行测量，即用一组作业或刺激来引起被测者的行为反应，然后根据这种测量的结果来间接地表示和解释所要测量的心理特征或认识发展水平。这便是大家熟悉的“考试”。所用的一组作业或刺激就是考试所用的一套试题，被试者就是考生。要求他们作出的行为反应就是答题，测量的结果用考试成绩表示。例如，可用“每秒读的字数”量度学生阅读的熟练程度。

2. 不同考试的分数缺乏可比性

虽然，教育测量可以借助考试给要测的属性指派一个数字（分数），但是从考试得到的分数，其含义是不明确的，而且不同考试的分数也缺乏可比性。设某生语文期末考试得80分，对80分这个成绩，可作出不同的解释：既可能说明该生已经成功地完成学习任务，也可能说明他尚有知识缺陷需要补习，既可能表明该生成绩比其他同学优秀，也可能表示该生处于中等水平，甚至还可能表示该生在班上成绩较

差。若该生的数学成绩为85分，也不能认为该生的数学比语文得好，也许他的语文成绩名列前茅，而数学成绩却居于下游。即使同一个考生同一科目的两次考试，其成绩也难以比较，如期中语文成绩为75分，也很难说期末比期中有“进步”。因为两次考试的难度可能不同。

考试的分数之所以含义不明确和缺乏可比性，就是因为它们缺乏统一的参照点和单位。

除非刚出生的婴儿，一般人的知识和能力是没有绝对零点的，即使某考生的外语考试成绩为零分，也不能说他的外语水平为零，因他在另一次外语考试中可能得10分。可见，通常考试分数的零点（参照点）不仅是相对的，而且也不确定，是由试卷编制者根据需要通过调整试题的难易程度确定的，不同的考试分数的参照点是不同的。任何一种考试都不能测到知识和能力的绝对数量，只能是相对于某种教学目标（如小学水平、中学水平）或在团体中的地位（如名次）。

用于量度考试成绩的单位在不同的考试中也往往是不相同的。例如，在当前的高考中，语文和数学采用120分制，生物为70分制，其他各科为百分制。表面看来，各门考试的分数在名义上都是以1分为单位，但是其分值是不相等的。在70分制里相当于满分的 $1/70$ ，在百分制里相当于 $1/100$ ，而在120分制里相当于 $1/120$ ，显然前者大于后者。即使各门考试都采取百分制，由于难度不同，其分值也是不相等的。

由此可见，在教育测量中，不同考试的分数因其参照点不同，单位也不同，它们既无法进行比较，也不能简单地直接相加。就如同要比较3千克和5斤的两个物体谁轻谁重一样，只有把它们变成统一单位（千克），即3千克和2.5千克

后，才能比较其轻重，也才能将其重量相加得5.5千克。

3. 教育测量的客体是人

教育测量的客体是有思想、有情感、有个性的并且生活在现实社会中的人。在考试中，考生处于被测量的地位，特别是在高考这样重大的考试当中，考试结果又与考生的某种利益相联系，考生心理紧张在所难免。有人会感到不安，有的人又会满不在乎，有的人会异常努力。因此，临场时考生种种心理的、生理的和社会的非智力影响是不容忽视的。

正是由于教育测量具有上述特点，就使得教育测量比其他测量要复杂得多，影响测量结果的误差也得多。

四、考试的误差来源

考试的误差主要来自三个方面：试卷编制、考试实施和考生本身。

1. 来源于试卷编制的误差

(1) 在短短2—3小时的考试中，是不可能考核某门科目的全部内容的，只能抽取部分内容编制试题。因此，试题的取样是否适当是误差的重要来源。若试题取样不均，且数量太少，就无法代表和覆盖学科的全部内容，考生也许遇到他拿手的试题，也可能遇到的是未准备的试题。这样考试成绩就受机遇的影响，而不能真正测出考生的实际水平。

(2) 试题的题意不清，措词模棱两可，答题要求不明确等。

(3) 试题的难度过大，导致考生大量猜测，难度太小亦测不出考生的水平。

(4) 规定时限虽可以鼓励考生加速答题，但若时限太短会使考生仓促答题，而引起不稳定的考试结果。

(5) 同一试卷的几种“副卷”在编制的实践中很难达到“等值性”要求，亦可成为误差的来源。

2. 来源于考试实施的误差

(1) 考试环境的偶然变化，如太阳直射，四周嘈杂，通风或降温措施临时出问题，以及发生一些不能预料的意外干扰事件，如阴天的照明灯突然停电，计时表出现故障，临时发现试卷或答题卷的印刷错误，有人迟到或生病，主考人处理作弊事件等，都会对其他考生引起不安和骚乱。

(2) 大规模的考试往往是成千上万，乃至几十万人同时举行的，主持考试的人员很多。主考和监考人员能否严格地执行考试规则，是否严格掌握时限，对考生的态度是否和蔼可亲 and 耐心等也会影响考试成绩。

(3) 评分标准掌握不当，评分难于做到完全客观都会产生误差。另外，人工登分的差错和累错分数也会发生误差。

3. 来源于考生本身的误差

这是最难控制的误差。即使有了编制优良的试卷、标准化的考试实施和记分程序、理想的考试安排，由于考生本身在以下几方面的因素也会引起误差。

(1) 考试动机会影响考生答题的态度、反应的速度。如果某人和大多数考生的动机不同，就会引起测量误差。

(2) 临考前或考试中的竞技状态、焦虑心情（一种紧张的不愉快的情绪体验）的影响。若过分紧张，即所谓“上场昏”，就不能充分发挥水平，甚至连熟悉的内容也答不出。反之，满不在乎也会影响考试成绩。

(3) 考生的应试经验、答题技巧、对指导语的理解程度或细心与否都会产生误差。

(4)生病、疲劳等生理因素的影响。

五、考试必须标准化

综上所述,教育测量属于间接测量,它的测量客体是人,测量的目标是人的知识、能力、性格等,因此要比物理测量复杂得多。前面我们从单位、参照点、工具和法则四个方面论述了物理测量的标准化问题。教育测量误差的来源之多及难以控制的程度,是物理测量不能相比的。因此,在教育测量中,标准化的问题尤为重要,而且难度较大。作为教育测量手段的考试必须从各个方面解决好标准化的问题,以便提高测量的有效性和可靠性。

首先,作为测量工具的试卷在编制时必须采取一系列的标准化措施,如设计考试蓝图,制定并公布考试大纲,对试题进行试测和分析,建立题库等,这些都是为了控制作为测量工具的误差,使其准确可靠。

其次,照标准化程序实施考试,以减少因实施条件的差异而引起的误差,这便是实施法则标准化的问题。

最后,建立“常模”,将考试中直接得到的原始分数转换为标准分数,也就是为了解决教育测量中的参照点和单位标准化的问题,使考试结果能真正反映考生的实际水平,达到考试目的。

§ 1-2 标准化考试的特点

一、考试标准化的含义

考试是个系统过程。包括试卷编制,组织施测,阅卷评分,分数解释等环节。所谓标准化考试就是使考试的各个环

节严格按照一定要求组织的，因而能够严格控制误差的一种考试。

1. 试卷编制方面

毫无疑问，对所有考生都必须用内容相同的一份试卷施测。标准化考试的试卷由专家们根据考试大纲编制，以保证能够真正测出所欲测量的知识和能力。其中的每一道试题都要经过预测和统计分析，删去难度不当、鉴别力小的试题。在拼配试卷时，试题的难度和排列顺序要得当，以符合被试者的心理特点。此外，标准化考试还需预备若干套等值的备用试卷。

2. 考试施测方面

标准化考试必须制定详细的施测规则。严格地讲，若考生不是在相同条件下进行考试，他们的考试分数就不能进行比较。因此，在施测规则中对试卷的印刷，考场的设置，试卷的收发手续，监考人的态度，考生的注意事项，测验的指导语，时间限制，意外事件的处理等均应有明确的规定，尽量避免环境及各种其他误差因素对考试成绩的影响。

3. 阅卷评分方面

标准化考试必须实行客观评分，唯有客观评分，才能够把分数的差异归于考生的差异。标准化考试大多采用能客观记分的试题。如果在必须采用主观性试题时，则应订出详细的评分规则，使评分尽可能地准确客观。

4. 分数解释方面

为了使从考试直接得到的原始分数有意义，同时使不同考试的原始分数能够进行比较，标准化考试必须提供“常模”或“定标”等资料作为解释分数的依据。

综上所述，标准化考试不同于教师自编的课堂测验。具体地说，它是以考试大纲为依据，采用标准化的系统程序编制和实施，按一定规则评分，并提供常模等资料解释分数的一种考试。

二、标准化考试必须具备的条件

1. 有效性

考试的有效性是指能测出它所要测量的事物的程度，可用“效度”进行衡量。这是考试应该具备的最基本条件。教育测量属于间接测量，要使考试具有较高的效度并不是一件轻而易举的事。但是无论如何，标准化考试必须具有较高效度（例如，美国的某些标准化考试要求效度达到0.40以上），这是衡量考试质量的重要指标之一。

2. 可靠性

考试的可靠性是指它能够稳定地反映出考生的实际水平的程度，一般用“信度”来衡量。一次考试如果没有信度，就好比用橡皮尺测量长度一样，会得到不同的结果。一般说来，可靠性差的考试，也不可能具有较高的效度。因此，标准化考试必须具有较高的信度（例如，美国的某些标准化考试要求在0.90以上）。

3. 鉴别性

一种考试若能把水平高低不等的学生恰当地区别开来，便是一种鉴别力高的考试。难度是指试题的难易程度。难度不当，将直接影响鉴别力的提高，过难或过易的试题都不能区别学生的实际水平，同时也难获得较高

4. 实用性

考试的实用性是指考试的实施、评分、分差解释等方面