

第一章 教育评价的历史与发展

研究教育评价的历史，无论是对高等教育的评价，还是对中等教育的评价或初等教育的评价，都是有意义的。研究教育评价的历史，无非是要人们了解其过去，但这并不是目的，真正的目的在于研究现在，探索未来。

教育评价学作为一门学科，同其他学科一样，也是经历过一个由产生到发展的过程的。由于教育科学的不断发展，其门类和分支越来越多，教育评价学就成了其中的一门重要学科。但教育评价可以说是随着学校教育的产生与发展而逐步发展起来的。

由于教育评价学是一门新兴的学科，所以在某些基本理论问题上仍然存在着一定的分歧，这正是新兴学科的一个特点。正因为如此，还应对它的历史发展进行必要的探讨。许多教育评价理论工作者已经从不同角度阐述了各自的见解，这对教育评价的历史研究是有参考价值的。

教育评价经历了从主观评价到测量、从测量到评价这样一个发展过程。其中一些基本问题，是开展教育评价实践活动应该有所了解的。

第一节 教育评价的源流

研究历史问题，必须尊重史实，必须进行必要的考证，根据文献资料来证实、说明，作出科学的判断。研究教育评价的历史，当然也要遵循这一原则。但是，由于每个人所依据资料的不同，可能得出不同的结论，这也是正常的现象。对同一资料，不同的人也可能做出不同的判断。所以，在学术上要重视对不同观点的研究。

研究教育评价的历史，既要考证国外的文献资料，又要考证本国的文献资料，不能因为我国的评价理论研究起步晚而变成“被遗忘的角落”。就我国的情况而言，应对科举制度及陶行知的教育评价思想进行必要的研究。

一、中国的科举制度与教育评价

追溯评价活动的历史，人们就会联想到古代的一种习俗。很多民族都重视这样一种礼仪活动：当孩子达到一定年龄时，必须从长者那里接受各种训练，主要是生活技能和劳动技能等方面的训练。只有经过这种训练而具有必备能力的人，才能得到大家的认可，取得成人资格。这就是原始部落中进行能力测定的活动。

评价是客观存在着的。“评价”一词最早出现在中国北宋时期。《宋史·戚同文传》记载：“市物不评价，市人知而不欺。”其原意在评论货物价格，而现在泛指衡量人物或事物的价值。即便从“评价”一词的原意来看，由于价格是价值的货币表现，也应当理解为评定价值。所以说，“评价”这一概念中国自古就有，并非

是从英文翻译过来的。

中国是考试的故乡，历朝历代都有考选制度，通过考试来选拔贤才。教育评价，就其历史发展来说，源于我国的隋朝。从对人才的选拔和录用上看，我国隋朝科举制度的影响是深远的。这一观点得到国内外的专家和学者的赞同。例如，日本的教育评价专家在谈到教育评价的源流时认为：从体系上看，使古代国家具有划时代的合理性的，是公元 6 世纪中国隋朝所制定的完备的科举制度。

关于科举制度的起始年代有三种说法：581 年、587 年、606 年。实际上，581 年是隋文帝（杨坚）建立隋朝的元年。隋文帝为了废除为世族所垄断的九品中正制，于开皇七年即 587 年，开考两种科目：志行修谨科，用于选拔品行出众的人；清平干济科，用于选拔才华出众的人。隋炀帝（杨广）在开皇二年即 606 年，增设进士科，开科取仕，即通过科考来选拔官吏，一般认为这是科举制度正式建立的标志。

科举，这是了解中国历史的人所熟知的。科举，就是分科取仕，即设科考试，再根据科考的成绩来录用官吏。所谓科举制度，是指从隋以后各封建王朝设科考试选拔官吏的制度。这个制度是以才能取人，主要看文才，打破了以门阀取人的九品制。科举制的常设科目有进士科、明经科、明科（律学）、明学科（书学）及明算科等，还有一些非常设科目。科举制虽历经隋、唐、宋、元、明、清各代，考试科目自宋代以后有所变化，但进士科却是历代都沿用的。

唐朝规定，通过进士考试后还要经过吏部考核。考核的标准有四条：其一是“身”，即仪表，标准是体貌丰伟；其二是“言”，即言论，标准是言词辨正；其三是“书”，即书法，标准是楷法遒美；其四是“判”，即文字逻辑，标准是文理优长。

宋朝规定考核在考试之前进行，标准分为五等，其要求是：一等学识优长，阅理精纯；二等才思赅通，文理周率；三等文理俱通；四等文理中平；五等文理疏浅。

元朝的科举分为文武两科，但重视经文，不许自由发挥，严防思想反抗。为了加强对官吏的监察，元朝将全国划为 22 道监察区，形成从中央到地方的严密监察系统。

明朝的科举也分为文武两科，但朝廷重文轻武。自明代起正式实行八股文取仕的办法，考试分为三级：一试为县试，二试为府试，三试为隐试。

清朝继续实行科举制，考试主要分四级：童试、乡试、会试和殿试。其中童试又分三级，即县试、府试和院试，是科举考试的初级形式。清朝还规定，三科以上考试不中的举人，挑选其中一等的任知长，二等的任教职。^①

科举制度的突出特点是：制度齐全，方法完备，等级分明，标准严格。科举制度成为封建地主阶级为本阶级选用人才的工具，这是由当时社会的经济、政治条件所决定的，也是地主阶级政治代表为限制门阀贵族权力，吸引人数众多的中小地主及知识分子参与政权而对人才考选制度的改革。这个制度在当时的历史条件下还是非常进步的，所以历经隋、唐、宋、元、明、清各个朝代，直至 20 世纪初的清末（1905 年）才在资产阶级的维新运动中被废止。这是我国古代历时最长的选拔人才的制度，共持续了 1300 多年。

国外有的学者认为，科举制度与教育是毫无关系的。我们认为，这种观点是没有根据的。其实，科举制度与教育是有关的，而且相关度也是比较大的。

陆红军：《人员测评与人事管理》，河南人民出版社 1987 年版 第 27—28 页。

事实之一：北宋中期，王安石实行变法的目的是为了富国强兵，缓和阶级矛盾，挽救封建统治的危机。此时，办学形式是把学校分为外舍、内舍和上舍，从外舍升入内舍和从内舍升入上舍，均须经过考核。只有进入上舍，才有资格参加科举考试，合格者得以授官。

事实之二：明朝规定，凡参加科举考试的人，都必须是学校培养出来的人，其他人一般不能参加科举考试。

事实之三：唐朝规定，参加科举考试的人，先由六学（国子学、太学、四门学、律学、书学、算学）和二馆（弘文馆、崇文馆）以及地方州县学校选送至尚书省，地方选送请求报考的自学士子至尚书省，均须经过审查才能参加科考。

由此可以看出，科举制度作为一种选拔考试制度，非常重视从学校培养出来的人员中选拔优秀者。从教育的目的上看，任何统治阶级都依靠学校为其培养人才，为维护本阶级的利益服务。

二、科举制度对西方国家的影响

中国的科举制度对西方国家公务员制度的建立，发生过深刻的影响，而这种影响又是以多种形式起作用的。凡在中国任过官职的外国人或到过中国的传教士，都对科举制度表现出浓厚的兴趣。

意大利旅行家马可·波罗于 1275 年沿着丝绸之路来到中国，在元朝任官 17 年。其间，他借工作之便，深入到中央和地方的政治活动中去，对中国的文官制度非常熟悉。他在《马可·波罗游记》一书中对中国的公开考试、择优任官的制度大加赞扬，使西方人有机会了解东方文明的精华。这本游记的出版时间不详，如果推算到 1292 年，那么距今已有 707 年的历史了。

明朝时期，继马可·波罗之后来到中国传教的是利玛窦。他在中国生活了 28 年，结交的人物非常多，交往较密切的士大夫有 160 余名；就人物阶层而言，有一品官 14 人、皇室成员 140 人、进士 10 人。利玛窦在中国的长期生活中，对中国文化产生了极大的兴趣，他把中国的文官制度作为向西方介绍中国的重要内容。

葡萄牙传教士克鲁兹的《中国游记》是 1569 年出版的，该书对中国通过科举制度选拔出来的官员倍加称赞。葡萄牙修士胡安·贡萨雷斯·德万多萨，1583 年出版了《伟大的中国》一书，其中第 14 章详细介绍了科举制度。这本书在罗马出版以后，很快被译成多种文字，广为流传，引起欧洲人对中国科举制度的重视。

英国在 1570—1870 年的 300 年间，用英文出版的有关中国文官制度、政治制度的书籍达 70 种之多。其中极力称赞中国科举制度的有《文官考试制度》、《中国札记》、《中国的历史与现状》等书。这个时期，英国驻华使节也把中国的文官制度作为向本国政府汇报的内容。尤其是在 1855—1870 年英国文官制度的形成时期，英国著名刊物《绅士》、《伦敦》和《雾》等杂志先后介绍中国录用官员的程序及方法，极力主张英国政府实行中国式的文官考试制度。

美国的文官制度也受中国科举制度的影响。1866 年 10 月，北京同文馆馆长马丁（W. A. P. Martin）曾在波士顿给东方学会作过题为“中国的竞争考试”的报告。1867 年 10 月，《北美评论》杂志盛赞中国的科考取仕。1868 年 5 月，罗得岛州议员托马斯·詹科斯（Thomas Jenkes）在向国会提交的报告书中，曾辟专章论述中国的文官制度。1870 年，斯皮尔（W. Spear）在其著作《中国

与美国》中也称赞中国的文官制度。^①

中国的科举制度对日本的影响也很大。例如，7世纪的《大宝律令》中就采用了这种办法，决定仿唐科举，建立选拔官吏的制度。明治维新之后，文官分为亲任、敕任、奏任和判任四等，就是借鉴唐朝官制中的制授、敕授、旨授和判补。从明治维新以后到第二次世界大战之前的高等文官考试，以及现在的国家公务员高级职员考试制度，都残留着科举考试的影响。

科举制对各国公职人员的录用考试产生了重大的影响。从19世纪后期起，西方国家都陆续建立了文官考试制度。英国于1855年建立了“文职人员委员会”，采用竞争性的公开考试来招募文职人员。美国于1883年成立了“文职人员事务委员会”规定担任公职必须经过考试。意大利在宪法里规定，为了升入各种和各级学校或从学校毕业，以及获得就业资格，均须经过国家考试，国家行政机关的官吏也经考试选拔录用。

教育评价是随着学校教育的发展而发展的。在人类历史上，自从系统地实施教育的机关在社会上出现以后，可以说教育评价就出现了。在学校教育中，评价活动历来都是教师和管理者的常规工作。例如，入学招生、成绩评定、学年升级、毕业资格的认定等，都是从管理目的出发进行的评价活动。考试是评价的手段，考核制度在学校教育中占有重要地位。

在中国，两千多年前就已经有了关于考核制度的论述。我国最早的一部教育专著——《学记》中记载：“比年入学，中年考校：一年视离经辨志，三年视敬业乐群，五年视博习亲师，七年视

^① 李盛平、陈子明：《职位分类与人事管理》，中国经济出版社1986年版，第198—199页。

论学取友，谓之小成。九年知类通达，强立而不反，谓之大成。’^①意思是说，学生到了规定的年龄入学，国家每隔一年考查他们学业及操行成绩一次：第一年考查学生析句分段的能力和学习的志向；第三年考查学生是否专心学习和对周围的人是否和睦相处；第五年考查学生学识是否广博，同老师是否亲密无间；第七年考查学生研究学问的本领和识别朋友的能力。符合标准的就叫做小成。第九年，做到认识事物能触类旁通、闻一知十和政治上成熟，以及立场坚定不移。符合这一标准的就叫做大成。^②由此可见，我国早期学校对评价活动的内容、标准及时间已有明确的规定。这部教育经典著作对其他国家的影响也是很大的。日本的谷口武在《学记论考》中说：“像这样一本书对日本教育史所发生的影响，是极为罕见的。”

从教育评价的方法论来看，以前主要是采用口试法和面试法，而现代所用的笔试法则伴随着学校教育的普及而出现的一种新方法。我国素有“考试故乡”之称，考试经验很丰富。从科举开始将以口试或面试为主改为以笔试为主，使考试制度有了新的发展。欧美许多国家都相继借鉴我国的考试经验。例如，英国是一个热衷于考试的国家，笔试法在剑桥大学是从1702年开始被采用的，而“数学荣誉学位”考试则是在1747年到1750年间搞起来的。这种荣誉考试制度，1839年被称为是“一种以最少的劳动和时间的代价，去鉴别成绩的完全有效的方法”。笔试法在美国是1845年在波士顿市教育当局开始使用的，当时在中小学阶段推行笔试是作为一项教育改革来实施的。

根据以上所谈的内容，关于教育评价的源流可以得出这样

高时良：《学记评注》，人民教育出版社1983年版第1页。

高时良：《学记评注》，人民教育出版社1983年版第34—35页。

的结论：“评价”这一概念在中国的北宋中期就已经出现了，并非是从外国翻译过来的；教育评价的源流在中国，主要源于科举制度；现代教育评价的许多方法，应该说是这个源流的延续和发展。孙中山先生曾说：“现在各国的考试制度，差不多都是学英国的。穷流溯源，英国的考试制度，原来还是从我们中国学过去的。”卡特总统当政时的美国人事总署署长艾伦·坎贝尔教授说：“当我接受联合国的邀请来中国向诸位讲关于文官制度的时候，我是感到非常惊讶的。因为在我们西方所有的政治学教科书中，当谈到文官制度的时候，都把文官制度的创始者归于中国。”

第二节 教育测量运动

教育评价的发展史，是以追求其客观性而展开的。因此，探讨教育评价的由来与发展，必须对教育测量有所了解，因为测量是以追求客观的量化结果为目的的。

在教育过程中，教师必须对学生的学业成就做出考查和评定。那么，如何做好这个考评工作，并使其更好地发挥在教育中的作用，则是教育测量应该去完成的任务了。

教育测量要确定学生知识和能力发展变化的现状及差异，并把其结果用数量描述出来，就要通过考试或测验来解决。换言之，教育测量主要是通过考试来使学生的学业成就数量化的。这就是说，讨论教育测量就必然要涉及到考试，而考试又是学校的常规活动，因此可以认为自学校产生以来就有了教育测量活动。讨论教育测量，从考试的演变说起，无非是要从其历史的发展中了解和把握发展的趋势。

一、考试演变与教育测量

由于文化科学技术的发展和社会物质条件的提高，考试的内容和方法也在不断地变化发展。就方法而言，西周的考试方法仅有口试、实际操作和演示；两汉的察举制和魏晋南北朝的九品中正制以推荐为主，考试为辅，隋唐以后的科举制则是以考试的成绩作为录取的标准。

西汉和东汉处于我国封建社会前期，由于统治比较稳固，历时又长，形成了以察举制为主的人才选拔制度。察举制，主要是由中央或地方长官先推荐人选，然后经过考试，合格者授以官职。为了保证察举制的推行，统治者还规定对推荐者和被推荐者均给予赏罚。如被推荐者考核为优秀，推荐者有赏；有弄虚作假，治罪。如不举者，不胜任也，当免。这个时期的考试，分为对策和射策两种形式，对策是命题考试，射策是抽签考试，两者均为笔试。

魏晋南北朝时期的九品中正制，就是各地方政府设立中正官，负责向中央政府推荐人才。州中正官叫作大中正官，负责举荐小中正，而小中正举荐的人才需经大中正核实。中正官举荐人才依据家世、才能和德行三项内容，并评为九个品级：上上、上中、上下、中上、中中、中下、下上、下中、下下。九品中正制等级严明，具有一定的进步意义，但后来逐步变成了“上品无寒门，下品无士族”的门阀制度。

唐朝是科举制度的兴盛时期，制度完备，方法多样。这个时期的考试方法除了口试和演示以外，还有帖经、墨义、策问、诗赋等方法。帖经，是将儒家经典掩其两端，中间唯开一行，再裁纸为帖，帖盖三个字，让考生写出这三个字，类似现在的填空题；墨义，是取儒家经典中的句子让考生对答，或对答本句含义，或

对答下一句，类似现在的名词解释或简答题；策问，是以历史典故、重大理论问题或现实社会问题为题，让考生分析并提出自己的主张及解决问题的办法；诗赋，是选儒家经典中的主题思想，以前人诗句或景物为题，规定韵脚，让考生按照声韵、格律作诗，类似现在的命题作文。评阅试卷时，只在卷面批上“通”或“不”一个字即可。但清朝评卷采用的是五个等级，即用“○、△、\、一、×”五个符号来表示。

综观考试的历史，人们可以看到：考试是随着人类社会的形成而产生，随着社会的发展而演变的。在漫长的历史长河中，考试形式的发展变化是十分缓慢的。虽然考试内容不断变化，但考试方法的变化却不大，基本局限于口试或面试。直到我国建立科举制度之后，才出现了由以口试为主到以笔试为主的重大变化。这里所说的“为主”，并不排斥使用其他的考试方法，实际上以笔试为主兼用其他方法的考试，一直沿用到今天。

考试作为一种制度，以其独立的体系被广泛地应用于社会生活的各个领域。从考试的历史发展看，分为社会考试和学校考试两大系统。社会考试的职能，一是遴选人才，二是评价学校教育质量；学校考试的职能，一是评价学生的学业成就，二是评价教师的教学质量。

从学校考试来看，其基本任务是通过笔试或口试来考查和评定学生的学业成就。但是，哪一种方法更为客观呢？这是教育专家和学者一直在研究的重要命题，也是教育实际工作者广为关心的话题。一般认为，笔试比口试优越。其理由是笔试的客观性和信赖性都比较强，而口试则相反，其客观性和信赖性都比较差。国外有的学者正是从这种观点出发，开始批判口试法的主观性，积极推行和普及笔试法。持有这种主张的代表人物是美国的贺拉斯·曼(H. Mann)和怀特(E. E. White)。他们是从

1845年开始推行笔试的，并提出以下理由：

第一，借助于笔试，可以得到更多的学力资料。

第二，关于学力的记录，可以保存下来。

第三，对每个学生都使用统一的试题，既能做到一视同仁又能方便于组织实施。

第四，对特定的学生或教师能减少偏袒和歪曲的可能性。

通过考试的实践使人们看到，笔试法的客观性确实高于口试法，因而使教育评价向客观化迈出了一大步。这是评价理论研究和评价实践活动相结合所取得的成果。

但是，应该看到，笔试法与口试法相比，其客观性如何也是相对而言的。从教育实践来看，学校历来的评价方法都带有一定的主观性，即便是前面谈到的笔试法也不例外。对此有人做过试验：将同一份中学英语作文试卷发给 142 名教师分别评定，结果一致性最高的分数是 80 分，只有 13 名教师，占 142 人的 9.5%；而最低分是 50 分 最高分是 98 分，两者相差 48 分。究其原因，是无客观的统一标准，只凭教师的好恶之感去评分。由此不难看出，即使是笔试法也有不客观的地方，即有主观性的一面。为解决这个问题，就需要研究评定成绩的具体标准，力求提高评价的客观性，降低其主观性。

在教育评价的历史发展中，为谋求评价客观性而做过努力的有英国的菲希尔（G. Fisher）和美国的赖斯（J. M. Lise）。菲希尔在 1864 年以校长的身份，根据价值程度制定了评分标准，其评定分数为从 1 到 5，并把它用于对本校学生的答案及作品的客观评分。赖斯以巴尔的摩市教育长的身份，从教育研究的目的出发，对当时偏重于读、写、算的学校教育课程进行改革。为了取得可信的量化资料，赖斯对拼字、阅读、计算等学力检查进行了客观的测定。其中，1894 年设计的拼字测验被视作早期的

测量。他的这一举动在当时产生了很大影响，因而被称作是教育测量的先驱。

二、心理测验与教育测量

心理测验产生于社会需要，而心理测验的产生与发展又促进了教育测量的发展。

心理测验方面的早期探索者，是英国的差异心理学的创始人高尔顿 (F. Galton)。他的学术观点，深受达尔文生物进化论的影响，致力于个体差异与心理遗传的研究。高尔顿于 1884 年创建了人类测量实验室，1904 年又在伦敦大学创办优生学实验室。他研究个体差异的测量项目包括身高、体重、肺活量、拉力和握力、扣击的速率、听力、视力、色觉等。他认为形成能力的个体差异的原因在于遗传，从调查 1768—1868 年这 100 年间英国的首相、将军、文学家和科学家的家谱中发现，在 977 位名人中有 332 位杰出人士，而普通人中 4000 人才产生一位杰出人士。他还认为，人的所有特质都可以定量叙述，这是建立人类科学的必要条件。所以他就最早应用统计去处理心理学研究资料，以数量代表所测得的心理特征差异。高尔顿提出的心理测量工具，不仅为后来心理测验的研究工作铺平了道路，而且为教育评价工作的量化技术提供了重要条件。

美国心理学家卡特尔 (J. M. Cattell) 在莱比锡大学冯特心理学实验室从事个体差异研究，1891 年他来到哥伦比亚大学担任心理系主任，在那里创立了一个实验室，开展研究达 26 年之久。卡特尔的研究可归纳为六个项目，即反应时间、联想、知觉和阅读、心理物理学、等级评列法和个体差异。但他的研究是以心理测验和个体差异为主题的，而且重在测量基本的身体活动或感觉——运动反应，因而他的心理测验不同于智力测验。卡

特尔使心理测验走出了实验室，为应用心理学的发展开辟出一条新的道路。他编制的心理测验达 50 多种，1890 年发表了名为《心理测验与测量》的论文，对实验进行了总结。

20 世纪，心理测验在理论研究与实践相结合上又向前迈进了一步。在这方面做出突出贡献的是法国心理学家、智力测验的创始人比纳（A. Binet）。他在心理学方面的研究范围很广泛，除了探讨催眠术以外，还有变态心理、个别差异、无意识思维学。1889 年，他在巴黎大学创建了法国第一个心理学实验室。比纳在心理学方面最重要的贡献，是对智力的研究和对智力测验方法的创新。通过研究，他认为智力是极其复杂的能力，不可能用简单的方法去研究广泛而复杂的课题。正是从这一观点出发，他不断地探索智力测量问题。为了研究公立学校低能儿童班级管理问题，1904 年比纳受法国教育部的委派，和许多教育家、医学家组成了一个专门委员会，他是委员之一。他主张用测验的方法去辨别心理有缺陷的儿童。1905 年，比纳与西蒙合作编成了《比纳—西蒙智力量表》，后经几次修订，成为智力年龄表。这个量表有两大特点：一是项目种类多，可以测量智力的多方面表现；二是排列由浅入深，使用同一量表可以测量智力高低不同的儿童。

目前，世界上的智力测验虽然众多，但基本原理和主要方法都是建立在比纳的量表的基础之上的，因此，该量表对心理学和教育学的影响都很大。从心理学方面看，智力测验已成为心理研究、心理诊断的必备工具；从教育学方面看，智力测验引起教育家的注意，开始编制教育测验。所以，心理测验的产生与发展又促进了教育测量的发展。

三、教育测量运动

心理学的基本原理及其方法，深刻地影响着学校教育和评价方法。虽然教育测验与心理测验在内容上有所不同，但在方法上又是相通的，因此，教育测验和心理测验必然会相互影响、相互促进。

教育测量形成为一次运动，还是 20 世纪初的事，其代表人物是美国心理学家桑代克（E. L. Thorndike）。他在大学读书时就进行过动物学习心理实验，即小鸡走迷津的训练和动物联想过程的实验研究。桑代克的成就在于首创用实验法来研究动物心理，以严格控制的实验研究代替自然观察，为动物心理研究谱写了新的篇章。他创造了迷路网、迷箱和迷笼等实验工具，研究鱼、鸡、猫、狗等动物的学习。他认为动物学习的基本方式是通过尝试错误而偶然获得成功，从而在刺激与反应之间形成连结，即在感觉神经与运动神经之间形成连结。

桑代克把“连结”这一概念推广到人类心理，把动物研究技术用于对儿童和年轻人的研究。后来，他又进一步研究人类学习，提出在应用学习定律时要考虑个体差异，并从这一观点出发改进学校教育工作。1904 年，桑代克发表了闻名于世的论著《精神社会测量学导论》；介绍了统计方法及编造测验的基本原理。该书提出了“凡存在的东西都有数量，凡有数量的东西都可以测量”的基本观点，为教育测量奠定了理论基础，对教育测量学的建设与发展作出了巨大贡献，因而桑代克被称为“世界教育测量之父”。

1908 年，桑代克依靠其弟子斯通（C. W. Stone），完成了算数标准学力测验，此测验是最早的标准测验。1909 年，桑代克又亲自完成了写字尺度标准化，又称书法量表，1912 年编制了兴

趣测验,1915年提出了职业测验,1917年又编制了办事员的能力测验,等等。此间,其他学者也编制了许多测验,如语文量表、作文量表、算术标准测验、职业兴趣测验等。这样,作为教育测量的代表工具——关于各科教学的学力标准测验已经陆续完成,在大约20年的时间内先后共发表了1300种左右。这些测验应用于学校教育调查和大规模的教育测验,从而促进了教育测量的发展。这样的测量活动体现了桑代克的一种思想:“凡存在的东西都存在于某些数量之中。所谓完全了解它,就意味着了解其质,同时也了解其量。”经过桑代克和其他学者的共同努力,教育测量工作得到了蓬勃发展。后来,人们就把上述测量工作称为教育测量运动。

西方教育测量运动的发展,对中国也有较大的影响。1918年,俞子夷编制了小学国文毛笔书法量表,这是我国最早的标准测验。1920年,北京高等师范学校和南京高等师范学校建立了我国最早的心理实验室。同年,廖世承和陈鹤琴在南京高等师范学校开设了测验课程,并用心理测验对报考该校的学生进行考查,后来出版了二人的合著《智力测验法》。该书被认为是我国最早的介绍心理测量的著作。

1922年,美国教育测量专家麦柯尔(W. A. Mecall)应中华教育改进社的邀请来华讲学。在他的指导下,北京师范大学、北京大学、燕京大学、北京女子高等师范、东南大学等校开始编制测验,共编了40多种。麦柯尔对这些测验水平的评价是:“至少都与美国的标准相等,有许多种甚至比美国的还好。”中华教育改进社还于1923年组织了全国性的小学教育调查,涉及22个城市和11个乡镇,共测验儿童9.2万名。^①

郭述平、王景英:《教育测量》,东北师范大学出版社1988年版,第9页。

1924年，陆志苇发表了修订的《中国比纳—西蒙智力测验》。他与吴天敏于1936年又进行了第二次修订，1979年吴天敏进行了第三次修订，1982年出版了《中国比纳测验》。此外，中国还出版了一些教育测量方面的专著，如陈选善的《教育测验》、王书林的《心理与教育测量》、廖世承和陈鹤琴的《测验概要》、艾伟的《小学儿童能力测量》、孙邦正的《心理与教育测验》等等。

进入20世纪以后，主要是指从20年代到30年代，促进教育测量急速发展的因素是：

第一，自然科学迅速发展的影响。

回顾历史，物理学、化学、天文气象、医学等自然科学，在过去的几个世纪内都没有什么大的进步，但自从运用了测量的方法以后，得到了飞速的发展。这个事实对社会科学的影响很大，同时也促进了教育学和社会学等学科的发展。在这种情况下，教育家开始运用心理实验方法去进行教育实验，应用测量法去衡量教育成果。这说明，采用测量法而获得成功的自然科学，已成为教育测量的范例。

第二，个别差异研究的影响。

个别差异的研究，是就人的身体和精神等方面的能力、气质及性格去探讨其差别。这一研究的目的，不仅仅是为了了解学生，而更重要的是为了根据其差别进行教育。比纳和西蒙于1905年根据法国教育部的意见，为解决法国教育工作中所面临的低能儿童的教育问题，共同制订了智力测验量表。随后，这一量表逐步为社会普遍所利用，同时也引起了教育科学研究工作者的重视。

第三，教育统计发展的影响。

在进行个别差异研究的过程中，为了分析所收集到的关于身体特性和心理特性的资料，找出个别差异所在，就必须进行相