

全国高等院校教育硕士规划用书

全国高等院校教育硕士规划用书



化学教学测量与评价

李广洲 任红艳 编著



科学出版社

全国高等院校教育硕士规划用书

化学教学测量与评价

李广洲 任红艳 编著

第1版 (917) 目录

第一章 绪论 1
第一节 化学教学测量与评价的概念 1
第二节 化学教学测量与评价的作用 2
第三节 化学教学测量与评价的类型 3
第四节 化学教学测量与评价的原则 4
第五节 化学教学测量与评价的编制 5
第六节 化学教学测量与评价的实施 6
第七节 化学教学测量与评价的评分 7
第八节 化学教学测量与评价的反馈 8
第九节 化学教学测量与评价的总结 9
第十节 化学教学测量与评价的展望 10

第二章 化学教学测量的编制 11
第一节 化学教学测量的编制原则 11
第二节 化学教学测量的编制步骤 12
第三节 化学教学测量的编制方法 13
第四节 化学教学测量的编制工具 14
第五节 化学教学测量的编制评价 15
第六节 化学教学测量的编制总结 16
第七节 化学教学测量的编制展望 17
第八节 化学教学测量的编制附录 18
第九节 化学教学测量的编制参考文献 19
第十节 化学教学测量的编制索引 20

科学出版社

北京

内 容 简 介

本书共分八章,内容涉及化学学科高考的试题和能力层次要求,包括现代测量理论在内的考试质量评价理论和方法,定量测量数据的报道,化学课堂教学(内容处理及教学技能)评价、学生个体(认知和技能)发展评价以及教学实验结果评价,最后介绍使用 SPSS 进行相关分析、线性回归分析和因素分析等。本书内容围绕实际研究问题展开,示例多,语言简洁易懂,便于自学;各章均配有研究性阅读材料,适应不同层次教学。

本书可作为化学学科教育硕士、化学课程与教学论硕士研究生以及化学(师范类)专业本科生的教材或教学参考用书,化学相关专业硕士研究生、本科生进行自学的参考用书,以及广大中学化学教育工作者、教学研究及管理人員的工具性参考用书。

图书在版编目(CIP)数据

化学教学测量与评价/李广洲,任红艳编著. —北京:科学出版社,2014.1
全国高等院校教育硕士规划用书
ISBN 978-7-03-039581-8

I. ①化… II. ①李… ②任… III. ①中学化学课-教学研究-研究生-教材
IV. ①G633.82

中国版本图书馆 CIP 数据核字(2014)第 010992 号

责任编辑:丁 里 郭慧玲 / 责任校对:邹慧卿

责任印制:阎 磊 / 封面设计:陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京市文林印务有限公司 印刷

科学出版社发行 各地新华书店经销

*

2014 年 1 月第 一 版 开本:787×1092 1/16

2014 年 1 月第一次印刷 印张:16 1/2

字数:419 000

定价:45.00 元

(如有印装质量问题,我社负责调换)

前言

根据教育目标为中心的行为模式界说,教育评价是一个通过系统地收集信息,对教育目标及实现目标的教育活动进行分析和价值判断的过程,是实施管理、改进教育的有效手段之一。教育评价可以涉及教育领域活动的各个要素。评价对象可以是学校教育中的学生(学业成就、品德、体能及个性发展),教职员等教育者(课程、教材、教学等),还可以是宏观的教育体制、教育目标、教育发展乃至某地的教育水平等。教育评价的理论与方法最早形成于美国,俄亥俄州立大学的泰勒(Tyler)等学者的研究及成果曾在世界范围内产生了深远的影响。

要判断实际教育活动达到目标的程度,收集和分析各种有关资料信息是必需的。于是,评价与测量就有了联系。可以认为,测量和评价是同一个活动的两个不同的阶段:测量对事物属性赋值,评价则对事物作出价值判断。

鉴于教育测量与评价是教育科学体系的重要组成部分以及其原理、方法对于实际教育工作者具有的重要价值,教育部教育硕士教学指导委员会(教指委)于20世纪90年代在《在职攻读教育硕士专业学位指导性培养方案》中将“化学教育(学)测量与评价^①”列为教育硕士(学科教学·化学)的专业学位必修课;其后,教指委在《全日制攻读教育硕士专业学位研究生指导性培养方案》中又将该课程列为化学学科教学方向教育硕士的专业学位必修课。我国的绝大多数大学的课程与教学论(化学)硕士研究生的培养计划中均列入了学位必修课程“化学教育(学)测量与评价”。南京师范大学20世纪90年代招收了几届面向高中化学骨干教师的化学教学论研究生课程进修班,开设了54学时“化学教育测量与评价”课程,进修学员反映良好;1997年起招收的教育硕士(学科教学·化学)和课程与教学论(化学)硕士研究生的教学计划中都列入了学位必修课程“化学教学测量与评价”(3学分)。开设该课程的目的是使化学教育硕士和化学课程与教学论硕士研究生能够掌握有关的原理与方法并进行实际的化学教学测量与评价,合理地进行基本的统计处理;了解国内外化学教学研究的现状及发展趋势,熟悉常用的化学教学研究方法;培养硕士研究生严谨的科研态度,从而为后继的学位论文研究打下基础,为以后在实际工作中切实提高化学教育教学研究质量打下基础。

作者长期从事化学教育教学研究工作,从事教育硕士(学科教学·化学)和课程与教学论(化学)研究生的培养和课程教学也已多年。在近20年不同类型硕士层次“化学教育(学)测量与评价”课程教学实践的基础上,对所使用的基于实际研究工作的教学材料进行系统地梳理并加以补充,整合形成本书。目的是在方便教学的同时帮助广大化学教师、教学研究人员掌握有关的原理和具体方法,科学合理地进行化学教学测量与评价,从而揭示本质提高结论的科学水平,进而具备和国外同行进行交流的基础,有利于实现化学教育科研从初级的描述、解释到较高层次的预测和控制的跨越。当然也可以帮助新手顺利进入化学教育科研的大门。

化学教育(学)测量与评价所涉及的内容是非常丰富的。本书选取内容的指导思想首先是从实际出发。所谓实际,包括我国化学教育教学研究现状及实际需求,已经和将要从事这方面

^① 由于学校中的具体学科教学活动是整个教育的基本途径,而具体学科教学评价极具重要性,故在具体的化学学科教育教学研究领域,可以提“化学教学测量与评价”。

研究工作的人员的基础(尤其是统计基础)两个方面。具体选取时注意准确把握国内外相关研究现状的差距及发展趋势,有些内容如教学实验设计、多元分析(除因素分析外)、测验等值等在化学教育教学研究中的应用暂未收入;有些内容如题目反应理论(包括 Lasch 模型)等则只作介绍未提具体应用。其次是突出实用性,将测量与评价理论在化学学科教育教学实践中的应用放在首位,既重视基础理论,又突出研究方法和分析技术的使用。将原理、方法和技术的介绍和化学教育教学要研究的实际问题相结合(书中示例均出自作者的课堂教学实践和化学教育教学研究实践),主要章节围绕实际问题而展开。有些内容则以研究性阅读材料的形式给出,研读材料是章节基本内容的补充和延伸。再次,在结构体系上,从我国的高考至考试质量评价,从测量数据的定量处理到测量资料的定性分析,从课堂教学评价、学生发展评价再到教学实验效果评价,最后是 SPSS 应用介绍。全书共八章,测量与评价的原理方法及应用贯穿其中,但各章又相对独立,可以适应不同层次的教学对象。此外,在写作方面,作者充分重视研究示例的作用,语言简洁易懂,以便于读者自学。和化学教育教学测量是一种客观的对所要研究的化学教与学活动属性赋值的过程相比,化学教育教学评价是一种主体性活动,是评价者根据教学目的和按照一定的标准作出的价值判断的活动。也就是说,评价的结论受制于评价者的价值观。对于同一现象,不同的评价者可能会得出截然不同的结论。因此,本书较少涉及化学教育教学评价结论方面的内容。

本书第四章和第五章由任红艳撰写,第三章和第八章由任红艳、李广洲合写,其余内容由李广洲撰写,全书由李广洲统稿。

在成书中,作者参考和引用了一些国内外研究者的专著、文章和研究报告等,在此谨表衷心的感谢。但限于作者自身水平,仍会有从事这方面工作的同行的最新研究成果未能在书中得到反映,恳请有关的专家、同行及读者指正。

李广洲

2013年8月于伯明翰

目 录

前言

第一章 化学学科高考	1
第一节 高考与教育测量	1
一、高考的历史沿革	1
二、教育测量	3
第二节 化学学科高考与能力层次	5
一、化学学科高考改革回顾	5
二、化学学科高考的特征	6
三、考试的分类	7
四、化学学科高考的能力和层次要求	8
第三节 标准化考试及试题命制要求	11
一、标准化考试	11
二、考试大纲	12
研读材料:2012年普通高等学校招生全国统一考试大纲(化学)	12
三、标准化试题命制的基本要求	14
第四节 化学学科高考的试题	16
一、客观性选择题	16
研读材料:猜测校正公式	20
二、主观性化学计算题	21
研读材料:相对评分法	21
三、限制性填空/简答题	21
四、信息给予题	23
第五节 考试结果的整理与描述	24
一、考试的教育功能	24
二、考试结果的收集与整理	26
研读材料:数据的类型	27
三、考试结果数据的描述	30
第二章 考试的质量评价	36
第一节 考试与教育目标	36
一、教育(教学)目标与考试目标	36
二、教育目标分类	37
研读材料:《普通高中化学课程标准》描述目标水平的词语	39
第二节 考试结果的有效性	39

一、效度的概念	40
二、内容效度	40
三、积差相关系数	42
研读材料:相关系数的平均数	46
四、效标关联效度和结构效度	47
第三节 考试结果的可靠性	48
一、内部一致性信度	48
二、评分者信度	50
三、等级相关系数	50
研读材料 I:信效度的理论意义及两者关系	53
研读材料 II:利用肯德尔和谐系数计算评分者信度	55
第四节 考试的项目分析	55
一、难度分析	56
研读材料:多重选择题难度的校正	56
二、点双列相关系数	59
研读材料:二列相关系数	60
三、区分度分析	60
研读材料 I:考试的信度与试题区分度的关系	62
研读材料 II:目标参照性考试的试题分析	62
第五节 考试理论的进展	62
一、经典测验理论	62
研读材料:概化理论简介	65
二、题目反应理论介绍	66
研读材料:拉什模型简介	69
三、认知诊断评价与测量的结合	70
第三章 测量结果的报道	71
第一节 概率与正态分布	71
一、概率	71
二、正态分布	73
研读材料:二项分布	76
第二节 百分等级和百分位数	80
一、百分等级	80
二、百分位数	84
研读材料:从原始数据资料求中位数以及中位数与均数的比较	85
第三节 转换分数及应用	86
一、Z 分数	87
二、多科成绩的总评定	88
三、T 分数	89

研读材料:原始分数正态化	91
第四节 等级计分及等级的数量化	92
一、正态分布计分	92
研读材料:教学班级的均分和标准差控制	94
二、质量等级的数量化	95
第四章 测量资料的定性分析	98
第一节 化学教育科研中的定性方法	98
一、定性分析的特点	98
二、定性分析的过程和注意点	99
三、定性研究与定量研究的比较	101
第二节 访谈法与出声思维法	103
一、访谈法	103
二、出声思维法	104
第三节 编码技术与问题行为图	107
一、编码技术	107
二、问题行为图技术	109
第四节 概念图技术	112
一、概念图及其发展	112
二、概念图的结构和类型	115
三、概念图的评估方法	117
第五章 化学课堂教学评价	122
第一节 教学目标确立的评价	122
一、教学目标的特点	122
二、教学目标的表述	123
第二节 教学内容处理的评价	125
一、同课异构的“同”与“异”	125
二、不同视角的同课异构	127
研读材料:“奇妙的二氧化碳”同课异构的评价案例	130
第三节 教学技能使用的评价	132
一、导入技能	133
二、结束技能	139
三、提问技能	142
研读材料:课堂提问技能的评价案例	147
第六章 学生个体发展评价	152
第一节 学习过程的质性评价	152
一、学习档案袋	152
二、Z管理图	155
第二节 个体发展趋势的量化评价	157

一、一元线性回归分析	157
研读材料:回归系数的实际检验过程	162
研读材料:剩余标准差	164
二、化学学习进步倾向的定量评价	165
第三节 化学实验操作技能测评	167
一、动作技能领域的教育目标分类	167
二、化学实验操作技能教学目标学习水平分类	168
三、化学实验操作技能的测量与评价	169
第四节 化学学习中学生情感测评	172
一、情感领域的教育目标分类	172
二、化学教育中情感目标分类	173
三、化学教学中情感领域目标建构	174
第七章 教学实验结果评价	177
第一节 总体水平的估量	177
一、标准误差	177
二、总体平均成绩的区间估计	178
三、 t 分布	179
研读材料:其他总体参数的置信区间	180
第二节 两对照集体水平的比较	181
一、假设检验	181
研读材料:两类错误	184
二、两均数差异比较	184
第三节 多个集体水平的同时比较	188
一、 F 分布	188
二、方差分析	190
研读材料:方差分析的条件	193
三、多个集体水平的同时比较	193
第四节 两因素方差分析介绍	194
一、 2×2 实验设计	194
二、双因素方差分析的一般格式	197
第五节 调查计数与比率的评量	198
一、比率的区间估计	198
二、两比率差异的比较	199
三、计数数据的评量	202
第八章 SPSS 的使用介绍	205
第一节 SPSS 基本操作	205
一、基本操作	205
二、数据的描述与转换	207

第二节 均值检验	210
一、单一样本均值检验	211
二、两独立样本均值差异检验	212
第三节 相关分析	214
研读材料:偏相关的计算公式	216
研读材料: Φ 相关的计算公式	219
第四节 线性回归分析	220
一、线性回归分析原理	220
研读材料:标准化回归系数	223
二、用 SPSS 进行多元线性回归分析	225
第五节 因素分析	228
一、因素分析原理	229
研读材料:高师化学系课程分析	232
二、用 SPSS 进行因素分析	234
主要参考书目	239
附录	241
附表 1 正态分布表	241
附表 2 t 值表	245
附表 3-1 F 值表(单侧检验)($\alpha=0.05$)	246
附表 3-2 F 值表(单侧检验)($\alpha=0.01$)	248
附表 4 χ^2 值表	250
附表 5 积差相关系数显著性临界值	251
附表 6 等级相关系数显著性临界值	252

第一章 化学科高考

办教育一定要讲求效果。人们如果想要了解教学方针和日常教学活动中学校以及教师所采取的措施在教育对象——学生身上所产生的实际效果,或是客观评价学生的知识、能力水平并预测学生能否在将来的某种学习或工作中获得成功,进而选拔人才或安排合适的能发挥其特长的工作,就要进行考试或测验。

测验是一个应用较广的概念。所谓测验是指在对教育对象进行定量化测定的过程中,由一定数量的项目试题组成的工具,借助这种工具可以测量人的某项品质(包括能力、兴趣、性格等)或学习掌握程度。测验以试卷为主要特征。考试作为一种术语,多用于教育上较正式的被社会所承认的场合下的测验,如“高考”不称“高测”,“公务员考试”不称“公务员测验”。此外,考试还与教学要求密切相连。考试之前一般有复习准备阶段。

考试制度是社会制度的一部分。因为高考是实现社会分工的重要环节,所以高考成为我国越来越多的考试中影响最大的考试。随着高等教育大众化时代的来临,每年一次的高考更是成为我国的“举国大考”。目前,我国经济社会的发展对人才的规格以及学历的重视程度较以前有增无减,高考的标准已经成为普通高中教学的强势“指挥棒”,使得学校一切教育教学活动皆以高考为中心、围绕高考来运作。“考什么,就教什么!”在很大程度上影响着普通高中生的学习,牵制了教育的目的。正因为如此,有必要深入地了解我们的“国考”。

第一节 高考与教育测量

作为一种大规模的教育考试制度,普通高等学校招生全国统一考试(National Matriculation Tests, NMT, 简称高考)创立于1952年。“文化大革命”期间停止运行11年(1966年到1976年),1977年才得以恢复,迄今已运行了数十年。高考制度的建立开辟了中国高校招生考试史的新纪元,是体现历史发展方向的重大改革,在世界高校招生史上占有重要的地位。中国的高考制度建立以来,在促进中学生努力学习、提高民族文化水平、维护社会稳定和公平等方面起了巨大的作用。经过高考的公平竞争,为高等学校选拔了千百万的人才。仅在1977年到2007年的30年间,高考就为国家选拔了3600万人才。

一、高考的历史沿革

1952年是中国高校招生考试史上一个非常重要的年份,是我国新旧招生制度的分水岭,标志着“高考时代”的开始。

1952年6月12日,教育部发布了《关于全国高等学校一九五二年暑期招收新生的规定》,明确规定自该年度起,除个别学校经教育部批准外,全国高等学校一律参加全国统一招生考试,采取统一领导与分省、市、自治区办理相结合的招生办法。为了便于组织统一招考工作,成立了全国高等学校招生委员会。

1952年8月15~17日,举行第一次全国统一高考。每个科目考试时间为1小时40分钟,每个单位时间考两科,国文、数学、化学、中外史地、物理、政治常识、生物、外国语(俄、英)每

个考生必须参加所有 8 个科目的考试。17 日安排加试笔试和术科。报考文法财经等院校或系的,政、国、外、史 4 科分数之和占 60%,数、理、化、生 4 科分数占 40%;报考理工农医等院校或系的采计分数的科目比例则正好相反。实行全国统一命题,制定统一的参考答案和评分标准,各大行政区招生委员会组织考试、评卷。

全国第一次参加高考的考生只有 5.9 万人,录取 5.32 万人;60 年后的 2012 年,全国参加高考的考生近 900 万人,录取人数达到了 600 余万人。

1977 年,高校招生恢复统一考试制度,当年报考人数约 570 万人,录取人数约 27 万人。自 1977 年以来的 30 多年期间,高考在许多方面都实行了改革,高考科目、内容、次数、招生录取体制、命题方式(分省)以及技术环节等都经历了多次变革。有学者认为,自 20 世纪 80 年代以来,高考至少有 14 次改革。若把各地的试验也加入,则有 22 次。这些改革可以分成以下几类^①。

持续至今有成果的改革(9 个):1984 年,艺术类等特殊专业招生实行统考+单考;1985 年,高考进行标准化改革,以最大限度控制考试误差,同时招生制度实行不收费的国家计划招生(大多数)和收费的国家调节招生(少数)并存的“双轨制”;1994 年到 1997 年,招生并轨、缴费上学;1999 年(4 个),网上录取,高校扩招,考试科目变为“3+X”,英语加试听力;2001 年,取消考生年龄“不超过 25 岁”和“未婚”的限制;2003 年,高考时间由每年的 7 月提前至 6 月。

消失成为历史过程的改革(5 个):1981 年的“六七模式”——文科考 6 门、理科考 7 门;1977 年的部分省市高考预选;1991 年“三南”(湖南、云南、海南)科目改革;1985 年开始的会考;1994 年,会考基础上的高考科目 3+2。

经历曲折和反复回到原点的改革(4 个):1984 年的保送生制度——1999 年起规定保送生都必须参加综合能力测试;2000 年的春季高考——目前多数省市已经放弃;20 世纪 80 年代中后期采用的标准分计分方法;2002 年广西的本专科高考分开考——只实行了一年。

继续进行尝试的改革(5 个):2003 年开始的自主招生;1985 年始于上海、2004 年推广的分省命题;2005 年始于上海的高职高专单独招考;2006 年始于复旦大学、上海交通大学的面试录取;2007 年宁夏等四省(自治区、直辖市)实行的基础教育课程改革后的首轮新高考[2008 年又增加了江苏等 11 个省(自治区、直辖市)]。

总体来看,技术层面的改革,如高考时间调整、考试标准化、网上阅卷和录取普遍得到肯定,而涉及高考次数、科目、内容、招生考试主体等制度层面的问题时,争议和反复就会不断出现。实践证明,指导理论正确、先进的改革在实践中不一定是可行的,勉强实行只会失败,只有在条件具备的时候,才能顺利推行。经历过“文化大革命”和废止高考的混乱局面,经过最近数十年的反复比较,虽然绝大多数人认为高考应当继续改革但也认为高考是目前最好的方法。“全面考核、推荐保送”,理论上十分美妙,但其弊端却是谁也解决不了的。因为在一定意义上说,高考是看学生本身的才学本事,推荐则往往复杂得多。

中国人发明考试这种公平竞争的办法,就是用来排除人际关系对于选才的干扰。这种选才方式有利于平民百姓。农村中学师生与一些大城市的专家学者对高考的看法是很不相同的;不同的高考经历和家世背景的人对高考的看法也很不一样。高考可以为较低社会阶层子

^① 周珣,原春琳. 30 年风雨:我们见证的改革和试验. 中国青年报,2007-6-27. 转引自:刘海峰. 高考改革的回顾与展望[J]. 教育研究,2007(11):19-24.

女(如农村的孩子)提供通过自己奋斗实现改变生存状态的重要机会。

高考是根据我国社会的要求(主要是教育活动的需要)而进行的有组织的选拔人才的活动,当然它也会反作用于社会,尤其是教育。高考通过评价教育结果引导着教育过程进而牵制了教育目的,从目前的情况看,高考教育功能的发挥(对社会的反作用)有凌驾于教育目的之上而对其制约之势。这是需要正视并加以研究解决的,历史上科举制度走向反面的教训值得汲取。

中国是考试的发源地,现代的高考仍然有着科举制度的影子。公元前140年,汉武帝就确定了“察举制”,从而正式诞生了考试。所谓“察举”就是经过考察后予以举荐。公元606年(隋炀帝时代)科举制度始见雏形;公元621年,唐高祖使科举制度正式登上历史舞台。之后它以其鲜明的特点、强大的生命力和作用力伴随着历代封建王朝活跃在历史舞台上,对中国社会的政治、经济,特别是对教育产生了重要的影响。在实行科举制度的历史阶段中除皇帝、贵族等世袭者外,从中央到地方的绝大多数官员都是科举出身的,也就是说,中国历史上有1300年的时间主要是由科举制度选拔出来的人员管理着。

然而,由于历史的原因,这种系统的、有目的的选拔人才的考试被当成一门科学来研究却是在科举制度逐渐被介绍到西方以后。

二、教育测量

20世纪初的西方,资本主义的大生产迫切需要大量熟练工人与各种专门人才,心理学特别重视个别差异和学习问题的研究,统计学的进步又提供了必要的数学工具。因此西方在引进中国的考试方法以后,便逐渐加以改造,把心理测验理论和统计学、概率论等数学方法引入考试中,重视命题质量分析,把考试内容从中国的重文轻理改为文理结合、全面测量,并且特别注意把考试与现代教育学和现代科学技术结合起来,因而使对考试的研究发展成为一门科学——教育测量学。

教育测量一般是指对教育现象进行定量化测定,主要研究对教育或训练效果进行测量的原理和方法。通俗意义上的教育测量仅指对学习结果——知识和技能的测量,也可以说主要测量学生在教育教学影响下的心理活动。教育测量学作为一门独立的专门研究考试和测验的教育科学产生于20世纪初的西方,二三十年代兴起,四五十年代定型,六七十年代得以迅速发展。它以现代教育学、心理学和统计学为基础,运用各种测试手段和现代技术,根据教育目标或教学计划,对教育现状、教育效果以及教育对象的学业成就、智能、品格、兴趣、态度、学术倾向等进行科学的测定和评价。教育测验活动的鼻祖是美国的教育心理学之父桑代克(Thorndike),他是当时美国心理测验运动的领导人。桑代克于1904年出版的《心理与社会测量导论》首次系统地介绍了编制测验的原理和用统计方法分析命题质量,他还用科学方法编制了首批标准化的教育测验学科量表。桑代克所说的“凡物的存在必有其数量”和另一位教育测量专家麦柯尔(McCall)补充的“凡有数量的东西都可以测量”这两句话构成了教育测量学赖以生存并发展的哲学基础。目前,世界各国都极为重视对教育测量的研究。

据有关文献介绍,国外对教育测量的研究是极为重视的。例如,美国在第二次世界大战以后,特别是在20世纪50年代后期以来,为大力提高教育质量以便保持其“科技第一”的优势,除制定法案外(如1958年的“国防教育法案”和1965年的“中、初等教育法案”)还动用联邦政府的巨款来贯彻这些法案,1975年佛罗里达州以州立法的形式要求举行严格的高中毕业班统

考,以保证毕业生的质量。后来,这种做法被很多州响应。纽约州议会通过的“测验的真实性法案”中规定,测验必须具备信度和效度资料。这样一来,许多教育工作者和多数的学校就不得不对自己的教育效果作出严肃的评价,重视测验的科学性和对测验的资料分析,从而促进了教育测量的发展。

国外对教育测量的重视还表现在有专业的教育测量队伍,从组织上保证了科学考试制度的实施。例如,英国的剑桥大学考试委员会(University of Cambridge Local Examination Syndicate, UCLES)创建于1858年,现在世界上有一百多个国家设有考点,每年有数百万人次参加考试。它于1993年11月起陆续在我国开考剑桥商务英语证书(Cambridge Business English Certificate) BEC1、BEC2以及CEIBT(Certificate in English for International Business and Trade, BEC3)。第二次世界大战后,不少国家和地区也开始建立专管考试的机构,专司编制、组织、评定考试的职能,如美国的教育考试服务社(Educational Testing Service, ETS)就是世界上成立较早的、颇有名气的一个民间考试机构,它是由美国教育理事会(American Council on Education, ACE)、卡内基促进教学基金会(Carnegie Foundation for the Advancement of Teaching, CFAT)和高等学校入学考试委员会(College Entrance Examination Board, CEEB)联合创办的。教育考试服务中心成立于1947年,现在全世界每年都有近千万考生参加它所设计和主持的近百种不同的考试,它的一万多考点分布在世界上176个国家和地区。教育考试服务中心所主持的考试有大型的教育考试,也有各种职业性考试,还有根据用人单位需要而设计的各种临时性测验。例如,专门测验外国留学生英语水平的TOEFL(Test of English as a Foreign Language),专门测验研究能力的研究生资格考试GRE(Graduate Record Examination),专门测验学生能力的美国大学入学考试SAT(Scholastic Assessment Test——学术能力测验)^①。这个公司专业性很强,内部分工细致,设有各种考试的命题和审题委员会、科研所、统计中心、印刷所、对外联络办公室、法律顾问处等。

20世纪20年代,教育测量学随着西方的科学技术被介绍到我国,其后得到过一些发展,但后来又停顿较长的时期。改革开放后的80年代初,国内开始翻译并出版了一些教育测量方面的书籍,也曾出版过一个内部刊物《教育统计与测验文选》。1988年11月,“中国教育学会教育统计与测量研究会”成立。

随着教育及心理测量理论的介绍引进,针对我国实际情况的教育测量与评价(特别是关于高考)研究得到了进一步开展,并取得了较为丰硕的成果,如北京语言学院汉语水平考试中心于1986年研制推出了第一套真正意义上的汉语水平考试(Hanyu Shuiping Kaoshi, HSK)系统,用以测试外国留学生的汉语水平。目前在国内30余个城市、国外数十个国家设有汉语水平考试考点。

20世纪80年代后期,我国教育部设立考试中心,各地区则设立教育考试院,负责编制并组织全国(或该省的)普通高等学校入学考试等正规考试。这对促进我国关于教育测量与评价的研究——特别是在保障极富特色的中国高考制度变革遵从现实理性、符合中国国情和社会发展阶段方面起到了重要作用。但国内的有关研究和实践在于考试目的,考试命题的标准和要求,以及对考试进行的估量方面,在考试的评分及统计分析的方法方面以及在先进测量理论的实践应用方面仍有待于进一步的工作。

^① 1926年至20世纪末,一直使用 Scholastic Aptitude Test;1995年开始用 Scholastic Assessment Test。

第二节 化学科高考与能力层次

考试是一种有意识地测量人的个体差异的活动。任何考试都有其特定的目的。化学科高考的目的是为普通高等学校选拔新生,对考生德、智、体全面考核,为择优录取新生提供学科的考试成绩(分数)。由于我国高考招生的规模受到国情、国力及社会需要的制约,还做不到满足每个人的不同愿望,因此必须从合格、水平相近的高中毕业生群体中再择优,选拔那些基础扎实、能力强、潜力大的最佳学生。所以就性质而言,化学科高考是选拔考试,不是水平测试。

一、化学科高考改革回顾

从1952年全国开始实行统一招生考试制度以来,化学就作为理、工、农、林、医等专业考生高考的主要考试科目之一。不同时期的化学科高考(考查目标、主要特征、卷面分值等)和该阶段整个高考(改革)的大环境密切相关。

1952年至1965年的化学科高考,主要以考查学生对化学知识的认知理解为基础特征,这一阶段化学试题非常注重书本知识的考查,对能力的要求较低,化学科高考试题主要集中在对化学概念、原理及关系的理解,以及一些主要实验方法的考查。

1977年高考得以恢复。在以后的十年中,化学知识的积累和技能的掌握被认为是化学学习的最重要目标,因此化学科高考提出了考查“双基”(基础知识、基本技能),通过“两个不超”(不超出教材范围、不超出大纲和考纲)落实高考命题的“两个有利于”(有利于高等院校选拔新生、有利于中学化学教学)。

1987年,高考正式提出要实现“以考知识为主”向“以考能力为主”的转变,注重“双基”和“基本能力”的综合考查(也有人称为“三基”)。化学科高考在重视考查考生掌握化学知识的熟练程度和覆盖面同时,更加注重考查学生对化学知识的内在联系、化学学科基本规律与方法的理解和运用程度,以化学教学大纲提出的“观察、实验、思维、自学能力”作为能力考查的中心目标,“能力立意”开始成为化学高考试题命制的重要原则之一(相对而言,1987年之前的试题命制则被认为主要是“知识立意”)。

相应地,化学科高考试题开始由知识型向能力型过渡。注重拓展和加深对化学知识的考查,注重灵活性、综合性的能力型试题的分值和比例逐渐提高。此外,题目变得小、巧、活,考试的知识容量大、覆盖面广,考试内容的连续性和稳定性、重点知识的重复性和发散性都较好。稍显不足的则是高考化学试题开始变得有些偏深偏难,直到20世纪90年代后期。当然,这些也是与1985年开始进行标准化考试探索的整个高考改革的大环境是密切联系的。

在此之前的高考中,化学和其他考试科目一样,卷面分值为100分。

1994年高考科目改革,“3+2”模式^①在全国正式推行。在这种背景下,化学科高考强化了“基础知识、基本技能、基本能力”的考查目标。考生对化学知识的自学能力,对化学问题分析、抽象、比较和推理的思维能力,对化学知识规律性的记忆能力,对化学实验的设计实验、分析现象、处理结果、得出结论的实验能力成为化学科高考落实“能力立意”的核心追求。“知识—能力”模式的化学科高考试题涉及的知识面广,能力考核的力度大、难度偏大,试卷考查功能清

^① “3+2”模式:语文、数学、外语3门为文、理科考生共同的必考科目,此外文科考生考历史、地理2门,理科考生则考物理、化学2门。每门150分,高考总分为750分。

晰——Ⅰ卷主要考查基础知识和基本技能、Ⅱ卷则着眼于考查基本能力(观察、实验、思维以及自学能力)。

2002年我国高考从沿用多年的“3+2”模式全面进入“3+X”模式^①。此外,从1999年起,高考也逐渐形成了“统一考试,分省命题”的格局。在新模式下,化学学科高考坚持“知识—能力”为核心的考查目标,注意考查考生在现实情境下解决实际问题的能力。高考试题命题制坚持“能力立意”,突出对化学学科主干知识的考查而不再强调学科知识覆盖面,注意学科内知识的综合和学科之间的综合^②,重视实验过程的考查,加强应用型考查,化学试题的设计与选材贴近生产与生活实际,关注社会热点与科技发展。

2007年以来,宁夏、江苏等15个先后实施基础教育改革化学新课程的省(自治区、直辖市)迎来了高考。总的来说,各地的化学学科高考在不同程度上都体现了基础教育新课程改革对高中化学教学的新要求:以理解化学的社会价值、生活价值和应用价值为核心,充分关注化学学科的核心观念、核心概念和核心方法。化学学科高考试题命题制又开始了从“能力立意”向“素养立意”的转变,突出科学探究、体现化学实验由验证到探究的重大转变,通过极大地丰富信息内涵,从而体现对信息获取、加工和迁移的考查的重视。

由于各省实际情况不尽相同,化学学科高考有理科综合卷(必考类)、理科综合卷(必考+选考类)和单科卷3大类,涉及的化学试题有十多套。总的来说,选择单科化学卷形式考查的则较少,多数省选择理科综合的形式考查,且其中化学卷面分值约为理科综合的1/3,较前有所减少^③。

二、化学学科高考的特征

选拔性质的普通高等学校入学考试(化学)有以下三个主要特征。

(1) 化学学科高考是常模参照性考试。高中毕业会考是以中学课程教学大纲为衡量标准的目标参照性考试,是成绩考试,它设定的标准并不以过线人数的多少而变动。而化学学科高考则是常模参照性考试,考生的成绩必须与当年当地应试群体的成绩相比较。高考后,从高分到低分,录满当年计划数额为止,因此每年全国各省市的最低录取分数线是不同的。化学学科高考除结果要有足够高的有效性、可靠性之外,还要求考试项目(试题)有一定的区分能力,其原因也在于此。

(2) 化学学科高考是注重考查能力的考试。学生在高中学习阶段结束之后,检验其获得的知识和技能称为成绩测验。会考虽然是国家承认的、由省(自治区、直辖市)组织命题和实施的大规模考试,但是从本质上讲,它仍属于成绩测验。它评价报告考生已经学会了什么和能做什么,而会考后化学学科高考则是侧重学习能力倾向测验(简称学能测验)。学能测验主要测量考生已有的和潜在的学习能力。因此,虽然学能测验不可能排除学习或训练及学科知识经验的影响,但其试题并不都是考查过去的教学内容掌握的程度,测试的重点常会集中在能够将这些内容迁移到广泛情景中的能力上。通过这种测验,可以预测考生将来能学会什么和从事某些活动成功的可能性。高考既然是选拔性考试,旨在选拔高中毕业合格的中学生中素质最好、基础最扎实、能力最强、潜力最大的学生,这些学生是将来得到深造机会和培养后其成功可能性

① “3+X”模式:语文、数学、外语3门为文、理科考生共同的必考科目,共450分;政治、历史和地理组成文科综合,物理、化学和生物则组成理科综合,综合均为300分。高考总分为750分。

② 理科综合测试于2005年采取分别列出理化生三个学科能力目标的方式,试卷中也不再设置多学科综合试题。

③ 王超文,彭蜀晋. 浅析建国以来的化学高考改革史. 化学教育, 2011, (1): 37-41.

最大的学生。显然,化学科高考属于这种学能测验范畴,应当向能力测试倾斜,或者说注重能力考查。当然测试中学生的能力仍然必须以化学知识和技能为载体来进行。

(3) 化学科高考是速度与难度兼有的考试。化学科高考选拔的考生,是准备继续升学的学生。速度测验的目的主要在于考查学生掌握学科知识和技能的熟练程度,同时,也可借此考查学生思维的敏捷性、严密性以及整体性。中学化学的陈述性知识较多,还有一些简单的程序性知识。如果要求高考对中学化学学习内容有一定的覆盖率,那么题目数量也就不可能太少。其次,高校教学进度比中学快得多,如果不设定时间因素考查,就容易误选一部分考生,这些学生可能在高校学习中遇到障碍。

化学科高考既然是旨在选拔,就必须能够区别不同水平的考生。要让那些能力强的考生得以发挥和表现自己的水平,试卷中就应当有考查不同能力要求的题目,个别项目试题的能力要求较高,有一定的难度,也就是必然的。这些难题并不期望多数考生都能够解答,它们只是为那些可能被录取到重点院校的少数学生设计的。现在的高考化学试卷和1977年恢复高考时相比,的确是难多了,但学生的整体水平也高多了。化学科高考的绝对难度应该适应全体考生的水平,在整体水平普遍提高的前提下,只有加大绝对难度,才能维持相对难度,从而有利于区分。高考是一场竞争,竞争的规律决定试卷的难度一定是考生目前所能达到的极限。考试是船,教育水平是水,水涨船高。

三、考试的分类

在实际考试中,通常需要将考试划分为若干类别。不同类型的考试,从考试设计、试题和试卷的编制等方面呈现出不同的特点和功能。

(1) 按考试的性质分类:可将考试划分为知识考试、技能考试、智力考试和专门能力考试。

知识考试着重考核学生对所学习知识内容的掌握程度的考试(又称形成性考试)。

技能考试着重考核学生对某项技能掌握程度的考试,多用于测试学生对实验设计、操作等实践性环节上技能的掌握程度。

智力考试是考核学生智力发展水平的考试。智力考试不按学科,而按假设的智力结构从各学科选取试题材料进行命题。

专门能力考试就是考核学生某种专门能力水平的考试,测试学生的学习能力——在一定时间内所获得的某种知识或技能的能力。

(2) 按考试的内容和范围分类:可以分为单科考试和综合性考试。

单科考试如学校的课程结业考试。考核范围包含几个学科课程有关内容的称为综合考试,如撰写毕业论文、毕业设计和学业答辩等。

(3) 按考试的实践要求分类:可以分为难度考试和速度考试。

难度考试是指考察被测者掌握知识的水平和深度,其试卷包括某学科不同难度的试题,由易到难排列,其中有些题目难度很大,但给学生充分的时间思考,以测量学生解答难题的能力。速度考试则是考察学生(被测者)掌握知识的熟练程度,其试题难度相接近,且一般都比较容易,但是题量多、时间紧,要求被测的学生在严格规定的时间内进行解答,主要以完成题目的数量作为成绩的指标。

(4) 按举行考试的时间来分类:可以分为预测性、形成性和总结性考试。

预测性考试往往在学习进行之前举行,其目的在于摸清学生对即将学习的新内容所具有的预备知识状况和认知能力状况。形成性考试多在教学过程中实施,目的在于检查教学是否