



王后雄教师教育系列教材

化学教育 测量与评价

Huaxue Jiaoyu
Celiang Yu Pingjia

王后雄 李佳 / 主编



北京大学出版社
PEKING UNIVERSITY PRESS



王后雄教师教育系列教材

化学教育测量与评价

主 编 王后雄 李 佳
副主编 胡志刚 姚如富 张文华 吴鑫德
编 委 (按姓氏笔画排序)
万 莉 王世存 文丰玉 邓 阳
李 娟 刘玉荣 孙建明 杨一思
沈久明 张世勇 陈迪妹 苑乃香
姜建文 袁振东 高 成 童金强
曾 艳 魏艳玲 满苏尔·那斯尔



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

化学教育测量与评价/王后雄,李佳主编.—北京:北京大学出版社,2013.2

(王后雄教师教育系列教材)

ISBN 978-7-301-22001-6

I. ①化… II. ①王…②李… III. ①中学化学课—教育评估—高等学校—教材②中学化学课—教育测量—高等学校—教材 IV. ①G633.82

中国版本图书馆CIP数据核字(2013)第016347号

书 名: 化学教育测量与评价

著作责任者: 王后雄 李 佳 主 编

责任编辑: 于 娜

标准书号: ISBN 978-7-301-22001-6/O·0911

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路205号 100871

网 址: <http://www.pup.cn> 新浪官方微博: @北京大学出版社

电子信箱: zyl@pup.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62767857 出版部 62754962

印 刷 者: 北京富生印刷厂

经 销 者: 新华书店

787毫米×1092毫米 16开本 22.25印张 500千字

2013年2月第1版 2013年2月第1次印刷

定 价: 45.00元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话:010-62752024 电子信箱: fd@pup.pku.edu.cn

☆ 本书获得华中师范大学国家教师教育创新平台理论创新研究项目资助

内容简介

教育测量与评价是当今世界教育科学研究的三大领域之一。化学教育测量与评价是根据化学教育的目的和标准,通过系统的调查、比较和收集资料,运用教育测量的原理和方法,对化学教育活动及效果进行质和量的描述或价值判断的过程。本书以新的教育理论和新的课程理论为指导,系统地介绍了化学教育测量的基本原理、方法及其在化学教育领域的应用。全书共十章,分别介绍了化学教育测量的基本原理和方法,化学教育测量的理论基础,化学教育测量的内容及要求,化学试题设计与编制,化学试卷设计与编制,化学教育测量质量的评价,中学化学题型解题研究,国际中学化学竞赛特点评述,国际化学高考测评改革等。

本书可作为高等师范院校化学专业的本科生、研究生及教师使用的教材,也可作为从事中学教学的化学教师、教研人员、考试评价、教师培训等教育工作者的参考书。

主编简介

王后雄,华中师范大学教师教育学院副院长,化学教育研究所所长,教授,博士生导师,兼任华中师范大学考试研究院院长,长江教育配套学习资源总策划。主要从事化学课程与教学论、化学教学诊断学、教育考试与评价等本科及研究生课程的教学和研究。主持国家及省部级科研项目 10 余项,著有《化学教学诊断学》、《新理念化学教学论》、《新理念化学教学技能训练》、《中学化学课程标准与教材分析》、《高中化学新课程教学案例研究》、《化学课程与教学论》、《化学方法论》等专著 30 多部,在 *Chinese Education & Society*、*Frontiers of Education in China*、《教育研究》、《课程·教材·教法》、《中国教育学刊》、《高等教育研究》、《教育发展研究》、《化学教育》等 SSCI、CSSCI 及核心期刊发表学术论文 200 余篇,其中 50 多篇被《新华文摘》和人大复印资料中心全文转载。曾获“全国劳动模范”、“人民教师奖章”、“全国教改‘十佳’教师”、“全国十大杰出青年教师提名奖”、“全国教育硕士优秀指导教师”等数十项荣誉称号及奖励,系享受国务院政府特殊津贴的教育专家。

李佳,华中师范大学化学教育研究所讲师,博士。主要从事化学课程与教学论、教师教育研究和教学。主持和参与国家及省部级科研项目 7 项,主编及参与高等师范院校教材编写 3 部,在《课程·教材·教法》、《中国电化教育》、《中国教育学刊》、《教育理论与实践》、《化学教学》等 CSSCI 及核心期刊发表学术论文 10 多篇,其中有多篇被人大复印资料中心全文转载。

前 言

《国家中长期教育改革和发展规划纲要(2010—2020年)》指出“必须坚持不断改革和完善考试和考试制度,使考试制度和命题技术有利于推进素质教育实施”。“建立基础教育质量标准体系以及评价、指导和监测体系,进一步改进教学和提升教育质量”。教育测量与评价是当今世界教育科学研究的三大领域之一。它在教育实践中具有很强的导向功能,在教育过程中具有重要的应用价值,当前已成为教育改革的难点和关键。教育测量与评价是以现代教育学、心理学、统计学和考试学为基础,运用各种测试方法和手段,对教育过程、教育结果、学业成就及受教育者的能力品格、学习能力倾向等方面进行科学测定和价值判断的学科,它是改进教育教学、实行科学管理和开展教育研究的重要工具。

化学教育测量与评价是根据化学教育的目的和标准,通过系统的调查、比较和收集资料,运用教育测量的原理和方法,对化学教育活动及效果进行质和量的描述或价值判断的过程。学习和运用化学教育测量与评价,有助于教师深入了解学生的学习状况,找出教学现状与教学目标的差距,恰当地做出教学决策,从而提高教学质量;有助于化学教育研究和管理工作者借鉴国际基础教育质量的通行评估指标,依据我国化学课程标准,监测化学基础教育的质量水平,通过反馈指导提高区域教育教学质量;有助于化学教育工作者科学认识化学教与学的规律,使化学教育教学实现从量性到质性的跨越,从经验到理论的升华。

当前我国正在进行的基础教育课程改革正在面临来自教育考试与评价的严峻挑战。由于“应试”导向,教师追求高、深、难的现象依然普遍存在;普通高中学业水平考试和学生综合素质评价制度亟待完善;考试命题如何有利于素质教育的导向及如何定期对命题进行评估的问题尚未解决;升学考试的命题质量及试卷质量存在失控现象;课程评价及教学评价尚未取得实质性突破;用评价改革引导“减负”任重道远;运用基础教育质量标准及评价体系监测我国基础教育质量尚未取得实质性进展等,教育测量与评价迫切需要回应我国基础教育所面临的困境和现实问题。化学教育考试与评价的变革与发展需要一线教师的积极参与,教师的测量理论水平和评价素质是教师专业化的重要组成部分。为此,本书从内容上体现时代需求和先进教育测量与评价理念,关注教育测量与评价实践中的问题解决,帮助教师提升研究意识和实践能力。

根据《教师教育课程标准(试行)》、《中学教师专业标准(试行)》精神,为编写出高水平教材,课题组对本书的定位与特色进行了多次充分研讨,并达成了基本共识。本书力图体现如下特色:

(1) 科学性。力图系统介绍化学教育测量与评价的原理、方法。包括目标参照测验的理论与技术,认知心理学的一些测量学新观点,对测验的开发、命题、实施、计分、结果解释、质量分析等都进行了详细的阐述,同时也关注到测量理论的新进展(项目反应理论、测验等值、题库建设等)和新的测验实施平台(电脑、网络等)给评价领域带来变化,对我国学者在测量学领域研究活动与成果也有介绍,力求反映测量与评价领域的当代特色和科学成果。

(2) 先导性。在继承传统教育测量与评价精华的基础上,突破传统教育测量与评价的局限,体现现代基础教育课程改革指导纲要的基础精神,吸收当前测量与评价领域的新成果,将现代测

量的最新理论与新课程的评价方法、评价模式(教学评价、学习评价、实验评价)等有机融合,较原有的课程有了一定的创新和突破,有利于读者更新教育评价理念,改进传统的评价方法和评价手段,提高教师专业化发展水平。

(3) 实用性。将教育测量与评价的原理、方法和技术与化学教育教学中要研究的实际问题相结合。如结合考试命题解读《课程标准》和《考试大纲》,从教育测量学的基本理论出发结合化学学科命题的特点,阐述化学中考、高考的目的和性质,考试的内容与要求,试题选材与题目构造,试题的命制与评价,学科能力与科学素养的考查,化学试卷的设计与编制等,凸显针对性和适应性,在化学教育实践中都具有很强的实用价值。

(4) 实践性。教育测量与评价是一门具有较强理论与实践意义的学科,在编写过程中力图使本书成为教与学的有效工具。本书从理论层面和技术层面对化学命题技术、命题特点进行研究,有利于推进素质教育的实施和实现教育评价的科学、公平;通过典型实例介绍各类化学试题的类型及解法规律,从而提高习题教学的质量;通过介绍中学化学竞赛大纲、试题特点及解法研究,力求使本书成为启迪思维、开阔视野、培养能力的学材。

本书的主编单位是华中师范大学,参编单位及作者有福建师范大学(胡志刚)、合肥师范学院(姚如富)、湖南师范大学(吴鑫德)、湖北大学(李娟)、河南师范大学(袁振东、刘玉荣)、江西师范大学(姜建文)、四川师范大学(万莉)、西华师范大学(文丰玉、高成)、山西师范大学(苑乃香)、西北师范大学(魏艳玲)、长江大学(童金强)、温州大学(陈迪妹)、信阳师范学院(沈久明)、赣南师范学院(张世勇)、黄冈师范学院(杨一思)、新疆师范大学(满苏尔·那斯尔)。参加本书编写的还有我校张文华、曾艳、王世存、孙建明、邓阳等老师。全书由王后雄、李佳负责体例设计并进行审稿和定稿工作。

本书的出版得到北京大学出版社、华中师范大学的大力支持,北京大学出版社于娜编辑为本书出版付出了辛勤的劳动,我校化学教育研究所部分研究生为书稿的校对做了大量的工作,在此一并表示感谢。本书编写时参考了国内外同行的研究成果和评价案例,在此致以诚挚的谢意!

由于我们水平有限和时间仓促,书中的缺点和不足在所难免,欢迎同行专家和广大读者批评指正。

王后雄
2013年元月

目 录

导言 化学教育测量与评价的意义	(1)
第一章 化学教育测量概述	(4)
第一节 化学教育测量的概念	(4)
第二节 化学教育测量的分类	(10)
第三节 化学教育测量的功能	(11)
第四节 化学教育测量的步骤	(13)
第五节 化学教育测量的发展	(14)
第二章 化学教育评价概述	(17)
第一节 化学教育评价的概念	(17)
第二节 化学教育评价的模式	(28)
第三节 有效的化学学习评价	(34)
第四节 化学教师发展性评价	(44)
第五节 化学教育评价的发展	(53)
第三章 化学教育测量的理论基础	(57)
第一节 教育统计知识	(57)
第二节 常模理论	(68)
第三节 真分数理论	(72)
第四节 实证效度理论	(74)
第五节 概括化理论	(77)
第六节 项目反应理论	(80)
第四章 化学教育测量的内容及要求	(87)
第一节 化学教育测量的目标要求	(87)
第二节 化学教育测量的知识要求	(96)
第三节 化学教育测量的能力要求	(100)
第五章 化学试题设计与编制	(121)
第一节 化学试题编制的基本原则	(121)
第二节 不同类型化学试题的功能	(123)
第三节 化学试题编制的基本方法	(141)
第四节 化学试题设计的案例分析	(161)

第六章 化学试卷设计与编制	(169)
第一节 化学试卷的基本特征	(169)
第二节 化学试卷的结构要素	(179)
第三节 化学试卷的组卷技术	(191)
第四节 化学试卷的评分标准	(198)
第七章 化学教育测量质量的评价	(207)
第一节 化学教育测量质量的评价概述	(207)
第二节 化学教育测量的信度	(210)
第三节 化学教育测量的效度	(214)
第四节 化学教育测量的难度	(217)
第五节 化学教育测量的区分度	(223)
第六节 化学试卷质量的评价	(228)
第八章 中学化学题型解题研究	(245)
第一节 化学选择题解题方法	(245)
第二节 化学填空题解题方法	(248)
第三节 化学简答题解题方法	(251)
第四节 化学图表分析题解题方法	(254)
第五节 化学信息迁移题解题方法	(257)
第六节 化工流程题解题方法	(261)
第七节 化学框图推断题解题方法	(264)
第八节 化学实验设计题解题方法	(269)
第九节 化学计算题解题方法	(273)
第九章 国际中学化学竞赛特点述评	(279)
第一节 化学奥林匹克竞赛简介	(279)
第二节 我国化学竞赛试题的特点	(286)
第三节 国外化学竞赛试题的特点	(288)
第四节 化学竞赛典型试题解题方法	(304)
第十章 国际化学高考测评改革	(315)
第一节 国外化学高考测评改革	(315)
第二节 我国化学高考测评改革	(331)
参考文献	(346)

导言 化学教育测量与评价的意义

教师只有具备了一定的测验能力才可能成为更优秀的教师。有效的测试将有助于改善教学的效果。

几乎每个人都能历数学校中开展的各种类型的测验,如期末考试、期中考试、单元检测、突击性测验(pop quizzes),以及(在公平情况下的)各种非正规的家庭小测验(mom quizzes)等。所有这些测验都有一个共同的特点:它们都表明教师想要了解学生到底掌握了哪些内容。更确切地说,教师之所以要开展测验,就是想要确定学生对他所教授的知识 and 技能的掌握程度。如果教师能够准确地评估学生目前的学习状况,那么之后就能够更有效地调整教学活动,以帮助学生更有针对性地获得提高和进步。研究者给出一个适用于教育背景的关于评价的操作性定义:教育评价是一种正式的方法,旨在判定学生在一些教师关注的教育变量上的状况。^①

思考与讨论

化学教师为什么需要了解教育测量与评价?

诊断学生的优势和不足

通过评价学生目前的状况,教师可以弄清楚两个问题:① 学生目前存在的薄弱点,就是今后教学的重点;② 学生已经掌握的知识和技能,也就是今后不需要再重复施教的地方。如果在教学之初就进行了这种诊断性的评价,那么对于教师制订教学计划是非常有帮助的。这种早期诊断通常被称为前测(preassessment),即发生在教师开展教学之前的评价。

监测学生的发展

课堂测评的一个重要功能在于它能够确定学生是否正在朝着教师预期的教学目标前进。当然,如果所有学生取得的进步都是令人满意的,那么教师的教学就不需要作出任何调整和改变。如果大多数学生的发展是令人满意的,只有少数学生落在了后面,那么教师可以针对个别学生,有步骤地采取一些具体化的补救措施。如果大多数学生都没有获得理想的发展,那么教师就应该大幅度修改所采用的教学方法了,因为显而易见,这种教学方法是不适用的。通过上述分析可以看出,用课堂测评来监测进步是一种历史悠久且合情合理的做法。

此外,监测学生的进展还有另一个重要的作用,那就是帮助教师利用评价的结果来构建一个对自身教学过程的形成性评价(formative assessment),这种评价旨在改善尚未成功的,但有余地补救的教学。

^① [美]詹姆斯·波帕姆,教师课堂教学评价指南[M].第5版,王本阳,赵婧,译,重庆:重庆大学出版社,2010:1-13.

评定等级

在大多数学校系统中,一项重要的工作就是在课程结束时或学年结束时对学生的学习成绩进行等级评定,这种评定的结果构成了学生人生履历中最初的成就记录,而且所有的记录都将伴随学生终身。因此,值得注意的是教师不要草率地去评定等级。不管我们喜欢还是不喜欢,学生的等级对他们而言都有非常重要的意义。

评定等级的最好方式就是收集与学生学业相关的各种证据,使教师在获得了足够的信息之后,再去决定学生应该归属于什么等级。

一般而言,如果教师收集的证据数量越多、种类越丰富,也就越容易对学生的等级作出更加明智的判断。

判断自己的教学效果

教师可以依据学生的表现来帮助自己评估教学工作的效果。假如针对一个为期三周的主题单元,教师设计了一些需要学生掌握的知识和技能要点。在教学之前,教师进行了一个简单的测试,结果表明:学生对该主题几乎一无所知。当单元教学结束之后,教师又开展了一个较详细的测试,结果反映:学生在单元学习的过程中已经掌握了大部分的知识和技能。

比较前测和后测的结果,不仅可以看出教师所教的学生已经获得了关于该主题的足够多的知识和技能,而且进一步证明了教师使用的教学方法是非常有效的。如果教师的教学方法确实能够促成预期结果的实现,那么这种方法就应该被坚持而不是被改变。

而与此相反,如果前测、后测的结果表明学生的进步微乎其微,学生在接受了教学之后掌握的知识和技能根本没有比教学开始之前增加多少,那么学生这种缓慢的发展就意味着教师在下一学期或是下一学年再次讲授相同的主题时,需要调整自己的教学方案。

上述四条中的每一条都直接关系到如何帮助教师作出正确的决策。当一位教师使用评价来判断学生的优势和不足时,评价的结果同时也可以帮助教师判定教学目标是否实现。当教师监测学生的发展时,评价的结果也可以帮助教师决定是否需要对正在进行的教学活动作出调整。当教师评价学生并对他们进行等级评定时,学生的表现可以帮助教师决定什么样的学生应该归入什么样的等级。最后,当一名教师使用前测、后测结果判断教学程序是否有效时,这个过程也可以同时促使他决定是否要大幅度修改目前的教学设计。



资料卡片

关于测验的自我检测

指导语

对于下列每一道题,请您使用如下选项作答:

SA=完全同意 A=基本同意 U=不确定 D=基本不同意 SD=完全不同意

以上回答没有对错之分,请您如实作答,用圆圈将您选择的答案圈出即可:

1. 教师对全班同学进行测试的主要目的是为了确定学生的分数等级。 SA A U D SD
2. 教师的教学应该以测试中可能涉及的知识和技能为主要内容。 SA A U D SD
3. 在课堂测试中,教师应该只设置客观性选择题。 SA A U D SD

- | | |
|---|-------------|
| 4. 学生的考分是衡量教师教学工作的标准之一,但除此之外,还有很多其他的有效标准。 | SA A U D SD |
| 5. 在学生完成课堂作业所需要的各项能力中,教师不需要测量学生的学习自信心。 | SA A U D SD |
| 6. 全国统一的标准化成绩测验不应该成为衡量教师教学成果的依据。 | SA A U D SD |
| 7. 教师真的没有必要去考虑他们课堂测试的信度问题。 | SA A U D SD |
| 8. 不可能找到一种合适的方式能够精确评判学生作文的质量。 | SA A U D SD |
| 9. 面对提高学生在重要测验中得分的巨大压力,教师可以采取任何办法来开展考前准备工作。 | SA A U D SD |
| 10. 重要的课堂测验应该在教师设计教学活动之前就编制完成。 | SA A U D SD |

自测题解析指南

对于第2题、第4题、第6题、第7题和第10题,请执行如下的评分规则:SA=5,A=4,U=3,D=2,SD=1。对于第1题、第3题、第5题、第8题和第9题,请执行如下的评分规则:SA=1,A=2,U=3,D=4,SD=5。可能获得的最高总成绩是50分,可能获得的最低总成绩是10分。当您读完这本书,您可能还想要重新检测一下(当然是在不能参考您前面作答的情况的前提之下)。

化学教育测量与评价是根据化学教育的目的和标准,通过系统的调查、比较和收集资料,运用教育测量的原理和方法,对化学教育活动及效果进行质和量的描述或价值判断的过程。学习和运用化学教育测量与评价,有助于教师深入了解学生的学习状况,找出教学现状与教学目标的差距,恰当地做出教学决策,从而提高教学质量;有助于化学教育研究和管理工作者借鉴国际基础教育质量的通行评估指标,依据我国化学课程标准,监测化学基础教育的质量水平,通过反馈指导提高区域教育教学质量;有助于化学教育工作者科学认识化学教与学的规律,使化学教育教学实现从量性到质性的跨越,从经验到理论的升华。

第一章 化学教育测量概述

学习目标

1. 理解化学教育测量相关概念。
2. 了解化学教育测量的不同类型。
3. 认识化学教育测量的功能。
4. 初步掌握化学教育测量实施的步骤。
5. 了解化学教育测量的发展阶段。

教育测量与评价是当今世界教育科学研究的三大领域之一,它不仅在教育科学体系中具有显赫的地位,而且在教育教学过程中具有重要而广泛的应用价值,在实践中具有强烈的导向作用,常成为教育改革的突破口。

随着科学技术的发展和进步,物理测量越来越精确,同时,人们也不断地尝试对人的知识、思维、想象和能力、学术水平、成就等心理特征进行测量。现代教育理论的发展,更加注意强调人的素质教育,强调发挥人的主观能动性,强调因材施教,要检验教育的效果,离不开对被教育者的评价,其中最重要的一环就是采用教育测量的方法来检测教育效果。

第一节 化学教育测量的概念

关键词

测量 教育测量 化学教育测量 参照点 单位 量表

为了正确而有效地使用化学教育测量这一检测化学教学效果或学生学业的工具,有必要认识测量、教育测量等基本概念。

一、测量

什么是测量(measurement)? 测量是人们认识自然界的一种方式,也是认识的一种方法。广义而言,测量就是根据法则赋予事物数量,即用一定的规则给事物属性指派数字或符号的过程即测量,这是迄今为止公认的测量定义。也可以说,测量是利用合适的工具,确定某个给定对象在某个给定属性上的量的程序或过程。作为测量结果的量通常用数值表示。

科学认识活动中的定量方法就是测量。卡尔纳普(R. Carnap)在他的著名的《科学哲学导论》中提出:“科学的概念如同日常生活的概念一样,可以方便地划分为三大类:分类的概念、比较的概念和定量的概念。而测量就是定量描述自然界、人类社会和人类自身规律性的一种方式,

较之分类和比较,更能清晰而精确地表述某一事物的属性。”

测量的定义包括三个特征:事物及其属性、法则、指派数字或符号。

事物是测量的对象或目标。如在测量溶液时,首先要确定的是要测量的溶液的属性,溶液的温度、密度、酸碱性等都是溶液的属性。在化学教学中,对学生的学业水平进行测量时,无法直接进行测量,但是可以从间接的角度去测量。例如,可以从学习的目标出发,编制相应的试卷来进行测试,这样就达到测量的要求。又如,要测量学生的实验技能、批判性思维,这些也是无法直接测量的,但是这些可以在人的具体活动和行为中体现出来,所以只能通过测量其外显行为或外在表现特征来推断。

法则是指测量所依据的规则和方法,是测量的关键。例如,对温度的测量可使用温度计,摄氏温度规定:在标准大气压下,冰水混合物的温度为 0°C ,沸水的温度为 100°C 。这样就形成了对温度测量的规则。对气体摩尔体积的测量,必须是在标准状况下,即 101 kPa 、 0°C 。使用好的法则可以得到比较理想的测量结果,而较差的法则只能导致不准确的测量结果。

指派数字或符号就是用数字或符号来代表某一事物或事物的某一属性的量。例如,张三在本次单元测验中得了87分,李四得了92分,我们说李四比张三多考了5分。数字本身没有意义,只是一种符号。用它来代表考生的测验成绩,这时它就变成了量化的数,可以对其进行解释和分析。

二、教育测量

(一) 几个基本概念

教育测量是根据教育目标的要求,按照一定的规则对教育活动的效果加以数量化测定的过程。^①它主要用于对学生精神特性的测定。由于教育测量主要是对学生精神特性的测定,对学生精神特性的测定比对物体的物理特性的测定要困难得多,主要原因在于学生精神时刻在发生变化,不易控制,而物体物理特性的测定无论在单位上还是在稳定性上都比对学生精神特性的测定优越得多。所以,教育测量与物质测量相比有它自己独具的特点。

关于测验的定义是广泛而复杂的。布朗(F. G. Brown)认为:“所谓测验,乃是对行为样本进行测量的系统程序。”^②此处系统程序是指测验在编制、实施和评分等方面都依据确定的规则。阿娜斯塔西(A. Anastasi)所下的定义更为完整,因而为大多数心理测验学家所接受:“测验实质上是行为样本的客观的和标准化的测量。”^③这一定义包含了三个基本因素:^④

1. 行为样本

一般来说,测验不可能涵盖某一行为领域中所有可能的行为表现,正如医生只抽取病人的几毫升血液来检验其血液特点一样,测验也只选取一组有代表性的行为来考查个体在相应行为领域的行为特征。当个体在某一测验中的反应很恰当地反映出测验所要测的东西时,该测验就为我们提供了有用的信息。因而可以说构成测验的行为样本是相应行为领域的一个有效的代表。

① 朱德全.教育统计与测量[M].重庆:西南师范大学出版社,1998:157.

② F. G. Brown (1983). Principles of Educational and Psychological Testing (ed 3)[M]. New York, Holt, Rinehart Winston, 57.

③ A. Anastasi (1982). Psychological Testing, 5th edition[M]. New York: Collier Macmillan.

④ 张敏强.教育测量学[M].北京:教育科学出版社,1998:18-19.

2. 标准化

标准化是指测验在编制、实施、记分及分数解释方面依据一套系统的程序。只有这样,测验才有统一的标准,使不同人的测验结果具有可比性。同时,可以减少无关因素对测验结果的影响,从而使之更为准确、可靠。

3. 客观测量的评价指标

测验的标准化是测验的客观而科学的测量所要求的。但是绝对的标准化是难以做到的,那么测验的客观性在什么程度上可为公众认可?这就牵涉对测验客观性程度的几个评价指标:① 题目质量分析,包括难度和区分度。这是筛选题目以构成一个好测验的基础。② 信度,指测验结果的可靠性程度。③ 效度,指测验结果的有效性程度,这是评价测验质量最重要的指标。

思考与讨论

测验与考试是怎样的关系?

在教育测量领域,经常出现“考试”一词,比如中学毕业考试、大学入学考试等等。考试与测验的差异主要体现在:

(1) 测验的范围较广,如心理测验、各种常识测验等。而考试一般指用于比较正式场合下的测验,比如说高考、自学考试等。有时二者可以互相同等,如既可说学年测验又可说学年考试;有些场合二者又不可替代,如高考不可说高测,心理测验不能称为心理考试等。

(2) 测验的目标一经确定,测验的内容一般在一定的范围之内变化,且测验可以经过相当一段时间的使用,逐步达到标准化过程,而考试则是目的性较强,会受时间、目的、地域等因素的影响,且考试的标准化较难达到。当今以此来划分测验和考试时并不严格,有些已经过严格标准化过程的测验仍被称作考试,例如 GRE(Graduate Record Examination)。

(二) 教育测量的要素^①

无论是物理特性的测量,还是精神特性的测量,都必须具有三大要素:参照点、单位和量表。

1. 参照点

参照点是计算事物数量的起点,又称零点。零点有两种:一种是绝对零点,如各种度量衡器上的零点;另一种是相对零点,是人定的零点,如温度的测量以冰点作为零点。在教育测量中,所使用的参照点几乎都采用人定的相对零点。相对零点的最大限制是不能以倍数的方式解释分数。例如,在教育测量中,甲学生的智商是100,乙学生的智商是50,我们不能说甲学生的智力是乙学生智力的2倍,因为没有真正的零智力的学生。同样甲学生的学科成绩是100分,乙学生的学科成绩是50分,我们不能说甲学生的成绩是乙学生成绩的2倍。这是我们在解释教育测量分数时,必须注意的问题。

2. 单位

物理测量的单位比较明显,如长度以米、厘米等为单位,重量以牛为单位,时间以月、小时等为单位。缺少单位,数量的多少就无法表示,数量分析也就无从进行。好的测量就必须具有好的单位。一个好的测量单位必须具备两个条件:一是有明确的意义,即同一单位在人们的心目中

^① 朱德全. 教育统计与测量[M]. 重庆:西南师范大学出版社,1998:165-167.

有同样的意义；二是有相等的价值，即单位与单位之间距离要相等。物理测量所使用的单位容易符合这两个条件，而教育测量使用的单位严格地讲不符合这两个条件。如以百分制为单位测量学生学习成绩，50分与51分之间价值差异不等同于99分与100分之间的价值差异。当然，教育测量必须具有单位，我们应当尽量改进教育测量的单位，使之科学化。

3. 量表

量表是参照点和单位的连续载体。它是测量的工具，是表示数量的方法。如天平是权衡质量的量表，尺子是度量长短的量表，而教育测量则是以文字试题、图形、符号、操作等方式来测量学生的学业成就、智力和思想品德等方面的发展水平。

在教育测量中所使用的量表有四种类型，即称名量表、顺序量表、等距量表和比率量表。

称名量表是一种最低水平的量表，这种量表与事物的分类相联系。它是用一定的数字对事物的类别加以标识。例如，用“1”和“0”分别表示男女学生在性别标识上的区别。这种数字只起到某种名称符号的标识作用，它只是对事物的属性进行归类以示区别，并没有高低序列的意思，更没有可加性，不能参与任何运算。运用称名量表，可以对各类别的频数进行累计，如计算参加测验的学生，有多少男生多少女生。对称名量表可运用的统计方法都属于对频数的统计处理，如百分比、 χ^2 检验等。

顺序量表是表示事物相对关系(如顺序关系)的数值特征的量表。如学生学业成绩的名次，可分为第一名、第二名、第三名等等。又如学生思想品德的等级，可用优、良、中、及格和不及格五个等级表示。顺序量表优于称名量表，因为它既能对事物进行分类，又能标志各类别之间的顺序关系。但顺序量表也存在缺点。因为它不能说明每一等级之差是否有相等的距离，没有表明测量单位的大小。因此，顺序量表不具有等距性，不能进行加减乘除的运算。但这种量表在教育测量中应用较广，如百分量表即属此种量表。教育测量应用顺序量表进行统计处理的方法主要有，中位数、百分位数、等级相关系数等。

等距量表最显著的特点是等距性，它对事物所作的数量划分是等距离的，是确定的，所以，它是比称名量表和顺序量表较高级的量表。如物理测量中的温度，用摄氏量表的刻度将冰水混合物的温度定为0度， 1.01×10^5 kPa下沸水的温度定为100度，其间作100等分，这样就以摄氏度为单位可以比较温度之间的差异了，而且这种量表上所表示的数值就可以进行加减运算，如可以求平均温度。由于等距量表没有绝对零点，它的零点是人定的，因此，用等距量表表示的数值虽可以作加减运算，但不能用倍数关系表示，不能进行乘除运算。等距量表虽有一些缺点，但在教育测量中却有实用意义。因为教育测量的结果都可以转化为等距量表进行比较，可以将原始测验分数转化为标准分数，进而可以将各种测验得到的不同单位的分数转换到同一等距量表上进行分析比较。等距量表在统计分析和处理时，应用比较广泛，许多统计方法都可运用。如计算平均数、差异量、相关系数等统计量，还可以运用Z检验、t检验、F检验等进行统计推断。

比率量表是测量中最高水平的量表。它既有绝对零点，又有相等的单位，因此，可以进行加减乘除运算。比率量表在物理测量中被广泛运用，但在教育测量中运用较少。因为它的使用条件极强，教育测量中的各种分数不易满足这一条件。

(三) 教育测量的性质^①

人们较易理解的测量是物理测量。物理测量一般能够获得比较准确和可靠的结果,也正缘于其直接测量的特征。与物理测量相比较,教育测量的对象是复杂得多的人。所谓千人千面,人的内在的心理亦是千差万别。人与人之间存在的差异,在心理学中被称作个别差异。教育测量永远无法直接测量,它只能通过检测心理现象的外显行为或外在表现特征来推知个体的心理能力和个性特点等。同时,教育测量很难排除一些无关因素的影响,诸如知识水平、教学条件、师资水平、情绪、健康状况、主试人导向等多方面因素都或多或少地影响到教育测量的结果,使之出现随机性或误差。

显然,教育测量的间接性、多元性和随机性注定它比物理测量要复杂困难得多。但是,美国心理与教育测量学家桑代克(E. L. Thorndike)和麦柯尔(W. A. Mc-Call)早就提出一个假说:“凡是存在必有数量,既有数量即可测量。”^②这就是说,事物的质可以转化为量来计算。

进行测量必须有相应的测量工具。显然,测量工具的性能好,测量的效果必然也好。物理测量借助于量尺、温度计、计时器、磅秤等工具,不同的工具又有不同的量纲单位,也就是说,它们各有不同的参照标准,这样才能使其测量结果具有可比性或可加性等性质。教育测量的主要工具是测验,旨在对教育效果进行科学的测量,因此,教育测量学研究的主要内容就是测验的编制和使用以及测验结果的评价。

教育测量所用的测验总是由一组题目组成,题目是构成测验的元素。好的测验必须是优良的题目的组合。比如一个用于选拔性目的的测验就应当可以把具有不同学业水平的考生区分开来,而若该测验中某一道题目所有的考生都得满分或都不得分,这一道题就失去了区分不同学业水平考生的效用。可见,选好题目是教育进行科学测量的一项重要工作。

物理测量应具有可靠性、客观性。教育测量亦同此理。一个好的测验应当在不同的人使用、不同的时间、地点施测同样具有一致性。要想确保这样的一致性,无论是物理测量,还是教育测量,都必须进行科学的测量。对教育测量而言,我们就要研究和讨论测验的可靠性和客观性,这就是测验的信度问题;我们还要研究和讨论测验反映测量目的的有效性,这就是测验的效度问题。这两个评价指标缺一不可,寻求和编制具有优良信度和效度的测验是教育进行科学测量的重要工作。

测验的结果一般都以分数或等级来表示。因此,测验分数的评定及比较、等级的划分以至对各个测验分数的解释等问题,也是教育测量学不可忽视的问题。

学校、社会对测验的要求是多种多样的,不同的测验目的有不同的测验要求,不同的测验又有不同的编制要求和不同的分数评定体系等。

三、化学教育测量

化学教育研究的对象主要有:化学课程、化学学习的原理和方法、思想教育和科学方法教育、智能培养、科学素养的培养、学业评价、化学教师、化学教学研究等。化学教育测量是将教育测量的一般原理和法则运用于化学教育过程中而自成体系的一门学科。它是教育测量的分支,是化学教育工作者的有力工具。

① 张敏强. 教育测量学[M]. 北京: 教育科学出版社, 1998: 1-3.

② [美]桑代克,等. 心理与教育的测量和评价[M]. 叶佩华,等译. 北京: 人民教育出版社, 1985: 3.

化学教育测量是一门应用性、实用性较强的学科,它是通过一定的方法和途径对化学教育目标、化学教育过程、化学教育结果以及影响化学教育的各种因素进行的一种测量活动,是有效实现化学教育目标和促进师生共同发展的不可或缺的手段。它在实践中不断地受到检验,而同时又不断地通过实践得到充实、发展和提高。

化学教育测量主要是针对狭义的教育即学校教育,即为了了解学生的化学学业发展和科学素养的培养,了解化学课堂的教学效果,了解化学教师的专业发展,了解学校的化学教育教学资源情况而进行的教育测量活动,它主要是针对化学课堂师生间的教与学的活动而进行的。化学教育测量有以下特点。^①

1. 对学生记忆化学知识水平的测量

化学知识大多是在前人大量实验事实的基础上综合概括而成的科学结论。尽管现代化学已经发生了重大变革,化学理论也有了很大的发展,但是化学学习仍然离不开对大量化学事实的记忆。特别是在中学教育阶段,很多重要的化学事实及定律、学说等很难让学生从根本上理解,对这一类知识的学习,必须依靠准确、牢固的记忆;学生能够接受的化学理论也必须以充分记忆的化学事实为基础。因此,化学学习对学生记忆化学知识的水平有着特殊的较高要求。根据化学学科的这一特征,化学教育测量需要精心选择那些有代表性的测量目标,以便能够全面地测量学生记忆化学知识的范围及其牢固程度。

2. 对学生理解和运用化学知识水平的测量

化学是研究物质的组成、结构、性质以及变化规律的一门学科,其研究对象已经深入物质的微观层次,往往是人的肉眼不能直接观察、其他感官不能直接把握的。为了让学生能够适应现代化学从基本上是从描述性的过渡到推理性的,从主要是宏观的过渡到微观的变革,在化学教学中,必须要求学生能够突破一定的思维障碍,学会根据宏观实验现象或化学事实把握其微观化学本质,真正理解化学知识;同时,学会经过由宏观到微观的抽象概括过程,来运用化学知识解决化学问题。因此,化学教学测量必须注意测量的深度。要为学生提供展开抽象思维活动的问题情境,测评他们在由宏观到微观的思维层次上理解和运用化学知识的水平,而不能将测评范围局限于化学知识的再认识或回忆的水平上。同时,也要考虑到学生思维发展的阶段性,不要超出他们的思维发展水平,以免挫伤他们的学习积极性。

3. 对学生化学实验操作技能水平的测量

化学是一门实验科学,培养学生的实验操作技能是化学教育不可缺少的重要目标。因而,对学生化学实验操作技能水平的测评必然是化学教育测量和评价的基本特征之一。实验操作技能属动作技能领域,其测评方法有着不同于认知领域的特殊要求。长期以来,忽视实验操作技能测评或是用纸笔测验代替实验技能操作考评,在一定程度上导致中学化学教学中产生违背基本规律、忽视实验教学、教学质量下降的恶果。所以,化学教育测量必须从化学教育的基本规律出发,认真研究测量学生实验操作技能水平的切实可行的有效方法。

4. 对学生化学学习兴趣与态度的测量

通过化学教育培养学生的化学学习兴趣与积极、主动的化学学习态度,是化学教育目标的重要组成部分。学生的化学学习兴趣与态度是直接指向化学学科本身的,对这种具体指向及其程

^① 周青,等.化学教育测量与评价[M].北京:科学出版社,2011:28-29.