

海峡两岸学术研讨会论文集

心理与教育测量

主编 张厚粲

副主编 马金科 孟庆茂



浙江教育出版社

前言

由中国教育学会教育统计与测量研究会、中国心理学会心理测量专业委员会和北京师范大学共同主办的海峡两岸心理与教育测量学术研讨会于 1995 年 10 月 22 日至 25 日在北京师范大学举行。

此次会议是海峡两岸心理与教育测量学界在中国大陆召开的第一次规模较大的研讨会。心理与教育测量学兴起于 20 世纪初,是心理和教育科学体系中相当重要的学科。它在中国的传播与发展几经波折。由于历史的原因,海峡两岸这方面的学术交流中断了近 40 年。近 10 年来,通过各种传媒,两岸心理和教育测量方面的信息交流逐渐增多,学者们的接触也日益密切。大家共同认识到:中华民族有着悠久的历史、教育历史,华人社会的心理与教育现象与西方社会相比,有着自己的特点;两岸彼此的独立研究,不利于我国心理与教育测量学的更好发展,难以达到最佳效果;长期的阻隔使两岸的学术用语出现了一些差异,如此下去,势必会影响两岸的学术交流。为了推动两岸学者的合作,促进具有华人特色的心理与教育测量学的发展,经过半年准备,经过论文征集和较为严格的筛选,两岸学者终于在北京欢聚一堂,畅所欲言地交流自己的研究成果。

莅临会议的心理与教育测量学界学者共 123 人,其中来自台湾省的学者 16 人。研讨会期间,有 83 位学者在大会和分组讨论会上作了精彩的学术报告,有 17 位学者委托他人作了书面发言。

研讨会组委会主席、北京师范大学张厚粲教授在大会上作了题为《心理测量学的研究与发展》的学术发言,从理论到实践对大陆心理测量进行了全面的总结。台湾师范大学吴武典教授代读了台湾师范大学何荣桂教授的论文《职业心理测验在台湾的发展》。国家教委考试中心马金科副主任作了题为《以教育测量学为基础的中国大陆大学入学标准化考试改革实践与发展概述》的学术发言。大会总结交流了两岸近年来在测验的应用领域所做的实际工作。在与会学者的共同努力下,大会为海峡两岸测验学者的交流与合作创造了良好的气氛,许多学者找到了合作研究的切合点:几十年来,我们使用的有代表性的量表都是修订西方的成形测验,与会学者们表示,要为编制适合我们中国人使用、具有中华民族特色的测验作出贡献,具有国际影响的测量工具必定会在两岸测验学者的共同努力下产生。

会议研讨内容广泛,涉及了心理与教育测量领域的各个方面,其中最大的特点是:理论上把握了现代的前沿水平,应用上更贴近实际生活,伸向社会的各个领域。代表们向大会提交的论文中,有关心理与教育测量理论方面的 28 篇,有关心理测验编制方面的 30 篇,有关教育测验编制与应用方面的 23 篇,有关心理测验应用方面的 20 篇,还有些论文从不同角度涉及社会生活。代表们的发言勾画出了心理与教育测量在人类生活中的全方位、多层次应用的画卷。讨论内容在纵向上涉及了婴幼儿、小学生、中学生到大学生、成人各年龄段,横向上在同一年龄段涉及学习、智力、性格、心理健康等方面;职业上不仅涉及了教育、公安、运动员、企业管理等职业,还涉及职业兴趣、能力倾向、群体人际关系等方面;在人的思维操作领域涉及问题求解搜索效率、阅读能力测评、学习障碍诊断等问题;在思想意识测量方面涉

及成就动机、生活信念、价值观等问题。

为了进一步促进海峡两岸心理与教育测量的学术交流,为了使更多的同仁了解研讨会的成果,同时也为了让全体与会学者以此次研讨会为新的起点,继续加强协作,深入钻研,我们从代表们提交的论文中选取了一部分有代表性的论文汇集成册,供大家研究参考。

本论文集的出版,必将进一步推动海峡两岸的学术交流和合作研究,必将推动心理与教育测量的研究与应用,并产生良好的社会效益。

主 编 张厚粲

副主编 马金科

孟庆茂

1997年3月

55.08/05

目 录

- [1] 心理测量学的研究与发展
张厚粲
- [7] 以教育测量学为基础的中国大陆大学入学标
准化考试改革实践与发展概述
马金科
- [20] 成就动机三种测量方法的比较研究
王 辉 时 勘 陈 为
- [30] 题目区分度的估计
张康举
- [36] 国外实证性因素分析及其应用研究的新进展
王 权 王秀玲
- [50] 多值记分考试项目反应理论的研究
——0—1 记分考试 IRT 的严格推广
李绍珠
- [56] 群体人际关系的概率模型
马国旗 华道云
- [66] 再谈测量集体进步幅度的方块分析法
李全柱
- [70] ACT 职业兴趣问卷修订报告
郑 波 龙立荣

-
- [81] 对价值观测验计分方式的研究
孟庆茂 黄良东
- [90] 高校课程考试科学化的探讨
曾桂兴 黄裕胜
- [100] 预测与难度控制
——“上海卷”在稳定难度过程中所进行的试验
胡江浩
- [109] 初中生问题求解搜索效率研究
余嘉元 杨 跃
- [117] 电脑自适应测验、解题辅导题库
孔祥斌 胡文俊
- [126] 标准参照测验分界分数确定方法的验证
肖 琳 王孝玲
- [141] 家庭关系测试与学习成绩的相关研究
——102 名中学生“家庭活动画”分析
张同延 陈 敏
- [147] 家长文化程度、职业和期望对中学生多项能力
发展的影响
徐跃玲 范存仁 林国彬
- [153] 测查心理特点 实施因材施教
——诊断小学生学习障碍的初步研究
杨志明 薛小强
- [163] 高考生心理卫生状况及其分析
章竞思 辛志勇
- [172] IRT 等级模型下目标参照性测验试卷组拼方
法和程序开发研究
戴海崎 谢旭升

-
- [180] CTT 题库建设的数学模型
卢正勇
- [187] 三种班级适应行为量表之编制及其应用
吴武典
- [210] 《学生性格量表(11~18岁)》(SPS)的信度与效度研究
张 锋 罗黎辉 王芙蓉
- [220] 《津医精神运动成套测验》(JPB)的初次标准化
王 栋 钱 明
- [227] 中文版庐-尼神经心理测验组筛检测验的修订
郭乃文 余丽桦
- [244] 高中毕业会考的问题与对策
侯光文 张发德
- [246] 高中会考标准及其建立
蔡建民
- [255] 化学高考预测效度研究
卫子光 韩家勋
- [271] IRT 与 CTT 对 1994 年高考数学试卷客观题部分的分析比较研究
边玉芳 戴凌云
- [278] 大学英语四级考试(CET4)通过标准的制定
朱正才
- [285] 心理测量在教育改造罪犯中的研究与应用
张安民
- [296] 《初中教师传记信息表》的编制
宋官东

- [308] 快乐体验之概念及其测量
郭俊贤 吴静吉
- [322] 《儿童心理综合量表》的修订
——《北京城区 4~9 岁儿童量表》的制订
周 谦 王建平
- [332] 跨文化管理行为研究的量具翻译问题
吴 岩 黄琪恩 Robert F Scherer
- [342] 企业中层管理人员社会愿望量表的编制
郭德俊 宋剑辉 许雪立 唐 京
- [354] 通用性向测验在中国台湾地区的修订与改进
之研究
林邦杰 黄国彦 修慧兰
- [372] 《儿童智力筛查量表》的编制和推广应用
傅根耀 汪文鋈
- [381] 小学生识字量测验的编制
王孝玲
- [389] 课堂教学结构因素对学习成绩影响的分析
——多变量增广混合逐步回归非线性数学模型
李 蔚 傅文瑞
- [397] 普通高考分数转换方案对比统计分析试探
陈庆良 鲁 直 王光翔
- [409] 关于考试改革与研究的几点看法
刘 昕
- [416] 小学生阅读能力测评研究
臧铁军
- [428] 《飞行动机量表》(FMS)的编制与效度分析
郑日昌 黄新宏 郭晨升

-
- [441] 计算机施测对心理测验的影响研究述评
韩布新 车宏生
- [450] EPQ 所测气质维度与 STI 所测神经特性之比较研究
陈国明
- [459] 电脑适性测验应用于学业性向测验之成效评估
张淑慧 范光辉
- [473] 对教学效能的图形测评
张 煜 李树珍
- [484] 目标参照测验项目难度的指标体系及估计方法
申亚权 张守臣
- [493] 舟山群岛中小學生创造能力特征研究
陈 敏 张同延 虞晋卿 郑曾祥
- [498] 测定学生心理活动的灰色途径
张晓晔 华 薇

心理测量学的研究与发展

张厚粲

北京师范大学心理学系,100875

大凡学科的发展必立足于社会的需求,同时还需得益于相关学科的进步,心理测量学的发展也不例外。

就世界范围而言,现代心理测量学的诞生是 19 世纪末的西方教育普及运动、两次世界大战中大规模的军队扩充和战后工业化大生产的结果。而早期测验的应用又引发了对心理量表的可靠性和有效性的学术探讨。许多早期测量理论家在积差相关的基础上,一砖一石地修造了经典测量理论(classical test theory)的大厦。进入 20 世纪 50 年代后,大规模国家级教育考试的广泛应用,使得经典测量理论的缺陷和不足日渐明显。同时,测验等值、题库建立等实际问题的出现,导致了对新理论、新方法的探索。Lord 等人于 20 世纪 60 年代提出并逐渐完善了一种新的测量学理论——项目反应理论(item response theory,IRT)。它的 3 个明显的优点,即对项目参数的估计独立于被试样本团体;对被试的能力估计独立于对他们实测的那组题目;可以根据被试的不同能力水平给出不同的估计精度——很快得到公认。之后随着统计方法的完善和计算机的普及,项目反应理论很快被应用于各类大规模教育考试、证书考试以及标准化心理测验

的编制等。有关的研究和应用在 20 世纪 80 年代末期非常盛行,达到了它的鼎盛时期。

进入 20 世纪 70 年代后,因为资格考试和证书考试在社会上日益流行,与标准参照测验(criterion-reference test)相关的理论和方法逐渐得到了测量学家的广泛重视。心理与教育测量的目的已经不再是仅仅区分被试的优劣,而在于评价被试对所规定的知识领域的掌握情况。由于这种与常模参照测验明显不同的目的,基于常模参照测验发展起来的经典测量理论和方法,包括题目分析的方法、测验信度的估计方法等,表现出了极大的局限性。这就需要测量学家针对标准参照测验的实际情况,创立和发展可用于这类评价活动的技术和方法。20 世纪 70 年代以来,Glaser、Popham、Berk 和 Hambleton 等人在标准参照测验领域中进行了一系列的研究工作。其中,Berk 于 1984 年主编的《标准参照测验指南》总结了此前十余年的研究成果,对编制和使用标准参照测验的各个环节作了说明,如标准参照测验的概念界定、测验题目的编写和筛选、测验切割线的确定、测验的信度和效度估计等。进入 20 世纪 80 年代以后,Hambleton 等人又对项目反应理论在标准参照测验中的应用进行了深入的探讨和研究。此外,有关标准参照测验的发展也得益于教育心理学对各学科的知识领域进行分析的结果。

测量学家也认识到,传统的客观形式的测验不能很好地评价人类的高级心理活动,而高级心理活动正是人们越来越关心的问题。因此工作表现评价(performance assessment)引起了大家的兴趣,而 20 世纪 60 年代末提出的概化理论(generalizability theory)为此类评价活动提供了良好的理论背景。此外,相关的研究还包括近年来比较流行的元分析理论(meta-analysis)、线性结构模型理论(lisrel)、决策模型理论等。

由于众所周知的原因,我国心理测量学的发展在大陆大约停滞

了 30 年。但从 20 世纪 70 年代末期开始,我们用不到 20 年的时间走完了西方心理测量学近百年的发展之路。中国的心理测量学家首先在部分高校开设了心理测量课程,恢复了专业人员的培养。在此过程中,还迅速消化和吸收了西方的测量理论,修订了一些国际上比较著名的标准化纸笔测验,如林传鼎、张厚粲修订的《韦氏儿童智力量表》、龚耀先修订的《韦氏成人智力量表》和《艾森克人格问卷》、刘永和与辽宁教科所修订的《卡特尔十六种人格问卷》以及宋维真与香港张妙青合作修订的《明尼苏达多项人格问卷》等。这些心理测验首先在精神卫生、特殊教育、学前和小学教育等领域为人们所接受,并被广泛应用。戴晓阳等(1993)的调查表明,从 1979 年至 1989 年,在我国临床心理学中使用的测验已达 80 多种,而使用频率最高的是智力测验和非投射性人格测验。

进入 20 世纪 80 年代中期后,测验的开发和应用工作有了一些新的特点。首先,由修订国外开发的测验发展到开始编制一些完全是针对中国人的测验,如在中国儿童发展中心支持下编制的 0~3 岁和 3~6 岁两个年龄组的《中国儿童发展量表》;其次,将测验的编制和修订工作拓展到了职业定向及人事选拔等多个方面,如戴忠恒等修订的《基本能力测验》和凌文轻等修订的《Holland 氏职业定向测验》等。此外,心理和教育测量的使用也逐渐扩大到了军事、人事等多个领域。

测量学家对我国考试制度进行了广泛的研究。20 世纪 70 年代末,北京师范大学张厚粲领导的研究小组就开始采用经典测量理论的方法,致力于大规模教育考试的研究。他们首先研究了普通高等学校入学考试,发现当时的考试存在着题目分布不合理、年度间难度差异较大、主观性试题过多、主试的评分者一致性系数较低等一系列问题。上述研究结果逐渐引起了国家有关部门及各专业学科教师的重视。故此于 20 世纪 80 年代中期,由国家教育考试机构组织心理和教

育测量学专家在国内举办了多次测量学、考试学普及和研讨班,使得考试的组织者们以及参加学科命题和阅卷的专家们都掌握了一些必要的测量学知识,并将此应用于我国的高等学校入学考试,明显地提高了考试的质量,并节省了大量的人力和物力。20世纪80年代中期以后,随着对教育测量学的进一步理解,也基于实践的需要,首先将高校入学考试与中学毕业会考从性质上加以区分,对它们各自的特点和规律开展深入研究,同时将标准参照测验的研究推广到高等教育自学考试以及各种职称和资格考试,并从20世纪90年代开始逐步采用较为先进的项目反应理论和概化理论来提高上述各类大规模考试的质量,取得了明显的社会效益。

此外,测量学家还积极参与相关的社会实践活动,如参与了军队的人员招募和安置、协助进行国家公务员考试,并为各类企事业单位进行了人员招募和选拔工作。

纵观近百年的测量学发展史,我们发现,当人们在谈论测量理论这一问题时,通常指的是如何应用数学方法对测验的可靠性和有效性进行分析和评价。这些在实践上固然重要,然而在重视测量学的统计理论的同时,人们大多忽视了测量内容的理论框架。例如,人们在不尽完善的智力理论基础上编制智力测验,在不尽完善的人格理论基础上编制人格测量。在这种情况下,测量学家编制出来的测验可能会过于依赖在编制过程中所采集的数据。这种方法也称为数据导向的方法(data-driven-approach)。这种思想方法作为测量学领域的主导流行了近百年。随着认知心理学的发展,在其思想影响下,心理学家越来越重视人类心理活动的内部加工过程及其知识结构,智力、人格及各学科心理学的研究均取得了新的突破。而有关上述心理品质的测量也开始出现理论导向方法(theory-driven-approach)的趋势,即在相对成熟的心理学理论的基础上完成测验编制的过程,并且将有关测量内容的理论框架也视为测量理论研究中的一个部分。在不

远的将来,测量内容的心理学研究和测量过程的统计学研究两者的有机结合,将给心理测量学的发展翻开新的一页。

与此同时,另一个需要引起人们注意的是情境测验和计算机模拟测验的发展和应用。标准化纸笔测验可以很好地度量人们的知识水平和认知加工过程,但不足以考察人们的复杂的个性特征和高级的认知加工过程。而情境测验和计算机模拟测验可以在上述领域极大地弥补纸笔测验的不足。就目前的文献来看,对情境的设计、主试的培训、结果的分析和综合等各步骤还缺乏明确的指导性的理论,有关研究还待发展。但我们相信,在今后几年内,这一领域将日益受到人们的重视,其研究也将不断深入,相关的理论必将成为心理测量学理论框架中一个有机的组成部分。

参 考 文 献

- [1] 陆震. 中国考试制度的演变及发展. 中国考试, 1991(1): 16~22
- [2] 张厚粲. 谈谈关于考试中的知识和能力问题. 中国考试, 1991(1): 36~38
- [3] 张厚粲, 丁艺兵. 心理测验理论及其发展. 教育研究, 1988(3): 67~71
- [4] 张厚粲, 张敏强, 陈富国. 教育测量理论与方法的应用研究. 中国考试, 1993(3): 14~20
- [5] 张厚粲, 龚文. 心理测量学的现状与发展. 心理学探新, 1994, 14(1): 1~4
- [6] 戴晓阳, 郑立新, Ryan J J. 心理测验在中国临床心理学中的应用以及与美国资料的比较. 中国临床心理学杂志, 1993, 1(1): 47~50
- [7] Feldt L S, Brennan R L. (1989). Reliability. In Linn R L

- (Ed.), *Educational Measurement* (3rd ed. , 105~146). New York:MacMillan
- [8] Messick S. (1989). Validity. In Linn R L (Ed.), *Educational Measurement* (3rd ed.). New York:MacMillan
- [9] Snow R E, Lohman D F. (1989). Implications of cognitive psychology for educational measurement. In Linn R L (Ed.), *Educational Measurement* (3rd ed.). New York:MacMillan
- [10] Sternberg R J. (1989). Cognitive psychology and psychometrics. In Hambleton R K, Zaal J N (Eds.), *Advances in Educational and Psychological Testing*, 367~393
- [11] Sternberg R J. (1992). Assessment of Intelligence. In Haywood H C, Tzuriel D (Eds.), *Interactive Assessment*, 234~267

以教育测量学为基础的 中国大陆大学入学标准化考试 改革实践与发展概述

马金科

国家教委考试中心,100816

在中国大陆以招收应届高中毕业生和未参加工作的高中毕业生为主的 1080 所普通高等学校,实行全国统一考试,即本文特指的大学生入学考试。国家教委规定,每年 7 月 7、8、9 日,全国同一时间举行考试。考试命题、考试实施、评卷、成绩报告及分数的统计分析工作由国家教委考试中心负责。运作体制分两级管理,在国家教委考试中心统一部署下,各省、自治区、直辖市专职招生考试机构负责考试的具体实施。考试的规模,1994、1995 这两年稳定在每年考生 250 万左右,每年录取新生 90 万,录取率在 36%左右。

上述考试在大陆俗称高考(以下简称高考)。自 1949 年以来,高考的形式由学校单独考发展到学校联合考、全国统考,中间又有 10 年时间废除了高校招生考试,最后又恢复全国统考,至今已经过 40 多年。它经过了曲折的发展道路,一直为社会各界关注,是中国大陆最敏感的社会化大型考试。由于它自身既有优势又存在弊端,在“文化大革命”一开始,就被“左”的路线夸大其弊端,错误地废除达 10 年

之久(1966~1976),这不能不说是一个深刻的历史教训。

本文不是就教育测量学的一般原理来进行论述,而是着重阐述该项考试是怎样以教育测量学为理论基础,用10多年时间进行标准化考试改革实验,取得初步成功并遇到了一些问题的情况,以求教于海峡两岸心理与教育测量学界的学者、专家们,把大家共同关心的大学入学考试办好。

一、高考实施标准化改革的背景

(一) 1977年恢复高考之后,对考试制度和办法都做了不少修订和完善工作。到1981年,社会各界对恢复高考都表示赞成,但同时也反映对它被错误地废止前存在的一些弊端改进甚少,改革的要求强烈。从教育测量学和教育学的观点来看,其主要弊端有:

(1) 考试缺乏明确、具体的要求,以教学大纲代替考试大纲,对考试内容、范围及知识能力的层次要求不够明显。

(2) 缺乏有效的控制试题难度的措施,致使有的学科年度之间的难度大起大落,对中学教学产生不良的“反拨”作用。

(3) 由于高度的保密要求,考试前缺乏试题征集、筛选环节。命题组织虽然比较严密,但多靠临时突击完成,致使题量较小、题型单一、知识覆盖面较小,影响测量考生的真实水平。

(4) 试卷的印制、包装、评卷、统计工作基本上都是手工完成,人力、财力消耗较大,也容易出错。评卷误差的控制手段也不完善。

(二) 考试制度虽然在中国已有1300多年的历史,中国是考试制度的发源地,为人类社会的进步作出了不可磨灭的贡献,但是不断发展和完善的教育测量理论、考试科学的最新成果却较少被有效的应用,与世界发达国家相比,还有较大差距。减少高考存在的弊端,缩小与国际水平的差距,明确高考改革的目标,已迫在眉睫。所以,高考