

化学教育测量

唐 力
况梦佛 编著
熊广星



广西师范大学出版社

化学教育测量

唐 力
况梦佛 编著
熊广星

广西师范大学出版社

化学教育测量

唐 力 主编



广西师范大学出版社出版发行

(广西桂林市育才路8号)

87458部队印刷厂印刷

*

开本787×1092 1/32 印张12 字数259千字

1989年2月第1版 1989年2月第1次印刷

印数: 0001—4000

ISBN 7—5633—0308—1/G·277

定价: 4.40元

内 容 提 要

本书是我国第一本关于化学教育测量的工具性专著。主要内容有教育测量概述，中学化学教育目标分类，化学测验科学化、标准化和化学教育评价等四部分。可供高等师范院校、教育学院化学专业作为教材或教学参考书，也可供各地化学教研员、中学化学教师进修用书。

序

测量是将人们观察到的事物给以数量化，从而显示该事物的属性的一种过程。教育测量同自然科学的研究测量一样，它们都具有数量化、单位的一致性、间接观察、相对性等特征。运用统计方法使原来彼此独立的数量产生新意义、新关系和新数量（如标准差、相关系数和 t 检验等）。

教育测量是教育科学中较晚兴起的一门学科。它的主要功能是提高教学质量和改善教学效率。它是通过诊断学习情况，了解教学效果，促进学习动机，改进教师素质，改善教材教法及测试结果（如信度、效度、验证、预测）等来实现教学目标和评价教学效益的。因此教育测量与评价是教学过程中不可或缺的重要环节。近几十年来，由于教育科研与教学发展的需要以及电子计算机的普遍使用，使这门学科得以蓬勃发展，有利于教学改革深化。

目前，在我国测试科研工作正在掀起，各种具有各自特点的“心理与教育测量”，“教育测量学”，“现代教育测量”等著作已陆续问世，惟学科教育测量著作至今尚未出现。唐力、况梦佛、熊广星三位同志编著《化学教育测量》，将教育测量的基本原理和技术应用于化学教育领域，形成一定体系，并由广西师范大学出版社出版，在填补化学教育测量与评价文献的空白上作出了贡献。这令我非常高兴，应竭尽绵力为之推荐并作序。

三年多以前，我曾为唐力同志主编的高中化学程序启发

教学试验教材作了介绍，认为程序启发教学实验是“一项有活力的试验”。现经三年多的教学实践证明，该试验的效果是较好的，已为许多化学教育工作者所公认；现在又从《化学教育测量》这本书中看到，这个实验将大大推动化学教育测量的建立，并回答了怎样运用各种测试手段和统计方法来检验教学实验产生的实际效果。它反映了近代教育测量由常模参照测验转向目标参照测验与评价体系上来，以促进学生德、智、体、美、劳的全面发展。

本书系抛砖引玉之作，尚未臻尽善尽美。深盼今后能有更多不同体系，各具特色的专著问世，促进我国化学教育事业更迅速地发展和提高。

李嘉音

1988年4月20日

前 言

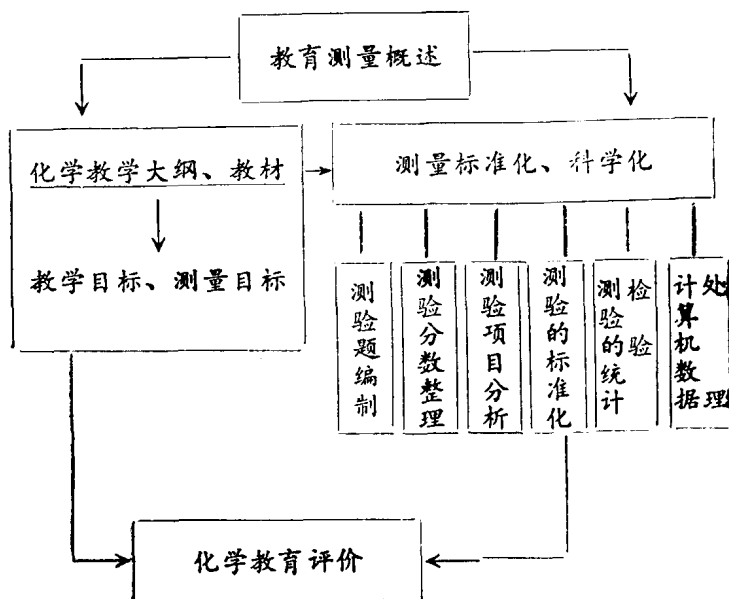
自1983年广西师范大学开展中学化学程序启发教学实验研究以来，编者深深感到怎样考核教学实验的效果，需要一套科学的程序和方法，这就产生了对化学教育测量的需要。随着教育体制改革的深入发展，教育管理要现代化，教育测验要标准化，教育评价要科学化，这些都需要借助于测量知识。1986年唐力同志尝试性编出《化学教学测量》十讲，并在广西师范大学化学系开出选修课；1987年况梦佛同志编出《描述统计基础知识及其在化学教育研究中的应用》，并在云南教育学院化学系讲授；熊广星同志担负着化学程序启发教学实验测量中的统计工作。在此基础上，我们编出《化学教育测量》讲义，并分别在两校（院）化学系试用，经修订讲义衍成本书。

本书旨在为化学教育测量的标准化、科学化提供基本原理和方法，力求切合实际，并适应化学教育和教学研究工作的需要。本书从我国化学教育的实际出发，分四个部分共九章内容，其组成结构体系如下：

第一部分，教育测量概述。包括教育测量的基本概念，测验的种类、功能、教育测验的产生和发展（第一章）。

第二部分，根据中学化学教学大纲和教材，确定教学目标 and 测量目标（第二章）。

第三部分，测量的科学化、标准化。包括测验题的编制（第三章），测验分数的整理、转化和解释（第四章），测



验质量的分析（第五章），测验标准化和教师自编测验（第六章），测验的统计检验和计算机源程序（第七、八章）。

第四部分，化学教育评价。阐明化学教育评价的作用、范围、内容，介绍几种教学评价模型（第九章）。

本书由唐力同志担任主编。唐力、况梦佛、熊广星三同志分工编著各章，最后由唐力统稿、定稿。

本书的编著工作得到化学教学法专家、华东师范大学李嘉音教授热心扶植，并为本书作《序》。

华东师范大学教育科学院教育咨询中心教育评价研究室副主任王钢同志和李嘉音教授主审了全书，谨表挚诚谢意。编著工作还得到广西师范大学、云南教育学院的支持和鼓

励，在此一并致谢。

在编写过程中，我们虽多次讨论，集思广益，但水平所限，错误之处在所难免，敬希从事化学教育和教育测量工作的同志及广大读者批评指正。

编 者

目 录

第一章 教育测量概述	(1)
第一节 教育测量的概念	(1)
第二节 教育测量的要素和种类	(6)
第三节 教育测量的功能和对测量的认识	(14)
第四节 教育测量发展简史	(18)
第二章 化学教育目标分类	(25)
第一节 目标分类概述	(25)
第二节 化学教育目标分类	(35)
第三节 单元教学目标的制订	(59)
第三章 化学测验类型和试题编拟	(68)
第一节 学业成就测验类型和评价	(68)
第二节 命题的原则和步骤	(74)
第三节 客观性试题编拟	(84)
第四节 非客观性试题编拟	(102)
第五节 实验操作技能的测验	(110)
第六节 自学能力的测验	(116)
第四章 测验分数的处理和转换	(124)
第一节 测验分数的初步整理	(124)
第二节 测验分数的统计特征量	(134)
第三节 测验分数中两列变量的相关量	(141)
第四节 测验分数的转换	(151)

第五章 测验质量的分析	(169)
第一节 常模参照测验的项目分析	(169)
第二节 目标参照测验的项目分析	(182)
第三节 测验的可靠性——信度	(185)
第四节 测验的有效性——效度	(193)
第六章 测验标准化和教师自编测验	(204)
第一节 标准化测验的要素和特征	(204)
第二节 标准化测验的基本步骤	(207)
第三节 教师自编测验	(209)
第七章 测验结果的统计分析	(218)
第一节 统计推断的基本概念和原理	(218)
第二节 单组测验的统计分析	(227)
第三节 两组测验分数平均数差异的显著性检验	(237)
第四节 分数分布的正态拟合检验	(246)
第五节 等级评定成绩的统计分析	(253)
第八章 测量数据的统计分析源程序	(267)
第一节 单组测验分数的描述统计分析源程序	(267)
第二节 两组测验分数比较的统计分析源程序	(282)
第三节 试题分析源程序	(297)
第九章 化学教育评价	(317)
第一节 现代教育评价简介	(317)
第二节 化学教育评价的内容、要求	(321)
第三节 S—P表教学评价	(329)
第四节 个人内差异评价	(344)
第五节 模糊综合评价法	(349)

本书主要参考文献	(357)
附录一 统计用表	(358)
附表 1 相关系数临界值表	(358)
附表 2 等级相关系数临界值表	(359)
附表 3 正态曲线分布表	(361)
附表 4 t值表	(365)
附表 5 χ^2 值表	(367)
附录二 CASIO fx—180p计算器使用说明	(369)

第一章 教育测量概述

测量 (Measurement) 是指人们对客观事物进行某种数量化的确定。随着科学技术的发展, 人们不但能对物体的长度、重量、温度、时间、空间、运动变化等进行越来越精确的测量, 而且对人的知识、技能、智力、兴趣、气质、性格、品德等精神特性, 也能进行测量。我国四个现代化建设, 需要培养、发现和选拔大量的各种人才, 需要借助于教育测量的基本原理和方法, 把精神特性转化为某种数量关系, 以加深对人类精神现象的了解, 服务于社会主义四个现代化。

第一节 教育测量的概念

一、什么是教育测量

教育测量 (Educational Measurement) 就是对学生知识的增长、能力的发展, 兴趣爱好、思想品德, 以及教育措施上许多问题按一定法则的数量化测定。这种测量是对人的精神特性的测定, 它远不如对物体的物理特性测定那样直接、精确、稳定, 但绝非不可测量。

美国心理学家桑戴克 (E. L. Thorndike 1874—1949) 在1918年曾说: “凡物之存在必有数量”, 美国测量学家麦柯尔 (W. A. Mocal) 于1922年提出: “凡有数量的东西

都可以测量”^①，这里“数量”表现出事物性质的差异，学生学业成绩有优有次，有的学生能回答较多的问题，有的却回答得少一些；有的人思维敏捷，反应快，有的人思维不敏捷，反应迟钝。这些差异，就为教育测量提供了可能和理论基础。

史蒂芬斯(S. S. Stevens)说：“就其广义来讲，测量是按照法则给事物指派数字”。例如，我们观察一个学生做实验，并询问他有关实验的问题，于是对该生实验原理的理解，操作要领的掌握，动作的娴熟，思维的敏捷程度，观察能力的强弱，实验结果的优劣程度等特性能作出粗略的评定，并把评定的结果用数字1—5表示出来，“1”表示不具备这些特性，“2”为稍有一些，“3”为中等，“4”为较多，“5”表示很多，这就是教育测量的一个浅显的例子。它符合史蒂芬斯的测量的定义。测量的定义包含了三个要素。

1. 事物及其属性

事物的属性和特征是我们测量的对象。上例用一个实验引起被试进行具体的认知与操作活动，以推测被试的实验知识和实验操作能力，显然这是对学生化学实验素质的一种间接测量，它反映学生个体的属性或特性。

2. 数字或符号

如前所述，测量是依照法则对事物属性作一定数字的指派，数字就是1、2、3……或A、B、C……它是代表事物特性的符号。

^①陈选善，教育测验，商务印书馆，2—8页(1934)

3. 法则

如何给事物指派数字？必须依据一定的法则或方法，如上例观察一个学生做实验的好坏或优劣程度，是根据该生对实验原理的理解等六个方面的好坏程度分别给以数字、等级，这就是法则。使用好的法则，可以获得好的测量，我们通常讲的标准化考试就是一种好的法则。这里法则包含了测量的内容、测量的步骤、评分的标准以及测验分数数量化的方法。测验编制的好坏，决定了教育测量的有效性和可靠性的程度。

二、什么是测验

测验 (Test) 是引起某种行为的工具，通常是指运用某些仪器 (药品)、试题来引起人们的某种行为，从而测定人们的某种特性。心理测验学家对测验下的定义很多，陈选善曾说：“测验是一个或一群标准的刺激，用以引起人们的行为，根据此行为以估计其智力、品格，兴趣学业等”。美国心理与教育测量学家布朗 (F. G. Brown) 说：“所谓测验，是对一个行为样组进行测量的系统程序”。这里“行为”一词是指被试对测验的测题所作的反应，以这一反应去推论被试的心理特性。例如有这样一个是非题：“在0.3摩尔/升的醋酸溶液中的 H^+ 离子浓度是0.1摩尔/升醋酸溶液的 H^+ 离子浓度的三倍。”被试对这个题目需要通过感知、回忆、思维、想象等心理过程作出反应。从反应结果，可推测被试能否分清强电解质和弱电解质，是否了解电离度跟溶液浓度的关系等，从而确定被试的心理素质和所达到的水平。

怎样理解测验“是对一个行为样组进行测量的系统程序。”呢？

“行为样组”是指一份测验卷所包含的题目样本，它应具有代表性，可据此去推论个体的特性。因此，要求测验题目的范围、重点、深浅程度以及数量都要符合一定的标准和要求，使测验卷所包含的题目，真正测到了所要考核的内容，那么通过测验所表现出来的行为，就能为改进教与学提供有用的信息。但如果我们仅仅根据被试对某个题目，或几个题目作出反应，来推断被试的心理特性，那是不全面的，因为这几个题目可能只是要考核内容的一小部分。为此，测验题目既要有一定数量，又要有一定质量，使题目具有代表性。有代表性的行为样组又是与题海战术不相容的，“题海”既违背测验的意义，又不符合教学的规律，影响学生的全面发展。

“系统程序”指的是测验的编制、施测、评分和分数的解释等方面都严格遵循一定的科学程序，以使测验准确、可靠，这样的测验又常称为标准化测验。标准化测验在确定测验目的，进行测验设计，编制测验试题，评价试题质量，筛选试题、配搭试题，测验的评分，对测验分数的解释，以及对测验实施环境、条件的控制等，测验的整个程序都实现标准化。使测验条件对所有被试都是同样的（详见第六章），其目的是保证测验的科学化，使获得的测验结果可靠（测验的信度）、准确（测验的效度）且难易适中（测验的难度）、有一定的鉴别力（测验的区分度）（第五章详细讨论）。

三、什么是化学教育测量

化学教育测量是运用教育测量的一般原理和法则于化学教育过程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一门必修课。

化学教育测量所运用的“系统程序”包括：化学教育目标的分类和教学目标的制订、测验类型的确定、试题编拟、测验分数的处理、转换和解释、测试的质量分析、统计检验、测验结果的评价等。它涉及到教育学、心理学、统计学、化学教学法以及化学专门知识。

化学教育测量又是一门应用性、实用性较强的学科。它必然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。

四、测量、测验、评价的联系和区别

测验是测量的工具。利用测验这个工具，对被试所掌握的知识、技能等方面，进行衡量取得的数量资料，对被试取得成就的客观描述，就是测量。测验比通常谈到的考试要广泛。测量后所获得的数量（测验分数）要作价值判断，就是评价。如某生学业成绩测验获得60分，60分本身并不能说明它有什么价值意义，只有用这个分数所反映的知识内容跟教学目标相比，相差多少，还存在什么问题，对存在的问题应该采取什么对策，才具有价值意义；跟他过去学习成绩相比是进步了还是退步了；跟其他学生相比是居上、居中，还是居下，都是对该生学业成绩的价值判断，即对该生学业成绩的评价。可见，测验、测量、评价虽然意义相近，但是有区别的。

评价不仅要对被试测量的结果进行好坏、优劣、高低的判断，还要了解诸如学习的兴趣、态度，学习的习惯，学习的意志，学习的环境，乃至健康状况，从定量和定性的结合上阐明对学习结果的影响，做出教育上的决策和给予必要的帮助。所以评价还具有诊断和治疗的功能。