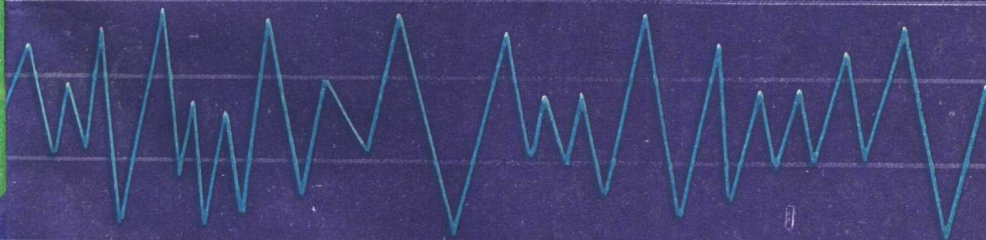


# 心理测量学讲义



华中师范大学出版社

---

---

# 心理测量学讲义

---

---

谢小庆 编著

华中师范大学出版社

1988·武汉

# 心理测量学讲义

心理测量学讲义

谢小庆 编著

\*

华中师范大学出版社出版

(武昌桂子山)

新华书店湖北发行所发行

湖北省通城县印刷厂印刷

\*

开本：850×1168 1/32 印张：10.75 字数：270 000

1988年7月第1版 1988年7月第1次印刷

印数：1—6 000

ISBN 7—5622—0123—4/B·04

定价：1.85元

本书系统地介绍了现代测验理论，阐述了心理测量的理论和方法，内容包括测验基本理论、题目分析、信度、效度、分数组合、分数解释、智力测验、人格测验和成就测验等。书中介绍了《教育管理与教育测量程序》的使用方法。各章配有自测试题。本书可以作为心理学、教育学、管理学、社会学、法学、医学等专业的本科生和研究生学习心理测量学的教材，亦可作为教育科学、社会科学、医学等与人的心理有关领域的科研工作者的参考书。

# 目 录

<b>第一章 绪论</b> .....	( 1 )
第一节 心理测量学的研究对象.....	( 1 )
第二节 学习和研究心理测量学的意义.....	( 4 )
第三节 心理测量的历史、现状和展望.....	( 12 )
<b>第二章 测验的一般理论</b> .....	( 24 )
第一节 测验的理论基础——特质理论.....	( 24 )
第二节 心理测量学研究的性质.....	( 31 )
第三节 测验定义.....	( 33 )
第四节 测验的编制.....	( 45 )
第一、二章自测试题.....	( 59 )
<b>第三章 题目分析</b> .....	( 64 )
第一节 难度.....	( 64 )
第二节 区分度.....	( 70 )
第三节 区分度与难度的关系.....	( 88 )
第四节 小样本资料的简单统计法.....	( 90 )
<b>第四章 信度</b> .....	( 95 )
第一节 真分数理论.....	( 95 )
第二节 信度定义.....	( 98 )
第三节 信度估计.....	( 101 )
第四节 因素分析.....	( 118 )
第五节 影响信度估计的因素.....	( 134 )
第六节 信度与测量误差.....	( 136 )
第三、四章自测试题.....	( 139 )
<b>第五章 效度</b> .....	( 144 )
第一节 效度概述.....	( 144 )

第二节	效标关联效度	(148)
第三节	内容效度	(169)
第四节	概念效度	(173)
<b>第六章</b>	<b>分数组合</b>	<b>(183)</b>
第一节	两个基本问题	(185)
第二节	组合分数的方法	(186)
第三节	各种组合分数方法的比较	(202)
第五、六章	自测试题	(206)
<b>第七章</b>	<b>测验分数的解释</b>	<b>(211)</b>
第一节	常模参考分数	(213)
第二节	结果参考分数	(226)
第三节	内容参考分数	(230)
第四节	分数解释中应该注意的若干问题	(231)
第五节	预测性测验分数的双重含义	(233)
<b>第八章</b>	<b>认知测验</b>	<b>(240)</b>
第一节	智力测验、能力倾向测验和成就测验	(240)
第二节	关于智力的一些理论问题	(243)
第三节	智力测验介绍	(252)
第四节	能力倾向测验	(265)
第五节	知识测验	(272)
第七、八章	自测试题	(276)
<b>第九章</b>	<b>人格测验</b>	<b>(281)</b>
第一节	人格测验的理论基础	(281)
第二节	自陈式测验	(284)
第三节	投射性测验	(295)
第九章	自测试题	(299)
<b>第十章</b>	<b>《教育管理与教育测量程序》简介</b>	<b>(301)</b>
第一节	概述	(301)
第二节	各模块介绍	(304)

附表.....	( 327 )
自测题答案.....	( 334 )
主要参考书目.....	( 336 )

# 第一章 绪 论

## 第一节 心理测量学的研究对象

心理测量学研究对心理属性的测量。心理测量学对心理属性进行定量化研究。

就象世上没有两片完全相同的树叶，世上也没有两个完全相同的人。人与人之间存在着差异，心理学中将这种人与人之间的差异称为个别差异。对个别差异的研究是心理学研究的一部分重要内容。人的个别差异不仅表现在高矮、胖瘦等生理方面，还表现在人的心理特征方面。有的人可以辨别出重量、声音、气味上的微小的差别，有的人却辨别不出；有的人可以辨认几十种不同的黑色，有的人却只能辨认几种；有的人诵读一遍就可以记住一首诗，有的人诵读许多遍也记不住；有的人喜欢热热闹闹，有的人喜欢一人独处；有的人的喜怒哀乐溢于言表，有的人却含而不露；有的人固执己见，有的人却惯于服从别人，等等。心理测量学致力于研究如何对人的这些个别差异进行合乎逻辑的测量。

就测量而言，心理测量与物理测量是一样的，都是根据一定的法则将某种物质属性表示成数字。然而，心理量的测量要比物理量的测量困难得多，它具有更大的间接性、多元性和随机性。对于重量、长度这些物理量可以直接用天平、尺子来测量，而对于智力、性格这些心理量则只能凭借测验来间接判断；对物理量的测量中可以尽量排除其它因素的干扰，仅仅测量所要测的因素，如重量、长度等。在心理量的测量中就很难做到这一点。在智力测量中，不可避免地要受到知识水平、教学条件、情绪、健康状况、主试人等多方面的影响；在物理量的测量中，可以得到相对



稳定的结果，在心理量的测量中则有很大随机性，某个学生在这次考试中得高分，在另一次类似的考试中得低分是常见的事情。

考察一个人的“阅读速度”，是一种比较简单的心理量的测量，但是并不容易，如果考察单位时间内朗读出的文字数量，就包含了口齿表达能力，并不一定反映一个人的真实阅读速度。如果是默读，又无法把握他实际阅读的数量。此外，在阅读速度的测量中，很难排除受测者的知识经验、内容材料的特点等影响。对于诸如言语能力、智力、性格等更复杂的心理特点的考察和测量，就更为困难。

“工欲善其事，必先利其器”。进行测量必须要有一定的测量工具。工具越好，测量的效果也越好。物理测量的工具是温度计、尺子、计时器等等。心理测量的主要工具就是测验。因此，心理测量学研究的主要内容就是测验的编制和使用。

测验总是由一组题目组成的，题目是构成测验的元素。选择好的题目是编制好的测验的前提。一个选拔性测验的目的是将不同水平的受测人的水平区分开，如果一道题目所有的人都得了满分或是所有的人都未能得分，那么，这道题就失去了区别受测人水平的效用，就不是一道好的题目。如果受测人在一道题上的得分与整个测验中的总分之间不协调，这道题上得高分的人的测验总分反而低。那么，这道题也不是一道好题目。什么样的题目是好题目？怎样选择好题目？这是我们将在第三章讨论的问题。

一项测验应该具有可靠性、客观性。一把尺子如果你用来量某物是那么长，我用来量也是那么长，那么，这把尺子就是可靠的。如果你量是一个长度，我量又是一个长度，那么，这把尺子就是不可靠的。测验也是同样，一个好的测验应该在不同人使用、不同时间施测之间具有一致性。今天，我们运用的许多测验的可靠性都较差。对同一份试卷，两位不同的评分人给出差异很大的评分，这是常有的事情，一个人在类似的两次测验中一次考得好、一次考得糟，也是常有的事情。这样的测验就是不可靠

的。我们将在第四章中讨论测验的可靠性，即测验信度问题。

可靠的测量不一定有效。在实际运用中许多测验的有效性不高，所谓“高分低能”就是测验有效性不高的集中表现。正是由于测验的有效性不高，才会出现“高分低能”，倘若测验的有效性较高，就只会是“高分高能”和“低分低能”。一个测验不仅应具有可靠性，还应具有有效性。我们将在第五章中讨论有效性，即效度问题。

一个测验往往包括许多分测验。例如，高等学校入学考试就包括语文、数学、物理等多个分测验。《韦克斯勒儿童智力量表》包括12个分测验。当一个测验包括多个分测验时，将分数怎样组合才算合理呢？是否应该平均加权？今天，我国的高等学校入学考试在根据总分录取时，是将各门功课的成绩平均加权的。在进行这种平均加权时，包含着这样一种假定：各门考试的成绩是等值的，即语文考试的10分与数学、历史、地理等考试的10分等值。然而，我们有什么理由认为语文的10分与地理的10分等值呢？根据广东省教育科学研究所对广东省高考考生成绩的一项研究，如果按照标准分数（即根据考生在全体考生中的位置将原始分数进行换算而得到的一个相对分数）对各门成绩进行组合，那么，一些原来总分在录取分数线以下的考生则应该被录取，而一些原来总分在录取分数线以上的考生则应该被淘汰。怎样对各个分测验的分数进行组合才算合理呢？这是我们将在第六章讨论的问题。

当我们经过测验而最终得到了测验分数的时候，并不是每个人都能正确地对这个分数做出解释，都能正确地理解和应用这些分数。例如，如果你在智商测验中得到了智商为110的分数，这个分数意味着什么呢？如果你的孩子或你的朋友在某次考试中得了90分，这个分数又意味着什么呢？他的这个分数既可能是全班最高分，也可能是全班最低分。在第七章中，我们将讨论分数的解释问题。

从不同的角度，可以对测验进行多种分类。从内容上，可以将测验分为认知测验和人格测验。认知测验主要考察受测者表现在认知方面的心理特点，包括知识水平和能力水平。认知测验又可以根据心理特点的稳定性划分为智力测验、能力倾向测验和知识测验。智力测验着重考察受测者较稳定的认知能力，对受测者的一般能力和较长期的发展潜力作出评价，可以被用来诊断受测者在智力上的缺陷。《斯坦福—比奈儿童智力量表》、《韦克斯勒成人智力量表》等即属此类。知识测验主要考察受测者表现在认知方面的较容易改变的特点，考察受测者的知识水平和受教育程度，着重对受测者的现状进行评定。学校里的教学考试，语文、数学等分科考试，即属于此类测验。能力倾向测验是介乎于智力测验和知识测验之间的一种认知能力测验，它既考察受测者较稳定的认知能力，也考察受测者的知识水平。高等学校入学考试、研究生考试等，即属此类。人格测验着重考察一个人的性格、气质、兴趣、爱好等方面较稳定的特点，《明尼苏达多相性格问卷》、《卡特尔16种人格特质问卷》等即属此类。在第八、九章中，我们将分别讨论认知测验和人格测验中的一些特殊问题，并介绍一些重要的测验。

最后，在第十章中将介绍《教育管理与教育测量程序》在电脑上的应用。

## 第二节 学习和研究心理测量学的意义

作为一种测量人的心理能力的工具，测验在现代社会生活中起着十分重要的作用。随着“任人唯贤”的风气越来越浓，随着教育中愈益重视“因材施教”，各类测验考试应运而生，被应用于各个领域，不仅学校通过考试来选拔学生和检查学生的学习情况，而且，工厂、机关、企业、商店等各条战线也开始通过各类考核来检查工作和选拔人员。考试不仅被用来作为选拔的依据和

标准，而且被用来促进各种学习和检查学习成果。考试发挥着什么样的作用呢？下面，我们分几个方面来讨论。

## 一、教育方面

笔者认为，当前学校教育中存在的一个重要问题就是学生的学习兴趣和内在动力被挫伤。学习和创造的动机有两种，一种是基于好奇心之上的对世界进行探索的兴趣乃至理想和信念，这是内在的动机，一种是对现实的功利的追求，这是外在动机。那种最有建设性的、最稳定的动机力量主要来自于前者。许多科学伟人之所以取得成就是出于对所研究问题的执着的追求，而非主要出于功利的考虑。人本来是好奇心很强的。但是在学校教育中，这种好奇心却逐渐地被分数、升学、就业、名誉等功利的考虑所取代，内在的动机力量越来越弱，出于功利的外在动机越来越强。在这种磨灭好奇心和破坏内在学习动机的过程中，考试发挥着十分重要的作用。追求考试分数成为学生学习的最直接、也是最主要的动力。

当前学校教育所存在着的另一个重要问题就是学生的创造性被破坏和磨灭。勿庸赘言，在推动科学发展和丰富人类文化财富方面人的创造性是十分重要的。只要与幼儿接触就会发现，他们经常提出一些非常奇特的问题和产生一些非常新颖的想法。不幸的是，儿童固有的这种创造性在学校中很快被扫荡了，他们被迫去模仿别人的所做、所说、所想。在学校中，他们习惯于接受已有的结论而不是去怀疑这些结论，他们热衷于去记忆这些已有结论而不是去思考这些结论。在这种磨灭创造性的过程中，考试又担负着重要的责任。根据心理学中有关记忆的研究，人的活动性记忆（即可以随时提取的记忆）容量是很有有限的，大量的记忆以非活动性记忆的形式保存在头脑中，需要凭借一定线索才能提取，然而，我们的考试却强迫受测者将许多完全不必要的内容包括在活动性记忆之中，浪费了受测者许多精力和时间。考试过

后，曾经背下来的东西很快忘记是常有的事。让那些曾经在高等学校入学考试中取得好成绩的讲师、教授们去参加高校入学考试，他们往往很难取得好成绩，因为许多知识“忘记了（即保存在非活动性记忆中）”。是他们水平不高吗？并不是他们的水平不高，而是高校入学考试并不充分反映水平，其质量不高罢了。

笔者认为，教育的第一职能是发展学生的人格，其核心是品质，其次才是发展学生的智力，再次才是传授学生以知识。今天，我们的许多学校教育恰恰是本末倒置了，注重传授知识而忽视发展能力，更不重视发展学生人格。在这种教育状况下，不仅学生知识面窄，实际分析问题和解决问题能力差，独立思考能力差，而且缺乏社会责任感，缺乏理想和抱负，审美能力低，兴趣狭窄，在这种对学生健全人格发展的干扰中，考试亦发挥着重要的作用。

这种以考试为轴心的学校教育受到国内外许多有识之士的尖锐的批评。本世纪最伟大的科学家爱因斯坦在《自述》中写道：

“人们为了考试，不论愿意与否，都得把所有这些使头脑偏离主要目标的废物统统塞进自己的头脑，这种强制的结果使我如此畏缩不前，以致在我通过苏黎世大学的最后的考试以后有整整一年，对科学问题的任何思考都感到扫兴。”他继续写道：“现代的教学方法竟然还没有把研究问题的好奇心完全扼杀掉，真可以说是一个奇迹，因为，这株幼苗除了需要鼓励以外，主要需要自由，要是没有自由，它不可避免地会夭折。认为用强制和责任感就能增进观察和探索的乐趣，那是一种严重的错误。我想，即便是一头健康的猛兽，当它不饿的时候，如果有可能，用鞭子强迫它不断地吞食，特别是人们强迫给它吃的食物是经过选择的时候，也会使它丧失其贪吃的习性的。”①

被诺贝尔奖金获得者生物学家梅多沃称为有史以来无与伦比的最伟大的哲学家的卡尔·波普尔在谈到学校教育时说：“我们

正在惊人地浪费着我们的时间，尽管我们老师受过很好的教育，而且都尽全力使学校成为世界上第一流的，他们教的许多东西是极其令人厌烦的——几小时、几小时毫无希望的折磨——对我来说已经不新鲜了。”②

著名物理学家、诺贝尔奖金获得者杨振宁在谈到考试时说：“我接触到很多第一流的物理学家，他们很能动脑筋，很会做实验，却不善于应付各种考试，如果仅凭考试取人，这些人才就可能被埋没。……如何想办法不浪费这样的人才是一个很重要和很紧迫的问题。”③美国著名数学家王浩认为，从小学到大学考试太多，学生为了准备考试而疲于奔命，一切为了考大学，学生死记硬背应付考试，没有充分消化和独立思考的时间，长此下去，学生头脑就比较迟滞了。④

我国著名教育统计学家陈一白先生在1982年写道：“当前问题的严重之处，还不在于考试办法不健全，而在于许多人因循守旧，对这种传统的考核办法的严重不合理之处缺乏认识，或是讳莫如深，以致成绩考试始终成为教育工作中最薄弱的环节。……这些早为科学准则所否定了的过了时的考试办法，我们今天仍在普遍使用，并且视为天经地义丝毫不容置疑的可靠量具，岂非怪事。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。”⑤

如此视之，考试乃教育的大敌。否！教育之于考试，恰似“鱼儿离不开水，瓜儿离不开秧”，须臾不可离之。

首先，要想发展教育事业，就必须不断地改进教育体制、教材和教法。好的体制、教材、教法要通过教育效果来体现，而教育效果又经常要间接地通过考试成绩来体现。例如，“集中识字”与“分散识字”、“先读书后识字”和“先识字后读书”，哪种方法更好呢？在学校期间，只能通过学生的测验成绩说明。心理学和教育学的研究是教育发展的基础和依据，心理学和教育学之于教育就好像生理学和解剖学之于医学。在心理学和教育学的研

究中，测验是一种十分重要的工具；在有关教育方面的定量化研究中，所应用的数据大多是测验成绩。儿童心理学着重研究不同年龄人的心理特点的不同，教育心理学着重研究不同教育环境下人的心理特点的不同。而体现心理特点的，一般都是各种各样的测验。

其次，教育的一条基本原则是“因材施教”。欲因材施教，必了解自己的学生。了解的方法主要有两种：一是基于经验之上的主观直觉，一是借助于测验的客观测量。前者是前科学的方法。在对温度的了解中，人的最初是凭借经验之上的主观直觉，后来发现，这种直觉是很不可靠的，很大程度上受到主体状态的影响，如对同样温度的水，手凉时会感到热，手热时会感到凉。于是，人们开始使用测量工具——温度计，用客观测量代替了主观直觉。对诸如能力、智力这样一些心理量的测量也是同样，主观直觉可以在一定范围内使教师了解自己学生的特点，但是这是靠不住的，很容易受到教师的好恶、偏好、情绪等的影响。因此，为了更准确地、更客观地了解自己的学生，使用测验是必不可少的。

从教育控制论的角度来看，教育过程可以被看作一种教师对学生学习进程的影响和控制的过程，而测验为这一过程提供了反馈信息。没有反馈就无所谓控制，因此，教育过程是离不开测验的。在程序教学（机器教学）中，测验是非常重要的一个环节。

在了解那些心理上有缺陷的儿童，对他们做出早期诊断并进行早期矫治方面，测验更是具有特殊的功效。这也正是始于本世纪初的测验运动的最直接的原因。

测验不仅可以帮助教师了解学生，还可以帮助学生了解自己。测验可以在一个人的学习安排和专业选择方面提供有价值的参考依据。

第三，测验是一种选拔人员的手段。今天，由于教育设施的有限，大量的测验用于选拔目的，如高中入学考试、高校入学考

试、研究生考试等。如果将发现人才的人称为“伯乐”，那么，测验就是伯乐手中的放大镜、皮尺或X光机，这些工具将提高发现人才的可靠性和科学性。

此外，在教育中，测验还担负着对学校管理人员、教师的选拔和工作评定的任务。

总之，教育离不开测验。我国在“文化大革命”期间已经有了惨痛的教训，取消测验考试，意味着教育的取消。

至此，我们已将读者置身于一种深刻的矛盾之中：一方面，测验是教育的大敌，危害不浅；另一方面，测验是教育的生命线，须臾不可离之。何如？答案很简单：改进测验，提高测验的质量。借助于数学工具，主要是统计学工具，可以有效地提高测验考试的质量。心理测量学的一个重要内容就是研究如何利用数学工具来提高测验考试的质量。计算机技术的发展，为这种努力提供了广阔的前景。一方面，数学工具促进了测验的改进和提高，另一方面，在改进测验的过程中，心理测量学家们又创造了新的数学工具。今天，几乎每一个工科和理科大学生都需要学习的数理统计学在其早期发展中是受到心理测验运动的推动的；同样，现已广泛地被应用于地质、气象、生物等许多学科的因素分析方法，也是由心理测量学家提出和发展起来的。

当今，“片面追求升学率倾向”是对普通教育的一个重要的干扰，遭到了社会的广泛批评。由于追求升学率而造成的对于学生全面发展的损害，很大一部分是由存在于教育之中的竞争造成的。这种竞争是存在于整个社会中的竞争的一部分，有其深刻的社会根源，从根本上讲，是由劳动主要是一种谋生手段而不是基本需要这一社会现实所决定的，是由社会的工业化这一趋势所决定的。这种现实的改变决非一朝一夕的事情，甚至不是一两代人的事情。在这种现实下就必须主要靠竞争来激发动力、保持效率。从这个意义上说，这种损害是不可避免的，是我们为了社会进步所必须付出的代价。但是，造成这种损害的另一个重要原因



是现行考试的不合理。在同样的竞争环境中，不同的考试可能造成颇不同的效果。合理的、好的考试不仅可以为高一级学校选拔出最有培养前途的学生，而且可以促进低一级学校的学生去努力丰富自己，改善自己的智能结构。相反，不合理的考试只会使学生死记硬背、浪费精力和时间。这种由于考试不合理所引起的损害是可能避免的。避免的途径就在于：借助心理测量学的研究来改进考试。

## 二、工业方面

测验不仅仅被运用于教育方面，而且被应用于广泛的领域。在现代工业生产中，测验发挥着不容忽视的作用。

现代工业社会的一个特点是高速，火车、汽车、飞机等速度越来越快；另一个特点是复杂化，工业设备的结构和功能越来越复杂，一个新式战斗机的驾驶员或一个核电站的控制操作人员，必须同时监控许多仪器。这样，人的可靠性问题就日益尖锐，人的失误就可能造成非常重大的损失。美国的三里岛核电站污染事故和苏联的切尔诺贝利核电站污染事故，都与操作人员的失误有关。于是，在工业设计时，就必须考虑人的可靠性问题，从而达到最好的效果，从而将产生事故的可能性降至最小。在操纵室的仪表设计中，有水平式、垂直式、圆周式、半圆式、窗孔式等，哪一种方式最适于操纵人员辨认呢？这就需要设计一定的测验来对一组操纵人员进行测试，并比较测试结果。在这个测验编制和结果比较的过程中，就需要运用心理测量学的知识。

人的有些心理特点是可以培养而获得的，如操纵技能、知识水平等，有些心理特点则较难通过训练而改变，如注意广度、情绪控制能力、勇敢、反应速度等。而对于某些特殊人员，如飞行员，这些心理特点又是非常重要的。因此，通过测验来对飞行员、驾驶员等特殊人员进行选拔就有非常重要的经济效益。一般说，训练一名飞行员的费用相当于一架飞机的造价，或者相当于