



心理测验的 原理与应用

郭庆科 编著 孟庆茂 审校



人民军医出版社

People's Military Medical Publisher

心理测验的原理与应用

XINLI CEYAN DE YUANLI YU YINGYONG

郭庆科 编著
孟庆茂 审校



人民军医出版社

People's Military Medical Publisher

北京

图书在版编目(CIP)数据

心理测验的原理与应用/郭庆科编著. —北京:人民军医出版社,2002.7
ISBN 7-80157-512-1

I. 心… II. 郭… III. 心理测验—高等学校—教材 IV. B841.7

中国版本图书馆 CIP 数据核字(2002)第 008967 号

人民军医出版社出版

(北京市复兴路 22 号甲 3 号)

(邮政编码:100842 电话:68222916)

北京国马印刷厂印刷

桃园装订厂装订

新华书店总店北京发行所发行

开本:787×1092mm1/16·印张 17·字数 410 千字

2002 年 7 月第 1 版 (北京)第 1 次印刷

印数:0001~4500 定价:29.00 元

(购买本社图书,凡有缺、倒、脱页者,本社负责调换)

内 容 提 要

本书系统地介绍了心理测验理论的最新研究进展,详细阐述了经典测验理论、项目反应理论、概化理论、结构方程模型、标准参照性测验的基本内容,并有代表性地介绍了传统的和新编制的各类心理测验,另外还介绍了适应性测验、测验等值、项目功能差异等测验理论的实际应用领域。在测验理论部分,作者侧重于从心理学的角度进行原理的阐释,尽量避开了有关的数学知识。本书的显著特点是理论性和实用性并重,使具备相应统计学知识的教育学、心理学和相关专业的本科生和研究生都能掌握。本书可作为心理学、教育学、社会学、管理学、人力资源等专业的本科生和研究生的教材,也可作为在以上学科领域从事研究和实际工作的人员的参考资料。

责任编辑 王 峰 秦素利 姚 磊

作 者 简 介

郭庆科 1969年8月生,山东菏泽人,山东师范大学教科院心理系副教授,研究方向为心理测量学和心理统计学,现为北京师范大学心理系应用与实验心理学方向的博士研究生。作者多年来一直从事心理测量学的教学与科研,曾在《心理科学》等心理学专业刊物上发表论文十多篇,并参编了多本心理学著作。

测验法是心理学的重要研究方法之一,是鉴别个别差异的重要途径。心理测量学经过百年的发展,理论体系已日趋成熟。在当今心理测验的发展过程中,智力理论、人格理论、信息加工理论越来越受到心理测量学家的重视,先进的统计方法、计算机技术对测验的影响也越来越大。在测验理论研究领域,随着经典测验理论、项目反应理论、概化理论、结构方程模式的不断发展和完善,心理测验对个别差异的评量也日益科学和客观,在认知心理学、智力和人格理论的影响下,心理测量学的理论导向性也越来越明显,即不仅关注个别差异的表现,而且关注个别差异的实质。

在心理测量学理论繁荣的同时,心理测量学的应用也越来越广泛,目前已成为应用性最强的心理学分支学科之一。其范围涉及教育、医疗、经济、法律、政治、军事等几乎所有的社会领域。心理测量的最终目的是服务于社会,促进人力资源的充分开发和合理配置,从而提高人们的生活质量,达到社会的繁荣。

由于众所周知的原因,心理测量学在我国的发展还落后于西方,无论是理论研究还是实际应用都存在很多薄弱的环节。因此系统介绍心理测量学理论和方法上的进展,培养学科的后备力量就显得尤为重要。本教材即是针对这一目的编写的。本书较全面而且是深入浅出地介绍了项目反应理论、概化理论、结构方程模式、标准参照性测验、适应性测验、测验等值、项目功能差异等测验理论和方法上的近期成就,以及各类心理测验的最新发展,便于读者了解心理测量学的发展现状。

作者一直从事心理测量学的教学和研究,本书即是在他多年积累的教案的基础上加工而成的。在写作中作者查阅了大量中英文资料,虚心请教了本学科的很多专家学者,在学术上取得了较深的造诣。

作者能在攻读博士学位期间写出这样的专著,我作为导师是比较满意的。这本书很适合作为心理学、教育学、社会学、管理学、人才资源等专业的本科生和研究生教材和参考资料。

孟庆茂

2002年1月于北京师范大学

前 言

心理测量学又称心理测验法,是心理学专业学生的一门必修课和基础课。在心理学的众多分支学科中,心理测量学是相当独特的一门课程。学好这门课不仅需要心理学的基本理论知识,更需要统计学方面的知识。与这门学科关系最为密切的是心理统计学和人格心理学,另外,普通心理学、实验心理学、教育心理学、发展心理学、认知心理学等也与这门学科有直接关系。

心理测量学的根本任务是将个体心理特质的差异数量化,但人的心理特质又具有内隐的性质,因此对它的数量化只能借助个体可观察的行为才能实现。而人的行为又是极为复杂的,是众多因素影响的结果。这些因素有的是内在的,有的则是外在的,有的是规律性的,有的则是偶然性的。因此人的心理特质与其外显行为间常常不具有确定性的关系,也就是说,借助外显行为对心理特质的测量必然要引起误差。

以上这种理性认识,对客观地了解心理测量学是必要的。这门学科经过百年的发展,在理论和方法上已相当成熟,并在各行各业都得到了广泛的应用。实践证明,使用心理测验确实能提高鉴别个别差异的准确性和效率,如果使用得当,心理测验能帮助人们改善生活质量,提高工作效率,进而达到人力资源和社会资源的合理配置和利用。但同时还应该看到,当前所用的心理测验并不是完美无缺的,因此借助心理测验对人的心理特质的了解是间接的、不全面的,有时甚至是错误的。心理测量的结果只能是了解人类个别差异的一个参考,决不能用来决定一个人的命运。因此,专业人员在使用心理测验时,必须持有谨慎和科学的态度。接受测验者同样也应有清醒的认识,决不能迷信心理测验。

针对心理测量学中的误差问题,心理测量学家们通过两种途径试图加以解决。其一是借助统计学的方法,即借助越来越复杂的概率统计和数理统计方法从心理现象的偶然性中把握其必然性,心理测量学在这一方向上的发展称为心理计量学。随着统计技术的发展,项目反应理论、概化理论、结构方程模型已被专业人士越来越广泛地接受,心理计量学取得了有目共睹的进步。作为一个现代心理测量学家,掌握这些知识是必要的。对这一部分内容,作者是从心理学的角度作了概括性的介绍,相信读者能够掌握。书中有时不得不涉及到一些复杂的数学知识,可作为一般内容进行了解。只要能掌握心理测量学的原理就足够了,有关的

数学运算可以由计算机软件完成。

另一途径是加强心理特质个别差异的理论研究,主要表现在人格心理学、智力心理学、认知心理学、发展心理学等领域的研究。这方面研究的目的在于发现个别差异的实质,重在探讨心理特质的内涵,特质与特质间的关系,以及特质与外显行为间的关系。这些研究使新一代的心理测验体现出了更强的理论导向性。这一领域的研究是提高心理测验有效性和客观性的根本。因此本书在写作时不仅力求全面地介绍传统的和新编制的心理测验,更重视介绍其赖以产生的理论基础。

本书的一个重要目的是致力于构建现代心理测量学的学科理论体系,由于这门学科发展较快,以往出版的教科书很少能完全体现这一理论体系,因此本书的出版有助于读者了解这门学科的全貌。由于本学科的内容较多,在学习时,读者可根据自己的需要在内容选择上有所侧重。有志于心理计量学研究的人可以将重点放在前七章,对后五章的内容可不作过高要求。而有志于将测验用于实际工作或其他学科领域研究的人,可不将项目反应理论、概化理论、结构方程模型等部分的学习作为重点。

另外,标准参照性测验、测验等值、项目功能差异的研究在我国仍是薄弱环节,社会上对这些领域已经有了越来越多的要求,希望能引起读者的重视。

在读博士期间写成这样一本不算短的著作,完全是出于我对这门学科的钟爱。与多数读者一样,我一直对能测出人的智力和人格差异的心理测验着迷和情有独钟,因此硕士研究生期间选择心理测量作为研究方向。但随着学习的深入,我发现要得到一个科学的心理测验是颇为不容易的事情,必须首先掌握科学的方法。现在通过这本书将自己多年心得和体会与读者分享,很有一种欣慰感。希望本书能传达给读者这样一个理念:即心理测验是一个科学的工具,使用它时必须要有严谨的态度和科学的精神,掌握科学方法后的回报绝对是值得的。

在本书的写作和修改过程中,我的导师孟庆茂教授在百忙之中挤出时间审阅了全稿,并翻阅了大量的珍藏文献资料,对有争议的地方一一核实,导师的严谨学风令我受益匪浅,同时也是我继续前进的动力。我的博士同学刘红云老师帮我整理了书中有关数学原理的阐述,使我澄清了很多模糊的认识。没有他们的帮助,本书是难以完成的,在本书即将出版之际,谨向他们表示感谢!本书引用了很多作者的参考文献,他们有不少是我尊敬的老师,在此一一致谢!

由于作者水平有限,书中定有不少不妥之处,请不吝指正!

编者 郭庆科

2002年1月于北京师范大学

目 录

第一章 心理测验概述	(1)
第一节 心理测验的历史	(1)
第二节 心理测量与心理测验	(9)
第三节 测验的种类和功能	(15)
第二章 测验结果的解释	(19)
第一节 常模参照性测验及其分数的解释	(19)
第二节 标准参照性测验及其分数的解释	(26)
第三章 信度	(31)
第一节 经典测验理论的信度观	(31)
第二节 信度系数的计算方法	(34)
第三节 信度的作用与影响因素	(38)
第四节 概化理论的信度观	(40)
第五节 标准参照性测验的信度	(48)
第四章 效度	(52)
第一节 效度概述	(52)
第二节 内容效度	(53)
第三节 效标关联效度	(56)
第四节 构念效度	(65)
第五节 效度研究中的其他问题	(71)



第六节 结构方程模型在测验效度研究中的应用	(76)
第五章 项目分析	(82)
第一节 项目难度分析	(82)
第二节 项目区分度分析	(84)
第三节 项目分析的相关问题	(86)
第六章 项目反应理论	(90)
第一节 经典测验理论的局限与项目反应理论的发展	(90)
第二节 项目反应理论的基本理论体系	(93)
第三节 项目反应模型	(96)
第四节 项目与测验信息函数	(103)
第五节 参数估计	(107)
第七章 测验等值与项目功能差异	(114)
第一节 测验等值	(114)
第二节 项目功能差异	(121)
第八章 心理测验的编制	(132)
第一节 常模参照性测验编制的基本过程	(132)
第二节 测题编写技术	(138)
第三节 题库	(144)
第四节 标准参照性测验的编制	(147)
第五节 项目反应理论在测验编制中的应用	(150)
第六节 适应性测验的编制与实施	(154)
第九章 智力测验	(158)
第一节 智力理论与智力测验	(158)
第二节 个别实施的智力测验	(160)
第三节 团体实施的智力测验	(172)
第四节 智力测验的使用	(176)

第一章 心理测验概述

人与人之间的个别差异是尽人皆知的事实,人们生理上存在身高、体重等的差异,在心理上则有能力、学识、性格、气质、态度、爱好、价值观等的不同,比如有些人性格开朗大方,有些人则腼腆孤僻;有些人智力超群,有些人则智力低下;有些人品行高尚,有些人却品行低劣。客观地揭示出这些差异是非常重要的,教育者要根据学生的心理特征采用合适的教育方法,对他将来的发展作出预测,并对其职业道路提供必要的建议;人事部门只有了解了人们的能力和个性特征才能将他们安排到合适的部门;咨询人员只有在准确了解了来访者的心理后才能针对他的问题给以合理的建议,等等。这种了解人类个别差异的需要推动了测量个别差异的系统工具——心理测验的产生。

要了解心理测验的原理和应用,就要从心理测验的一些基本知识和基本概念入手。本章就从心理测验的发展史谈起,并介绍心理测验的概念、种类和用途。

第一节 心理测验的历史

一、中国古代的心理测验

虽然心理测验作为一门学科是近代才产生的,但人们对心理与教育测量的探索却可以追溯到很久以前。中国古代的学者早就注意到人与人之间在能力和个性方面的差异,并尝试使用有效的方法将不同心理品质的人加以区分。因此很早就产生了心理测量的思想,并开始了心理测量的实践。

中国古代的医书《黄帝内经》中曾把人分成太阴、太阳、少阴、少阳、阴阳和平五种,这一分类可看作心理测验的前奏。

孔子按照能力和知识将人分成上、中、下三等,并说“惟上智与下愚不移”,“中人以上,可以语上也,中人以下,不可以语上也”。这种划分包含了等级评定法的基本思想。

孟子曾对齐宣公说:“权,然后知轻重;度,然后知长短,物皆然,心尤甚,王请度之。”就包含了将人的能力、品格等心理特性加以量化的思想,这与心理测量学家们的信条“凡物之存在必有其数量”(桑代克,1918)、“凡有其数量的事物都可以测量”(麦克尔,1939)表达了同样的思想。

三国时刘劭著有《人物志》一书,将人分成圣贤、豪杰、傲荡、拘傥四种:“心小志大者圣贤之人也,心大志大者豪杰之雋也,心小志小者拘傥之人也。”包含了等级评定的思想。1937年

施罗克(Shryock, J. K.)曾将《人物志》翻译到美国,书名为《人类能力之研究》。刘劭认为“众人之察不能尽备”,只能“观其感变以审常度”。用今天的话来讲,就是说,人的所有行为是不能完全被观察到的,只能通过观察其有代表性的行为来预测其整个人格。这与现代心理测验中对“行为样组”的测量是一致的。

在测验的实践方面,中国也是远远早于西方的。公元前11世纪,西周采用“试射”的办法选拔文官,不仅测验射箭的命中率,还要观察试射者的行动是否合乎礼仪,举止是否合乎乐律。这可以视为人类历史上最早的心理测验。到了汉代,文官考试开始使用笔试,其内容包括法律、军事、农业、税收和地理。

至隋大业二年(公元606年),中国开始了科举取士的制度,用一系列严格的考试逐级选拔人才,这比18世纪欧洲运用笔录取取文官的制度要早一千多年。任何一个心理测量量表都要以中国的科举考试为开端。

二、心理测验在西方的早期探索

在西方,达尔文的进化论思想为个别差异的研究提供了框架。1796年格林尼治天文台的观察者们首先注意到了人与人之间反应时的差异,由此引发了关于人类个别差异的系统实验研究,促进了对感觉运动速度的测量。这时欧洲的民主思想和教育与医学中对弱智儿童的关注推动了智力测量的研究。

1. 冯特

冯特是实验心理学的创始人,他首次将实验方法引入心理学研究。冯特在实验中证实了个别差异的存在,并发明了测量思维敏捷性等方面个别差异的工具。早期的实验心理学和心理物理学为心理测验的发展奠定了基础。

2. 高尔顿

第一个直接推动心理测验产生的人当属英国生物学家高尔顿。高尔顿开创了个别差异心理学的研究,并引入了定量研究方法。1869年,高尔顿出版《遗传的天才》一书,提出了能力遗传决定论,在1883年出版的《人类能力及其发展的研究》一书中,高尔顿首次提出了“心理测量”的概念,列举了大量有关人类个别差异的资料,并多次重申个别差异可以测量的观念。

高尔顿认为所有智力行为都与人的内在感觉能力有关,因此运用心理物理学方法,通过测量人的运动反应时和感觉辨别能力就能测量到人的智力。在这一思想指导下,他进行了大量的人类测量学研究。1884年,高尔顿在伦敦国际博览会上建立一个“人类测量实验室”,此后的六年中,共搜集了9337人的生理、感知觉方面的个人资料,这些资料包括身高、体重、头颅大小、手背跨度、中指和前臂长度、握力、呼吸力、拉力、视力、听力、以及视觉和听觉反应时等。高尔顿曾试图运用他的学生皮尔逊发明的相关程序计算这些测量与各成就指标间的相关性,但结果并没有证实他的设想。

3. 卡特尔(Cattell, James M.)

卡特尔是推动心理测验产生的另一个重要人物,1890年,卡特尔发表了一篇有重要意义的论文《心理测验与测量》,第一次使用了心理测验这一概念,他认为:“心理学者不立根基于实验与测量上,决不能有自然科学的精确”,“如果我们规定一个统一的手续使在异时、异地得出的结果可以比较、综合,则测验的科学和实用的价值都可以增加。”这些观点可作为测验编制的重要指导思想。

卡特尔发展了高尔顿的方法,提出 10 项心理测验,包括如下一些内容:

- (1)握力测量,用握力计测肌肉力量;
- (2)动作速度测量;
- (3)触觉两点阈测量;
- (4)引起痛觉的最低点测量;
- (5)对辨别重量最小差别的能力的测量;
- (6)对声音反应时的测量;
- (7)说出混在一起的四种颜色的名称;
- (8)将 50 厘米的线平分;
- (9)对 10 秒时间判断的测量;
- (10)复述听过一次的字母。

卡特尔也认为生理能量与心理能量密不可分,因此对智力的测量也主要是对感觉辨别力、反应时、动作过程等的测量。卡特尔的学生维斯勒(Clerk Wissler,1901)进行了一项开创性的研究工作。他搜集了哥伦比亚大学和伯纳德大学 300 名学生的智力测验分数与学业成绩,试图以智力测验分数预测他们的学习成绩,这可视为对智力测验进行的最早的效度研究。但维斯勒的研究结果令人失望,因为二者的相关很低。例如,学业成绩与数字记忆测验的相关为 0.16,与握力测验相关为-0.08,与颜色命名相关为 0.02,与反应时相关为-0.02。与数字记忆的相关虽达到了显著水平,但其对学业成绩的预测能力却是微乎其微。

维斯勒的工作证明用简单心理过程测量智力的方法是错误的,改变了智力测验发展的方向,启发了以后比奈-西蒙智力测验的产生。

但现在看来,造成这一结果的原因一方面是因为卡特尔仅仅测量了一些简单的心理活动,另一方面则可归因于维斯勒的研究方法,维斯勒所选取的研究对象都是聪明的大学生,智力水平可能都很高,缺乏代表性,且学业成绩也未必能代表智力的高低(克龙巴赫,1984)。以后的研究重新证实了反应时与智力的相关(Jensen,1982;Sternberg,1985),且把反应时当作智力测验的一项重要指标。

冯特、高尔顿和卡特尔对心理测量的研究主要包括人的基本心理能力,如感觉辨别力、反应时、动作过程等,虽然没有得到预期的结果,但却为心理测量的产生打下了基础,树立了人的心理属性可以测量的理念。

4. 医学和精神病学家的的工作

1800 年以后,出于对心理机能异常者诊断和治疗的需要,一些医学和精神病学家开始了对智力障碍和心理变态患者的系统研究,并尝试使用有关测验对其进行鉴别,他们对心理测验的贡献不亚于当时的实验心理学家。

直到 19 世纪初,人们才认识到智力障碍和心理变态的区别。法国医生艾斯克罗(Esquirol, J. E. D.)是第一个对二者做出明确区分的人。他认为智力落后是与生俱来的,且难以改变,而心理变态则是成年后才出现的,可以通过一定的措施加以治疗。

艾斯克罗把言语能力作为诊断智力障碍的主要依据,他按照言语能力障碍的程度将智力落后分成三类:一类只会使用短语,一类只会使用单词,一类只会发出叫声,这与今天将智力障碍分成轻度、中度、重度的分类是相当的。艾斯克罗对比奈的影响是显而易见的,比奈的智力测验也以言语能力为主,且编制测验的最初目的也是为了鉴别智力落后的儿童。

另一位法国医生沈干(Seguín, E.)在教育和训练落后儿童方面做出了突出贡献。他曾是艾斯科罗的学生,后来又师从意太(Itard, J. M. G.)并继承了意太的训练方法。他开设了一个专门训练落后儿童的实验班,他所使用的感觉训练和肌肉训练的方法,至今仍在智力落后儿童教育机构中使用,而其中感觉辨别和运动控制的内容已成为现今操作智力测验和非言语智力测验的组成部分。

19世纪中期和后期的精神病学家,已开始使用标准化的测验来检测精神病人的知觉、记忆等高级认知过程。格拉赛(Grashey, 1885)使用移动的字、符号、图画等检查病人的再认,这是后来使用的记忆鼓的前身。1889年,里格(Coward Rieger)的测验中已包含了视觉认知、短时记忆、长时记忆、语句及动作模仿、造句等高级认知内容。

对心理测验的产生有重要影响的还有克雷匹林、艾宾浩斯等。克雷匹林(Kraepelin, E.)出于临床诊断的目的开始编制和使用测验,以区分正常人和心理变态者以及不同类型的心理疾病患者。克雷匹林及其他精神病学家侧重对推理能力、在持续性工作中意志努力的稳定性等复杂心理过程的测量,他们认为这对被试的实际行为有预测作用。艾宾浩斯认为智力的高低表现为综合能力的差异,他首创了语句填空测验测量学生的能力,测验结果与学生学业成绩有很高的相关。填充测验的形式被以后的心理测验广泛采用。

三、心理测验在西方的产生

法国心理学家比奈是第一个智力测验的创立者,他因此被称为心理测验的鼻祖。比奈(1857~1911)生于法国尼斯,早年学过法律、医学,以后致力于心理学的研究。从青年时期就对人们行为的个别差异及不同年龄阶段行为的发展变化产生兴趣。

比奈尝试了很多方法鉴别人们智力的差异,包括数学、回忆、暗示性、道德判断、触觉辨别力、心算、乃至比较头骨大小、手相等方法,还曾经用自己的两个女儿当被试进行了反应时测量。最后发现对测量智力最有意义的是解释物体、图画及故事,并发现具有良好判断力的儿童在注意力及词汇方面也有优越的表现,这使他相信一般能力的存在,将智力活动视为一个有机的整体。

比奈将智力定义为:选择并保持一个明确方向的倾向,为达到预定目标而做出适应的能力及自我评价的能力。比奈认为,智力的基本技巧包括预测一个行为或事件的结果,具体体现在以下几个方面:

- 检查自己行动的结果
- 对自己当前行为的监督
- 现实性测量
- 各种协调、控制性的行为
- 学习及解决问题

比奈开创性地发现智力与复杂的任务有关,尽管很难明确说明这些任务测量了什么,但被试在这些任务上的表现确能预测其智力行为。这些为他编制第一个智力测验打下基础。

1904年,法国教育部委派教育家、医学家组成一个委员会,研究公立学校低能儿童的教育方法,比奈也是委员之一,他主张用心理测验的方法鉴别智力落后的儿童。比奈与其助手西蒙合作,编制成了历史上第一个智力测验,并在1905年的《心理学年报》上发表文章《诊断异常儿童智力的新方法》介绍此量表,史称1905年量表。此后比奈又在1908和1911年对此量表进

行修订,产生 1908 年量表和 1911 年量表。

1905 年量表共包括 30 个由易到难的问题。特点是测验项目种类繁多,能测量智力多方面的表现。测验包括一些测量感知觉的题目,但语言部分占了很大比例。主要涉及推理、判断、理解等高级心理过程。测验适用于 3~14 岁的低能儿童,也可用来对正常儿童作某种程度的区分,以通过的题目数来衡量智力的高低。

1908 年,比奈和西蒙根据 1905 年量表的使用情况和问题对测验进行了第一次修订。1905 年量表主要是用来鉴别智力落后儿童的,测题偏易,不适用于正常儿童;1908 年量表删去了这些不适当的题目,增加了一些较难的新题目,使测题总数达到 59 个,测题按年龄分组。组别为 3~13 岁,共 11 个年龄组,每组的测题都保证该年龄组儿童有 80%~90% 的人能通过,使用智力年龄表示儿童的智力发展水平。该测验使用了现代智力测验的一些重要概念,如效率、常模等。

1911 年,比奈又对 1908 年量表进行了修订,这次修订版的变化不大,主要是删改了一些项目,并重新划分了年龄组,包括:删掉了 9 个旧项目,增加了 4 个新项目,使测题总数达 54 个,增加了 15 岁组和成人组,除 4 岁组仅有 4 个项目外,其余组项目数均为 5 个。不幸的是:比奈于当年去世,年仅 54 岁。他的去世是心理测验领域的重大损失。

四、心理测验在西方的发展

1. 智力测验的发展

比奈—西蒙量表问世后,立即引起世界范围内的广泛关注,并被翻译成多种文字,1916 年,美国斯坦福大学的推孟将比奈—西蒙量表进行了修订,形成著名的斯坦福—比奈智力量表。推孟对原量表进行了较大的改动,保留了比奈—西蒙量表的 51 道题目,自己又编制了 39 道,使测验总长度变为 90 题,适应的年龄组为 3~14 岁组儿童,另加普通成人组和优秀成人组,并首次使用了智商(Intelligence Quotient,即 IQ)的概念。斯坦福—比奈量表被广泛地应用于教育和临床领域,是最有影响的智力测验之一。此后它分别在 1937,1960,1972 和 1986 年被四次修订。

但斯坦福—比奈量表也存在一些缺点,突出表现在智商的概念上。美国心理学家韦克斯勒看到了这一缺点,并从 1934 年起着手编制智力量表。1939 年,韦克斯勒发表了第一个用于测量 16~60 岁成人智力的韦克斯勒—贝尔韦量表(W-BI),这个量表以后发展成为韦克斯勒成人智力量表(WAIS,1955)。韦克斯勒在 W-BI 的基础上于 1949 年发表第一个儿童智力量表(WISC),测量对象为 6~16 岁的儿童。1967 年出版韦氏幼儿智力量表(WPPSI),1974 年发表 WISC 修订本 WISC-R,1981 年发表 WAIS 修订本 WAIS-R,1991 年, WISC 第三版(WISC-III)出版。

韦克斯勒改进了智商的概念,提出了离差智商,这是一个巨大的进步,他还扩展了智商的概念,从言语、操作及总体三方面衡量智力的高低。

1938 年,英国心理学家瑞文(Raven, J. C.)出版瑞文标准推理测验,这是一个著名的非文字智力测验。既可用于个别测验,也可进行团体测验。1947 年瑞文又出版彩色推理测验和高级推理测验。

2. 标准化教育测验的发展

在教育测验方面,虽然在公元 606 年隋朝就开始了科举考试,但欧洲学校一直没有笔试的

传统,直至1702年,英国的剑桥大学才开始使用笔试录取学生。1845年,在美国著名教育家贺拉斯·曼(Horace Mann)的倡导下,麻省波士顿市教育委员会开始用笔试的办法考察该市所属学校的毕业生。这以前的教育测验都是以主观性试题为主,评分缺乏客观性。

客观的标准化教育测验的最初尝试者是英国格林尼治医院的教师费舍(George Fisher)。他深感对学生成绩的评定缺乏统一的标准,因此试图建立评定学生各科成绩的统一量尺。他搜集很多学生的书法、拼写、算术、语法、作文、历史、自然、图画、法文等科的成绩,汇编成《量表集》一书,作为衡量学生各科成绩的标准。费舍对收入量表集的学生成绩评定了等级,其他学生的各科成绩与这些标准相比较,就能判断其优劣等级。用今天的话说,收入量表集的学生就相当于常模样本组,而量表集就相当于常模量表。费舍的工作是教育测验历史上的重大进步。但是,《量表集》中学生的各科测验分数及其等级是费舍本人主观决定的,缺乏客观的依据。

1894年,莱斯(Rice, J. M)博士主张用统一的测验去评定不同学校学生的成绩。1895~1905年间,他先后编制了算术、拼字、言语等测验,测试了数万名学生,引起了教育界对教育测验的关注,莱斯因为在测验上的突出贡献而被称为教育测验的开创者。

1904年,美国心理学家桑代克(Thorndike, E. L.)出版《心理与社会测量导论》一书,这是关于测验理论的第一本著作,该书系统地介绍了统计方法及测验编制的基本原理,为测验的发展奠定了基础,并提出“事物的存在必有其数量”的著名观点。1908年桑代克发表书法量表,这是第一个书法量表,在教育测验历史上有重要意义。1909年,他又根据统计学“等距原理”为测验量表确定了单位。此外他还编写了拼写量表、作文量表、图画量表等。桑代克在测验理论及实践研究中的贡献使他成为测验运动的中心人物,被称为教育测量的鼻祖。

在桑代克的推动下,一些标准化教育测验相继出现。1908年,桑代克的学生斯通(Stone, C. W.)编制了算术测验,1911年,希莱格斯(Hillegas, M. B.)编制了作文量表,白根汉(Buckingham, B. R.)编制了拼写量表,柯蒂斯(Courtis, S. A.)编制了算术标准化测验。纽约的波士顿等大城市设立测验研究部,组织测验专家进行大规模的教育测验。这一时期客观的标准化测验得到了极大发展,特点是注重测验评分的客观化,并用常模对测验进行解释。

经过桑代克等人的早期探索,标准化教育测验逐渐走向成熟。早期的教育测验都是单科测验,如拼法、算术测验等,为适应调查的需要,于是出现了综合测验。综合测验能同时考查学生在几个主要学科上的学业成绩。1920年,宾特纳(Pintner, R.)发表的《测验汇编》,孟禄(Monroe, W. S.)和白根汉合编的综合测验是最早的综合测验。1923年,凯利(Kelley, T. L.)、鲁驰(Ruch, G. M.)和推孟编制了斯坦福成就测验(Stanford Achievement Test),已具备现代测验的性质,在当时非常流行。

1917年以前,教育测验主要限于小学科目,1917年后,用于中等以上学校的教育测验得到了发展。心理测验的理论和方法对教育测验产生了巨大的推动作用,越来越多训练有素的心理测量学家和教育测量学家投入到教育测验的开发和应用中,使教育测验得到深入发展。由此,教育测验进入了兴盛期。20世纪30年代后,教育测验不仅用于评定学生的学业成绩,而且还广泛应用于会计、律师、医生、商业、交通、军队等职业领域,进行人才选拔、安置、培训等。

为顺应教育测验发展的形势和社会各界对教育测验的需要,各国相继成立了专门管理考试的机构,组织专家编制、实施和管理测验。如美国的教育测验服务社(Educational Testing Service, 即 ETS),美国大学测验中心(ACT),英国伦敦职业考试中心,日本大学入学考试国家