

学校的理想装备

电子图书·学校专集

校园网上的最佳资源

学科现代教育理论书系·化学·化学教育测量和评价



前言

教育测量学作为对教育现象进行定量化测定的一门教育科学，从本世纪初建立以来，已受到世界各国的普遍重视。这门学科在人才选拔、因材施教、教育评价、教育研究等方面发挥着越来越大的作用。而化学教育测量和评价则是在运用教育测量和评价的研究成果逐渐生发、凝聚而成的一门年轻学科，可以说是在借鉴或移植教育测量和评价的原理、规则来具体研究化学教育现象，以获得定量化的认识而逐渐成型的。

从我国的教育和研究实践看，在 70 年代至 80 年代初，重新恢复教育测量和教育统计等专业课程及重组研究课题以后，随着教育的深入开展，化学教育界也开始重视科学的、定量化的测量和评价研究，经过一段试验改革的实践，开始积累了大量的、宝贵的数据和经验。近几年来，已开始有研讨化学教学目标的论著问世。应当承认，化学教育测量和评价的研究与化学教育的其他研究领域相比，仍处在资料不够充分、理论上停留在经验性概括的阶段，但其建立和发展的条件业已具备。“一切知识都是从经验开始，又终结于经验。”（爱因斯坦语）任何学科的建构如学者治学一样，总是从实践中取得经验，储宝撷英，积厚薄发，然后自成一家。基于以上认识和信念，我们在近六七年理论研究和教学实践的基础上，将不太系统的研究成果梳理成一个体系，并形成成为本书。

本书的重点在于介绍化学教育测量和评价的基本理论、方法和技术，讨论化学教育测评研究的重要课题及其研究成果和发展趋势；在写作时采用突出化学教学测评，尽力体现化学教学中“学”与“教”这一双边特征，以化学教育目标与化学教育测评标准为基点和起点，延伸、分解为实施测评的程序与方法，化学标准化考试放在本书的中间，最后以化学课堂教学评价为结尾。这种体例意在促使教育者（化学教师和研究工作者）努力从化学教育测验与测量程序中获得更多信息，以利于为改进化学教学、提高教育质量制定决策。“测验与测量之所以有用，是由于它允许而且被用来做出较好的决策。”（R.L.桑代克和 E.P.哈根语）我们认为，化学教育测量和评价是一门实践性很强的应用学科，一定要重视紧密结合化学教育实际，运用并充实、发展教育测量和评价的一般原理，要在实践研究上下功夫，努力去探求化学教育定量化测定的诸多规律性的结论，进而对改革和提高化学教育工作做出科学的决策。

全书由王建成执笔，刘知新统稿、审稿。我们尽力将文稿写得既贴近实际，又反映研究成果及发展趋势，既重视理论探讨，又注意可操作性。这一尝试的效果如何，还有哪些不足，尚祈海内同行不吝指教。

作 者

1995 年 10 月于北京师范大学

序

刘知新

化学教育作为科学教育的一个分支，以其特有的功能在实现第一流人才培育及普遍提高全体公民的文化科学素质这一使命中，起着重要作用。众所周知，化学教育正是在化学科学、技术与社会，以及学校教育、社会教育等这些大教育环境中，不断发展的。应当说，化学科学的进展与科技教育的发展推动了化学教育的繁荣和更新；学校教育与社会教育的扩展、改革和不断完善，为化学教育提供了培育良才的广阔天地。总之，社会的进步，科学技术的发展，教育的普及与提高，为化学教育的产生、发展和繁荣并发挥其多种教育功能提供了智能源泉与人才基础。应当强调指出：化学教育与其他学科教育一样，在大教育系统中发挥着维系事业兴衰、人才延续等多种功能。这些教育功能可概括为：简约有效地将人类的文明遗产传授给受教育者的传输功能；按社会的需要培养人才的塑造功能；用最经济的人力、物力和时间造就大批合格人才的高效功能；以及超前为社会的进步和革新培养适用人才的变革功能。

中华人民共和国的化学教育，自 1949 年至今，从基础教育、职业技术教育、高等教育到继续教育，不论在规模上还是质量上，都取得了令人瞩目的成就。全国各级各类学校的广大化学教育工作者为此付出了辛勤的劳动，创造并积累了丰富的教育教学经验，这是我国和世界教育科学的珍贵财富。但毋庸讳言，由于受历史的制约和人所共知的原因，我国化学教育理论研究工作起步较晚，人员较少，机构又不够健全，对于教育实践中提出的诸多重大命题，以及国际上普遍关注的某些学术研究前沿课题，尚未从理论上给出回答，或未从理论与实践结合的高度上进行深入探研。理论来源于实践，理论一经群众掌握就会变成巨大的物质力量。化学教育理论也是如此。本人作为一名老化学教育工作者，有幸从 50 年代中期就参加中国化学会组织的有关化学与化学教育的学术研讨活动，从 1979 年至今一直亲身参与中国化学会化学教育委员会及中国教育学会化学教学研究会（1983 年起始）计划和组织召开的多次全国高、中等学校化学教育经验交流会、学术研讨会、课程和教材研讨会、化学实验教学经验交流会等。每次会议的论文均有数十篇，甚至百多篇，其中不少优秀论文已在国内期刊发表。从国际上看，IUPAC（国际纯粹化学和应用化学联合会）从 1970 年开始，已加倍努力于谋求改进世界各国的化学教育，并与 UNESCO（联合国教科文组织）协同召开过 13 次 ICCE（国际化学教育会议），出版了若干部化学教育论文集和论著。另外，世界各国的化学教育家在各自的研究领域，都笔耕不辍地为国际化学教育理论做奉献。1981 年据 64 个国家的不完全统计，各国创办的化学教育（教学）期刊就有 168 种。文苑书林，浩瀚得很！但是由于各种原因，国际上大量论文、资料难以为我国广大化学教育工作者检索、利用；国内的诸多专题研究论文和著述，似多局限于就某些论题的研究或偏重于适应教材建设的需要，对于化学教育学——化学教育理论体系的几大构成，尚未见到系统论述的著作面世。

为了建构具有中国特色的化学教育理论，反映国内外当前的研究水平，以促进我国教育改革，面向现代化，面向世界，面向未来，广西教育出版社组织出版了这套《学科现代教育理论书系》中的化学现代教育理论丛书。本丛书各册的第一作者和统稿人都是在该领域学术有成的专家。全体作者均本

着理论联系实际的原则，力求从化学教育规律来阐释和探研有关的理论与学术前沿课题。当然，作为化学教育理论著述，本丛书也完全可以作为大学后继续教育或化学教育高级学位研修用书。各册（《化学教育史》除外）论述的重点虽侧重于基础教育阶段的化学教育理论问题，但是从教育规律的普遍适用性这一层面来看，这些结论对于大学或大学后教育，以及中专等化学教育实践，可供借鉴之处当不是个别的。

本丛书共 6 册，简要介绍如下：

《化学实验论》以辩证唯物主义认识论、自然科学方法论、现代教学论为指导，论述化学实验的构成、意义和作用，剖析各类化学实验及其功能，探研化学实验与发展学生思维的关系，从宏观与微观的视角揭示化学实验及其方法论的深刻涵义。

《化学课程论》从化学课程的设计与化学教材编制的现实出发，探研不同课程论思想在化学课程开发的实践中运行与演变的规律及趋向，阐释、论述我国化学课程、教材建设中的基本经验与理论问题。

《化学教学系统论》运用系统论的观点阐发、研讨化学教学的构成要素及其相关领域的原理或范型，从多视角考察、概括化学教学系统的结构和功能及其运作圭臬。

《化学学习论》从化学学习系统与学习原理的高度探研化学学习过程、模式和方法，对化学学习能力与学习机制进行剖析，探索深入开展学科心理研究的某些基本课题。

《化学教育测量和评价》基于化学教育目标论阐发和研讨化学教育测量和评价的基本理论、方法和技术问题，对认知、情感和动作技能领域的化学教育测评等作了新的探索。

《化学教育史》以历史唯物主义和辩证唯物主义为指导，对化学教育产生的历史背景、化学史各时期的化学教育的演进，以及近、现代化学教育的发展等进行了研究、概括。以史为鉴，明古鉴今。

本丛书写作注意了：科学性，力求准确、完整、系统；新颖性，取材努力反映时代气息，体现教育改革精神；实用性，各册在介绍有关理论和研究前沿的同时，均力求结合实例给读者以解决实际问题的思路与方法。

本丛书在成书过程中得到不少同行的关心，并参阅、借鉴了不少国内外学者的研究成果，在此一并表示诚挚的感谢！衷心希望本丛书面世以后能够得到化学教育界的专家和广大读者的关注与指教，祈使这套丛书在加快、深化化学教育的改革和发展，发展大学后继续教育和活跃化学教育学术研究等方面，发挥它应有的作用。

1995 年 10 月于北京师范大学

总 序

顾明远

师范院校中有一门必修课，叫做教材教法。它是一门培养教师技能的专业课程，但是历来不受人们所重视。在一些专业学科的教师、专家们眼里，似乎教材教法不过是剖析中小学的教学大纲和教科书，教会师范生如何去上好一堂课，没有什么学术性。他们认为，上好一堂课，保证教学质量的关键主要是有高的学术水平。这是一种误解。但是这种误解不是没有缘由的。原因之一是，这些专家们不懂得，教育既是一门科学，又是一门艺术，只有高深学问，不懂教育规律，没有掌握教育教学的艺术，课就上不好，或者事倍功半。原因之二是，过去的教材教法课确实存在着不少问题，它只分析现有的教材，不对学科、课程以及教育教学的规律进行研究。因此要解决这个问题，除了改变专家们的误解以外，更重要的是研究这门学科的发展，提高学科的理论水平。我认为，师范院校的教材教法不能只分析一门课如何讲授，更重要的是要研究、分析一门科学的发展历史和现状，以及其发展的内在逻辑，结合学生的认知特点，遵循教育规律，把它组织成一门学科。学科并不等于科学。一门科学要变成学校里的学科，需要经过一番改造。改造的理论就是一门学问，本身也应该是一门学科。这门学科是跨学科的，它既要研究某门学科的科学规律，例如数学教材教法既要研究数学教学规律，又要研究教育规律，要把两者有机地结合起来，从这个意义上来讲，教材教法的名称显得落后了。因此把它改为学科教学论或学科教育学是适宜的。

讲到这门学科还有一段历史，不得不讲一讲。我国学位制度建立之初，在教育类门类中就设有教材教法作为二级学科培养研究生，授予学位。但是它的评议因为涉及文理各学科，因此分散在文理各学科评议组中。由于教材教法主要是研究学科教学的理论，文理各学科评议组的专家们认为难以对他们评议。这样这门学科的授权问题就处于无人评议状态。1983年在国务院学位委员会召开第二届博士、硕士授权点学科评议组会议期间，我向当时教育学评议组召集人刘佛年教授提出，把教材教法的硕士授权点拿到教育学组来评议，并把名称改为学科教学论，以提高对它的学术要求，从而提高它的学术地位。这个提议得到刘佛年教授的支持和学位委员会的批准，并在以后专业目录调整时把教材教法正式更名为学科教学论。从此学科教学论有了较大的发展。至今全国已有硕士授权点19个，培养了硕士研究生数百名，出版的专著也有几十部，这是十分可喜的现象。

学科名称的更改是十分容易的事，但要把它发展成一门真正的学科并非易事。当时有人提出改为学科教育学，我们认为时机还不成熟，首先要把学科的教学理论研究好。教育学是一个更广泛的概念，它涉及到教育系统内部各个领域，而学科教学论主要涉及教育系统中教学方面的理论，即使把这部分研究透彻，成为一门学科也是不容易的。当然，有的学者愿意把它称为学科教育学，如果确已研究成熟，这无疑是对教育科学发展的一个贡献。

把教材教法改造成为学科教学论是一次理论上的飞跃。教材教法过去只是教育学中的一个部分。学科教学论则变成了教育科学中的一个重要分支学科。这种飞跃有没有根据，具备不具备条件呢？1988年我在为《语文教育学》写序时就说，已经具备了必要的条件。这是因为：第一，近几十年来教学论、课程论、心理学、教育测量学、教育评价学等学科有了新的发展，它为学科

教学论的建立奠定了理论基础；第二，我国改革开放以来引进了国外的各种教学理论，开拓了我们的视野，启迪了我们的思想；第三，我国有一批长期从事教材教法研究的学者，他们在师范院校有长期的教育实践，积累了丰富的经验，并且有较高的理论修养，这是建立学科教学论的组织基础。应该说，1978—1988年这门学科的建设是有成绩的，不仅培养了众多研究生和出版了多部专著，而且学科体系基本上建立起来了。更为可喜的是不少专家都在关心这门学科的建设。得到各学科的专家的重视是至关重要的。因为学科教学论这门学科毕竟是跨学科的，文理各专业学科是它的基础。

近些年来，许多学者把学科教学论又提高到学科教育学的高度来研究，这又是一次飞跃。学科教育学不仅要研究学科的教学理论问题，而且要从教育学的基本原理出发，从培养人的高度来讨论学科教育的问题。它不仅要揭示学科教学的教学规律，还要揭示学科教学培养人的规律。学科教育学不仅要讨论该门学科如何设置课程，如何编制教材，如何选择教学方法，如何组织教学，更重要的是要分析本门学科在培养人的整体工作中的地位和作用，并从这个角度出发研究课程、教材、教法，研究它与其他课程的关系，与学校中其他教育活动的关系等等。

广西教育出版社组织全国学科教育理论工作者和实际工作者编写一套大型丛书《学科现代教育理论书系》，我认为正是时候。这刚好是十多年来的一次大总结，大检阅。证明学科教育学这门新兴学科已经在中国大地上成长起来。我当然不可能通览这套丛书，但是从编辑出版计划中的书目可以看到，它涉及语文、数学、物理、化学、外语等中学教学计划中的主要学科，每门学科又分教学论、课程论、学习论、实验论、教育测量和评价等专著，有的学科还著有教学艺术论及其他更细的内容，真是丰富多彩。作者群中有老一代的学科教育学专家，也有年青一代的学者。我认为，这套丛书的意义，不仅在于它总结了十多年来我国学科教育学研究的成果，而且在于它展示了学科教育学发展的广阔前景，在于它培养了年轻一代学者。这是从教育理论战线上来讲的。至于对我国教育的实际来讲，这套丛书的出版一定有利于我国广大教师业务水平的提高，有利于教育质量的提高。我预祝出版的成功。

1996年春节

出版说明

这套丛书，从 1991 年 3 月出版第一批第一本《数学学习论》算起，至今已有 6 个年头了。如果从 1988 年年初开始数学教学理论丛书的组稿活动算起，则有 9 年之长。如今，数学、物理、化学、语文、外语，五个主要学科的教学理论丛书，已配套成龙，每个学科 6 本共 30 本，取名为《学科现代教育理论书系》。洋洋洒洒几千万字，构成了基础学科的基本理论研究，也构成了我社的基本骨干工程和基本的教育理论出版特色。

以近十年的时间建构一整套力求具有中国特色的教育理论丛书，其间的曲折、甘苦，自然一言难尽。但从反映教改成果、服务教学改革来看，又当义不容辞。从建构教育出版社的出版个性、出版文化来考虑，更有深刻意义，有重大价值。在改革开放的新历史时期，出版社靠什么来支撑？靠什么去竞争？靠什么求发展？用什么作奉献？答案可以有很多，对策可以开列不少。但根本的应少不了这么两条：一靠骨干工程，二靠名牌精品。骨干工程是出版社的战略布局，名牌精品是出版社的灵魂生命。两者的完美结合，构成了出版社的质量、信誉、知名度和文化品位，它是出版社存在的基础，竞争的手段，持续发展的后劲，文化积累的主体，向人民奉献优秀文化的根本保证。

本着这样的认识，这样的追求，我们出版了这套丛书。当然，还有另外几套别的系列。

我们期待着读者的鉴定。

我们迎接着市场的检验。

我们也渴望着教育界、理论界的支持。

我们将一如既往地努力，千方百计奉献更多的精品，给教育，给民族，给将来。

广西教育出版社

本书内容提要

本书在一般教育测量和评价理论、方法与化学学科特点密切结合的基础上，全面地介绍化学教育测量和评价的基本理论、方法和技术，以化学教学测量和评价为重点，讨论化学教育测量和评价研究的重要课题及其研究成果和发展趋势。写作上既重视理论探讨，又贴近实际，增加可操作性，为读者开展化学教育测量和评价活动及进行有关研究提供有益信息。

第一章 化学教育测量和评价概述

化学教育测量和评价是在一般教育测量和评价的原理、方法应用于化学教育领域的基础上逐步发展起来的。化学教育测量和评价既要接受一般教育测量和评价理论的指导，又有其不同于其他学科测量和评价的特殊性质。因此，本章将首先简要介绍教育测量和评价的产生和发展的历史进程，在此基础上讨论现代教育测量和评价的基本理论和方法，然后讨论化学教育测量和评价的基本特点及研究的现状与发展趋势。本章介绍的有关化学教育测量和评价的一般理论与方法，对以后各章内容的讨论都具有指导意义。

第一节 现代教育测量和评价的意义

一 教育测量和评价的产生和发展简介

1. 教育测量和评价的产生

教育测量和评价的产生与对学生学力和学业成就的检测活动的发展有着紧密的联系，其历史源远流长。早在我国的西周时期，就初步建立了学校教育制度，开始了教育测评。到公元 606 年，我国隋朝开始实行了科举制。国外学者也认为，中国古代的科举考试是教育测评的最初萌芽。但是由于种种原因，本世纪以来，我国的教育测量和评价的研究一直未能发展起来；而在西方，其中主要是在美国，自 19 世纪后半叶以来，教育测量和评价得到了迅速的发展。

19 世纪上半叶以前，学校考试一般都缺乏标准和应有的客观性，其基本方法是对学生逐个口试。随着学生人数的增多，这种考试很难继续实施了。1845 年，美国波士顿文法学校首先引入了书面考试。但这时学生成绩评定的客观标准问题仍然未能得到解决。1864 年，英国教师费舍尔（George Fisher）收集了许多学生的成绩样本，编制了第一本《量表集》，作为度量学生成绩的标准；1897 年，美国的莱斯（J.M.Rice）博士发表了他对 20 个学校的 16000 名学生所作的拼字测验的结果；1905 年，第一个智力测验量表——法国的《比纳-西蒙量表》发表。1916 年，美国斯坦福大学心理学教授特曼（L.M.Terman）发表了对比纳量表进行修订后的更完善的智力测验量表——《斯坦福量表》，首次引入了智商的概念，使智力测量有了比较科学的计算方法，标志着心理测验已达到了比较成熟的阶段。

1904 年，美国心理学家桑代克（R.L.Thorndike）发表了《心理和社会测量学导论》，介绍了心理统计方法和编制测验的基本原理；1909 年，桑代克又编写了用于书法、拼字、作文、图画测验的标准测量工具，使教育测量开始走上了科学化的道路。因此，桑代克被称为教育测量学的鼻祖。1918 年以后，教育测量的使用范围逐渐从小学发展到中等以上学校，许多大学也开始设立了教育测量学课程。至本世纪 20 年代末，教育测量已发展到全盛期，教育测量在学力检测与教育成就的量化、客观化、标准化方面取得了很大的成绩。

随着教育测量运动的发展，其弱点也逐步暴露出来。由于当时的测验都是围绕着教科书编制的，测验内容只要求学生记忆教材的知识内容，未能包括对学生的社会态度、兴趣、情感的检查，不能反映学生的全面发展和实际需要，因而它很快引起了人们的不满和批判。

1929 年，美国经历了一场严重的经济危机，教育的社会效果问题越来越多地受到人们的关注，这就直接推动了以教育的社会价值为依据的教育评价活动的发展。从 1934 年始到 1942 年止，美国俄亥俄州立大学教授泰勒（R.W.Tyler）受卡内基基金会的资助，进行了历时 8 年的课程与评价研究，这就是教育评价发展史上著名的“八年研究”。在对以往的课程和测验设计进行了尖锐批评的基础上，泰勒提出了一套以教育目标为核心的课程和测验编制原则，试图以此把社会的要求、学生的需要反映在课程与测验中。为了把这一思想与早期的测量区别开来，泰勒和他的同事正式提出了教育评价的概念：“在本质上，评价过程乃是一种测量课程和教学方案在多大程度上达到了教育目标的过程”；即认为教育评价就是衡量教育活动达到教育目标程

度的一种活动，测量是它的手段。根据泰勒的理论进行的评价活动能够获得关于目标完成情况的信息，有助于发现存在的问题，改进教育工作。所以，这一理论受到了较为广泛的欢迎。

2. 教育测量和评价的发展

泰勒等人的教育测量和评价观点以及相应的“目标评价模式”是教育测评历史上第一个比较完整的、影响较大的理论。自30年代以来的几十年中，这一理论一直占据着主导地位。为了发展泰勒评价模式的实施技术，美国教育心理学家布卢姆(B.S. Bloom)等人提出了教育目标分类学理论，对完善教育评价理论起了重要作用。但随着测评实践和理论研究的发展，这一理论也逐渐暴露出一些根本性的缺陷。泰勒理论的核心是目标，但由于教育目标概念与教育结果的概念紧密相关，这就容易使人们将更多的注意力集中到对教育结果的评价上。因此，泰勒评价更多地被看作是一种对结果的总结性评价；然而，形成性评价可能有着更重要的作用。此外，人们还认识到，目标本身的合理性也需要评价，对那些预期目标以外的教育效果也不能排除在评价之外。

在对泰勒评价观进行分析和批判的基础上，又有人提出了新的教育评价观。克龙巴赫(L. J. Cronbach)修改了泰勒的评价定义，将评价定义为：“为作出关于教育方案的决策，收集和使用信息。”他强调：“评价能完成的最大贡献是确定教程需要改进的方面。”斯塔佛比姆(D. L. Stufflebeam)给评价的定义是：“为决策提供有用信息的过程。”并强调：“评价最重要的意图不是为了证明(prove)，而是为了改进(improve)。”在这一观点的基础上，他提出了著名的CIPP(Context-Input-Process-Product)模式，即把对背景、输入、过程、结果四个方面的评价结合起来的一种评价模式。受斯塔佛比姆等人的影响，美国教育评价标准委员会给评价的定义是：“评价是对某些现象的价值和优缺点的系统调查。”

二 对现代教育测量和评价概念的认识

从上述教育测量和评价概念的产生和发展历史看，首先是对学生学力和学业成就的评定需要，导致了教育测量的产生；初期的教育测量逐步发展为标准化测量，提高了测量的科学性和客观性；又随着社会对全面发展的学生的需要，教育测量开始走向教育评价，并从只注重结果的总结性评价，逐步发展为同时注重改进教育过程的形成性评价。西方的学者更强调评价为教育决策提供信息的功能，甚至拒绝把评价看作是一种价值判断过程，试图在更广阔的范围内给出评价的定义。

我们认为，现代教育测量和评价是在对教育目标、教育过程、教育结果（现实的或潜在的结果）以及影响教育的各种因素做出量（或质）的记述的基础上进行的一种价值判断活动。教育系统中的一切因素都是测评的对象；教育测量就是对评价对象的现状、属性与规律做出客观的、定性或定量的描

瞿葆奎主编：教育学文集·教育评价，人民教育出版社1989年版，第160页。

瞿葆奎主编：教育学文集·教育评价，人民教育出版社1989年版，第164页。

瞿葆奎主编：教育学文集·教育评价，人民教育出版社1989年版，第301页。

瞿葆奎主编：教育学文集·教育评价，人民教育出版社1989年版，第298页。

陈玉琨编著：教育评估的理论与技术，广东高等教育出版社1987年版，第26—28页。

述，这种描述的结果是对测评对象的某些属性分配数值；教育评价就是在教育测量的基础上，对评价对象做出价值判断。

这里，我们已经将教育测量和评价看作是两个密切联系、又相互区别的概念。测量和评价是同一活动的两个不同阶段。测量是一种事实描述过程，只对事物的属性赋值；而评价是一种价值判断过程，是在测量的基础上对事物的好与差、对与错、优与劣的判断。不过，测量只能是在评价需要的指导下的事实描述，否则，测量没有任何意义；而评价必须是在测量基础上的价值判断，否则，评价结果就是空中楼阁，毫不可信。正是由于教育测量与教育评价之间的密切联系，我们将两者合在一起，统称为教育测量和评价，或简称为教育测评。

教育测量是对测评对象进行客观的描述，尽可能地反映测评对象的本来面目，因而使测评活动带有客观性；然而，尽管教育评价是在客观事实的基础上进行的，但评价结论必然要与评价者本身对评价对象的主观认识密切相关，必然要反映评价者的主体需要和愿望，这就使教育测评活动不能不带有评价者的主体意识。可见，教育测量和评价实际上是一种客观性与主体性高度统一的活动。

三 教育测量和评价的地位和作用

为了更全面和深入地理解教育测量和评价的意义，有必要再从系统分析的角度，看教育测量和评价在教育系统中的地位和作用。

由于教育活动是“社会、心理和控制三方面的统一”，这三方面的因素可以构成教育系统的三个最基本的子系统，即目标系统、行为系统和控制系统。

教育目标是教师（教育者）和学生（被教育者）通过教育活动应当协同达到的教育质量标准，是指导教师和学生进行教学活动的基本依据。教育过程各个阶段教育目标的有机组合，就构成了教育目标系统，是教育目的的具体化。目标系统体现了教育系统中社会方面的成分，从方向、任务和内容三方面，决定了教师的教和学生的学。

行为系统是指教师和学生通过心智和信息的相互作用而实现教育目标的过程以及影响这一过程的各种心理因素、教育手段等所构成的整体。它包容了教育系统中心理方面的成分。教师和学生为实现教育目标而进行的相互作用就是通过这一系统而实现的。

控制系统是指由控制教育过程朝着实现教育目标方向发展所必需的各种教育手段和方法的最佳组合。该系统包容了教育系统中控制方面的成分。由于一切控制过程都是按照一定的目标对系统在状态空间中的各种可能的状态进行选择，使系统达到或趋近于被选择状态的过程；在选择之前，首先要确定在不同时间的教育系统的空间状态，而这一过程，实际上就是教育测量和评价过程。所以，教育测量和评价可以看作是该系统中的主要因素。

上述三个子系统的有机结合，即构成了一个完整的教育系统。教育活动过程则可以看作是由这三个方面的因素相互作用而构成的既有区别而又相互联系的三个阶段，即目标的定向阶段、目标的控制实现阶段和目标教学结果的测量和评价阶段。在不同的活动阶段中，教师和学生既有各自不同的活动

方式，又相互配合、协调统一。教育系统的结构及其活动进程可如图 1-1 所示。教育活动进程具体体现为教师的教授进程与学生的学习进程，两者密切联系；而教育活动进程的三个主要阶段又分别与教育系统的三个子系统相对应。

总之，教育目标为教师和学生提供了一致的活动方向；教育测量和评价为防止偏离这一方向提供了有效的反馈控制手段；两者共同作用于教育过程，保证了教师的主导作用和学生主体地位的和谐统一，最后达到教育目标，实现教育系统的整体功能。

从上述对教育系统的简要分析不难看出：

(1)教育测量和评价是教育系统中不可缺少的组成部分。换言之，教育测评并非教育系统可有可无的，也不是教育系统外部强加的。只要有教育过程存在，就必然伴随着教育测评，否则，教育过程就不完整。认识到这一点，就可以增强我们开展教育测评活动的自觉性。对教育测评不需要讨论要不要的问题，而是应当研究如何进行教育测评、使之符合教育系统功能的要求的问题。

(2)教育测量和评价的主要目的不仅是鉴定，而且还有改进，从某种程度上说，改进比鉴定更为重要。因为教育的根本目的并不是对被教育者做出鉴定，而是要帮助他们全面发展。教育测评作为控制教育过程的主要手段，必须为改进教育过程提供必要的反馈信息，这是教育测评直接的、主要的功能；对学生学习成绩优劣、好坏的区别，与教师和学生的动机有紧密的联系，因而很重要，也是调控教育过程所需要的信息，但其毕竟不能直接而是间接作用于教育过程。认识到这一点，有助于我们积极、主动地利用和开发教育测评的全部功能，提高教育的质量和效率。

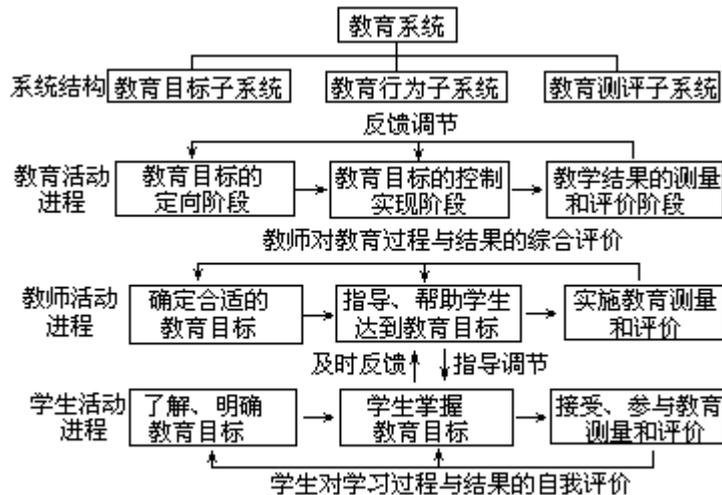


图 1-1 教育系统的结构及其活动进程

第二节 教育测量和评价的基本功能、实施程序与主要原则

一 基本功能

1. 定向功能

教育目标体现社会需要和学生全面发展的需要；教育测评是以教育目标为依据，判断教育系统的功能是否实现。判断教师教学水平的高低，主要看他是否有效地帮助学生达到教育目标；判断学生学习成绩的好坏，也是主要看他达到教育目标的程度。这就促使教师和学生在进行教育活动时，必须以教育目标为准绳，保证教育过程朝着目标指引的方向发展。

2. 鉴定功能

通过教育测量和评价，能够对学生的学习成就做出鉴定，选拔社会所需要的各种人才；也能够对教师的教育水平做出鉴定，为人事决策提供依据。

3. 诊断功能

通过教育测量和评价，调查了解或验证教育过程中可能存在的各种问题，并诊断问题存在的原因，为制定解决问题的策略提供依据。

4. 调节功能

教育测评所获得的大量信息，不仅可以用于选拔人才和人事决策，还能够用于对影响教育系统的各种因素之间的协调，使之更恰当地相互配合，以利于教育过程的优化。根据测量和评价信息，教师可以调整自己的教学内容、方法、程序，以利于学生达到教育目标；学生可以调整自己的学习方式，提高学习效率；教育管理部门可以制定管理措施、提高管理水平。

5. 激励功能

教育测评是对教师的劳动成果和学生的学习成果的一种鉴定与评价，能够给人带来精神上的满足，也会对人产生压力或动力，提高教师和学生的教与学的热情，激励他们以更多的精力投入教育活动。

6. 教学功能

教育测量和评价除了能够为协调和控制教育过程提供信息之外，本身也可以作为帮助学生达到教育目标的一种有效手段。测评的内容一般都是教学的重点；在测评过程中，学生对这些重点内容会进一步地记忆、思维、强化，从而会巩固和发展已有的学习成果。由于激励功能的存在，教育测评的这种教学功能有时甚至比其他的教学手段更加有效。

7. 交流功能

只有在参加教育测量和评价活动的有关人员测评的标准有一致认识的前提下，测评活动才能顺利展开，并获得预期的成果。统一认识的过程就是一个同行交流的过程。在总结教育经验时，利用教育测量和评价对教育活动的全面、系统和客观的描述、评价，有利于使经验更为客观、科学，因而也更有说服力，更便于经验的交流和推广。

二 实施程序

1. 准备阶段

(1)确定测量和评价目的。即解决为什么进行测评的问题。测评目的是测评活动的出发点。虽然教育测量和评价兼有多种功能，但教育的具体目标不同或是在教育过程的不同阶段，测评的侧重点并不相同。测评对象的确定和测评方法的选择，显然必须符合测评目的的需要。

(2)设计测量和评价指标体系。就是按照测评所依据的标准和要求，具体规定将哪些因素作为测评对象，同时确定各个被测评因素的相对重要性程度；测评因素构成了测评指标的内容，测评因素的相对重要性程度则表现在指标的权重系数上。所有测评指标及其权重系数的集合，即构成了测评的指标体系。指标体系的设计不仅是准备阶段，也是整个测评过程中的关键性工作。

(3)选择测量和评价方法。即根据测评目的和设定的指标体系的具体要求，选择搜集信息的方法和分析信息得出评价结果的适当方法。方法确定之后，就可以设计出测评的具体方案、步骤，准备所需要的测量和评价工具、表格、资料等，确定参加测评工作的人员并使他们明确自己的职责。

2. 搜集测量和评价信息阶段

按照设计好的指标体系和测评方案，系统、全面地搜集所需要的信息资料。教育测量和评价中常用的搜集信息资料的方法有测验法、观察法、谈话法、问卷调查法、个案研究法等。

3. 分析、判断阶段

经过第二阶段所得到的信息资料，一般都是比较庞杂、分散的，难以从中直接得出评价结论。因此，需要运用教育统计学原理，尽可能地利用电子计算机等现代化手段，对原始信息资料进行去粗取精、去伪存真的分析、处理、加工，在此基础上，形成综合判断，获得评价结果。

4. 测量和评价结果的综合利用阶段

按照测评目的组织实施的测评活动，其评价结论主要应当满足预先的需要；由于教育测量和评价具有多种功能，其结果可以综合利用。一般来说，教育测量和评价结果可用于教育目标系统的优化、教学过程的改进、提供教育管理决策的依据和对测评过程及测评方法的改进等几个方面；还可以用作其他相关教育测量和评价活动所需资料的一个部分。

三 主要原则

教育测量和评价的主要原则，就是在进行教育测评活动时所必须遵循的最基本的行为准则。

1. 目的性原则

测评活动必须有明确的目的，不能为测评而测评。测评的标准、方法、步骤都要根据测评目的的需要来确定。测评目的越明确，测评者和被测评者就越容易在测评目的指导下统一认识、协调合作，测评活动的效果就越好，效率就越高。一定要防止那种无明确目的和计划，或只凭长官意志、主观臆断的滥评现象。

测评活动的目的要正确。教育测量和评价作为教育系统的重要组成部分，以寻求教育活动的最优化方式作为自己的出发点和归宿。因此，一切教育测评活动的最终目的，都要有利于教育目标的实现。为此，设计测评方案 and 实际进行测评活动时，要注意使学生积极参与测评过程，指导和帮助他们进行自我测评，利用测评结果激发他们的学习积极性，并培养学生良好的学习习惯和方法。测评要有利于教师改进教学过程，提高教学质量，还要注意保护和提高教师的教学积极性。

2. 科学性原则

教育测量和评价的全过程都要科学、合理，所采用的测评指标、方法、

步骤等都有一定的理论或实践基础，由此而得出的评价结果也是科学、合理、符合实际的。

要以教育目标作为测评的客观标准或确定测评标准的依据；根据教育目标确定的测评指标体系一旦确定就不能随意改动；对同一类测评对象要采用同一测评标准，而对不同的测评对象必须采用不同的测评标准；测评过程搜集的信息要尽可能及时、准确、丰富、全面；对测评信息的统计、分析，要尽可能采用有一定理论和实践依据的、合乎逻辑的方法；要注意教育测评活动中客观性和主体性统一的特点，尽可能防止主观因素或错误的价值观对评价结果的干扰和影响。

3. 实践性原则

教育测量和评价要有益于教育实践活动，其本身也能够经得起教育实践的检验，并在教育实际中切实可行。

在满足目的性原则和科学性原则的前提下，教育测量和评价活动不应该增加教师和学生太多的负担，测评的方法、手段要与实际的教育环境和条件相称。要尽可能少地占用教师和学生的时间、精力及物力、财力，但要尽可能多地获得测评的实际效果，即实现测评活动本身的最优化。实践性原则还要求，在相同的条件下，用相同的测评方式所得到的测评结果是相同的，也就是，测评活动具有可重复性，这样的测评结果才是真正可靠的。这实际上是教育测评的科学性原则在教育实践中的具体体现。

目的性原则保证了教育测评对教育系统的有效性；科学性原则保证了教育测评活动本身的可靠性；实践性原则保证了教育测评的可行性。因此，教育测量和评价活动必须遵循这三个最基本的准则。

第三节 化学教育测量和评价研究

一 化学教育测量和评价的特点

化学教育测量和评价的基本性质与一般教育测量和评价是一致的。由于化学学科在其学科内容上区别于其他学科，在化学学科教育目标上必然有其特殊的要求，因此，以化学教育目标为依据的化学教育测量和评价在测评的对象、内容和标准上也表现出一定的特点，并要求有与之相适应的测量和评价方法。

我们认为，可以从以下几个方面来认识化学教育测量和评价的特点。

1. 对学生记忆化学知识水平的测量和评价

化学是一门以实验为基础的学科。很多化学知识都是前人通过无数次的实验，在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，化学学科还不能像数学、物理等学科那样有比较强的概括性，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，对很多重要的化学事实及定律、学说等很难让学生从根本上加以理解，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，化学教育测量和评价需要精心选择那些有代表性的测评目标，以便能够全面地测评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和评价

由于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到物质的微观层次，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运用化学知识解决化学问题。

根据化学学科的这一特征，要求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和评价

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，由于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，化学教育测量和评价必须从化学教育的基本规律出发，认真探研测评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和评价

通过化学教育培养学生的化学学习兴趣与态度，应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度的测量和评价，显然带有一定的学科特点。

5. 对化学教师实验教学水平的测量和评价

以实验为基础既是化学的学科特征，也是化学教学的基本特征。这一特征反映在化学教育测评上就是：在测量和评价教师的教学质量和教学水平时，要以教师的实验教学水平为重要内容，采用各种切实可行的测评手段，考察教师能否通过实验教学，让学生亲身体验探索科学规律的活动；能否结合实验事实和实验过程，让学生认识化学概念和理论的形成过程并理解和掌握它们；能否结合典型的化学史实，让学生了解化学科学的发展进程；能否让学生通过实验并运用已学到的知识去解决问题，使他们在科学态度、科学方法以及分析和解决问题的能力方面得到培养、训练。

二 我国化学教育测量和评价的研究现状及发展趋势

我国化学教育测量和评价的研究和实践，是伴随着教育测量和评价理论的研究和发展而逐步开展起来的。教育测评虽然在我国源远流长，但本世纪以来却未能得到发展。本世纪 30 年代，西方教育测量理论曾被介绍到我国，中国学者王书林、陈选善分别撰写了《教育与心理测量》、《教育测验》，在国内有一定影响。抗日战争开始后，教育测评研究中断。1949 年 10 月以后，由于众所周知的原因，教育测量和评价研究也未能得到应有的重视。改革开放以来，国外的研究材料开始陆续被介绍到国内来。随着人才培养和教育体制改革的需要，教育测量和评价的理论、方法受到了越来越多的重视，教育测量和评价的研究和实践在国内逐步展开。现阶段，化学教育测量和评价研究主要是将一般教育、心理测量和评价理论引入到化学教育中，结合化学学科特点，在理论研究和实践探索的基础上，逐步建立化学教育测量和评价的理论和方法体系。现将近年来化学教育测量和评价研究与实践的主要内容及其发展趋势概述如下。

1. 测量和评价功能的开发和利用

化学教育测量和评价改进教学活动的功能已经日益被认识和开发。形成性测评与总结性测评相结合，并注重形成性测评，是近年来测评功能开发的显著特征。

形成性测评主要是在教育活动进程中，为帮助实现教育目标、改进教育过程而进行的测评活动；总结性测评是在教育活动的某一阶段结束之后，为了解活动效益而进行的测评活动。这两个概念目前在化学教育实践中使用相当广泛。为加深对它们的理解，现将两者的特点及区别列于表 1-1 中。

表 1-1 形成性测评与总结性测评的区别

	形成性测评	总结性测评
目的	改进教学过程	总结教育成果
作用	了解教学过程中的问题或缺陷, 提供改进信息	鉴定教育成就, 提供决策信息
内容	短期、小范围中的、水平层次较低的教育目标	长期、较大范围内的、含高水平层次的教育目标
结论	学生达到教育目标的状况, 存在的问题及可能的原因	学生的学习成就及相互间的差异
方法	单元测验	学期或学年考试、毕业考试等
标准	目标参照	常模参照

从过去只注重对学生的成绩进行考核与评定, 开始走向充分利用测评结果改进教学过程, 测评活动的效益有了明显提高。运用布卢姆等人提出的掌握学习策略, 按照制定明确具体的教学目标、根据目标组织教学、适时进行形成性测评这一套教学策略组织实施的化学教学改革实验, 已经在较大范围内取得了较好的效果。教师开始注意引导和帮助学生积极参与测评过程, 让学生学会自我评价, 主动发现自身存在的 learning 问题, 自我改进, 自我激励。总结性测评的结果也不是仅用于对学生的成绩评定, 而是注重发现其中有利于改进教学过程的信息。

目前, 化学教育测量和评价的功能还有待于进一步开发。例如, 测评的诊断功能应用得还不够广泛; 对测评的教学功能还不能有意识地加以应用; 对形成性测评与总结性测评怎样更好地结合、相互匹配等问题, 还有待于深入研究。

2. 化学教育目标系统的建立和完善

化学教育测评要以化学教育目标为客观标准, 建立化学教育目标系统是化学教育测评的必然要求。

中学化学教学大纲所规定的化学教学目的与要求, 还不能作为目标系统使用。过去, 大纲中提出的“初步了解”、“了解”、“掌握”、“熟练掌握”等表示教学要求的词汇, 并没有作明确的界定, 不同的教师就会有不同的理解; 大纲中规定的培养能力方面的要求, 也未能像对知识教学的要求那样明确、具体。新编《九年义务教育全日制初级中学化学教学大纲》已经对化学知识和化学实验技能的水平层次作了比较明确的说明。

为了建立和完善更加明确具体的化学教育目标系统, 国内的许多学者也就化学教育目标的学习水平分类提出了很多好的见解; 不少新编的化学学习指导书、习题集、练习册, 也都给出了比较明确、具体的学习目标。这些研究和实践, 在一定程度上, 帮助教师和学生明确了某一具体教学阶段的目标, 因而有利于提高教学质量; 同时, 目标的明确、具体化, 也有利于化学测验的编制和对测验结果的准确评价。

目前, 对化学教育目标系统的研究, 还多集中在认知领域, 对于情感领域和实验技能领域内目标的研究比较缺乏; 对目标系统的内容、形式等方面还未能取得共同认识。这些将是今后化学教育测量和评价研究的重点课题之一。

3. 测评方法和技术的改进

教育测量的理论、方法和技术逐步用于化学教育测量。1987年，广东省进行了化学标准化考试试验。在全国高考化学试题中，客观题的比例在1978年以前是0%，而到1988年以后，已经基本稳定在70%以上。国内已经出现了一些建立在计算机上的化学试题库。在测量理论指导和标准化考试方法的影响下，常规教学中的测验编制开始程度不同地应用教育测量和标准化考试的测验编制技术，测验题中主观随意性下降，教师的命题水平提高。

教育统计方法和电子计算机技术逐步用于对测量结果的统计、分析过程。对学生测验成绩的统计已经不仅仅满足于平均分、及格率、优秀率等传统数据，标准差、相关系数、标准分等统计量数也开始用于学生成绩的描述；借助电子计算机，能够对测验的信度、效度、难度、区分度等进行统计计算，甚至进行因素分析，使通过测量所得到的信息更加丰富、全面、系统、准确，提高了评价结果的可靠性。至90年代，光电阅读技术已经在全国范围内用于高考阅卷，大大提高了高考阅卷工作的效率和评分的准确性。

目前，化学教育实际中测验的改进主要是集中在认知领域，而对情感领域和动作技能领域化学教育目标的测量还缺乏行之有效的工具；教育统计方法的应用还不够广泛，针对化学教育领域的统计分析方法尚未形成；化学教育评价还缺乏代表性好、适用范围广的常模资料或其他能被广泛接受的评价标准。这些也将成为今后化学教育测量和评价研究的努力方向。

三 现代化学教育工作者的测评素质要求

根据教育测量和评价在教育系统中的地位和作用，了解和掌握教育测量和评价的基本理论、方法、技术，并将它们适当应用于教育活动之中，应作为教育工作者的基本素质之一。对于现代化学教育工作者而言，有必要从以下几方面来培养和提高自己的测评素质。

1. 学会正确地运用或制订化学教育目标

在选用或被要求按照现成的化学教育目标系统进行教学活动时，要能够全面了解该系统的制订依据、分类体系、内容结构；在要求教师根据教学大纲和教材内容组织教学活动时，要能够自己制订出适合具体教学条件的合理的教育目标系统。要能够根据目标系统确定某一具体化学教学活动的目标，利用目标的定向功能，组织起合理、有效的化学教学活动。这一素质在当前化学教学逐步实行“一纲多本”的形势下更为重要。

2. 学会正确地选用或编制化学教育测量工具

要能够根据测评目的，选择或编制合适的测量工具。在选用现成的测量工具时，要能够根据测验所提供的效度、信度、难度、区分度等指标，看其是否符合自己的测验目的和是否适合被测学生集体；在自编测验时，要能够按照测量理论的基本要求，广泛取样，保证测验的代表性，尽可能使测验内容与测量目标一致，以保证测验的质量。

3. 学会正确得出科学合理的评价结果

要能够对测验本身的质量做出正确评价，在测验质量基本满意的前提下解释测验结果。要能够对学生的原始测验分数进行整理，并做必要的统计计算；确定出解释测验分数的标准或依据，综合各个方面的测量信息，对学生个体或集体的测验成绩做出合理评价。

4. 学会全面开发测评功能

要能够根据测评信息和评价结果，对化学教育目标系统、化学教育活动过程、化学教育测量工具等各个方面提出改进措施，注意保证使评价结果一定有利于教育目标的实现，防止或克服测评活动可能产生的副作用。能够引导和帮助学生利用测评结果进行自我评价，激发他们的学习动机，提高学习效率。

第二章 化学教育目标与化学教育测评的标准

化学教育目标不仅是化学教育测评的客观标准，而且是指导教师教学和学生学习的方向性依据。因而制订科学、合理、可行的化学教育目标系统是十分重要的，这也是近年来化学教育研究和实践的热点问题之一。

本章将讨论关于化学教育目标的概念、内容与类别、学习水平分类、制订的过程与方法等一系列理论和实践问题。

第一节 化学教育目标概述

一 化学教育目标的概念

目标是从人的需要出发所规定的行为目的，是人们争取达到的某种理想结果的标准、规格或状态。教育目标是人类社会根据自身的需要确定的教育活动的标准、方向和要求。

化学教育作为人类教育活动的组成部分，必然有其特定的目标。具体地说，就是通过化学教育过程，预期学生产生的思维、情感和行为的变化的方式。由于化学教学活动是整个化学教育活动的基本途径，因此化学教育目标又常常被称为化学教学目标。

化学教学大纲规定了化学教学的目的，这是对整个化学教学活动所作的一种原则性的规定或要求。但是，对比较复杂的化学教学活动来说，仅有像化学教学目的这样的比较笼统的、原则性的规定，不能满足各个层次具体化学教学活动的需要。要使化学教学大纲中规定的化学教学目的和要求落实到整个化学教学活动的各个方面，就必须对各个层次的具体化学教学活动水平都作出具体的规定。譬如，需要对每个学年、每个学期、每个单元、每个课时、甚至每个课时中的每个教学环节的的教学活动都作出具体的规定，才能真正落实化学教学的目的和要求。这种对某一种具体的化学教育（教学）活动所作出的具体规定或要求，就是化学教育（教学）目标。

因此，我们认为：化学教育（教学）目标是指化学教育（教学）活动的主体在具体的化学教育（教学）活动中所要达到的预期结果和标准，其主要内容是学生的思维、情感和行为的变化的方式。

可以说，化学教学目的和化学教学目标都是对化学教学活动所提出的要求和规定，但是，两者涵盖的内容宽窄不同，其区别在于：

第一，化学教学目标是化学教学目的的下位概念，是特殊和一般的关系。化学教学目标只对某一具体的化学教学活动起指导作用；而化学教学目的则要对整个教学过程，包括各层次的教学活动都起指导作用。在一系列的教学活动中可以提出一系列的教学目标，而该系列的所有目标都要受到教学目的的制约，该系列的总体效应就是教学目的。

第二，化学教学目的具有稳定性和指令性，化学教学目标具有灵活性。化学教学大纲作为化学教学的指令性文件，所规定的化学教学目的体现了社会的意志及化学教学的客观要求，教师必须遵照执行而不可随意变更，具有一定的强制性；而化学教学目标则可以根据具体的教学实际情况来制定，且可以由教师根据教学需要加以调整、变更，是带有一种教学策略意味的规定，具有较大的灵活性。

通常所说的化学教育目标有狭义和广义两种界分。狭义的化学教育目标与某一具体的化学教育活动相联系，如某一堂化学课的教学目标；而广义的化学教育目标指的是化学教育目标系统。化学教育目标系统由一系列大小不等、但存在递进关系的各个具体的化学教育目标组合而成，它包括化学教育的总目标、学年目标、学期目标、单元目标、课时目标等各个层次，各个下属层次的目标都是其上位目标的具体化。化学教育目标系统中各层次教育目标之间的关系如图 2-1 所示。

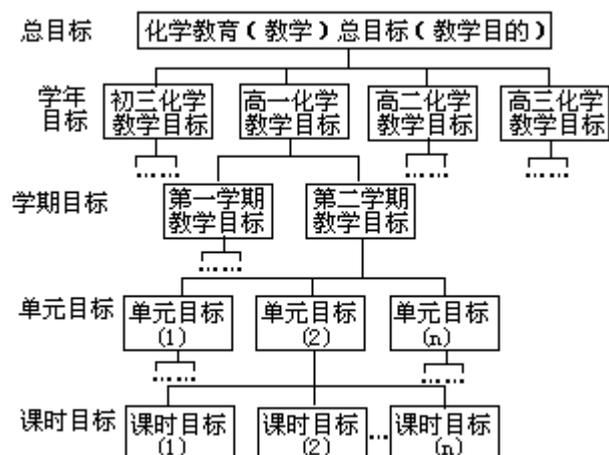


图 2-1 化学教育目标系统中各层次教育目标之间的关系

二 化学教育目标与化学教育测评标准

测评标准是人们在测评活动中应用于测评对象的价值尺度和界限。化学教育测评的标准及制定标准的依据是化学教育目标。这是因为：

(1)以化学教育目标作为化学教育测评标准,是实现化学教育测评功能的基础。

进行化学教育测评的目的,就是为了实现化学教育目标。化学教育目标反映了人们从事化学教育活动的目的和意愿,当它被当作测评的标准或依据时,才能够使目标起着引导和制约教学,对学生的发展起着调节和控制作用,并最终决定和支配着学生的发展方向和达到教育目标的程度。化学教育目标作为化学教育测评标准,在决定了测评活动方向的同时,保证了化学教育目标对化学教育活动的指导作用。

目标作为观念形态的价值意识反映了人们的需要,而需要是激发人的有意识活动的基本动力。当化学教育目标和化学测评标准一致时,化学教育目标就能够成为化学教育活动中教师和学生的一种共同期望,这种期望将在实际的教与学活动中产生激发和激励情绪的作用。当测评标准和教育目标相分离时,教育目标的这种激励作用就得不到保证和强化。

为了实现化学教学测评的功能,无论是诊断性测评、形成性测评还是总结性测评,都要以化学教育目标为标准。诊断性测评要发现学生缺少哪些实现化学教育目标所必需的基础和条件,以便采取适当的调节措施;形成性测评要了解学生达到教学目标的程度,并分析未能达到教学目标的原因,以采取有针对性的措施;总结性测评则要根据学生达到教学目标的程度,给学生以合理的成绩评定,以便激励先进,鞭策后进,激发学生的学习动机,推动教学活动的不断展开。

应当指出,当测评标准偏离教育目标时,测评活动不仅不能促进教育目标的实现,甚至会干扰、破坏教育目标的实现。例如,有些学校以片面追求升学率作为评价标准,违背化学教学的基本规律;放弃以实验为基础的化学教学原则,为对付高考,以画实验代替演示实验,以背实验代替学生实验,这显然是和化学教育目标背道而驰的。结果,在相当大的程度上造成学生高分低能,严重影响了学生的全面发展。

(2)化学教育目标能够作为化学教育测评的标准或作为制定测评标准的依据。

化学教育测评的标准包括最基本的两方面内容，一是化学教育活动的效果，可称之为效果标准；二是化学教育活动的效率，即一定时间内化学教育活动效果的大小，可称之为效率标准。

社会对化学教育的需要、化学教育活动主体的利益以及化学教育活动的特性及规律，都可以体现于教育目标之中；教育目标本身原本也具有价值尺度的性质和功能，能够作为测量与评价标准的基础。化学教育测评实质上就是将测量对象在某一时刻的态度与教学目标所确定的状态进行比较并做出价值判断，因而，化学教育目标可直接作为化学教育测评的效果标准。而化学教育测评的效率标准又以效果标准为前提和依据。

以化学教育目标作为测评标准的基础或依据，能够满足各种化学教育测评活动的需要。例如，测评学生的化学学习质量，要以学生对化学教育目标的达到程度为依据。无论是测验的信度、效度检验，还是对试题难度、区分度合理程度的判定，都要根据测验目标与化学教育目标的一致性程度来判断；根据化学教育目标的要求检验学生的学习效果，就不仅要看学生掌握化学知识的程度，还要看学生掌握化学实验技能、解决化学问题的能力，以及科学方法与态度的发展水平等各个方面。只要提供可测量的化学教育目标，就可以评价化学教育目标所要求的各方面内容；而以化学教育目标作为学生学习质量测评的依据，则可以保证测评的全面性、科学性。在测评教师的化学教学质量时，要以教师是否有效地帮助学生实现化学教育目标为主要依据。无论是教学内容的确定、教学方法的选择，还是在备课、上课、批改作业及课外活动等各个环节的行为表现，其测评的依据都是看教师在这些方面的活动是否有利于化学教育目标的实现。教师的教学水平、教学态度、思想修养等教师的素质特征也需要依据它们在实现化学教育目标过程中的实际作用来测评。

(3)化学教育目标作为测评标准，能够保证化学教育测评的客观性。

化学教育目标一旦确定之后，就可以独立于人们的主观意识之外而存在。测评者和被测评者可以同时参照化学教育目标检测化学教育测评本身，从而成为一种“外在的客观标准”，在一定程度上避免了测评者主观因素对测评过程的影响，保证了测评过程的客观性。同时，由于教育目标的这种客观性，测评活动的主体能够对测评标准取得共同认识，从而消除彼此之间由于主观认识的偏差而造成的测评误差，并且能够在对教育目标取得共同认识的基础上，交流有关教育活动的信息，更好地实现对教育活动的反馈调控。

三 化学教育目标设计的基本原则

作为化学教育测评的基础或依据，化学教育目标的设计要符合下列原则：

1. 目标系统和教学目的一致性原则

化学教育目标实质上是化学教学目的的具体化、行为化、操作化。因而，它必须完整地、充分地体现教学目的，既不能遗漏化学教学目的的任何一方面要求，更不能与化学教学目的相违背，从而与化学教学目的保持高度的一致。要使化学教学目的存在于系统化了的、由各个具体目标构成的目标系统之中，在实现具体化学教育目标的过程中，通过目标之间的相互联系和配合，最后实现化学教学目的。

一致性原则要求目标系统尽可能全面、完整，但这并不等于说，要把对

教学的所有要求都全部集中于目标系统中，这实际上是不可能的，也是毫无意义的。在制定目标系统时，为了能够把握影响教学目的达到的主要因素可以有意识地忽略一些虽然影响但属于次要的因素，这也并不违背系统的整体完备性要求。

一致性原则还要求系统内各具体目标之间也具有 consistency，即都符合教学目的而不能相互矛盾或冲突，否则，两条相互矛盾的目标中，至少有一条和教学目的相违背。

2. 系统内各具体目标的相互独立性原则

化学教育目标系统和化学教学目的的一致性原则，决定了系统内各个目标之间必然存在着紧密的联系。但是系统内各目标之间必须是彼此独立的，而不能相互重叠；不存在因果关系，不能从一个目标导出另一个目标，这就是对化学教育目标系统内的独立性要求。

独立性原则的意义是明显的。如果目标之间不能相互独立，那么，必然有一些是多余的。多余目标的存在不仅会无意义地加大测评工作量，降低测评工作的效率，而且还会因为重复目标被重复评分，使评价结论发生偏差，从而降低整个评价工作的科学性。

然而，目标间的包容关系并不影响独立性要求。有些目标分别处于不同的水平层次，高层次目标可能包含了低层次目标。例如，“能够配平氧化还原反应方程式”这一目标，必然包含了“正确判断氧化还原反应”这一目标。这两个目标都可以容纳于目标系统之中，但要出现在不同的学习阶段。高层次目标和低层次目标相比，增加了新的要求，它们之间仅存在包容关系而并不存在重复关系，因而并不违背独立性原则。像这样设立符合学生不同层次的认知发展水平的目标，有利于学生循序渐进地达到化学教学目的要求，符合化学教学本身的特点。

3. 目标的具体性原则

建立化学教育目标系统的主要意义之一，在于使化学教学目的具体化，使之成为化学教学评价赖以进行的、直接而明显的比较标准。为此化学教育目标必须符合具体性原则，具备可测性和可比性。

可测性要求用操作化的语言定义目标所规定的内容，使之可以通过实际观察而得以直接测量。教育测量和物理测量之间虽然有本质上的差别，对人的内在行为不可能像物理测量那样直接和准确。但是，人的内在行为总可以通过各种外在行为而表现出来。建立教学目标，就是要设法发现学生的外在行为和内在发展水平之间的某种必然联系，通过观察、了解学生的外在行为表现而实现对其内在发展水平的测量。

可比性要求目标要反映出测评对象属性中共同的东西，并规定出一条相应的尺度。测评对象同质是比较的前提；规定相应的尺度，是进行判断的依据。但由于外显行为和内在发展水平之间联系的复杂性，同质的判断与尺度的规定都是十分困难的。这正是化学教育测评研究迫切需要解决的问题之一。目标的可比性一般体现在系统中各目标的相互关系即目标的权重上。

4. 目标系统的可接受性原则

可接受性原则要求目标系统既要符合评价对象的实际状况，又要使依据目标系统而进行的教学测评切实可行，以切合当前的教学实际。

根据可接受性原则，一方面化学教学目标要和目标使用对象的一般教学水平相适应，要求过高或过低都是不可行的；说明化学教育目标的方式当然

是使用对象一般易于接受的，要避免难以理解的专门术语。另一方面，化学教育目标要便于实际进行的评价活动，依据目标系统可以搜集到足够的评价信息，搜集信息及处理信息的方法和手段与教学实际（如师生花费的教学时间、必要的物质条件等）相适应。

不切实际的教学目标形同虚设，还会给评价活动带来困难甚至影响评价结论的科学性，并挫伤评价活动参加者的积极性。这种情况是应当加以防范或纠正的。

第二节 化学教育目标的类别

化学教育目标作为化学教育活动的预期结果和标准，有着十分丰富的内涵，而内涵不同的各类目标有着不同的性质特征及其相应的评价方法。因而，从化学教育的目的出发，了解化学教育目标的各种类别，对于我们科学地制订化学教育目标，并选择合适的评价手段和方法是十分必要的。

一 认知目标、情感目标和动作技能目标

中学化学教学大纲规定了中学化学教学的目的、要求，这无疑是我们制订中学化学教学目标的指导和依据。根据大纲的规定，化学教学不仅要传授化学知识，培养解决化学问题的能力，还要对学生进行辩证唯物主义教育和爱国主义教育、培养学生的实验操作技能等等。不同内容的要求，其性质及相应的评价方法也有所区别。因而，为了使化学教育目标能够全面反映大纲对学生不同方面的要求，需要对目标在内容上做出合理的划分。

借鉴美国教育心理学家布卢姆等人的教育目标分类学理论，可将化学教育目标从内容上划分为以下三类：

1. 认知领域的化学教育目标

布卢姆等人提出的认知领域包括有关知识的回忆和再认，以及理智技能和能力的形成等方面的目标。布卢姆认为，技巧或技能+知识=能力。这里的技巧或技能是指操作的模式和处理问题的综合性技术。理智能力是知识与理智技巧和技能的组合。在解决需要理智能力的问题时，要求学生组织或改组问题；认识什么材料是适用的；记住这些材料；并在问题情景中利用这些材料。

我们认为，认知领域可以理解为涉及掌握知识和运用知识解决问题的能力等方面的目标集合。就化学教育而言，则包括学习和掌握化学知识、培养分析和解决化学问题的能力等方面的目标。大纲中规定的“使学生比较系统地掌握化学基础知识和化学基本技能，初步了解它们在实际中的应用；培养和发展学生的能力”等，都属于认知领域的化学教育目标。

常规的化学测验或考试，其测评目标一般都属于认知领域。在这一领域进行测评的常见问题是，偏重于对化学知识的回忆或再认而不能全面地反映学生运用化学知识分析和解决化学问题的能力水平。因而，怎样更好地测评学生的能力水平及其变化，仍然是这一领域的测评亟待研究解决的问题。

2. 情感领域的化学教育目标

克拉斯沃尔(D.R.Krathwohl)和布卢姆等人将情感领域描述为：注重情感、情绪或接受与拒绝程度的目标。在文献中，这类目标有许多是用兴趣(interests)、态度(attitudes)、价值观(values)、欣赏(appreciation)和适应(adjustment)等术语来表示的。克拉斯沃尔等人对情感领域未能像对认知领域那样做出比较明确地说明或界定，而只是给出了一个大致的范围，对上述几个常用的术语常常也有不同的理解。

布卢姆等编，罗黎辉等译：教育目标分类学·第一分册——认知领域，华东师范大学出版社1986年版，第36页。

克拉斯沃尔和布卢姆等编，施良方等译：教育目标分类学·第二分册——情感领域，华东师范大学出版社1989年版，第5页，第24页。

我们认为，学生伴随着化学学习而发生的全部情感方面的变化都属于情感领域的化学教育目标，如辩证唯物主义观点教育、社会主义和爱国主义教育、科学态度培养等等，都包括在这一领域之中。目前，情感领域目标已经越来越多地引起广大化学教育工作者的注意。

情感领域化学教育目标的测评，是一个较新的领域。对学生情感方面的测评，不能像认知领域那样要求学生给出最高表现，也没有什么标准答案；常规的测验或考试方法，几乎完全不适用；影响测评结果的因素也相当复杂。为了促进情感领域化学教育目标的达到，必须尽快研究和建立可靠和有效的测评方法。

3. 动作技能领域的化学教育目标

布卢姆等人将动作技能领域描述为：注重某些肌肉的或运动的技能、对材料和客观的某种操作的目标，或某些要求神经肌肉协调运动的目标。哈罗（A.J.Harrow）认为，人类一切可观察到的随意动作都属于这一学习领域。该领域行为的独到之处就是：它们都是可观察到的学习者表现出来的、教育工作者旨在使学生掌握的随意动作和动作形式，这些动作或形式是特定课程之教育目的的一个不可或缺的部分。

我们认为，学生对化学实验的观察和操作技能，是动作技能领域化学教育目标的主要内容。我们知道，化学实验的观察和操作技能与对实验原理知识的掌握和运用紧密相关，而后者一般属于认知领域的目标。因此，为了避免重复，我们可以限定：该领域的化学教育目标主要是与化学实验有关的学生外在行为变化。

动作技能领域化学教育目标测评也是以学生的最高表现为对象，但必须以学生的外在行为表现（不包括纸笔操作）为评价对象。现行中学化学教学中的实验操作考试就是该领域目标测评的一种典型表现。

二 到达目标和方向目标

到达目标和方向目标是由日本学者提出的一种教育目标分类法。所谓到达目标，就是把要学生掌握的知识和技能作为固定的目标提出来，也就是那些通过教育活动使学生确实能够完成的目标，反映了要使每个学生都能实际达到的、对教育成果的最低限度的要求；所谓方向目标（或期待目标），就是把期待学生具备的能力或学力作为方向提出来，或反映通过教育活动对学生发展的期待和基本方向，它是一种理想的、对教育成果的最大期待。到达目标是实在的、确定的，根据到达目标组织的教学内容是明确的，所以，对学生到达目标程度的评价可以是“绝对评价”；而方向目标给出的只是学生的努力方向，目标本身是向远处无限延伸的，根据方向目标组织的教学内容是比较模糊的或不确定的，与之相应的评价只能是“相对评价”。

我们认为，对化学教育目标也需要作类似的划分。如“对学生进行爱国主义教育”、“培养学生的思维能力”等目标，都不是一节课或几节课能够

克拉斯沃尔和布卢姆等编，施良方等译：教育目标分类学·第二分册——情感领域，华东师范大学出版社1989年版，第5页。

哈罗、辛普森等编，施良方等译：教育目标分类学·第三分册——动作技能领域，华东师范大学出版社1989年版，第26—27页。

万勇：到达目标与目标到达度评价，外国教育资料，1983年第6期，第21—34页。

到达的，而是需要相当长的化学教育阶段，甚至要与其他学科共同作用才能使达到目标，应属于方向目标；而对于某一化学知识的初步理解和掌握、对一种实验仪器的模仿操作等，可以通过一个具体的教学活动帮助学生达到目标要求，如：“记住水的分子式”、“学会配平化学方程式”、“学会酒精灯的使用”等等，就属于到达目标。

在实际教学中，由于未对目标做出上述两种区别，往往会产生两类错误。一类是到达目标规定得明确、具体，而方向目标却模糊甚至没有，只注重具体化知识、技能教学，而忽视了能力培养和思想教育等重要的教学要求；另一类是不了解方向目标实现的长期性，企图在一节课中就实现方向目标，或是对两类目标不加区别，在某一节课的教学目标中也列出如“培养学生的观察能力”这样的内容，并在类似的课中都这样列，不仅不能落实，显得空洞、无力，而且使目标内容庞杂、繁琐。

可见，对化学教育目标做出方向目标和到达目标的划分，并分别研究这两类目标的到达途径，有利于发挥教育目标的指导作用和目标的全面落实。

三 完成目标、提高目标和体验目标

这三类目标的本来意义是：完成目标：要求完全掌握特定的具体知识和能力的目标。在目标到达时，能对其状态做出明确表述，并能根据与这种表述的比较确认目标到达的程度。提高目标：要求朝着某一方向提高或加深的目标。基本上是通过个人的自我比较（和以前比……）或与他人比较（和谁比……）的方式才能把握的具有进步、提高、深化性质的目标。体验目标：不是以学生的某种行为改变作为直接目标，而是以发生特定的内在体验为目的的目标。把在学生的智力和心理成长中具有重要意义的接触、感动、发现等体验作为教学-学习活动的直接目标时，就是这种体验目标。

我们认为，这种分类是以实现目标的方式为基本依据的。就化学教育来看，完成目标是学生在某一化学教育活动结束时即刻完成的目标，如对某一元素化合物知识的记忆、对某一化学概念的理解等；提高目标可看成是在完成目标的基础上，学生通过自己的努力，使已经达到的目标进一步发展、深化，如对元素化合物知识的运用、对化学概念的深入理解和系统化，等等；体验目标的概念尤为重要，根据这一认识，可以将某些较难表现在学生行为变化上的教学要求具体化，从而指导教师的教学行为，如“结合某一化学知识进行爱国主义教育”这一目标，可以通过体验目标——要使学生受到教师爱国主义情感的感染或熏陶——作为教学-学习活动的直接目标。对体验目标的评价，可通过对教师的课堂教学行为的直接观察等方式来进行，如作为化学课堂教学评价的重要内容，促进化学教育目标的全面实现。体验目标的明确化，对当前化学教学中比较薄弱的情感领域目标教学，有着特殊的意义。

四 教授目标和学习目标

教育目标作为教学过程中师生共同的努力方向，协调着师生双方的活动，学生为实现目标而学，教师为帮助学生实现目标而教。这里，教师和学生的活动目标都以教育目标为指导，似乎是完全一致的。但是，在实际教学过程中，在师生头脑中观念存在着的目标并不完全相同。即使客观规定的教

育目标要求师生共同遵守，但由于师生有着不同的知识基础，在教学过程中的地位和作用有着明显差异，因而，师生对同一教育目标会有不同的理解，实际上，教育目标对师生的教学和学习行为要求显然也是有区别的。所以，有必要区别作为指导教师教学行为的教授目标和作为指导学生学习行为的学习目标。

教授目标和学习目标实际上只是教育目标的进一步具体化。两类目标的制订都必须以教育目标为依据。教授目标通常是为学生达到课堂教学的完成目标而设计的，重在保证教师的教学行为能够指导、帮助学生达到课堂教学目标；而学生的学习目标不仅要包括课堂学习目标，还要包括课后通过自己的努力而实现的提高目标，最好还能包括达到课堂学习目标所需要的知识基础的准备目标。教育目标中的某些内容可以不直接或明确出现在学习目标中，如思想教育等方面的要求，只是明确规定在教授目标中，对学生而言，可称之为一种隐性目标，这样，可能更有利于这类目标的实现。

第三节 认知领域化学教育目标的学习水平分类

一 认知领域化学教育目标分类的意义

我们已经知道，化学教育目标按其内容可以分为认知、情感、动作技能三大领域。但是，化学知识教学及知识学习过程中有关能力的培养，毕竟是化学教学内容的主体。因而，认知领域的化学教育目标也是化学教育目标的主要内容，并且是制订其他领域目标的基础。所以，要制订化学教育目标首先要解决认知领域化学教育目标的制订问题。

在化学教学实际中，对知识教学的要求还是比较具体和明确的。制订认知领域化学教育目标的关键，是要解决能力培养要求的具体化问题。

能力是指人们能够顺利完成某种活动的心理特性，可以看作一种能够将知识运用于新问题或新情境的本领。具备知识和能够运用知识，是能力这种心理特性所不可缺少的两个要素。“无知必定无能”，因为不具备某种知识，就谈不上对它的运用；“有知未必有能”，因为虽然具备了某种知识，但不知如何运用，仍然不能解决新问题。可见，知识和能力之间，存在着不可分割的密切关系。

能力的核心是人们为达到特定目标而组织和改组材料的心智活动过程。知识学习过程实际上也是一个复杂的智力活动过程。没有发展到一定水平的智力，就很难正确、有效而迅速地掌握知识，所以，智力发展水平是知识学习的基础；反之，智力的发展或能力培养又是在知识的学习过程中实现的。我们认为，可以将知识学习和能力培养看作同一学习活动的两种不同指向：当学习指向学习者内部时，表现为知识的掌握；当学习指向学习者外部时，则表现为学习问题的解决，即知识的运用或能力培养。因而，在教学过程中，知识学习和能力培养应当统一实现。

化学教学中能力培养要求，可以体现在对化学知识的学习水平的要求上。学生的能力或智力发展水平可以通过观察学生能否在不同的水平上运用这些知识解决问题的方式来测量，而学生运用知识的水平又是和他们掌握知识的水平紧密联系的。学生掌握知识的水平越高，运用知识的可能性就越大，其能力发展水平就越高。知识的掌握水平一般是和知识学习的难易、深浅联系在一起，所学习的知识越深或越难，要求学生进行的智力活动越复杂，对学生的智力发展水平的要求也就越高。因此，当学生掌握了较深或较难的知识后，我们可以认为他具备了相应的智力或能力发展水平。

这样，认知领域化学教育目标对知识和能力的双向要求，可以统一体现在对化学知识教学的要求之中。知识传授的要求，体现在知识的内容上，能力培养要求体现在知识的学习水平上。能力培养要求的具体化，通过确定知识的学习水平而得以实现。那么怎样确定知识的学习水平呢？对这一问题的回答，正是讨论认知领域化学教育目标的学习水平分类的意义所在。

二 国外两种教育目标分类学理论的特点

布卢姆等人提出的认知领域教育目标分类学理论，影响大，应用广。该理论的主要特点是：

(1)以外显行为作为教育目标分类的统一基点。该理论尽可能用意义较明显的行为动词来描述各个学习水平。由于外显行为是可观察、可测量的，因

而这种分类有利于确定和描述可观测的教育目标，从而有利于进行较客观的教育测评。

(2)以行为的复杂程度作为划分教育目标类别的依据。该理论的六种学习水平，从“知识”到“评价”，其代表的行为复杂程度依次增大，使分类表现出层次性；由于学习水平是按行为由简单到复杂的递增连续性划分的，任何两类相邻的学习水平之间相互连续，如“知识”和“理解”连续、“理解”和“应用”连续等等，使分类表现出连续性；由于复杂程度高的行为可以包含复杂程度低的行为，反映到学习水平分类上，表现出累积性，也就是，后继类别的目标总是包含了前继目标。如“应用”是在“知识”和“理解”基础上产生的；没有“知识”和“理解”，便不可能有“应用”。

由于该理论具有上述特点，使该理论具备了作为一种科学的分类理论的最基本的特性。按照这一理论对教育目标进行学习水平分类，比较容易达到目标系统设计的具体性原则和独立性原则的要求。但是，我们不能将该理论原封不动地套用到认知领域化学教育目标分类上。这是因为，该理论只是从一般教育目标分类出发，并未深入讨论各个具体学科教育目标的特点；这一理论和我国化学教学的实际还有一定距离。例如，根据我国中学化学教学大纲和教材，可作为“评价”水平的知识很少；有些知识只要求学生大概了解而并不要求准确记忆，即知识水平要低于“知识”；将“分析”和“综合”列为两个不同类别，不易为广大中学教师所接受，也给对目标的分类说明、知识分类、命题测量等带来一定程度的困难。所以，按照目标系统的可接受性原则，还需要对该理论的分方法进行分类方法进行一定程度的改造，使之切合我国化学教育的特点。

美国教育心理学家加涅(R. M. Gagné)主要是根据各种学习结果所需要的学习条件的异同而提出其教育目标分类学理论的。按照加涅的观点，不同种类的学习结果需要不同的学习条件。例如，学习者要学得概念，必须先具有辨别能力；要学得规则，必须在能够确定一类事物或现象的关系和属性的基本条件下才能够实现。虽然在加涅教育目标分类的主要类别(语言信息、智慧技能、认知策略)之间，不存在层次关系，因而各类目标之间没有层次性、累积性和连续性，但在智慧技能的各个亚类(概念、规则……)之间，却具有明确的层次性和累积性。根据加涅的理论，我们可以分析教学中各类知识的学习所需要的内部条件，其中包括相应的能力培养要求，从而确定各类知识的学习水平。加涅通过长期研究，已经逐步揭示了各种学习结果和各种学习条件之间的关系，其中有些部分可以直接为我们所借鉴。智慧技能类中各个亚类之间的层次划分，也可借鉴于知识的学习水平划分。此外，该理论还可以指导我们设计到达教学目标的教学过程。

三 中学化学教育目标的学习水平分类

许多化学教育工作者曾经对此提出过见解，如：

泛读、识记、理解、应用、综合；

识记、理解、应用、分析综合、探究；

初中：了解、理解、掌握；

上海市教科所：初中化学学习质量评价手册(内部资料)，1986年。

曾灼先等：广东省1987年化学高考标准化的设想和做法，化学教育，1987年第5期，第15页。

高中：识记、理解、应用、分析、综合、评价；
识记、简单应用、综合应用、创造应用；
识记、理解、简单应用、综合应用；
知道、领会或学会、应用和掌握、综合运用、创新；
识记、理解、简单应用、综合应用、创见；
对“理解、掌握、熟练掌握”做出行为界定；

.....
上述各种分类方法，实际上并无本质的差别，都可以看作是和布卢姆的认知领域教育目标分类学理论一脉相承的。

在对化学教育目标进行学习水平分类时，必须处理好概括化和具体化的矛盾。分类层次太少，会使目标过于概括、粗略、笼统，也就失去了目标分类的意义；但分类层次太多，又会使目标分得太细、琐碎，给目标的制订及其测量带来较大的困难，也会脱离当前教学的一般水平和条件。

我们根据对各种教育目标分类学理论的理解和对我国中学化学教学实际的认识，提出中学化学教育目标的学习水平分类如表 2 - 1。

唐力等编：化学教育测量，广西师范大学出版社 1989 年版，第 37—58 页。

广州市中学化学能力培养教改试验组：中学化学培养目标的制定和实施，化学教育，1986 年第 5 期。

裘谷永等：杭州市 1986 年高中招生考试化学命题改革试验报告，化学教育，1987 年第 5 期。

吴俊明等：试论中学化学教学目标的分类，扬州师院学报，1987 年第 12 期。

北京市朝阳区化学教研室：明确教学目标改进评价方法 提高教学质量，化学教育，1986 年第 6 期。

胡学增：关于现代教学测量与评价理论若干问题的思考，上海教育，1988 年第 7—8 期。

表 2 - 1 中学化学教育目标的学习水平分类说明

1. 知道：阅读教材或其他资料，对知识形成大致印象
(1) 查找：找出知识在教材或资料中的位置
(2) 概述：说出学习材料的大意
2. 记忆：记住学习材料的基本内容，注重知识“是什么”
(1) 辨别：辨别事实材料、概念、化学用语、实验现象及有关数据的正误
(2) 回忆：记住概念的完整定义，准确描述化学事实，正确书写化学用语
3. 理解：记住学习材料的要点，明确所学知识和其他知识的联系，了解知识的来龙去脉。不仅知其然，并且知其所以然
(1) 转换：当知识“改头换面”后，能抓住其实质；或把握同一类知识的不同方面及其内在的一致性
(2) 解释：说明事物的原因，回答“为什么”
4. 简单应用：掌握知识的应用范围和适用条件，能够举一反三，学会运用某一种知识解决一些简单的化学问题
(1) 直接应用：模仿知识的应用方法。在判明问题情境适合于知识的应用条件后，直接代入有关规则解决问题
(2) 间接应用：问题情境并不直接适合于知识的应用条件，而需要经过简单的分析、推理，对原有的知识、情境进行适当转换，然后解决问题
5. 综合应用：掌握知识之间的区别和内在联系，使知识系统化。在此基础上，综合运用多种知识，解决一些较为复杂的化学问题
(1) 归纳：根据问题的要求，对所学知识进行分析、比较、归纳，使之系统化
(2) 推理：根据自己的观点，综合所学知识得出新的认识，然后以新认识为依据解决问题

化学教育目标的学习水平分类与布卢姆认知领域教育目标分类之间，各个水平层次的对应比较如图 2 - 2 所示。

化学教育目标分类	布卢姆教育目标分类	
知道	知识	学习水平由低到高 ↓
记忆		
理解	理解	
简单应用	应用 分析	
综合应用	综合 评价	

图 2-2 化学教育目标分类与布卢姆教育目标分类的对应比较

图 2-2 化学教育目标分类与布卢姆教育目标分类的对应比较

通过教学研究实践我们体会到，目前各种学习水平之间的界限尚未达到泾渭分明的程度，相邻的两个学习水平之间有一定程度的交叉、覆盖，或者说，界限比较模糊。要使这些界限更为清晰，还有待于深入研究。对学习水平的亚层次的探索，则是这类研究的一个方向。我们将五种层次的学习水平

又进一步细分为十个亚层次，是对化学教育目标具体化的一种尝试。学习水平的更细致、具体的划分，是与学习心理过程的深入研究交织在一起的。学习行为被描述得越具体，学习水平的划分就可能越细致。随着学习心理研究的深入，学习水平层次还可以进一步细分。如图 2 - 3 所示。



图 2-3 学习的层次与学习水平分类的关系

第四节 化学教育目标的制订

一 化学教育目标的结构框架

在制订化学教育目标时，要根据教学大纲、教材、学生、教学条件等具体情况，考虑到化学教育目标在内容、类别、学习水平上的不同，并要符合教育目标制订的具体原则。从上述要求出发，试提出化学教育目标的结构框架。其要点如下：

(1)以化学知识和实验技能教学的内容及其学习水平分类为目标系统的主要载体，情感领域目标要结合知识和技能教学的具体内容制订。

(2)以化学知识结构为目标系统结构的依据，即目标的前后序列服从化学知识的结构序列。目前，主要是以化学教材中的知识序列为目标系统的结构依据；实验技能领域目标的结构序列服从知识结构序列。

(3)要根据化学知识结构和教学评价的要求，划分合适的教学单元，确定单元内所含知识、技能点的学习水平，作为学生的到达目标，指导每一个知识、技能点的学习；整个单元对学生在能力和情感方面的要求，作为方向目标提出，指导全单元的学习。

(4)教授目标和学习目标分列。学习目标要涵盖单元内知识、技能、能力学习的主要内容，包括准备目标、（课堂）完成目标、（课后）提高目标，使学生明确学习要求及其计划安排；教授目标要包含学习目标，并在此基础上，围绕课堂教学的完成目标，确定对教师的课堂教学行为的具体要求，以保证完成目标的实现。

(5)情感领域的目标，以体验目标的形式，只出现在教授目标中，要求教师在知识、技能教学的同时，采用适当的方法，引起学生内在的情感体验，以达到情感教育的要求。

二 化学教育目标的具体制订

1. 教学单元的划分

在划分教学单元时，需要注意以下两点：

(1)各个单元的教学内容要有相对完整性。

(2)单元教学内容及其所对应的单元教学时间要有利于形成性评价的实施。

为了使单元教学内容有相对完整性，要考虑到化学学科的知识结构特点及其教学要求。可以将化学教材中的章或节作为单元划分的基础，这样能够利用现有教材的知识结构，便于单元划分，提高工作效率。为便于进行形成性评价，单元教学的时限要适中。当单元学时过长时，关于教学或学习过程的信息会由于遗忘等原因而损失，不利于及时反馈；而单元教学时间太短时，势必会增加评价的次数，很可能导致教学效率的下降。单元教学时数可以有一定的弹性，以便具体教学过程的实施。根据我国中学化学课程设置的实际状况，可以4个教学周的教学时数之总和作为一个单元教学时数的基数，为了使单元教学内容有相对完整性，允许有一个教学周时数的弹性变化。

研究表明，以现行统编中学化学教材作为单元划分的基础是可行的。该教材是建国后40多年来中学化学教学经验的结晶，比较严格地体现了现行中学化学教学大纲的要求；它以物质结构理论为主线，较好地反映出现代化学科学的结构特点。由于该教材的结构体系比较合理，教材中每一章的知识内

容都有相对独立性，且每章的教学时数划分也基本上与形成性评价的要求相一致。因此，可以将此教材的一章作为一个现成的教学单元；对于其中个别教学内容较多、需要4个教学周以上的章，可以分为两个单元。这时，单元的划分又是以教材中的节为基础了。

2. 确定单元内各个知识点的学习水平

对于如何确定化学知识点的学习水平，目前尚无较明确的标准。我们认为，一般要考虑到以下因素：

(1) 该知识点在整个知识结构中的地位、作用及其发展。

(2) 该知识点在教学过程中对培养学生能力的作用，即该知识点的智力价值。

(3) 学生接受该知识点的学习基础及其他教学条件，即学生能够达到某种学习水平的可能性。

(4) 该知识点在解决常见化学问题中的作用或使用的频度，即该知识点的实用价值。

按照认知领域化学教育目标分类的五种学习水平，那些只需要学生了解而不必记忆、需要时知道从何处查找的知识，如物质的一般物理性质（熔点、沸点的具体数值等）、一般的化工流程、一般化学史知识等，可以划到“知道”水平；属于“记忆”水平的可以是那些要求学生较准确地记忆或再认，但并不理解（或是难以理解）的知识，如重要的化学用语、物质的物理性质、一般化学事实等类知识；属于“理解”水平的是那些要求学生弄懂、弄通、知晓其来龙去脉的知识，如一般化学概念、原理等；属于“简单应用”水平的是那些经常在各种化学问题中出现的、要求学生掌握其基本运用条件的知识，如一些重要元素化合物的化学性质等重要的化学事实、化学理论、化学原理等；属于“综合应用”水平的是那些应用相当广泛、可以成为不同知识之间的联结点，或是处于某些知识系统的中心的知识。对那些能够培养学生正确的思维方法、开拓思路、激发兴趣的知识点，要划到较高层次。通过考察和分析一些较重要的化学测试题，如高考化学试题等，发现那些出现在难度较大、综合性较强、对于解题十分关键、且出现频率较高的知识点，也将它们划到较高的学习水平之中。

为分散难点，对某些学生难以一次就达到较高学习水平的知识点，可以在目标系统中分级出现，开始出现在较低水平，以后逐步过渡到较高水平。

从学生的接受能力和教学条件出发，在不同的教学系统中，同一知识点可以划到不同的学习水平之中。如重点中学和一般学校在知识水平上应有所区别，后者在某些知识点上要比前者低一个层次或一个亚层次。

3. 确定方向目标

根据单元内知识内容、学习水平及其教学过程的具体特点，规定结合本单元知识的学习着重发展学生能力和进行思想教育的基本要求，以指导整个单元的教学过程。

4. 确定准备目标、完成目标和提高目标

在知识点及其学习水平确定之后，化学教育目标的主体——认知领域目标就已经确定了，也就是明确规定了学生的学习任务。但是，这并没有规定到达目标的具体过程。还需要指明教师和学生在学习过程中的更具体的行为目标，使他们明确自己的任务，并相互配合，以利于提高教学效率。

准备目标是学生在课前需要掌握的、作为课堂学习的基础的、过去的学

习目标。制订准备目标的目的是为了指导学生的课前预习。准备目标的内容应是那些作为课堂学习基础的、比较关键的知识点及其需要达到的学习水平，不仅是化学学科的知识，也有可能涉及与化学学习密切相关的其他学科的知识；准备目标也可以是学生通过自学就能够达到，因而不需要经过课堂学习的那些目标。

完成目标是通过课堂教学过程帮助学生达到的目标。这些目标是从全部目标中挑选出来的，以它们作为课堂教学目标，有利于教师发挥其对教学过程的主导作用，帮助学生理解知识、发展能力，提高学习效率。

提高目标是那些需要学生在课后通过进一步努力才能达到的目标。这些目标为学生的课后复习、提高指明方向。

5. 确定体验目标

根据课堂教学目标中所包含的化学知识的具体内容，确定教师必须采取的行为方式，以引起学生一定的感情体验，从而培养学生正确的兴趣、态度和某些基本观点。

第三章 化学教学测试题的编制

测量与评价是教学系统不可缺少的组成部分，测量又是评价的依据。为了全面、准确地获得教学反馈信息，必须掌握科学的测量方法。在教学过程中，最常用的测量方法是命题测验。本章首先讨论化学教学测量的基本原理，在此基础上，讨论化学测试题编制的基本原则及一般技术要领。

第一节 教学测量的基本原理

一 教学测量的一般特点

所谓测量，实际上是一种比较过程，是通过将被测物体与参照物进行比较，从而对被测物体赋值，以说明物体的某种属性的过程。我们可以把被测物体与参照物放在一起直接比较，也可以通过标准物体而进行间接比较。我们比较熟悉的物理测量一般也是一种间接比较过程。例如，为了测量桌子的长度，就可以用米尺来完成。米尺在测量过程中起到了两个作用：作为一种测量工具，确定桌子的长短；作为一种测量尺度，为测量提供了标准物——以米为单位的测量尺度。

所谓教学测量就是应用一定的测量工具，收集各种和教学有关的信息、资料，为教学评价提供依据。它尽管比物理测量复杂得多，但两者从本质上看却有类似的原理。美国教育心理测量专家桑代克和麦柯尔（W.A.Mecall）早就提出过一个著名的假说：“凡是存在的就有数量，既有数量即可测量。”教学中常用的测验或考试，就是一种测量工具；表示测量结果的分数就是一种测量尺度。测量的基本过程是：用一组标准刺激物——测验题——去激发学生的反应；根据学生的答题情况对学生的反应质量赋值——评分。与一般物理测量相比，教学测量具有以下特点：

1. 测量尺度有较程度的不确定性

教学测量的尺度一般是无形、抽象的，不像物理测量那样容易客观地确定和把握。在教学测量中，对同一对象的测量，不同的测量者使用的尺度可能并不相同，甚至同一测量者对同一测量对象在不同时刻进行测量时，其测量尺度也会不一致。由于教学测量的尺度一般都存在于测量者的头脑中，很难对其进行可靠性检验。

2. 测量的间接性

我们实际上并不能直接测量学生内在的掌握知识和能力发展水平，而只能通过测量学生的有关外在行为表现来实现我们的测量目的。这类似于用温度计测量物体的温度，是通过水银柱的高度变化而间接地推定物体的温度。当然，学生的外在行为表现与其内部发展水平之间的关系要比水银柱高度与物体温度间的关系复杂得多。

3. 测量的代表性

在进行测量时，我们不可能、也没有必要测量事物的全部属性，而只是选择那些有代表性的、反映事物本质的那些属性作为测量目标。在教学测量中，要确定那些反映教学目标到达度的代表性属性或学生的典型行为表现，要比确定代表桌子长度的属性困难得多。

4. 测量单位的近似性

作为物理测量的单位有两个基本条件：一是意义确定，二是单位等值。教学测量中的单位只能近似地满足这两个条件。常用的考试分数作为测量的单位，1分并不像1克那样有确定的意义；80分与90分的差异也不等同于50分与60分的差异；10分之差不能像10毫升之差那样意义明确。

对教学测量的基本特点的了解，可以帮助我们在进行教学测量时，保持清醒的头脑和采取正确的态度。既要看到对学生内在特性进行测量的现实可能性，因而不能全盘否定教学测量的意义和作用，又要看到教学测量与物理测量的本质区别，了解其局限性，而不要盲目、机械地使用测量结果。

二 教学测量的基本步骤及要求

1. 明确所要测量的属性

在进行任何测量之前，首先必须解决为什么测量和测量什么的问题，即明确测量的目的和内容。

测量内容是由测量目的决定的。为了达到测量目的，我们要明确哪些属性对测量是适合的和必要的。要测量的应是那些与测量目的密切相关的属性，避免那些与测量目的无关的属性。这样，不仅可以节省人力、物力、财力，还有利于测量结果的正确性。当与测量目的有关的属性不止一个时，要把有关属性全部包括在测量内容之中，并予以适当权衡。

在教育测量中，当测量目的确定之后，需要测量的属性却不能像物理测量那样容易确定。物质的某些被测量的属性，如体积、质量、温度等，一般已经被确立、公认或公理化了；而教育测量的对象是学生的某种心理属性，对这类属性的认识，常常不像对物理属性那样有一致的看法。例如，在进行智力测验时，首先要回答：什么是智力？哪些行为可以看作是智力水平较高的行为？……对这些问题，至今尚未统一认识。因而教育测量的第一步，是要在测量者的认识基础上，根据测量目的的需要，对所要测量的属性下一个比较清楚、精确和能为人们所接受的定义。

根据化学教学目标的要求，认知领域化学教学测量的目标是学生掌握知识的程度和能力发展水平两个方面，这主要通过测量学生“能否在不同的学习水平上运用知识解决问题”这一属性来实现。对这一属性的认识将随着教育心理学、教学测量实践的深入而逐步发展。

2. 确定显示属性的方法

在明确了所要测量的属性之后，测量的第二个步骤就是寻求一套能够分离出我们所要测量的属性，并使之显示出来的方法。比如，为了测量桌子的长度，我们可以把直尺（测量工具）与桌子较长的那一边（测量对象）进行比较（实施测量），从而显示出桌子的长度（属性之一）。这是一套最简单的测量操作方法。由于各种物理测量工具及其操作的结果是一致的，可以经过再次操作被验证，因而已被普遍接受。

教育测量的操作方法要取决于我们所接受的待测属性的定义。实际上，在对属性定义时，往往已经考虑到能够引入或展示属性的方法，并从可以实施操作的角度构成属性定义，即操作定义。当定义比较含糊或不能引入显示定义的操作方法时，对该属性的测量就可能有五花八门的标准，所得到的结论自然难以一致。设计出和属性定义一致的、能够比较清楚和精确地显示被测属性的科学的测量操作方法，是教育测量研究的主要课题。

在化学教学测量中，为了考察学生是否能够在不同水平上运用知识解决问题，可以采取的一种测量方法是：通过精心设计的测试题，给学生提供一个问题情境，这一情境既涉及到知识的具体内容，又涉及到运用知识的难度水平。学生在此情境中进行一系列的思维操作，最后给出操作结果，并显示出其外在行为——给出问题的答案。若答案正确，则可以认为学生掌握了该知识并达到了相应的学习水平。学生在试卷上给出的一系列答案，即可作为教学测量的结果。

3. 属性数量化

在确定了显示属性的操作方法之后，测量的第三个步骤就是把操作结果

用数量表示出来。就桌子的长度来说，我们测量之后要得出的结果是多少米等，这一结果就是桌子长度的数量化。米是代表长度的基本单位，也是物理测量的标准物之一。对学生心理属性的测量，也要有类似的标准物，它们被称之为量表。教育、心理测量量表有以下五种类型：

(1) 名义量表。

又称分类量表。它只对被测变量作定性描述，与分类有紧密联系。例如，学生对选择题的答案只按对或错加以区分；对学生的性别只能按男女分类；对中学可以按照省属重点、市属重点、区重点、一般中学等分类。这些都可以用名义或分类量表描述。

(2) 位次量表。

又称等级量表。它是用以反映事物的相对关系（顺序关系）的数值来表示的量表。该量表既无相等单位，也无绝对零点，只能将变量排出一个顺序，而不能指出其间的差别大小。对该量表的数值可以按次序进行统计，但不能进行加减运算。一般能力等级、品质等级、喜爱程度等，都可以用该量表来描述。教学中常用的“甲、乙、丙、丁”、“A, B, C, D, E”、“优、良、中、差”等都属于等级量表。

(3) 等距量表。

在这种量表上，单位间的距离是确定的，对变量可以进行加减运算，但由于变量不具有绝对零点，不表示倍数关系，因而不能进行乘除运算。常见的温度、各种能力分数、智商、教学中的考试分数等都属于等距量表。

(4) 比率量表。

这类量表既有相等单位，又有绝对零点，因而可以对变量进行代数运算。常见的身高、体重等都属于比率量表。教学中使用的标准分也具有比率量表的性质。

(5) 模糊量表。

这种量表是用模糊数量的方法来描述那些存在着亦此亦彼的中间过渡状态的事实，用以反映事物处于某种状态的程度。如描述一位教师的教学态度，可以先确定教学态度表现的几种状态，并用数值来代表这些状态。然后，根据被评教师的实际表现，看其接近某种给定状态的程度，再赋予其适当的数值。

在进行教学测量时，要根据测量目标及其被测属性，选择适当的量表使属性数量化。对使用不同量表描述或记录的测量结果，必须使用与量表性质一致的统计分析方法，才能得到正确的结论。

三 对测量工具的质量要求

1. 可靠性

测量工具的可靠性表现在：当使用测量工具进行测量时，不论是多次重复测量，还是由多人进行测量，其结果都是一致的，表明测量过程的一贯性、一致性、再现性和稳定性。物理测量工具的可靠性是明显的。例如，一把没有伸缩性的尺子，多次测量同一张桌子的长度，结果都可以是相同的。但是在教学测量中，由于测量尺度的不确定性，影响测量过程的因素也比较复杂，类似的没有伸缩性的尺子却不易找到。实际上，任何对个人的行为样本进行的测量，必然或多或少地包含着某些测量误差。因而，没有一种教育测量工具是绝对可靠的。

反映测量工具可靠性的指标称为信度。当然，不具备必要信度的测量工具是没有使用价值的。

2. 有效性

一个测量工具，要能够真正测出所欲测量的某种属性，这样的测量工具才是有效的。例如，要测量桌子的长度只能用尺子；但是再精确的尺子也量不出物质的质量。这就是说，尺子对测量长度是有效的，而对测量质量却是无效的。教学测量工具同样存在有效与否的问题。如化学测验中就不能有过多、过难的数学计算，否则，测验结果说明的不是学生的化学学业成就，而是数学成就。

反映测量工具有效性的指标称为效度。在教学测量中，越是能够抓住测量目标并紧扣测量属性，测量工具的效度就会越高，有效性就越好。

第二节 命题计划的设计

一 命题计划的意义与作用

对命题者而言，命题计划可看作是一份工程蓝图，是编制试题时的指导和依据。如果命题计划制订得准确、合理，那么，只要命题者严格按照命题计划编制试题，就可以保证试题的质量。

在化学教学实际中常常可以看到这样的现象：命题者在编制试题前有着不同程度的心中无数，命题过程中难免有一定程度的主观随意性；有的命题者在试题使用之后还不能较清楚地回答试题的测试目的和编制依据等命题的基本问题。这就很难保证试题质量了。我们知道，“任何测验都只能是与课堂上所涉及的内容有关的学生行为的样本。”实际上，我们不可能测验在课堂教学中涉及的全部内容，而只根据测验的具体目的去选择那些最重要、最有代表性的教学内容来进行测试。试题是由“项目”——需要学生逐个回答的各种类型的问题——构成的。命题的主要过程实际上是选择或编制合适的项目并将它们合理组合的过程。要保证试题有良好的效度，关键是要能够保证选择或编制出的项目所涉及的测试内容确实很重要，真正能够代表那些我们所欲测量的教学内容。这就必须在选编试题之前，精心设计命题计划。

一份好的命题计划有以下两个最基本的作用：

(1) 保证试题是所要测量的教学内容的代表性样本，且能够反映出各个部分内容之间的相对重要性，以便试题取样适当，提高测量的效度；

(2) 规划代表不同的知识内容和学习水平的各类测试项目的比例分配，在保证试题效度的同时，保证试题的难度合理。

二 设计命题计划所要解决的主要问题

1. 测验的主要目的是什么？

化学教学中常用的测验，按其测验目的可以分为预测测验、诊断性测验、形成性测验和总结性测验四类。预测测验（或预备测验、安置测验）是为了对学生进行合理的分班、分组、安排教学内容和进度，在讲授新知识之前，先对学生应具有的有关学习基础进行测验；诊断性测验是为了发现学生在学习中的问题，以便对症下药，改进教学；形成性测验的目的在于了解学生在教学过程中知识、能力、思想品德等方面的形成情况，及时掌握反馈信息，以便辅导学生学习，弥补教学中的不足，改进教学工作；总结性测验是在某一门学科或某一学习阶段结束后，为了对学生的学习成果进行总结、鉴定时所进行的测验，通常是在学期或学年末进行。测验的目的不同，选编项目的侧重点也应相异。

2. 怎样确定试题的题量？

试题的题量或测试内容的多少，与给定的测试时间有关。测试时间一般是在命题之前根据测验目的确定的。形成性的单元测验可用 15 分钟左右，总结性的学期或学年考试多为 90—120 分钟。测试时间要和测验目的一致，同时又要考虑到教学效率；时间过长，会占用过多的教学时间而影响教学任务的完成；时间过短，势必会限制试题的题量而影响测试目的的达到。

在测试时间确定之后，试题题量的确定依据是：学生回答试题所需要的平均时间和给定的测试时间一致。这里的平均时间是指：测试者所希望的一定比例的学生完成试题所需要的时间。我们认为，对形成性测验，要让90%以上的学生完成试题，以免损失过多的教学反馈信息和影响学生的学习积极性；对总结性测验，应使80%以上的学生完成试题，这样尽管对部分学生而言，题量过大，但对总体学生而言，可以使试题有足够的题量，保证试题对测试内容有必要的覆盖面，同时提高考试的效率。

在估计学生的答题时间时，需要考虑到以下三个基本因素：

(1)学生的实际水平。

学习水平较高的学生，知识掌握得好，解题能力强，因而答题速度快，需要的答题时间就短；反之，学习水平较低的学生需要的答题时间则较长。

(2)试题本身的难度。

这里的难度是指答题的思考难度。难度较高的项目所占的比重较大，所需要的答题时间越长，如回答属于“应用”水平的项目所需要的时间显然要比“记忆”水平的项目更长一些。但项目的难度要受到学生练习效应的影响，如学生通过练习已经了解了项目的解法或记住了项目的答案，就可能把命题者按“应用”水平设计的项目难度降低到“记忆”水平的项目难度了。因此，在估计项目难度时，要注意练习效应的影响，比如，尽量增大试题的样本数或尽可能多地采用新样本。

(3)阅读试题和书写答案所需时间。

这是指除解题思考所用时间以外，用于阅读试题和书写答案所需要的时间。如果项目以较长的文字叙述或以图表等形式表述时，就需要较长的阅读时间；需要数学计算的项目要比仅涉及单纯文字的项目答题时间要长一些；回答论文式项目显然要比回答选择题需要的时间更长。

3.怎样确定各类项目的比例？

试题是由项目构成的，项目质量决定着试题的质量。就一个具体的测试项目而言，它实质上是命题者给学生提供了一种运用化学知识解决化学问题的特殊情境。因而，它既涉及到一定的化学知识内容，又包含着运用知识解决问题的水平。

在题量大致确定之后，为了使项目的选择符合测试目的，要确定代表不同的知识内容、不同的学习水平或难度的各类项目的比例，使之能够对教学目标或教学内容有较好的代表性，试题的总体难度适合于被测学生。

对形成性测验而言，要以单元教学目标为依据，先要保证每个教学目标都有一个测验项目与之相对应，在此前提下，再考虑那些在教学过程中发现的、需要通过测验进一步了解和确认的问题，如估计学生在某些教学目标上的到达程度可能不够理想，或估计学生在学习过程中可能会出现错误等等，以便在试题中有意识地增加有关项目的比例。

对总结性测验而言，由于测试的范围更广泛一些，所以，确定各类项目的比例时显得更为复杂一些。代表不同知识内容的项目比例，可依据该知识内容所需要的教学时间在测试范围内所有知识内容所需要的总的教学时间中的比例来确定。也就是说，达到某个教学目标所用的教学时间越多，代表该目标的测试项目所占的比例也应越大。由于教学目标是分单元、按章节制订的，所以，可以把某章节相对应的教学时数与测试涉及所有章节教学的总时数之比，作为代表该章节教学目标的项目应占的比例。以教学内容所需教学

时间的比例作为确定代表该教学内容的测试项目在试题中应占的比例的依据，比较客观、合理，可以有效地保证测试内容取样的代表性。

测试项目的难度与项目涉及的知识的水平有关。一般而言，学习水平越高，运用知识解决问题的难度也越大。因而可以根据项目代表知识的学习水平来估计项目的难度。在形成性的单元测验中，各种学习水平的项目比例已经由项目代表的教学目标确定，不需要另行设计；对总结性测验而言，涉及的教学内容较多，需要对测试目标进行抽样，这就需要考虑不同难度水平的项目比例问题了。由于总结性测验要求具备对学生的学业成就进行鉴定并区别优劣的功能，因而在项目分配上，既要涉及一般水平的学习能力，还要有一定比例的、涉及较高学习水平的项目。我们建议，如果教学目标按照“知道、记忆、理解、简单应用、综合应用”五种学习水平分类，其中“知道”水平的教学目标一般不作为测量目标，代表其他四种学习水平的项目可按表 3-1 分配比例。

表 3-1 总结性测验不同学习水平(难度)

学习水平	难度(通过率 (%)	比例分配 (%)
记 忆	> 85	15
理 解	50 — 85	35
简单应用	15 — 50	35
综合应用	< 15	15

- 说明：1. 难度比例按能够基本正确地回答该类项目的学生数所占参加测试的学生总数的百分比计算。
2. 比例分配按该类项目的总分值所占试题满分值的百分比计算。

这里，我们已经将项目所对应的学习水平作为估计其难度的依据了。尽管这种估计可能不是十分准确，但要比仅仅凭主观经验估计更好一些。我们可以通过测试后的统计分析，对原来的估计不断进行调整，以至于逐步达到在下次测试时对项目难度估计到比较准确的程度。

三 命题计划的构成与使用

在明确代表不同知识和不同学习水平的各类项目的比例之后，就可以构成命题计划，并将其以命题蓝图的形式表达出来。表 3-2 就是这种蓝图的示例。表中的每个数字代表其所在章（横向对应）和应有的学习水平（纵向对应）的某一类项目的比例（按项目的分值计算）。从表中可以看出：第一章的知识内容在试题中出现的比例是 30%（第 1 行、第 5 列所对应的数值）；“记忆”水平的项目在试题中的比例是 16%（第 4 行、第 1 列所对应的数值）；而第一章知识内容中涉及“记忆”水平的项目比例是 5%（第 1 行、第 1 列所对应的数值）。

有了命题蓝图，可以使我们有目的、有计划地选择或编制项目。在选择或编制项目时，先确定项目所代表的知识内容，再估计项目的学习水平或难度，然后与命题蓝图逐项对应。这样，项目取样能够有较好的代表性，从而可以保证试题的效度。

设计命题蓝图的技术要点如下：

命题原则 1：明确测验的目的，使试题的内容和形式符合测验的目的和要求，能够测出所要测量的知识和能力。

命题原则 2：试题内容的取样要有代表性，即要对测量目标有充分的覆盖面，且各个部分内容的比例要适当。

第三节 不同类型项目的编制

一 项目的类型

项目类型，又称为题型。不同类型的项目有不同的性质特点，在测验中的作用也不一样。因此，命题者必须对不同类型项目的性质特点及其基本的编制要求有比较明确的认识。

化学教学测验中常见的几种题型分类如图 3 - 1 所示。



图3-1 化学教学测验中常见的几种题型分类

这种分类的标准是给出答案的形式。论文式项目的答案完全是由答题者给出的，命题者几乎不作限制，因而答案是最不确定的，几乎没有标准答案，而只能有答案要点，评分时受评分者主观因素的影响较大，所以又被称之为主观性项目；客观式项目的答案范围完全被命题者所确定，答题者只能在给定的范围内选择答案，因而正确答案是确定的，不同的评分者可以得到完全相同的评分结果，评分比较客观；限制式项目答案的确定性介于论文式项目和客观式项目之间，其答案的确定程度取决于项目的具体编制方法，如填空题，其答案可以是一个单词和符号，也可以是好几个单词或句子。但限制式项目的答案不能像论文式项目那样，基本上由答题者个人所确定，而要在一定程度上受命题者或项目本身的制约。

不同类型的项目有不同的性质特点。桑代克等人总结了各类题型的应用范围如表 3-3 。

表 3-3 各种测验类型优缺点一览表

测量要表	论文式	限制式	客观式
1.能测量解决新问题的能力	++	+	++
2.能测量组织、综合能力	++	+	--
3.能测量独到见解或对问题的创新能力	++	+	--
4.能把在某个题材范围内所具有的特殊才能与一般的写作、拼写和语言应用能力区分开来	--	-	++
5.具有诊断的潜在价值目标	--	-	++
6.能够对教学目标进行抽样检查	--	-	++
7.能够对教学内容进行抽样检查	--	-	++
8.没有猜测答案的机会	++	++	--
9.不同的评分人能评出相同的分数	--	-	++
10.能准确地区分考生的能力水平	--	-	++
11.能由未经训练的人员或机器评分	--	-	++
12.能够迅速评分	--	-	++
13.命题所需时间不多	+	+	-

从表 3-3 可以清楚地看出，客观式项目的优点明显多于论文式项目和限制式项目。客观式项目答题方式最简便，便于在短时间内测量较多的教学目标，取样较广泛，评分客观，可用机器代替人工阅卷；但构成试题所需要的项目较多，因而命题所需要的时间、精力较多；由于客观式项目提供了确定的被选答案，学生只能在其中做出选择，所以它难以测量学生的综合、创新能力，也提供了猜测答案的机会。客观式项目的缺点所在正是论文式项目的优越之处。但论文式项目的缺陷也是明显的，客观式项目正是在对论文式项目所存缺陷的认识基础上发展起来的。一般认为，客观式项目和论文式项目各有短长，限制式项目的优、缺点也介于两者之间。因此，过分地推崇或排斥某一类型项目的做法都是不可取的。在教学测量中，要灵活地选用各种不同类型的项目，使之符合测量目的。

在化学教学测量中，上述各类项目都被使用。其中最具有代表性、也是最为常用的项目有三类，即选择题，填空题，简答题。

选择题是最典型的客观式项目。它除了具有便于广泛取样、评分客观等优点外，通过对学生项目反应情况的统计分析，可以得到较多的教学反馈信息。这类项目已被越来越普遍地使用着。是非题可看作是二选一的选择题；配对题也可以很方便地改造或分解成为选择题。从测量要求看，填空题和简答题还是必要的。简答题可以代替论文式项目满足化学教育测量的大部分测量要求，其不足部分可以采用考试之外的其他形式，如让学生做化学小论文等作补充。对于那些要求学生准确地回忆化学事实、化学用语等测量目标，填空题有着其他题型不能替代的测量功能。

二 选择题、填空题及简答题的编制技巧

不论是何种类型的项目，其表达和指导语句的文字都要清楚，要使学生明确需要他做什么、怎么做、以何种形式给出答案等；文字要尽可能地简明、扼要（当然不能遗漏重要的解题条件），尽可能降低阅读难度，防止学生产

生歧义。即：

命题原则 3：题意要清楚，文字要简明。

不论是何种类型的项目，其答案都要有可靠的依据。答案应当是确定的，但并不排斥有一个以上的正确答案，特别是对填空题和简答题更是如此。不过，在正确答案不止一个时，必须对答案的可能范围有清楚的了解。这样，一是为了保证评分的客观性，二是为了保证测验的教学作用，不至于因为答案有问题而对学生产生不良影响。即：

命题原则 4：应有不至于引起争论的正确答案或对答案范围有清楚的认识和把握。

1. 选择题

(1) 题干中要包含尽可能多的项目内容，备选答案要尽可能简短。

例 1 欠佳题

二氧化碳是..... ()

- (A) 一种不易溶于水比空气重的气体
- (B) 一种无色无味的易溶于水的气体
- (C) 一种能溶于水比空气重的气体
- (D) 一种黄绿色有刺激性气味的气体

修改后的题为：二氧化碳气体..... ()。

- (A) 不易溶于水比空气重
- (B) 无色无味易溶于水
- (C) 能溶于水比空气重
- (D) 黄绿色有刺激性气味

(2) 要尽可能避免否定式陈述。

当题干使用否定式陈述时，要求学生在各个备选答案中选择其中的内容错误的答案，才能获得项目的正确答案。这样往往会使学生答题困难，而困难原因和测量目标之间并不存在紧密联系，即使学生答案正确，并不能因此而断定学生真正掌握了正确内容，无法给教师提供学生掌握正面知识的信息。不过，有时为了让学生发现自己的错误，如过分地肯定一切或否定一切、忽视某些例外情况等，可以用否定式结构，以引起学生的注意。

(3) 尽可能增大备选答案的迷惑性，降低猜中正确答案的概率。

要使备选答案属于同一类性质。如备选答案的长度相似；叙述上不采用特殊的修饰词；剔除那些有明显错误的答案；避免在题干中使用与正确答案中相类似的词等。

多使用一些似是而非的迷惑答案。那些似乎合理的答案往往来自于教师从学生的作业、课堂回答问题和平时的观察了解等方面得到的有关学生学习过程中常见错误的信息。因此，平时要注意积累有关信息。

避免使用“以上皆是”或“以上皆不是”作为备选答案。学生只需要发现有一个错误选项，即可排除“以上皆是”这一备选答案，从而增大了猜中答案的机会；使用“以上皆不是”作为备选答案，相当于采用了否定式的陈述，仅能测量学生识别某些错误的的能力，但识别错误不能代表掌握了知识。

增大备选答案的数目。例如将备选答案从 4 个增加到 5 个。

2. 填空题

(1) 空出来的应是那些带关键性的词。

填空题的空白要求学生用测验范围内某个重要事实或重要的词来填写，

避免那些次要、枝节的内容。

例 2 欠佳题

每_____物质都含有阿伏加德罗常数个微粒。

修改后的题：

每摩尔物质都含有_____个微粒。

(2)空白不宜过多，也不宜出现在句首，以便学生理解题意。

例 3 欠佳题

_____元素包括 F, Cl, Br, I。它们的_____都有 7 个电子。

修改后的题：

卤族元素包括_____；它们的最外电子层上都各有_____个电子。

(3)要避免一空多答；当答案是数量时，要规定数量单位和数字的精确度。

例 4 欠佳题

将铝片和铜片浸入到硫酸铜溶液中组成原电池，则该电池的正极是_____，负极是_____。

该题可以有多种答案，如正极可填：铜片，发生还原反应的极，析出铜的一极等。

修改后的题：

将铝片和铜片浸入到硫酸铜溶液中组成原电池，则在该电池中铜片是_____极，铝片是_____极。

例 5 欠佳题

能够和 0.5L1mol/L 盐酸完全反应的氢氧化钠的量为_____。

修改后的题：

能够和 0.5L1mol/L 盐酸完全反应的氢氧化钠的量为_____g (或 mol)。

(4)避免可能的暗示。

例如，空格的长短可能会提示答案字数的多少，因此，要使空格的长度相等。

(5)避免那些只需要机械记忆就能回答的项目。

除少数需要考察学生对重要的概念、定义准确记忆的项目以外，要避免照搬教材上的原句。

3. 简答题

(1)应主要用以测量较高学习水平的目标。

由于简答题不便于广泛取样，评分时主观因素影响较大等缺点，一般尽可能不用这种题型；但有些测量目标用选择、填空难以实施测量，如需要考察学生对化学实验过程的掌握，或需要了解学生进行化学计算的过程和步骤时，就可以利用简答题这种难以为选择或填空题替代的特殊功能。一般简答题只用于那些涉及较高学习水平的目标，如“简单应用”以上水平的目标。

例 6 欠佳题

什么是氧化剂？什么是还原剂？

这样的项目只能考察学生对概念的记忆，不需要用简答题型。

修改后的题：

试从电子得失的观点说明什么是氧化剂？什么是还原剂？

或：

物质中某元素_____电子被_____，该物质就是氧化剂。

(2)要对答案的范围有明确、具体的限制。

这样可以保证在一定时间内让学生完成项目，并使答案有一定的可比性。

例7 欠佳题

怎样制取硝酸？写出制备过程并说明反应条件。

由于硝酸的制备可以有多种途径，使该项目有多种不同答案；即使要考察学生对知识的分析、综合能力，也需要对答案范围有所限制。

修改后的题：

怎样从氨氧化制硝酸？写出制备过程并说明反应条件。

(3)计算题以测量学生是否掌握计算原理和方法为主要目标 数学运算不宜过难、过繁；当不需要了解计算过程时，计算题可用选择题型代替简答题型。

例8 简答题型的计算题

0.1mol/L 氢氧化钠溶液 100mL 和 0.5mol/L 的盐酸 50mL 相混合，求混合溶液中氢氧化钠的物质的量浓度和氯化钠溶液的物质的量浓度。

例9 选择型型的计算题

0.1mol/L 氢氧化钠溶液 100mL 和 0.5mol/L 的盐酸 50mL 相混合，混合溶液中氢氧化钠的物质的量浓度为_____mol/L 氯化钠溶液的物质的量浓度为_____mol/L。

(A)0.025 (B)0.05 (C)0.0125

(D)0.01 (E)0.0166

(4)要提供可能出现的各种答案和相应的评分要求。

简答题的评分标准要细至各种可能的答题步骤，每一步骤都有相应的给分点，对各种答案也要有相应的处理办法，以保证评分的客观性。

(5)要避免让学生在给定的项目中任意选择作答。

如果让学生在给定的项目中任意选择作答，会使学生成绩失去可比性，也增大了学生押题或猜中答案的可能性，导致更多的测量误差。

第四节 不同学习水平项目的编制

化学教育目标规定了知识的不同学习水平。在测量不同学习水平的教育目标时，项目给出的问题情境要与学生在学习相关知识时所要求的学习水平相同，让学生在适当水平的问题情境中活动，根据其反应，判断其是否达到目标要求。这就要求项目的测量水平必须和相应教育目标的学习水平一致。

一套与教育目标直接对应的测试项目，可以进一步体现教育目标对学生的行为变化的具体要求。因而，设计出能够准确体现教育目标的测试项目有着更为重要的意义。只有教育目标和相应的测试项目结合在一起，才能保证教育评价的顺利进行。这一要求是项目编制过程中的指导思路，也是项目编制的主要技术关键。

然而，要编制出不同学习水平的测试项目，并能使其与相应的教育目标紧密对应，却是一件相当困难的工作。这与教育目标的制订一样，也难免受到较多的主观因素的影响。这也正是化学教育测评需要研究的一个重要课题。

设计各种学习水平的测试项目都需要注意练习效应对项目学习水平的影响。当问题材料已经为学生所熟悉时，材料所提供的情境只能让学生回忆出问题的答案，而不能使学生按照命题者所希望的学习水平活动。此时项目的学习水平只能是“记忆”。因此，要遵循下列命题原则：

命题原则 5：要尽可能采用对学生而言是较新的问题材料。

虽然项目所对应的学习目标是学生所学习过的，但可以采用与目标学习时不同的变式材料构成项目，特别要注意的是，不要从学生都可能找到的教学参考资料、习题汇编等材料中原封不动地照搬习题作为测试项目。

一 记忆水平的项目

记忆水平项目所涉及的目标，一般是要求学生能够记住所学知识，而并不要求学生理解。测量时主要看学生能否回忆或再认出有关知识。

例 10 填空：地壳中含量最高的元素是_____，其次是_____。

填空题是编制记忆水平项目最常用的题型。学生一般只有能够准确回忆出所测量的知识内容时才能给出这类项目的正确答案。记忆水平的项目有时也可用选择题来设计，例如：

例 11 选择：高锰酸钾的化学式是..... ()。

- (A) $KClO_3$ (B) KI
(C) $KMnO_4$ (D) K_2MnO_4

回答这一类项目时，学生不需要准确回忆而只要能够辨别或再认出有关知识即可给出正确答案。因此，这类项目的测量目标属于低于“记忆”的“辨认”水平。当采用选择题测量记忆水平的目标时，学生往往会从题中的备选答案中得到正确答案的提示。选择题在测量记忆水平目标时的这种暗示效应很难避免，因而一般不采用选择题测量记忆水平的目标。

记忆水平的项目常用于对重要的化学用语（化学式、化学方程式等）、某些代表性元素及其化合物的重要物理性质、用途、有关化学事实等知识的测量。

二 理解水平的项目

理解水平的项目测量学生是否了解有关知识的由来及其主要特征。这类项目所给出的问题情境，与目标学习时的情境基本相同，只是在某些非本质方面有所变化，要求学生能够识别那些经过改头换面的知识，抓住其本质特征；或是在抓住知识的本质特征的基础上，解释有关事物或现象，说明其原因，回答“为什么”之类的问题。

例 12 选择：下列关于 2mol/L 氢氧化钠溶液的叙述中不正确的是

..... ()。

- (A) 1L 溶液中含有 2mol 氢氧化钠溶液
- (B) 把 2mol 氢氧化钠溶于 1L 水中所形成的溶液
- (C) 把 2mol 氢氧化钠溶于水，使其体积成为 1L 而得到的溶液
- (D) 把 80g 氢氧化钠溶于水，使其体积成为 1L 而得到的溶液
- (E) 溶有 40g 氢氧化钠的 1L 溶液，将其一半体积的水蒸发去之后所得到的溶液

该项目的测量目标是对“物质的量”概念的理解。

三 简单应用水平的项目

简单应用水平的项目测量学生能否将所学知识应用于新的问题情境之中，能否把握一些重要的化学概念、定律、原理和公式等知识的适用范围及条件。在回答这一类项目时，要求学生能够判断问题情境和有关知识之间的适用程度，然后直接运用所学知识解决问题；或是能够对原有知识或问题情境作适当转换，使之相互适应，进而解决问题。

简单应用水平的项目所给出的问题情境一般不同于目标学习时的情境，但所涉及的知识点是单个的，不需要学生综合知识或抓住知识之间的相互联系，因而解决问题时所需要的分析、转换、推理等过程相对比较简单。

例 13 填空：要配制 0.5mol/L 的硫酸溶液 200mL，需要 1mol/L 的硫酸 _____ mL。

该项目要求学生在理解物质的量浓度概念的基础上，运用“配制前后溶质的量不变”这一原理解决问题。

例 14 选择：下列各种微粒中，只能作还原剂的是

..... ()。

- (A) Cl_2 (B) S (C) S^{2-} (D) H^+ (E) O_2

该项目要求学生根据单质或化合物中元素的化合价，判断其得失电子的可能性。

例 15 选择：制备下列气体时，不能用浓硫酸的是

..... ()。

- (A) HCl (B) HF (C) H_2S (D) SO_2 (E) CO_2

该项目要求学生运用浓硫酸的氧化性和给出的五种气体被氧化的可能性等知识解决问题。

简单应用水平的项目常用于测量一些代表性元素或化合物的化学性质等重要化学事实、重要的化学概念及理论等方面的知识。

四 综合应用水平的项目

综合应用水平的项目测量学生能否根据解决问题的需要，通过分析、推理、概括、综合，抓住知识之间的相互联系，重新组织材料，找到解决问题

的途径。综合应用水平的项目所给出的问题情境一般都是全新的，解决问题所涉及到的知识点多在两个或两个以上。

例 16 有 4 瓶失去标签的无色溶液，分别是氯化钠、硝酸钠、碳酸钠和溴化钠。请用实验方法进行鉴定，写出实验步骤、现象和有关的化学方程式。

该项目要求学生综合应用 4 种化合物的特性知识，并比较它们之间的区别和联系。

由于综合应用水平的项目涉及到多个知识点，如果对项目进行“双向”分析，应当如何确定其所属的知识点呢？可以从以下两个方面考虑：根据项目中主要的、在解决问题过程中起决定作用的知识点决定；根据知识学习的顺序，由最新学习的知识点确定。

第五节 项目编排与试题的确定

一 编排项目合成试题

编排项目之前，先要对项目再次进行选择。根据命题计划的要求，所有测量目标都应有编制好的项目与之相对应。为了保证试题的质量，避免项目之间的相互影响，每个测量目标最好有2—3个项目与其对应，以便在合成试题时对项目作进一步挑选。

在合成试题时，需要遵循以下两个原则：

命题原则6：各个项目必须彼此独立，不能相互影响，每个项目都不能为其他项目提供暗示或正确答案的线索。

由于编制项目时无法考虑到项目之间的相关，有时某一项目的表述内容有可能成为解答另一项目所需要的信息，如一个项目要求学生回忆的分子式或化学方程式，正好出现在另一项目的表述文字中；或是某一实验题给出的实验内容，正是另一道计算题所需要的计算原理，等等。这是合成试题时首先要注意避免的。对选择题，要注意使代表正确答案的符号（数字或字母）出现顺序呈随机状态。有的教师为了评分方便，让正确答案以某种固定的顺序或规律出现，如“A, B, C, D; A, B, C, D……”；或是让正确答案总是出现在某一固定序号上，如第三或第四个备选答案经常为正确答案，等等。这类答案排列的模式容易被部分学生猜出，不仅会降低考试结果的客观性和可靠性，还可能对学生产生不良影响。

命题原则7：试题中项目编排要便于学生作答。

为了保证试题的效度，试题的编排形式不应成为影响学生答题的因素。因而，试题中项目的编排方式要尽可能便于学生作答。为符合这一原则，可从以下三方面入手：

(1)要把类型相同的项目编排在一起。不同类型的项目要求学生的答题方式也不相同。把类型相同的项目编排在一起，不仅可以只用一套指导语来说明许多同类项目的答题方式，更重要的是还可以减少因变换答题方式而对学生答题的干扰。

(2)在同一题型的一组项目中，要把内容相近的项目编排在一起。这样做的意义在于增强试题总体的紧凑感，便于学生集中思想，提高答题效率；也便于教师集中了解学生对某一知识块的学习状况。

(3)要将项目按照从易到难的顺序编排。当试题覆盖面较大、项目较多、项目在难度上有一定梯度，且测验时间较紧时，项目从易到难编排，对于鼓励学生积极回答问题，特别是鼓励低年级学生或是那些知识基础和学习能力较差的学生持续努力答题而不至于中途放弃答题，有着比较好的心理效应。另一方面，把难度较大的项目排在最后，有利于保证测验的效度。这是因为，项目难度大，学生答对的可能性小，即使由于时间紧迫来不及作答，对测验结果也不会有多大的影响。

二 项目权重（分数）的确定

项目合成试题之后，需要确定每个项目的权重或分数。一般是先估计每个项目的平均权重（用规定的试题满分除以项目总数），再对照命题计划，并根据各类项目的特点适当加以调整。同类项目的权重应相同。填空题可按照项目中所含有的空格总数确定填空题的总权重，每个空的分数相同；选择

题可先确定每个项目的权重，再根据项目总数求出选择题的总权重；简答题可以根据每个项目中所包括的得分点确定项目权重，每个得分点的分数相同。为了便于评分和记分，尽可能使每个项目的分数都是整数，必要时，可以通过项目总数的增减来达到这一目的。

在一般教学测试题的编制过程中，有一种比较常见的观点是：根据项目的重要性程度来确定项目权重，即项目越重要，其所占的分数也应越多。这一观点有值得商榷之处。因为显然存在的问题是：项目的重要性又依据什么来确定呢？我们很难从主观上判断，项目所对应的测量目标哪些是重要的而哪些是不重要的。根据项目所涉及的知识内容、项目难度以及完成项目所需要的时间来确定项目的分值，都很难避免主观因素的影响。因此，我们认为，根据项目的类型特点和项目总数确定项目的分值可能更客观一些。使同类项目的权重基本相同，如果我们认为需要增大或减小某一类测试目标的比重，可以通过增大或减小相应项目的数目来调节，而不必人为地增大或减小某些项目的权重，以避免命题过程中主观因素的影响。

三 完成试题双向细目表

根据命题计划编制项目、再由项目合成试题之后，还需要将命题计划和试题联系起来，以便检验命题计划的执行情况，同时，说明试题中每个项目的意义，即项目和测量目标之间的一一对应。这就需要完成试题双向细目表。

下面就以表 3-4 为例，来说明双向细目表的结构及其所表示的意义。

表 3-4 高一化学第一学期“期末考试题”双向细目表(示例)

	记忆	理解	简单应用	综合应用	合计
第一章	一、 1. (3)	二、 7. (2)	二、 11. (2)		28 %
	一、 2. (3)	二、 8. (2)	二、 12. (2)		
	二、 6. (2)	二、 9. (2)	二、 13. (2)		
		二、 10. (2)	三、 34. (6)		
第二章	二、 14. (2)	二、 15. (2)	二、 18. (2)	三、 30. (6)	32 %
		二、 16. (2)	二、 19. (2)		
		二、 17. (2)	二、 20. (2)		
		三、 28. (6)	三、 29. (6)		
第三章	一、 3. (2)	二、 21. (2)	二、 25. (2)	三、 33. (6)	40 %
	一、 4. (2)	二、 22. (2)	二、 26. (2)		
	一、 5. (4)	二、 23. (2)	二、 27. (2)		
		二、 24. (2)	三、 32. (6)		
		三、 31. (6)			
合计	18 %	34 %	36 %	12 %	100 %

1. 双向细目表的结构

该表格在形式上与命题计划类似，但方格中所表示的是每个项目的题号及其分值，例如，“一、 1. (3)”表示第一大题的第一小题，其分值是 3 分。

2. 双向细目表所表示的意义

(1)说明试题的构成。该试题共由三种类型、34 个项目所组成。第一种

类型的项目（可能是填空题）共 5 个（一、1~5）；第二种类型的项目（可能是选择题）共 22 个（二、6~27）；第三种类型的项目（可能是简答题）共 7 个（三、28~34）。

(2)说明每个项目的测量目标(或对应的教学目标)及其权重(或分值)。例如：项目（一、1），它所对应的第一章教材中属于记忆水平的教学目标。只要对照单元教学目标，即可了解该项目测量目标的具体内容。该项目的权重为 3%（即全卷满分如果是 100 分，则该项目的满分就是 3 分）。

(3)说明在试题中各章知识内容以及属于各类学习水平的项目所占的比重。例如，从表中纵行的“合计”栏中可看到，第一章的知识内容在试题中的比重为 28%；从横列的“合计”栏中可看到，属于“简单应用”水平的项目在试题中的比重为 36%。

(4)说明试题中各个知识块(章)内所涉及到的属于不同学习水平的教学目标的比例。例如，将第一行、第一列的方格中所有项目的分值相加，得其值为 8，即表示第一章属于记忆水平的测量目标在试题中所占的比重为 8%。这就可以便于我们对照命题计划，考察试题的实际构成是否符合命题计划的要求。

双向细目表是评价试题质量的重要依据，也是指导学生对测试结果进行自我评价时不可缺少的工具。因而应有下列命题原则：

命题原则 8：试题必须配有双向细目表。

四 试题“等值复本”编制

对于形成性的单元测验，常常需要有两套符合同一份双向细目表、测试目标完全一致而测试项目又不相同的试题；对于总结性的学期或学年考试，往往也需要两套等值的试题，俗称“A,B 卷”，以防在施测过程中出现不测。这就需要编制试题的“等值复本”。在实际命题过程中，一般是精心编制第一套试题作为“源本”，然后依据“源本”对项目进行逐个“复制”。现将两种常用的复制方法举例介绍如下：

1. 改变提出问题的方向复制项目

例 17 源题：下列微粒中第一电离能最大的是..... ()。

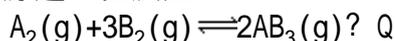
(A)Na (B)Mg (C)Al (D)K (E)Si

复制题：下列微粒中第一电离能最小的是..... ()。

(A)Na (B)Mg (C)Al (D)K (E)Si

2. 向同一内容的另一分支提出问题来复制项目

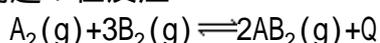
例 18 源题：在反应



达到平衡时，下列各曲线中符合勒沙特列原理的是

..... ()。

复制题：在反应



达到平衡时，下列各曲线中符合勒沙特列原理的是

..... ()。

五 预试和试题的最后确定

试题编制完成之后，在正式使用之前，需要经过预试。因为命题者对测试所需时间的估计、对正确答案的看法及答案范围的估计、对评分标准的规定等有可能不够准确、合理。经过预试，就可以提前发现类似的问题，从而对试题作最后的修正，以免将命题者对试题的主观估计方面的错误带到正式考试之中，造成无法弥补的损失。特别是对那些规模较大的、参加学生较多的考试，预试应是命题过程中不可缺少的步骤之一。

预试的对象最好是与测量对象相同或相近、但又不属于测量对象的样本，如为本校学生编制的试题，可以将同类其他学校、相同年级且教学进度基本相同的教学班的学生作为预试的对象；在某些对试题质量要求不高的情况下，为方便起见，也可以请同行教师或命题者本人作为预试对象。如果预试结果与命题者事先的估计不一致，就需要认真分析其原因，然后根据分析结论对试题作必要的修订。在化学教学测量中，经常会在对某些实验现象的描述或对某些元素化合物的制备过程等方面出现争议或问题，如对实验现象描述得不够准确，或制备过程和方法不尽合理等等，有时甚至在高考试题中也出现类似问题。对此，命题者在编制试题和对预试结果的分析过程中要予以充分注意。

由于预试具有保证试题质量的重要作用，应有下列命题原则：

命题原则 9：试题在正式使用之前要经过预试和预试后的修订。

第四章 化学试题质量分析

化学试题质量分析是对测试结果做出正确评价的前提。只有试题质量符合要求的测试结果，才能作为评价的依据。同时，通过试题质量分析，能为进一步修改试题提供依据，有助于化学试题库的建立，还可以帮助我们发现试题编制过程中有哪些成功的经验和失败的教训，从而有助于我们提高命题的技能、技巧。

本章将讨论有关化学试题质量分析的问题，包括试题总体的质量分析和项目的质量分析。

第一节 试题质量的初步评价

一 测验分数的分布状态

通过对测验分数进行初步整理，作出简单次数分布图，即可根据分数的分布状态对试题质量做出比较直观而初步的估计。

图 4-1 是一种常见的学生成绩次数分布直方图。该图直观地表示出：以 5 分为一个分数段，每个分数段所含分数的次数或百分数（频数）。如果找出每组的组中值和该组次数对应的点，并将相邻各点依次连接，即可制成次数分布多边图（图 4-2）。如果分数的个数较少，次数分布多边图往往呈不规则的折线形式；如果分数的个数无限增多，则组数增多，组距缩小，分数的次数分布图就会趋向于圆滑曲线，即分数的次数分布曲线。

常见的考试分数的次数分布曲线可如图 4-3 所示。

图 4-2 学生成绩次数分布多边图

在已知被测学生总体的学习成绩分布呈正态的前提下，可以根据图 4-3 所表示的几种分数分布状态，对试题质量作出比较粗略的评价。

(1) 正态分布。说明测试结果与学生的实际状况一致，各种难度的项目比例合理。

图 4-3 常见的考试分数的次数分布曲线

(2) 正偏态分布。说明试题难度偏高，难度较大的项目比例偏大。呈这种分布的试题有利于将成绩优秀的学生和中等程度的学生区别开，但不利于将中等程度的学生和成绩较差的学生区别开。

(3) 负偏态分布。说明试题难度偏低，难度较低的项目比例偏大。呈这种分布的试题有利于将成绩较差的学生和中等程度的学生区别开，但不利于将中等程度的学生和成绩优秀的学生区别开。

(4) 双峰型分布。说明试题存在两极分化现象，即难度偏高的和难度偏低的项目较多，而中等难度的项目偏少，项目难度的分布缺乏梯度，不够合理。呈这种分布的试题可以区别中等程度的学生，但不利于区别出成绩优秀的学生和成绩较差的学生。

(5) 平坡型分布。说明试题中各种难度的项目比例接近，梯度较大。呈这种分布的试题区分度较高，但分数之间的差异偏大。

(6) 陡峭型分布。说明试题中同等难度的项目较多，梯度偏小。呈这种分布的试题几乎不能将不同程度的学生区分开，分数分布过于集中。

二 平均分

平均分是表示一组分数的集中趋势的一种统计量数，也是教育测量中使用最为广泛的一种统计指标。平均分的一般计算

$$\text{公式是： } \bar{X} = \frac{\sum_{i=1}^N X_i}{N} \quad (\text{式 4-1})$$

式中， \bar{X} 为平均分； X_i 为学生 i 的测验分数； N 为参加测验的学生数。

平均分的基本作用在于，它可以作为一组分数的典型代表。除平均分之外，能代表一组分数的其他统计量数还有中数、众数等。所谓中数是指在一

组按照大小顺序排列的分数中位于正中间的那个分数；所谓众数是指在—组分数中出现次数最多的那个分数，也就是分数的次数分布曲线最高点所对应的分数。与中数、众数相比，平均数是一组分数的最好代表。这是因为平均数具有下列基本性质：

(1) 平均分是由全部分数求出的。

(2) 每个分数与平均分的离差之和等于零，即 $\sum_{i=1}^N (X_i - \bar{X}) = 0$ 。这说

明平均分是各个分数的中心。

(3) 当分数分布呈正态时，平均分即位于分布曲线的中点并和曲线的最高点相对应。

(4) 各个分数与平均分的离差的平方和比各个分数与其他任何集中量数的离差的平方和都小。因此，它是计算其他统计量数时使用最多的数据之一。

根据平均分的数值，可以直接判断试题的平均难度，看其是否与预期的试题难度一致。平均分过高或过低，说明试题过易或过难，不适应被测学生的实际水平；只有当平均分与预期吻合或相当接近时，测量目的才有可能实现，这时，试题质量才可能符合测试目的的要求。因而平均分可用于对试题质量做出初步评价。

三 标准差

—组测验分数，只有平均分是不足以表明其分布状况的。如图 4-3 中“平坡型”分布和“陡峭型”分布，其差异是十分明显的；但这两种分布的平均分却可能完全一样。因此，要说明—组分数的基本情况，只有反映分数集中趋势的集中量数不行，还需要有反映分数离散情况的差异量数，两个量数集合起来，才能说明—组分数的基本特点。

标准差就是描述—组分数离散情况的最常用、最可靠的统计量数；标准差和平均分一样，也是计算其他统计量数时常用的基本数据之一。

标准差用符号 S 或 σ 表示。其计算公式为：

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}} \quad (\text{式4-2})$$

式中各个符号的意义同式 4-1。

在利用标准差数据对试题质量做出初步评价时，主要是看学生测验分数的离散情况是否与学生学习成绩的实际分布一致。如果一个教学班学习成绩比较整齐，测验结果分数分布应当相对集中，此时，分数的标准差要小才符合学生的实际状况；相反，如果是一个学习成绩参差不齐的教学班，测验分数的标准差大才符合实际。只有当标准差表明的测验分数的离散情况与被测学生集体的实际状况一致时，试题质量才有可能合格的。因而可以用标准差对试题质量做出初步评价。

严格讲， σ 是指总体的标准差， S 指样本的标准差。符号 S 大小写各书多不统一，本书采用大斜体。

第二节 试题的信度

一 相关系数

相关系数是教育测评常用的一种统计量数。在利用信度、效度、区分度的指标评价试题质量时，也必须用到相关系数的计算公式。因此，在讨论试题的信度之前，有必要先了解相关系数的概念及其计算方法。

相关系数是表示两列变量之间相关程度的一种统计量数。所谓相关是指两类事物或现象在发展变化的方向上相互联系的状况，一般有以下三种相关状态：

正相关——两列变量的变化方向相同，即一种变量有增大或减小的变动时，另一变量也同时发生增大或减小的变动。例如，人的身高和体重之间就存在正相关。

负相关——两列变量的变化方向相反，即一种变量有增大或减小的变动时，另一变量却同时发生减小或增大的变动。例如，练习次数和错误出现次数之间就存在负相关。

零相关——两列变量的变化之间没有明显联系，即一种变量有增大或减小的变动时，另一变量只作无规律的变动。例如，学生的学习成绩和其身高或和体重之间就是一种零相关。

相关是自然界或社会中常见的一种现象。要注意的是不能将相关关系和因果关系混淆，即当两列变量存在相关时，不能认为一种变量的变化是另一种变量变化的原因或结果。

相关系数不仅表示两列变量之间相关的方向，而且还能表示相关程度的大小，常用 r 表示相关系数，其取值介于 -1.00 — $+1.00$ 之间。当 $r > 0$ 时，表示正相关； $r = 0$ 时，表示零相关； $r < 0$ 时，表示负相关。 r 的绝对值越大（越接近于 1），表示相关程度越大； r 的绝对值越小（越接近于 0），相关程度越低。相关系数并不代表相关的比率或百分数，更不是单位相同的相关量的度量。

当两列变量都是等距变量且都来自于正态总体时，相关系数可用皮尔逊积差相关 (Pearson's product moment correlation) 公式计算，其公式如下：

$$r = \frac{\sum_{i=1}^N (X_i - M_x)(Y_i - M_y)}{N \cdot S_x \cdot S_y} \quad (\text{式4-3})$$

式中， X_i ， Y_i 分别为连续变量 X ， Y 中的成对数据； M_x ， M_y 分别为 X ， Y 的平均数； S_x ， S_y 分别为 X ， Y 的标准差； N 为被统计的成对变量的数目。

如果直接从两列变量的原始数据计算相关系数，可用下式：

$$r = \frac{\sum_{i=1}^N X_i Y_i - \frac{1}{N} \sum_{i=1}^N X_i \cdot \sum_{i=1}^N Y_i}{\sqrt{\left[\sum_{i=1}^N X_i^2 - \frac{1}{N} \left(\sum_{i=1}^N X_i \right)^2 \right] \cdot \left[\sum_{i=1}^N Y_i^2 - \frac{1}{N} \left(\sum_{i=1}^N Y_i \right)^2 \right]}} \quad (\text{式4-4})$$

这样求出的相关系数被称为积差相关系数或皮尔逊相关系数。在化学教学测评中，学生的考试分数常用百分制表示，其性质符合积差相关计算公式

的要求。在利用上式计算相关系数时，要注意两种变量要成对取值，成对值的数目一般要大于 30，且两列变量都呈正态分布，以保证计算出的相关系数稳定、可靠。

当我们所收集的数据不属于等距变量，例如，用等级变量(A, B, C, ...)或分类变量(及格或不及格)表示时，需要采用其他方法计算相关系数。这将在以后的篇幅中，结合其具体应用再作介绍。

二 信度的概念

信度是反映测量工具可靠性的指标。常见的试题信度多用相关系数来表示，即用同一组被测学生样本对两组项目的反应之间的相关，作为反映两组项目测量的一致性程度的指标，称作信度系数，简称信度。

信度实际上只是个理论构想概念。测量理论认为，学生在测试中所得到的实际分数(X)由有效分数(X_V)、无效分数(X_{SE})和随机误差(X_{RE})三部分构成，即：

$X = X_V + X_{SE} + X_{RE}$ (式 4-5) X_{SE} 是和测验目的相关的分数， X_{RE} 是和测验目的无关的测量误差(系统误差)， X_{RE} 是测量的随机误差，学生的真实分数是 X_V 与 X_{SE} 之和。对一个被测群体而言，测验分数的离散程度可以用变异数或方差(即标准差的平方)来表示；上述三种分数都有其相应的变异数。利用变异数具有可加性这一性质，可将测验分数(实际分数)的总变异数(S_X^2)分解为有效变异数(S_V^2)、系统变异数(S_{SE}^2)和随机变异数(S_{RE}^2)，即：

$$S_X^2 = S_V^2 + S_{SE}^2 + S_{RE}^2 \quad (\text{式 4-6})$$

信度系数(r_{XX})被定义为真实分数的变异数在实际分数的总变异数中所占的比例，即：

从式 4-6 可知，当总变异数 S_X^2 一定时，测量的随机误差 X_{RE} 越大，其变异数 S_{RE}^2 也越大，则真分数($X_V + X_{SE}$)的变异数($S_V^2 + S_{SE}^2$)就越小，这时，根据式 4-7 可知，信度系数 r_{XX} 的值也越小。由此可见，信度实际上反映了测量随机误差的大小。

由于测量随机误差的主要来源可能不同，因而计算信度的方法也就可能不同，计算结果就有不同的意义，或仅仅表示信度的不同方面。所以，在利用信度作为评价试题质量的指标时，必须注意所用信度的实际来源及其所表示的具体意义。将不同的信度系数进行相互比较是没有什么实际意义的。

三 几种常用信度的意义及计算方法

1. 再测信度

用同一份试题对同一组学生前后实施两次测试，计算两次测验分数之间的相关系数，即得再测信度。这种信度能表示两次测验结果有无变化，反映测验分数的稳定程度，所以又称为稳定性系数。

当测验分数具有等距变量的性质(如化学教学测验分数常用的百分制)时，再测信度一般用皮尔逊积差相关公式(式 4-3 或式 4-4)计算。

再测信度能够提供关于试题的测量结果是否随时间变化的信息。两次测验分数的误差变异主要来自测验条件和受测者身心状况的变化。再测信度

高，说明试题受学生状况和测验环境变化的影响小，可以认为试题是比较稳定的。但再测信度很容易受到练习和记忆的影响，前后两次施测间隔时间要适当。间隔时间太短，学生对第一次测验记忆犹新，练习影响大；但间隔时间太长，则学生身心发展与教学效果等足以改变测验分数的意义。

2. 复本信度

如果用两份等值平行的试题测量同一组学生，再求两次测验分数的相关，其结果就是复本信度。这时同样可以用式 4-3 或式 4-4 计算。

当两个测验同时连续施测时（为了抵消施测顺序的影响，可以使半数学生先做 A 卷后做 B 卷，使另一半学生先做 B 卷后做 A 卷），分数变异主要来自于两份试题在取样上的差别，即两份试题的等值程度，所以，复本信度又称作等值性系数。如果两份试题在不同的时间施测，则分数的变异与试题的稳定性和等值性都有关系，这是对试题可靠性的最为严格的检验，其数值也相对最低；这时所得到的信度称为稳定性与等值性系数，或再测复本信度。

复本信度的局限在于，由于复本之间的相似性，记忆或练习效应的影响仍然不能完全消除；编制两份完全等值的复本也是比较困难的。

3. 分半信度

在测验没有复本且只能实施一次的情况下，可将测验项目分成对等的两半，根据学生在这两半测验中所得的分数计算相关系数，即得分半信度。

这种信度考察的是同一试题的两半项目之间的一致性，所以可称为内部一致性系数，其意义类似于等值性系数，但不同的是，这里的两个等值复本是在测验后由同一份试题划分出来的。

计算分半信度先要对测验分半。不同的分半法可能会得到不同的信度值。为了使两半基本等值，可将项目按由易到难的顺序排列编号，然后按奇数和偶数序号将项目分半。要注意使那些性质相同、联系紧密的项目分在相同的一半，否则会使信度值偏高。

由于分半信度实际上只是半个测验的信度，测验越长、项目越多，两半分数的相关就越高。因此，对长度不同的测验，要用斯皮尔曼-布朗（Spearman-Brown）公式对其信度值加以校正：

$$r_{xx} = \frac{2r_{hh}}{1+r_{hh}} \quad (\text{式4-8})$$

式中 r_{hh} 为两半分数的相关系数； r_{xx} 为校正后的分半信度。

4. 同质性信度

这里的同质性是指试题中所有项目之间的一致性。若测验中各个项目得分相关较高时，不论项目的内容或形式有何差别，测验都为同质；相反，即使所有项目看起来都好像是测量同一特性，但项目得分之间的相关很低，测验仍然是异质的。测验越同质，同质性信度的值也越大，则说明试题的内部一致性程度越高。因而，同质性信度又称作内部一致性系数。

常用计算同质性信度的方法如下：

(1) 当项目采用二分法（项目只有通过或未通过两种分数）计时，可用库德-理查逊公式（Kuder-Richardson formula），即 K-R20 公式计算：

$$r_{K-R20} = \frac{K}{1-K} \cdot \left(\frac{1 - \sum_{i=1}^K P_i Q_i}{S_x^2} \right) \quad (\text{式4-9})$$

式中，K 为整个测验的项目总数； P_i 为项目 i 的通过率； $Q_i=1-P_i$ ； S_x^2 为测验总分的变异数。

(2) 当项目属于论文式类型，采用连续分数计时，可用克龙巴赫所创的 a 系数公式：

$$a = \frac{K}{1-K} \cdot \left(\frac{1 - \sum_{i=1}^K S_i^2}{S_T^2} \right) \quad (\text{式4-10})$$

式中， a 为信度系数；K 为项目总数； S_i^2 为项目 i 得分的方差， S_T^2 为试题总分的方差。

当项目以二分法计时，项目得分方差

$$S_i^2 = P_i \cdot Q_i \quad (\text{式4-11})$$

代入式 4-10 中，即得到式 4-9。可见，式 4-9 可看做是式 4-10 的一个特例。

现将信度估计的几种方法小结于表 4-1。从上述几种信度的具体意义看，似乎还没有找到一种完全适合于教学测验的信度估计方法。在利用信度评价试题可靠性时，要注意其局限性。

第三节 试题的效度

一 效度的概念

效度是指测验的有效性，即一个测验能够测出其所要测量的特性或功能的程度。效度或效度系数与信度一样，也是个理论构想概念。在测量理论中，效度被定义为：在一次测量中，与测验目的有关的有效变异数与测验总分变异数的比率，即：

$$r_{XY} = \frac{S_V^2}{S_X^2} \quad (\text{式4-12})$$

式中， r_{XX} 为效度系数， S_V^2 为与测验目的有关的有效变异数， S_X^2 为测验总分的变异数。从式4-12不难看出，在学生的实际分数中，与测验目的有关的有效分数的比重越大，则效度系数的值也应越大。

根据式4-6，不难理解效度和信度的关系。当总变异数 S_X^2 一定时，测量的随机误差 X_{RE} 越大，其变异数 S_{RE}^2 也越大，则真分数 $(X_V + X_{SE})$ 的变异数 $(S_V^2 + S_{SE}^2)$ 就越小；也就是说，效度要受到信度的制约，信度高的测验其效度也不可能高；但另一方面，信度高的测验其效度未必就一定高，这是因为，信度高时，虽然真分数 $(X_V + X_{SE})$ 的变异数 $(S_V^2 + S_{SE}^2)$ 大，但是系统误差的存在同样会导致效度下降（ S_{SE}^2 增大时 S_V^2 降低）。可见，信度只是效度的必要条件而非充分条件。

效度总是和一定的测验目的紧密相关的。一种测验总是为了测量某种特性或功能而编制的，判断测验效度的高低，就是看它能达到测验目的的程度。如果能正确、真实地测出所要测量的东西，那么这对一定目的来说便是效率高的测量；反之，则可认为该种测量的效度不高。显然，对某种目的来说是效率较高的测量，对另一种测验目的而言就很可能是效率很低的测量。离开测量目的谈效率是没有意义的。

二 几种常用的效率及估计方法

1. 内容效率

内容效率是指根据测验目的确定的测验内容所引起的预期反应达到测验目的的程度。就化学教学测量而言，测验目的是考察学生达到教学大纲所规定的教学目的和要求的程度，测验内容应是对化学教学大纲所规定的全部教学内容的代表性取样。如果测验的内容取样代表性好，则学生对测验内容的行为反应可以代表教学所引起的学生的全部行为变化，如对化学知识的记忆、理解和应用等，该测验就可以正确地测量出学生的学习成绩，那就可以说，该测验有较高的内容效率。

但是，要用有限的几个或几十个项目代表全部教学内容是比较困难的。因而，测验项目的代表性有大有小，测验的内容效率有高有低。教学测验具备较高内容效率的条件是：一要有定义完好的内容范围；二要使项目对教学内容有较好的代表性。一般而言，按照第三章所介绍的命题步骤和原则设计出的化学试题是可以满足这两个条件的。命题计划设定了测验内容的范围及各部分的比例；双向细目表表明了每个项目所代表的测量目标。所以，当双向细目表与命题计划一致，且项目能够与双向细目表中的测量目标一一对应时，基本可以保证试题的内容效率。

判断一份化学试题的内容效度如何，可以采用以下步骤：

考察测量的目的、测量目标的分类体系，了解试题取样的依据和材料的来源；

考察试题的双向细目表，看其双向细目的划分及各个部分的比例是否符合化学教学大纲和考试目的的要求；

逐个考察每个项目，包括项目内容、参考答案、评分标准等，并与双向细目表对照，看项目能否代表其测量目标；

综合上述各项的考察结果，对试题的内容效度作出判断。

这种方法实际上是一个逻辑分析过程，以此方法确定的内容效度也可称为逻辑效度。此外，还可以用再测法和经验法判断试题的内容效度。再测法是在教学过程前后分别用同一份试题或等值复本进行两次测验，若后一次测验分数大大高于前一次测验分数，则说明测验内容和教学内容一致，试题有较高的内容效度；反之，内容效度就低。经验法是用包含不同年级教学内容的试题测验各个年级的学生，考察不同年级的学生的总分和在代表各个年级教学内容的各项目上的反应情况，如果学生的测验分数和项目通过率随年级增高而增高，则说明项目代表了不同年级的教学内容，测验有较高的内容效度。

用内容效度考察试题的有效性，是目前比较合适且应用较多的方法。但由于内容效度尚未找到较理想的数量化指标，因而妨碍了试题评价信息的交流和各测验间的相互比较。

2. 构想效度

构想效度就是指测验对预先设立的某一理论上的概念、构想或研究特性的实际测量程度。考察和研究构想效度的目的是要回答下面的问题：一个测验要测量的理论构想是什么？测验对该构想测量到何种程度？在测验分数的总变异中有多少来自要测的构想？

要判断一个测验的构想效度，通常需要经过以下三个基本步骤：

确定能够解释学生在测验上表现的理论构想；

根据理论构想推演出各种关于学生测验成绩的假设；

用逻辑分析、统计分析等方法来收集证据、验证假设。

例如，中学化学教育目标（认知领域）的学习水平分类，实际上也是一个理论构想，我们可以通过构想效度来评价某种分类方法的合理性。假设：当项目的测量水平与测量目标（即学习目标）的学习水平一一对应时，学生在属于同一学习水平的项目上的反应基本相同，而在属于不同学习水平的项目上反应基本不同，那么，学生在属同一学习水平的项目上的得分应存在高相关，而在不同水平层次的项目得分应具有低相关。根据这一假设，我们可以对所有项目或有重点的选择部分项目，计算其两两之间的相关系数，考察它们之中高相关的是否属于同一学习水平、低相关的是否属于不同的学习水平，从而验证我们的理论假设，进而评价学习水平分类的合理性。

我们还可以通过对项目难度的分析来评价关于化学教育目标学习水平分类的理论构想的合理性。假设：不同学习水平的项目应有不同的难度水平；属于高层次学习水平的项目难度大，而低层次项目的难度小；学习能力不同的学生对水平不同的项目反应也应不同：高分组学生在低水平项目上的反应差别小，而在高水平项目上的反应差别大；低分组学生却恰恰相反，只在低水平项目上的反应有差别而在高水平项目上的反应几乎相同。根据上述假设，可以将项目按其难度大小进行分组分析，也可以分析高分组学生和低分

组学生对难度不同的项目的反应情况，对假设进行检验，从而评价学习水平分类的合理性。

当对测验结果的分析检验不能证实构想时，可能会有多种解释，例如：

构想本身有错误；测验的构想效度低，即测验没有能够测量构想；检验假设的程度、方法不够合理，等等。尽管构想效度目前还存在着许多局限性，但它毕竟提供了一种研究理论构想的具体方法，使教育测量成为发展教育理论的有效工具之一。

3. 效标效度

效标效度是指测验分数与效标的相关程度。所谓效标，就是检验测验有效性的一种参照标准。效标常用一种公认比较可靠或权威的测验结果表示。这实际上就是用一种已知的且认为其“有效”的测验结果去检验另一个新测验的有效性。之所以不直接用效标测验去代替新测验，往往是因为新测验可能比效标测验更为简单、易行。

化学教学中常用学生的实际高考成绩与模拟高考试题得分之间的相关来检验高考模拟试题的有效性。这里应用的就是效标效度的检测方法，高考就成了模拟考试的效标。

由于效标效度采用测验分数和效标分数间的相关系数来表示，可以对被检测的测验进行定量化的分析比较，相对于内容效度而言更为客观，其意义也较直观，易于被理解和接受。但是，效标效度在实际化学教育测量中应用得并不多，这主要是因为目前尚未找到比较合理的、可以满足化学教育测量各种要求的效标。除高考外，还没有其他比较权威的化学考试，而高考却不能作为一般化学教学测验的效标。这是因为两者之间在内容、性质上都有很大差别。高考是为高校录取新生服务的选拔性考试，而教学测验多是为教学服务的形成性评价的一部分；高考试题要在中学化学的全部教学内容范围内取样，而教学测验往往只就某一教学阶段的内容范围取样。在教学实际中，有时用教师对学生的等级评定作为某种测验的效标，但这样的效标难免会带有较多的主观因素的影响。某些心理测验如智力测验、国外的一些学科水平测验等，与我国目前的化学教学实际尚有较大的距离，也很难作为化学教育测验的效标。因此，寻找化学教育测验的合理效标，是当前化学教育测评研究的重要课题之一。

现将三种效度的意义及其估计方法小结于表 4—2。

表 4-2 效度的类型、意义及其估计方法

类型	意义	估计方法
内容效度	测验内容反映测验目的、要求的程度	逻辑分析
构想效度	测验分数能够用理论构想加以解释的程度	提出假设和验证假设
效标效度	测验分数与效标分数之间的相关程度	计算两种测验分数之间的相关系数

第四节 项目质量评价

试题是由项目构成的，如果项目质量不高，试题的质量不可能很高。因而项目质量是试题质量的前提。但是，试题的质量不高，却不能因此而肯定构成试题的所有项目质量都不高。在质量较差的试题中，也有可能包含一些质量较高的项目。所以，项目质量评价既是试题质量评价的重要组成部分，又是对项目进行修订、筛选、积累或建立题库的必要条件。

一 项目的难度分析

1. 难度的意义及其计算

难度是指项目的难易程度。反映项目的难易程度的数量化指标叫做难度系数，简称难度。项目越难，说明能够正确回答该项目的学生越少，或对该项目学生能够正确回答的成分越低。对项目难度的计算，通常有两种方法：

(1) 用通过率计算难度。

当项目以二分法计分（答对得分、答错不得分）时，难度一般用正确回答项目的人数与参加测验总人数的比值为指标，即：

$$P = \frac{R}{N} \quad (\text{式4-13})$$

式中，P 为项目难度；R 为答对该项目的人数；N 为参加测验的总人数。

(2) 用项目得分的平均值计算难度。

当项目是用连续分数计时，难度一般用参加测验的全体学生在该项目的平均得分与该项目的满分的比值为指标，即：

$$P = \frac{\bar{X}}{W} \quad (\text{式4-14})$$

式中，P 为项目难度； \bar{X} 为该项目的平均得分；W 为该项目的满分。

用式 4-13 或式 4-14 计算所得的难度值与项目的实际难易程度正好相反。难度值越大，项目反而越容易。比如， $P_1=0.9$ ，此项目难度小； $P_2=0.1$ ，此项目难度大。为了解决这一矛盾，有人提出将式 4-13 或式 4-14 改为

$$P = 1 - \frac{R}{N} \quad (\text{式4-15})$$

$$P = 1 - \frac{\bar{X}}{W} \quad (\text{式4-16})$$

这样， $P_1=1-0.9=0.1$ ； $P_2=1-0.1=0.9$ ，与项目的实际难易程度一致。不过，式 4-13 和式 4-14 是传统的难度计算公式，应用已十分广泛。本书的难度系数采用式 4-13 和式 4-14 计算。只要我们理解其意义，就不会造成认识上的混乱。

2. 项目难度对测验结果的影响

(1) 对分数分布形态的影响。

试题中所有项目的平均难度就是试题的难度。难度不同的项目比例不同，试题的难度就不同，分数的分布形态也会因此而不同。比如，所有项目的难度都是 1，全部学生都会得满分；如所有项目难度都是 0，则全部学生都得零分。在这两种情况下，所有分数都集中在两个分数上（100 和 0）。可见，过易或过难的测验，会使测验分数相对集中在高分段或低分段，这时，分数的离散程度小（标准差小）。

如果难度接近 0 的项目比例较大，得低分的学生较多，分数分布就会相对集中于低分段，则分布呈正偏态；如果难度接近 1 的项目比例较大，得分高的学生较多，分数分布就会相对集中于高分段，则分布呈负偏态。当项目的难度成梯度出现，即从 0.1—0.9 的项目比例相同或接近时，项目的平均难度接近 0.5，此时分数的分布形态就最接近正态。

(2)对测验的鉴别能力的影响。

当一部分学生能够正确地回答问题而另一部分学生不能正确回答时，两部分学生之间就形成了相互比较。在测验中，学生间相互比较的可能性越多，就越有利于对学生进行鉴别。项目的难度不同，提供这种相互比较的可能性的程度也不相同。例如，当参加测验的学生数为 100 时，某项目的 $P=0.2$ ，说明有 20 个学生答对了该项目，80 个学生答错了该项目，这时，学生之间相互比较的可能性为 1600 次 ($20 \times 80=1600$)；如果 $P=1$ 或 $P=0$ ，学生之间就没有相互比较的机会 ($100 \times 0=0$ 或 $0 \times 100=0$)。不难看出，当 $P=0.5$ 时，项目能够提供给学生相互比较的机会最大 ($50 \times 50=2500$)。所以，项目的 P 值越接近 0.5，对学生的鉴别力也就越大；同理，项目的平均难度即试题的难度越接近 0.5，试题的鉴别力也越强。

3. 项目难度的评价

一个项目的难度值应该多大才合适？对此，并没有一个确切的数值为其答案。虽然项目难度在 0.5 时，其鉴别力最强，但这并不意味着项目难度在 0.5 时质量最高。评价项目难度是否合适的依据，只能是测验的目的。

在平时的形成性测验中，与化学教育目标相对应的测量项目都需要编入测验中，尽管这些项目在前置性测验中难度值可能为 0，而在后置性测验中难度值可能为 1，也不能将其舍弃而人为缩小测验取样的内容范围；在总结性测验中，需要对学生的学习成绩做出区分，要求测验的平均难度达到或接近 0.5，但为了保证取样的代表性，应当使各个难度水平的项目都有分布，且比例适当。有时，还需要根据测验的特殊目的来配置难度不同的项目比例。例如，我们要挑选 10% 的学生参加化学竞赛，项目难度应相对集中于 0—0.20 之间，使项目的平均难度控制在 0.1 左右。这样做虽然会使大部分学生不能完成测验，但能够达到选拔优秀学生的测验目的，项目难度的配置仍然是合理的。如果我们希望一次测验既要选拔最高水平的学生，又要找出那些需要进行个别辅导的学生，就需要使难度分别为 0.1，0.5，

0.9 的项目各占 $\frac{1}{3}$ 左右。对于高考命题，在要求高考兼有为高校选拔合

格新生和评价中学教学质量双重职能时，这就不能只是简单地使项目难度都保持在 0.5 左右了，而需要精心配置项目的难度比例，使考试的及格率与录取率正好相切。这样才能既有利于高校选拔新生，又可避免不良的社会效应。

要对测验中的项目难度比例进行合理配置，必须在测验正式实施之前就对项目难度做出准确判断。要注意的是，仅仅依靠主观经验来确定项目难度是不够的，因为项目难度不仅取决于编制水平，还与被测学生的具体情况有关。一个本来很容易的项目，可能会因为学生由于某种原因没有学习过有关知识而变难；而一个很难的项目，也可能由于学生测前做过有关练习而变易。对被测总体中的某一部分学生来说是容易的项目，可能对另一部分学生来说是较难的项目。这就说明， P 值所反映的仅仅是项目的相对难度而非绝对难度。正因为如此，预测是十分必要的。

二 项目的区分度分析

1. 区分度的意义

区分度是指测验试题或项目对学生的实际水平的区分程度。具有良好区分度的试题或项目，实际水平高的学生应该能够通过或得高分，实际水平低的学生不能通过或只能得低分。所以，测验的区分度又叫鉴别力。

测量理论对测验的区分度十分重视，称之为判断测验是否有效的指示器，并作为评价试题和项目质量、对项目进行筛选的主要依据。从区分度本身的意义就可以看出，当测验是以选拔为主要目的时，只有试题和项目具有良好的区分度，测验才会有一定的效度。区分度和测验的信度也存在着密切的关系。有人通过研究发现，测验的信度随项目的平均区分度的提高而增长，且信度增长的速度较区分度增长为快。因而，提高试题和项目的区分度是达到理想的测验信度的一个有效途径。

项目的区分度与难度的关系，前面的讨论已经涉及。我们知道，在难度接近 0.5 时，项目的鉴别力接近最强，也就是区分度值接近最大。区分度相同的项目其难度值可能不一样。在这种情况下，P 值小的项目所能区分的是高分组的学生，P 值大的项目所能区分的是低分组的学生。为了对全部参加测验的学生都能够有所区分，需要测验中的全部项目都有必要的区分度，且难度不同的项目比例也要适当。

在任何测验中，无论测验的具体目的如何，提高项目的区分度总是对改进测验质量有利。

2. 区分度的计算

(1) 用内部一致性系数计算区分度。

这里的内部一致性是指学生在某项目的得分与其所得测验总分的一致性。用内部一致性系数计算出的区分度的意义在于：项目的区分度高，说明总分较高或测验合格的学生在该项目上通过或得高分的可能性也高；而总分较低或测验不合格的学生在该项目上通过或得高分的可能性也小。项目的区分度低则恰恰相反，在该项目上通过或得高分的学生反而总分低或测验不合格，而在该项目上不能通过或得低分的学生反而总分高或测验合格。在标准化或较大规模的测验中，多采用相关法计算项目的得分与测验总分的一致性，表征项目的区分度。

根据项目的计分方法和测验总分或结果的表示方法的不同，可采用以下几种方法计算项目的区分度。

积差相关法。

当项目和试题总分都采用连续分数计时，可用式 4-4 来计算项目的区分度。用变量 X 表示学生在某项目上的得分，用变量 Y 表示学生的测验总分，其积差相关系数即可代表该项目的区分度值。

点二列相关。

当项目以二分法计分、测验成绩以连续分数表示时，可用点二列相关公式计算区分度。其计算公式如下：

$$r_{pb} = \frac{\bar{X}_p - \bar{X}_q}{S_i} \sqrt{pq} \quad (\text{式4-17})$$

式中, r_{pb} 为点二列相关系数, 即项目区分度; \bar{X}_p 为答对项目的学生总分平均数; \bar{X}_q 为答错项目的学生总分平均数; S_t 为全体学生的总分标准差; p 为答对项目的学生所占的人数比率; $q=1-p$ 。

二列相关。

当测验总分为连续分数, 项目也以连续分数计分, 且学生的项目得分接近正态时, 可以将项目得分按给定的“通过分数”分为通过和不通过两类, 也就是将连续分数人为地变成二分法计分的分数。这时, 可用二列相关计算项目的区分度。其公式如下:

$$r_b = \frac{\bar{X}_p - \bar{X}_q}{S_t} \cdot \frac{pq}{y} \quad (\text{式4-18})$$

式中, r_b 为二列相关系数, 即项目的区分度; \bar{X}_p , \bar{X}_q , S_t , p 和 q 的意义同式 4-17; y 为正态分布下答对的学生百分比所在位置的曲线高度。

四分相关。

当项目分数和测验总分都是连续分数而又都被人为二分时, 可用四分相关计算区分度。其计算公式为:

$$r_t = \cos \left(\frac{180^\circ}{1 + \sqrt{\frac{ad}{bc}}} \right) \quad (\text{式4-19})$$

式中, r_t 为四分相关系数, 即项目区分度; a, b, c, d 的意义如下列四

		测验总分	
		及格(+)	不及格(-)
项目分数	通过(+)	a(+ +)	b(+ -)
	不通过(-)	c(- +)	d(- -)

格表所示:

用 r_t 估计相关时, 一般要求 $a+b+c+d > 200$ 。当 a, b, c, d 四项次数接近时, 估计值最可靠; 而当 a, b, c, d 中某项次数过少、甚至为 0 时, 计算结果无意义。

ϕ 相关。

当项目分数和测验总分都是二分变量时, 可用 ϕ 相关计算项目的区分度。其公式如下:

$$r_\phi = \frac{ad - bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \quad (\text{式4-20})$$

式中, r_ϕ 为 ϕ 相关系数, 即项目区分度; a, b, c, d 的意义同式 4-19。

(2) 用极端分组法计算区分度。

用极端分组法计算区分度, 是通过比较两个极端效标组(高分组和低分组)在同一项目上反应的差异来估计项目区分度的。其计算公式如下:

当项目用二分法计时:

$$D = P_H - P_L \quad (\text{式4-21a})$$

式中, D 为区分度; P_H 为高分组通过该项目的人数比例; P_L 为低分组通过该项目的人数比例。

当项目以连续分数计时：

$$D = \frac{\bar{X}_H - \bar{X}_L}{H - L} \quad (\text{式4-21b})$$

式中，D为区分度； \bar{X}_H 和 \bar{X}_L 分别为高分组学生和低分组学生在该项目上的平均得分；H和L分别为该项目的最高得分和最低得分。

高分组和低分组的划分，是先将被测学生按其总分由高到低的顺序依次排列，然后取其上端一部分为高分组，取其下端一部分为低分组。对于极端分组的人数应占多大比例，要视被测学生总体的具体情况而定。如果测验总分的分布符合正态，最适当的比例是高分组和低分组各占27%；如果分数分布较正态分布平坦，高、低分组人数比率要略高于27%；一般情况下，其比率介于25%-33%之间即可。

用极端分组法计算项目的区分度也被称作鉴别指数，其意义明确，具有计算简便、易于理解的优点，但所得结果不够精确，通常只用于一般教学测验中。

3. 项目区分度的评价

(1) 相对评价。

由于区分度值有多种计算方法，不同方法的计算结果并不一致；同时，区分度值还受到被测学生的状况如样本大小、样本的同质性、样本的分组方法等因素的影响。因而，项目的区分度值与其难度值一样，也具有相对性。所以，我们在评价项目的区分度时，要注意区分度值的来源，区别对待，而不能任意比较其值的大小，并以此评价项目质量。目前，我们还很难规定出一个统一的定量标准作为评价项目区分度的依据。一般而言，在被测学生总体状况相同或基本类似的前提下，按照同一种方法计算的区分度值可以直接比较：区分度越大者，项目质量越高。无论测验的目的是什么，在反映项目质量的其他指标相同的条件下，应尽可能选用区分度值大的项目。研究表明，尽管采用不同方法计算出的区分度值不同，但在根据区分度值决定项目取舍时，结论却是完全一致的。可见，通过对项目的区分度进行相对比较而挑选项目，其结论是基本可靠的。

(2) 显著性检验。

用内部一致性系数作为项目的区分度值，还可以通过对相关系数的显著性检验来评价项目质量。每种方法的计算结果都有相应的显著性检验方法。一般是将计算所得的区分度值与相关系数显著性临界值比较，如果检验结果是相关系数值达到显著性水平，就说明项目得分与测验总分之间确实存在着内部一致性（或一致的可能性相当大），项目的区分度水平是可以令人满意的；但如果检验结果未达到显著性水平，那么，项目得分与测验总分之间的相关可能是由某种偶然因素造成的，项目的区分度尚未达到测验所要求的水平。在一般教育或心理统计、测量学著作中都可以查阅到相关系数显著性临界值表。

三 选择题的反应模式分析

通过对项目的难度和区分度评价，能够为我们提供评价和选择项目的依据或线索。对那些难度不合理、区分度较低的项目，可以根据学生对选择题各个备选答案的选择率，对选择题的质量做出评价，并按以下6种情况，分析选择题存在的问题及其原因：

(1)如果难度值接近于 1.0,说明项目中正确答案过于明显,或为学生提供了某种暗示,需要增大错误答案的迷惑性。

(2)如果某个迷惑答案几乎没有被选择,说明该选项不具备迷惑性,错得过于明显,除增加了阅读项目的时间外,不起任何其他作用。一般来说,如果迷惑选项的选择率低于 3%,该选项就应删除或修改。

(3)如果某错误答案的选择率高于正确答案,就可能是编制时确定的正确答案有误,或教师在教学中发生了错误。这时,需要找到问题的原因,或是修改答案,或是纠正教学中的错误。

(4)如果规定选择题的正确答案只有一个,而高分组学生的选择却集中于两个答案,且两者的选择率相近,说明该项目可能存在两个正确答案,或是在某种意义上原定正确答案以外的另一个答案也存在着一定的正确性。这时需要仔细分析项目内容,重新确定答案。

(5)如果高分组对正确答案的选择率与低分组相近甚至更低,那就说明该项目的测量内容与学生的实际水平无关。该项目在修改试题时应删除。

(6)如果某项目各个被选答案的选择率几乎相同或未作答的人数较多(并非由于答题时间不够),说明该项目过难或题意不清,学生无法作答或只能凭猜测作答。这就需要适当降低项目难度或使题意更加清晰一些。

比较难度和区分度不同的选择题,可以发现:那些难度适中、区分度较高的选择题,在学生对被选答案的选择率上,表现出以下两个共同的特点:

(1)正确答案的选择率最高;

(2)迷惑答案的选择率相同或相近。

我们可以将这两个特点称之为选择题的最佳反应模式。因为,如果不能同时符合这一模式,就必然会出现一种或多种上述选择题可能存在的 6 个问题。这一模式也可以作为我们评价选择题质量的依据之一。

第五章 化学测试结果的分析与评价

在测验质量可靠的前提下，学生的测验分数就成了教师评价学生学习成绩的依据。但是，孤立的测验分数是没有任何意义的。这就需要教师掌握一定的评价标准，正确地解释测验分数的意义，对学生的学习成绩作出公正而合理的评价，并从测验结果中发现那些有利于改进教学的有益信息，进而提高教学质量。

第一节 学生集体测验分数的评价

在实际化学教学中，常常需要教师对自己所任教的班级的学习成绩作出评价。这种评价经常采用横向比较和纵向比较两种方式。横向比较如与本校的平行班级或外校的同年级学生进行比较；纵向比较如与前一个教学过程的测验成绩比较。横向比较是为了发现本班在同类学生总体中的相对位置，而纵向比较是为了发现本班学生的变化。

一 根据分数分布估计测验成绩

在已知参加测验的全体学生（总体）分数分布接近正态的前提下，可以根据被评学生集体的分数分布状态估计该集体的测验成绩。当学生集体的分数分布呈现图 4 - 3 所表示的几种典型形态时，可对学生集体的学习成绩作如下估计：

(1) 正态分布。说明样本的学习成绩正常。心理学研究表明，人的智力水平符合正态分布；常规班级教学中，学生的学习成绩一般也呈正态分布，即大部分学生处于中等水平，优秀学生和学习困难的学生比例都比较小。

(2) 正偏态分布。说明样本中成绩较差的学生偏多，成绩优秀的学生偏少。

(3) 负偏态分布。说明样本中成绩较差的学生较少，成绩优秀的学生较多。这是一种较为理想的成绩分布。

(4) 双峰型分布。说明样本中存在两极分化现象，即一部分学生成绩优秀而另一部分学生学习困难。这是最不利于组织班级教学的一种分布形态，一般需要采取某种管理措施如重新编班，来改变这种不利局面。

(5) 平坡型分布。说明样本中好、中、差的学生比例接近，学生间的差异也偏大。

(6) 陡峭型分布。说明样本的测验成绩比较集中，差异较小或比较整齐。

二 根据平均分和标准差估计测验成绩

平均分是教学实际中应用最为普遍的一种统计量数，它是代表性最好的集中量数。平均分可以反映出学生集体测验成绩的平均水平。我们可以比较被评学生集体和参加测验的学生总体之间的平均分差异，学生集体的平均分越高，说明该集体的测验成绩越好。在根据学生集体分数的分布形态估计成绩时，我们并不知道该集体在总体中的位置；有了平均分数据之后，由于平均分总是位于分布曲线的中部，据此就可以判断被评学生集体在总体中的大致位置了，使我们对学生集体成绩的估计更加细致、全面些。

在教学中，有时需要根据几个学生集体的平均分计算其总平均分。我们不能只是简单地取几个平均分的平均值，而要依据下式：

$$\bar{X}_i = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_k \bar{X}_k}{n_1 + n_2 + \dots + n_k} \quad (\text{式5-1})$$

式中， \bar{X}_i 为总平均分； n_1, n_2, \dots, n_k 为各学生集体所含学生数； $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ 为各学生集体的平均分； k 为被统计的学生集体的个数。

标准差是代表性最好的差异量数。标准差为我们估计分数的离散情况提

供了数量化指标。标准差越大，说明分数间的差异越大，分数分布越不均匀；相反，标准差越小，说明分数分布越集中，大部分分数集中在平均分左右。因此，标准差越小，平均分对学生集体成绩的代表性越好；反之，标准差越大，平均分的代表性越差。可见，在用平均分描述学生集体成绩时，需要同时用标准差来说明平均分的代表性大小。

在教学中，有时需要根据几个学生集体的标准差计算其总标准差。我们同样不能只是简单地取几个标准差的平均值，而要依据下式：

$$S_t = \frac{\sum_{i=1}^k n_i S_i^2 + \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}_t)^2}{\sum_{i=1}^k n_i} \quad (\text{式5-2})$$

式中， S_t 为总标准差； n_i 为学生集体的人数； s_i 为学生集体的标准差； \bar{X}_i 为学生集体的平均分； \bar{X}_t 为总平均分； k 为被统计的学生集体的个数。

在用平均分和标准差估计学生集体的成绩时，一般都同意平均分越高，成绩越好。但是当标准差大小不同时，哪种情况学生集体的成绩更好呢？这就要看评价者所依据的教学质量标准是什么了。例如，有人认为，标准差大，说明学生之间差异较大，可能会有成绩突出优秀的学生存在，有利于培养尖子；但也有人认为这样不利于面向全体学生组织教学，总有部分差生存在。标准差小，学生比较整齐，有利于班级教学，但也可能缺乏突出优秀的学生，等等。根据不同的评价标准，就会得出不同的结论。这里要涉及教育观念等较深层次的问题，不拟在此展开讨论。

三 不同学生集体的测验分数比较

利用学生的分数分布、平均分、标准差估计学生的学习成绩，仍是比较粗略的。例如，两个班的考试平均分相差3分，我们能否断定其中一个班肯定要好于另一个班呢？对这类问题的回答，就需要进行统计学上的差异检验了。下面我们通过举例介绍教学中常用的一种平均分差异的检验方法。

1. Z 检验

如果学生总体是正态分布或接近正态分布，样本容量较大 ($N \geq 30$) 时，可用 Z 检验方法检验两个样本的平均数差异的显著性。

检验的具体步骤是：

计算 Z 值。

Z 值的计算公式为：

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (\text{式5-3})$$

式中， \bar{X}_1 ， \bar{X}_2 分别为样本1、样本2的平均分； S_1 ， S_2 分别为样本1、样本2的标准差； n_1 ， n_2 分别为样本1、样本2的容量。

确定显著性水平。若 $\alpha=0.05$ ，则 $Z_{0.05}=1.96$ ；若 $\alpha=0.01$ ，则 $Z_{0.01}=2.58$ 。 $Z_{0.05}$ 和 $Z_{0.01}$ 都为定值。

比较 Z 和 Z_α ，作出判断。如果 $Z \geq Z_\alpha$ ，可以在 α 所确定的水平上断定，

样本 1 和样本 2 的平均分存在显著差异；如果 $Z < z_{\alpha}$ ，则说明样本 1 和样本 2 的平均分无显著差异，或者说，这种差异是由抽样误差造成的。

例 1A, B 两学校的高中一年级各 100 名学生参加了由区教育局组织的化学统一考试。A 校的平均分为 78 分，标准差为 9 分；B 校的平均分为 76 分，标准差为 8 分。问两校学生的化学成绩有无显著差异？

计算 Z 值：

$$Z = \frac{78 - 76}{\sqrt{\frac{9^2}{100} + \frac{8^2}{100}}} = \frac{2}{1.204} = 1.66$$

确定显著性水平 $\alpha = 0.05$ ， $Z_{0.05} = 1.96$ ；

作出判断： $Z = 1.66 < Z_{0.05} = 1.96$ ，A, B 两校高中一年级学生的化学成绩并无显著差异。

2. t 检验

如果样本容量较小 ($N < 30$)，可用 t 检验方法检验两个样本的平均数差异的显著性。

检验的具体步骤是：

计算 t 值。

为了简化计算，可用以下公式计算 t 值的近似值。

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (\text{式5-4})$$

式中， \bar{X}_1 ， \bar{X}_2 ， S_1 ， S_2 ， n_1 ， n_2 的意义同式 5-3。

确定显著性水平。 t_{α} 值要随自由度 df 值而变。根据 α 和自由度 df 值，查 t 值表，确定 $t_{(df)\alpha}$ 值。 $df = n_1 + n_2 - 2$ 。

比较 t 和 t_{α} ，作出判断。如果 $t > t_{\alpha}$ ，可以在 α 所确定的水平上断定，样本 1 和样本 2 的平均分存在显著差异；如果 $t < t_{\alpha}$ ，则说明样本 1 和样本 2 的平均分无显著差异，或者说，这种差异是由抽样误差造成的。

第二节 学生个体的测验成绩评价

一 以常模参照评价学生的测验分数

以常模参照评价学生的测验分数，就是把学生的原始测验分数转换为常模分数，进而指出其在某一群体中的相对位置，以此作为评价学生学习成绩的依据。这里的参照群体也称常模团体。被参照的常模团体，应与被评价的学生特质相同，而且要对具有同类特质的所有学生总体有很好的代表性。常模分数就是以常模团体的一般水平（如平均分）作为参照标准的导出分数。

在化学教学测评中，我们可以根据测验的目的、内容和被测学生的特点，取得我们所需要的不同常模。例如，教师要确定某个学生在其所在班级中的学习状况，就可以将班级学生的平均水平作为班级常模；如果教师要比较某个学生的化学成绩在全区或全市同级学生中水平的高低，那么，全区或全市同级学生在化学考试中的平均水平就是全区或全市的化学成绩常模。

在化学教育测评中，最常用的常模分数是百分等级分数和标准分数。

1. 百分等级分数

百分等级是指在常模团体中，某分数以下的人数的百分比。百分等级分数是把常模团体的全体学生分为 100 份，按被评学生的原始分数换算出其在常模团体中的相对位置后所得到的一种导出分数，常用 P_R 表示。例如，某学生在一次化学考试中的原始成绩转化为百分等级分数 $P_R=75$ ，则表明在全体学生中有 75% 的人低于该生的分数。 P_R 越小，表明所处的相对位置越低，成绩越差。

百分等级分数的意义十分明显，计算也比较简单。无论测验分数的分布是何种形态，都可以用百分等级表示某学生在团体中的相对位置。可以用百分等级分数来比较一个学生两次测验的成绩，也可以比较两个团体的测验成绩。与教学中常用的排名次相比，百分等级分数更具有一般意义。因此，它在实际教学中的应用十分广泛。但在使用百分等级分数时，要注意其以下特点：

百分等级属于顺序量表，百分等级分数的差异与测验分数上的差异不成比例，其单位是不等的，因而不能进行代数运算，这给对分数的进一步分析带来了困难。

百分等级只表明学生在团体中的相对位置，并不表明学生掌握所测内容的百分比。百分等级和测验内容答对率不是一回事，某位学生的百分等级只取决于该生的分数与其他学生分数的关系，而与全体学生测验成绩的好坏无关。常模参照的特点即在于此。

百分等级对测验的分数变化反应不敏感。一方面，原始分数的很大变化不能在百分等级上引起相应的变化，这在两端分数中表现尤为明显。例如，第一名和第二名、倒数第一名和倒数第二名之间原始分数可能相差很大，但百分等级却只差 1。另一方面，在分数分布的中部，尤其是接近正态分布的测验分数，中间分数分布比较密集，往往分数的较小变化会引起百分等级的较大变化。例如，有 100 名学生被测总体，其中百分等级从 45 到 55 的两名学生的原始分数可能只差 3.5 分。

2. 标准分数

标准分数又称 Z 分数。其计算公式如下：

$$Z_i = \frac{X_i - \bar{X}}{S} \quad (\text{式5-5})$$

式中, Z_i 为学生 i 的标准分数; X_i 为学生 i 的原始分数; \bar{X} 为原始分数的平均数; S 为原始分数的标准差。

从式 5 - 5 不难看出, 标准分数是以标准差为单位度量原始分数离开其平均分数的量数, 它可以表示某个原始分数在团体中所处的位置, 即该原始分数在平均分之上多少个标准差, 或是在平均分以下多少个标准差。

标准分数具有以下性质:

一组标准分数的平均分为 0, 标准差为 1。

标准分数的绝对值表示某一原始分数与平均分的距离; 标准分数的正、负表示原始分数是在平均分之上还是之下。

标准分数的分布与原始分数的分布相同。如果原始分数的分布是正态或接近正态, 则标准分数的分布范围在 -3 到 +3 之间。

在原始分数的分布是正态或接近正态时, 标准分数和百分等级分数有对应关系。如果知道某学生的 Z 分数, 可以直接查正态分布表, Z 值所对应的正态曲线下面积比例乘以 100, 便是该分数的百分等级; 如果已知百分等级, 也可查正态分布表找到其对应的 Z 值。

可见, 标准分数是一个以绝对零点为参照点、以标准差为相等单位的等距量数。不管原始分数的平均数、标准差如何, 相同的标准分数表示在分布中处于同样的相对位置。标准分数具有可比性, 还具有可加性, 可以直接合成运算。因此, 我们在对学生的不同考试成绩进行比较时, 或是综合学生的各科考试成绩对学生做出评价时, 应当将学生的原始分数化成标准分数后再合成或比较, 而不能直接使用单位不等、参照点不同的原始分数进行合成或比较。

由于 Z 分数有负数和小数, 给以后的统计分析带来不便, 也与日常教学中使用的评分形式不相一致, 美国教育测量专家麦柯尔建议将 Z 分数转换为 T 分数。其转换公式如下:

$$T = 10Z + 50 \quad (\text{式 5 - 6})$$

式中, T 为 T 分数; Z 为标准分数。

转换成 T 分数以后, 标准分数就不再有负数和小数了。 T 分数的平均数为 50, 标准差为 10。由原始分数转换而来的 T 分数若大于 50, 则在平均数之上; 若小于 50, 则在平均数之下。在原始分数分布为正态或接近正态时, T 分数的最大值为 80, 最小值为 20。

T 分数实际上是由标准分数派生出的标准化分数。在麦柯尔思想的基础上, 人们为了使用的方便, 又派生出一些其他的标准化导出分数。如美国大学入学考试委员会所采用的 CEEB (College Entrance Examination Board) 分数, 就是一种类似于 T 分数的标准化导出分数。其计算公式为:

$$\text{CEEB 分数} = 100Z + 500$$

著名的 TOEFL 考试, 是按下式报告考试成绩的。

$$\text{TOEFL 分数} = 70Z + 500$$

3. 常模表

常模表是一种呈现常模资料的方法, 是用来作为比较标准的一种分数量

表。为了使各种考试分数可以相互比较，可以把各种考试分数通过百分等级或标准分数将其等值化，并放到统一的量表上，这就构成了常模表。利用常模表，对任何原始分数，不需要进行计算，就可在表上直接查到它在常模团体中的导出分数，从而明确原始分数的意义。

常模表一般有三个组成部分：一是原始分数量表，二是与每个原始分数对应的导出分数，三是有关常模团体的描述资料。请看表 5 - 1 给出的示例。

表 5-1 × × 中学高二化学学年末测验常模表（示例）

原始分 数量表	文科班			理科班		
	百分等级	标准分数	T 分数	百分等级	标准分数	T 分数
100				100	2.82	98
99				99	2.60	96
98				98	2.38	94
97				97	2.16	92
96				96	1.93	89
95				95	1.71	87
94				94	1.49	85
93				93	1.27	83
92				90	1.04	80
91	100	2.20	92	89	0.82	78
90	99	2.00	90	87	0.60	76
...

二 以目标参照评价学生的测验分数

以目标参照评价学生的测验分数，是根据预先规定的教学目标，看学生对教学所要求的知识、技能的掌握情况来确定测验分数的意义，评价学生的测验成绩。

常模参照评价虽然相对比较客观、应用也相当广泛，但其自身也有一些难以克服的缺陷。由于常模参照分数着重指出被评学生在集体中的相对位置，不管学生如何努力或学习成绩的好坏，都被排成名次，总有名列前茅和名落孙山之分，容易对学生产生不利的心理影响；常模参照分数不能明确指出被评学生达到教育目标的程度，也不宜于诊断学生的学习困难；当参照的常模不同时，常模参照分数就不能相互比较。目标参照评价是一种绝对评价，它不是将被评学生与其他学生或常模团体进行相对比较，而是与事先规定的、独立于常模团体之外的教育目标进行比较。学生可以通过目标参照分数，发现自己在目标到达度上的进步，以增强学习动机；由于目标参照分数直接和教育目标相关，教师可以利用它诊断学生的学习困难；只要被测学生所追求的教育目标相同，就可以相互比较他们的目标参照分数。因而，目标参照分数可以弥补常模参照评价的不足。

测量项目是与其测量目标对应的。以目标参照评价学生的测验分数，就

表中样本取自该校 253 名（其中文科 101 名，理科 152 名）学生的化学测验分数。文科： $\bar{X}=75.5$ ， $S=5.0$ ；理科： $\bar{X}=75.5$ ， $S=4.5$ ； $T=10Z + 70$ 。

必须明确达到教育目标的具体行为标准。依据不同的测量目标，要求学生达到的行为标准可能有多种。常用的目标参照评价的标准有三种，相应的有三种目标参照分数：

1. 掌握分数

掌握分数是说明学生是否达到教育目标所要求的最低标准而设置的一种目标参照分数。在实际教学中，我们有时并不需要对学生进行详细区分，而只需要知道学生对一些基础知识和技能是否掌握。为此目的而设计的测验可称为掌握测验。对掌握测验的计分，只需要定出一个可接受的最低标准即可。一般采用二分法，如及格——不及格，通过——不通过等等。如学生达到最低标准，则计为及格，说明其已经掌握了所需求的基础知识和技能，从而可以进入下一个单元的学习；如达不到最低标准，则计为不及格，说明学生未能达到教育目标的最低要求，一般不能直接进入下一单元的学习。

掌握分数的确定标准，是根据测验目的和要求的教学实际确定的。在化学教学实际中，常常以60分为及格标准，如果学生的测验分数达到或超过60分，则认为学生达到了教学目标的最低要求。我们认为，从掌握学习的要求看，一般以75%—85%的正确反应作为及格或通过的最低标准比较合适。

掌握分数常用于那些能够确定出可能接受的最低标准的资格考试或成就考试。如高中毕业会考，一些发放职业资格证书、执照的考试等。

2. 正确百分数

正确百分数是指学生答对项目所占的百分比，即：

$$\text{正确百分数} = \frac{\text{答对的项目数}}{\text{项目总数}} \times 100\% \quad (\text{式 5-7})$$

正确百分数的意义十分明确，说明学生对所要求的教育目标的掌握程度。正确百分数越高，说明学生对目标掌握得越好。与二分法计分的掌握分数相比，正确百分数可以对学生起到进一步的区分作用。正确百分数的计算也很简单，因此，其应用也比较广泛。

在应用正确百分数时要注意，测试项目必须是有限的测量目标的代表性取样。测量目标是教育目标的具体化、行为化；项目又是测量目标的代表性体现，这样，正确百分数才能代表学生掌握教育目标的实际程度，否则，正确百分数是毫无意义的。同时还要注意，对于取样范围不同、难度不同的考试，正确百分数的意义显然不同；即使在同一次测验中，两学生的正确百分数相同，但他们正确回答的项目不同，其分数的意义也是不相同的。

3. 等级分数

当我们用等级评定的方法来表示学生的某种知识、能力、品质等方面的发展水平时，所用的分数就是等级分数。特别是对学生的某种素质，比如思想品德、化学学习兴趣和态度、化学实验操作技能等，难以用学生回答项目情况来准确评定时，就可用等级分数来表示学生的水平。

所谓等级分数，就是把要测量的某种能力、技能或品质分为若干个等级，每个等级赋予某一分数。比如评定某项操作技能，分为5个等级，分别赋予5分、4分、3分、2分、1分；对每个学生的实际操作分别评定出适当的等级，并给予相应的分数。

当常模参照考试与目标紧密联系时，目标参照分数与常模参照分数之间并无严格的界限。有时掌握分数或等级分数就是同时按照被测团体达到教育目标的平均水平确定的。因此，可以把两种分数结合起来使用，在确定掌握

分数或等级分数时，既要按照教育目标本身的要求，又要依据被测团体对教育目标的平均到达度。

三 评价学生成绩时应注意的问题

1. 明确评价标准，要正确地解释或比较分数

在解释测验分数之前，要有合理、明确、可被接受的评价标准。同一个测验分数，按照不同的评价标准，所得出的结论可能完全不同。对于不同测验的分数由于测验的内容、难度不一，所得的分数就不能直接比较，因为它们没有统一的量表。当两次测验分数可以用同一个标准或等价的量表等值化时，才能够相互比较。例如，对同一学生集体实施两次化学测验，所得到的分数都可以转换成百分等级分数或标准分数进行比较。

2. 要考虑到测验误差

任何测验，其信度和效度一般都可能不足，都要受到测验误差的影响，测验结果并非是完全可靠的。因此，在解释测验分数时，不能脱离学生的实际及测验的具体环境条件，应考虑到测验误差的存在。

测验误差与测验的信度有关。信度高则误差小，反之则误差大。所以，不要把测验分数看成是一个确定不变的值。由于误差的存在，测验的实际获得分数总是表示一个区间。因而在解释测验分数时，最好是用学生的实际得分加、减一个测量标准误来表示学生的真实成绩，即

$$X_v = X_r \pm s_{\text{meas}} \quad (\text{式5-8})$$

式中， X_v 为学生的真实分数； X_r 为学生的实得分数； s_{meas} 为测量标准误。

测量标准误与测验的信度有关，即

$$s_{\text{meas}} = S\sqrt{1 - r_{XX}} \quad (\text{式5-9})$$

式中， S 为测验分数的标准差； r_{XX} 为测验的信度。

为此，我们在比较两个学生的测验分数时，不能仅仅看分数的绝对差异，还要看它们的真实分数区间。只有两个分数区间完全不重合时，我们才能确信这两个分数之间真正存在差异。

3. 不要一次考试定终身

任何一个测验都不能同时测出多方面的能力。一个测验的测量范围越广，项目之间的同质性就越差，测验结果的效度也就会越低；反之，测验项目的内在一致性越好，测验结果的效度也就越高。这就是说，越是好的测验，其测量目标就越集中。学生在不同方面的能力，只能通过不同的测验才能测得。因此，我们不能凭借一次测验就评定学生各个方面的能力，而要综合那些能够反映学生能力的各方面材料来全面地对学生作出评价。

第六章 化学标准化考试

标准化考试近年来已经得到了越来越多的关注。通过了解标准化考试的基本理论及其实施程度，并将标准化考试的一些原则、思想应用于化学教育测量，有助于提高化学教育测量的科学化水平。本章首先介绍标准化考试的基本理论，然后讨论我国化学标准化考试的一些主要问题。

第一节 标准化考试的基本理论

标准化考试是在工业发展进入第二次浪潮时，在社会劳动标准化、雇佣办法标准化的影响下，随着教育方面的课程设置、学生选拔及智力测验的标准化而产生和不断发展、完善的。对何谓标准化考试，迄今尚未有一个公认的确切定义。我们可以通过对标准化考试和一般考试的比较来认识标准化考试的特点；再通过对标准化考试的基本程序和要求了解，较全面地认识标准化考试。

一 标准化考试的特点

与一般的考试或测验相比，标准化考试具有以下特点：

1. 命题过程的标准化

标准化试题的编制要严格按照考试计划进行。在标准化试题的编制过程中，预测是必不可少的重要程序。凡编入标准化试题的所有项目，必须预先经过试测。根据试测结果，在对项目质量进行严格的统计分析的基础上筛选出那些难度适宜、区分度较高的项目构成试题。未经过预测的项目不能用于标准化试题中。这样，对试题的难度、区分度以及效度、信度的估计有预测结果作为比较客观的依据，而不像非标准化考试那样，仅仅根据命题者的主观经验来判断。

2. 施测过程的标准化

非标准化考试对施测过程的控制一般并不十分严格。但标准化考试对于如何向学生说明考试目的、怎样指导学生答题、对考试的环境、时间以及考试中可能发生的偶然事件的处理方法等等，都做了统一的规定，并要求在施测过程中严格执行，以保证标准化考试在统一的环境、标准下进行，从而保证公平合理地对待所有考生，保证考试结果的客观性，提高考试分数之间的可比性。

3. 评分计分方法的标准化

评分和计分方法是导致考试误差的一个极其重要的因素。标准化考试采用各种方法严格控制评分计分过程中的误差，如：统一规定详细的评分标准、阅卷采用流水作业方式、对评分计分结果多次复查、采用计算机阅卷等。

4. 分数解释的标准化

目前，标准化考试多用于较大规模的常模参照性测验，在解释分数的意义时，一般都以事先编制的常模为依据。

二 标准化考试的基本程序和要求

1. 制订考试大纲

考试大纲是对考试的内容、方法和要求的明确规定，是整个考试的根本依据，对考试起着法定文件的作用。考试大纲由考试的主持机构根据考试目的，经过慎重、广泛的调查研究后制订。考试大纲应提前一年向社会公布，以便考生和有关人员对各考试的各种要求心中有数，而且将考试置于社会的监督之下。标准化考试必须具备科学、明确的考试大纲。考试大纲的基本内容包括：考试的目的和性质、考试的内容和范围、考试的方法和形式、样题、评分与分数解释方法等。

2. 编制试题

一般由考试机构组织学科专家组成命题委员会和审题委员会进行这一工作。命题的基本程序是：

(1)拟定命题计划。通常借助于双向细目表实现。

(2)编写项目。编写时，项目的内容和水平要与双向细目表的要求一致。

(3)预测。预测的目的是：核定试题的难度和区分度；了解完成试题所需时间，决定试题长度；进一步审题，提出修改方案。组织预测时要注意：取样要有代表性；要有足够的样本参加；实施预测的过程和环境要力求与正式考试一致；要在预测时间上留有余地，使受试者都有机会完成试题，以保证每个项目都能得到预测。

(4)项目分析。对预测结果进行统计分析，确定项目的难度、区分度、选择题备选答案的适宜度等，对项目重新筛选和修订。对适宜的项目可存入标准化考试题库。

(5)拼题。根据考试的目的和要求，选择合适的项目，加以适当编排，拼成试题。拼题时要注意：试题要与考试大纲及命题计划一致；试题的长度要适宜；项目的编排顺序利于考生作答。

(6)编制试题副本。各份副本必须等值，即试题中对应项目的内容、难度、题型、数量、编排顺序等都相同，但项目不重复。

3. 实施考试

要做好考试前的试题保密工作。试题的印刷、传递、保管、发放等都按统一的标准控制。考场管理要标准化，一般要求：所有考场的物理环境和文化环境相同；主考人在考场中的行为统一、规范，以保证考生有相同的心理环境；分发试卷和收卷有统一规定，特别在有A、B卷时要有具体明确的要求。

4. 评分计分

要制订评分细则，明确给分标准，采用流水作业的方式，减少主观因素的影响。要有专人负责试卷的保管、分发、收回，统计阅卷工作进度；要组织试评，解决评卷中的疑难问题，做好抽查和监督工作，及时发现、纠正错误及宽严不当等不合理现象；分数的登录、累计、复查等都要有专人负责，严格管理。

5. 分数解释

对于以选拔为目的的常模参照考试，都应建立代表性好的常模，以此作为参照点，将卷面分数或原始分数转化为导出分数；对于以鉴定为主要目标参照考试，则应建立科学的比较标准。使分数的解释更为客观、明确、合理。

三 对标准化考试应有的认识

1. 标准化考试的概念

通过以上讨论，我们可以将标准化考试定义为：标准化考试是指根据统一、规范的标准，对考试的各个环节包括命题、施测、评分、计分、分数解释等都按照系统的科学程序组织，从而严格控制了误差的考试。

标准化考试的概念有三层含义：首先要有统一而规范的标准，在其约束之下，考试的各个环节才能有章可循，做到统一、规范；其次要有一套严格的科学程序来组织考试，以保证标准的贯彻执行；第三，通过严格的标准和科学的组织程序，最大限度地减少误差，保证考试的可靠性和有效性。

可见，标准化考试是一个系统化、科学化、规范化的测量过程。只有考

试中的所有环节都实现了标准化，即按标准确定考试的目的和计划，命题标准化，施测标准化，评分计分标准化，分数解释标准化等，才能称作标准化考试。只是个别或部分环节的某些标准化做法，不能代表整个考试的标准化，因而也不能称为标准化考试。那种单纯认为只要用客观题或选择题的考试就是标准化考试，或是认为只要采用标准分数计分或解释分数的考试就是标准化考试，显然是不全面的，也是不正确的。还有那种将未经预测的客观式项目组合起来就冠之以标准化试题集的做法也是不恰当的。

2. 标准化考试的长处和局限

与一般考试相比，标准化考试的长处表现在：（1）能够最大限度地减少考试误差，使考试结果比较准确、可靠；（2）具有统一的参照标准，使不同考试的分数之间具有可比性；（3）可以用来校准其他考试。

但是标准化考试并不是万能的。标准化考试只是将误差控制在较小范围，但不可能排除所有误差。那种认为只要实现了标准化，考试中的所有问题都能解决的想法是不切实际的。由于考试的目的不同，测量的特性也不一样，因而那种只凭借一次标准化考试定终身的做法也是不足取的。

由于考试理论、方法和手段是不断发展和完善的，标准化考试的实现，并不意味着考试方法发展到顶峰。目前，标准化考试的许多标准是建立在经典测验理论之上的，但经典测验理论仍有一定的缺陷，需要发展和完善，如效度、信度、区分度等指标是建立在一种比较理想化的假设之上，与考试实际难免存在偏差，从而影响了指标的准确性。因此，我们对标准化考试不可迷信。

虽然标准化考试无论在理论上和手段上还不够成熟，还有一定的局限性，但是就目前情况看，它毕竟提出了一些比较科学、可行的标准和措施，对考试误差进行了比较有效的控制，而且较好地建立了参照点和度量单位，使考试规范、科学。因此它仍然是我们应当认真加以学习、借鉴的考试方法。

3. 学习、借鉴标准化考试的意义

学习、借鉴标准化考试的主要意义在于，通过应用标准化考试的理论、思想、方法，发现和改革现行考试体系中的问题和弊端，提高考试的科学化水平。

应当看到，任何考试都有一定程度的标准化水平，没有任何标准的考试不能成其为考试。标准化只是一个相对概念，并无有无之别，而只有水平高低之分。在教学过程中，照搬标准化考试程序和方法是比较困难的，其实施效果和工作效率都可能不理想。但是，标准化考试的理论和方法可以指导我们比较系统、科学地组织考试过程，分析和解决考试过程中的问题，在可能的条件下，有目的、有重点地控制考试误差，从而提高化学教育测量的科学化水平。

第二节 化学标准化考试的实践

一 高考化学考试的标准化

我国自 1977 年恢复高考以来，传统考试的弊端越来越明显地暴露出来：命题由少数人突击进行，试题质量缺乏科学的评判指标，仅凭命题人员的个人经验编制试题，质量难以保证；考试各个环节的管理缺乏科学的标准；评阅试题缺乏有效的控制误差的方法，因而使考试的信度和效度受到了影响。如何提高高考的质量，迫在眉睫。当国外的标准化测验随着改革开放的浪潮被引入我国的时候，由于它可以为控制考试误差提供一种比较有效的方法，很快引起了广泛关注，标准化考试的原理和方法很自然地应用于高考改革。

标准化考试的理论和方法被逐步应用于化学高考试题中具体表现在以下几个方面：

1. 客观题的比例逐步增大

1978—1992 年高考化学试题三类题型所占分数的比例见表 6-1。

表 6-1 1978—1992 年高考化学试题三类题型分数比例

年份	选择题	填空题	简答题	年份	选择题	填空题	简答题
1978	0	8	92	1986	34	38	28
1979	0	20	80	1987	45	36	19
1980	17	26	57	1988	60	24	16
1981	30	10	60	1989	55	33	12
1982	25	14	61	1990	60	28	12
1983	30	4	66	1991	55	31	14
1984	21	18	61	1992	55	27	18
1985	30	30	40				

说明：题型按学生答题的具体方式划分。其中，填空题和简答题按要求学生给出答案的字数粗略划分；是非题归类于选择题。

1980 年起，选择题型被用于高考化学试题之中；从 1980—1985 年，选择题的比例一直徘徊在 30% 以下，试题仍以主观式题型为主；从 1986 年开始，选择题比例突破 30%，在 1988 年上升到 60%，其后基本稳定在 55% 左右，客观式题型的比例也稳定在 80% 左右。这从一个侧面反映出国内对客观式题型特别是选择题认识的变化过程（见图 6-1）。

随着客观题比例的增大，项目总数增大，单个项目分值降低，大的综合题趋向于小型化，使综合题的评分客观性增强，从而整个试题的客观性增强。

2. 试题的覆盖面增大

由于客观题比例的增大，项目总数增大，使试题中出现的知识点几乎涉及统编化学教材的所有章节。

图 6-1 1978—1992 年高考题型比例变化示意图

3. 命题方法向标准化靠近

试题的编程序和指导思想逐步纳入标准化考试的范围。如：根据教学大纲制订的考试目标比较明确，考查知识和能力的要求更为具体；命题以双

向细目表为依据，广泛征题，选点预测，逐步建立高考试题库。

4. 评分的客观性增强

试题分为 卷和 卷， 卷用计算机代替人工阅卷，不仅增大了评分的客观性，也大大提高了阅卷工作的效率； 卷由于填空题比例较大，总题数较大，评分较为简单、明确，从而使评分误差相对减少。

5. 公布考试大纲

1991 年 11 月，国家教育委员会考试中心公布了《普通高等学校招生全国统一考试化学学科说明》，明确说明了考试性质、考试内容、考试形式及试卷结构，并给出了题型示例和样题。

二 广东省化学标准化考试试点

为了探索我国高考标准化之路，国家教委将广东省确定为实行高考标准化的试验区，在 1985 年先进行了英语、数学两科的标准化考试试验；1986 年，试验扩大到英语、数学、物理三科；1987 年，又在广东省首次进行了化学学科标准化考试试验。

这次试验按照标准化考试的要求，并结合广东省和中国的实际，制订并比较严格地实施了符合标准化考试特点的试验工作程序，主要内容和特点包括：

(1) 制定考试大纲。

以化学教学大纲、统编教材和高考的目的要求为依据，确定考试的目的、范围、内容、方式、规格以及评分计分方法等。试题分一卷和二卷，一卷全部是四选一的选择題，满分 70 分，答卷用答题卡，由计算机评分计分；二卷全部为简答题或填空题，满分 30 分，由教师手工评分计分。考试大纲提前半年向社会公布。

(2) 编制命题双向细目表，如表 6-2 所示。

表 6-2 1987—1990 年广东省化学高考标准化中的细目表考查目标

考查内容 \ 考查目标	考查目标					合计
	识记	理解	应用	分析综合	探究	
基本概念与理论	1	13	9	7	2	32
元素化合物	3	5	6	5	2	21
有机化合物	1	5	3	4	2	15
化学计算		3	4	8		15
化学实验	1	6	2	6	2	17
合计	6	32	24	30	8	100

(3) 建立试题库。

在省内广泛征题，经指定专人评估或编入试题在观察点试测后，存入计算机。

(4) 命题与试测。

编制了三份测试题，在全省观察点试测，测后将难度低、区分度差(0.20 以下)的项目淘汰，保留合乎质量要求的题存入题库。

(5) 建立分数解释常模。

以观察点的三次试测结果分析为依据，估计和控制试题的难度(平均分)

和标准差，对学生的正式考试成绩用 T 分数报告。原始分数转换为 T 分数的公式是：

$$T=KZ+C$$

其中，Z 为标准分，由 $Z = \frac{X - \bar{X}}{S}$ 得到（X 为原始分数， \bar{X} 为平均分，S 为标准差）； $K = \max[S] + 1$ ，这里的 $\max[S]$ 为标准差的整数最大值；C 是常数，按平均分确定。实测 $\bar{X} = 59.0$ ， $S = 12.6$ ，所以 $K = 14$ ，取 $C = 60$ 。原始分数转换为 T 分数之后，考生的分数分布更接近正态，有利于选拔。

(6) 分析考试质量。

分析结果表明：一卷、二卷和全卷的考生分数分布都比较接近正态，稍呈负偏态；分数经 T 转换后，68 分以上考生人数百分比与该省高校新生录取率相同；试题区分度良好；试题所涉及的知识内容比例恰当，在认知领域里的能力考察层次分明、比例合理、项目难度分布适当。一卷题量稍多，难度稍大，答卷时间不够充裕，二卷则题量尚可，但难度不够，答卷时间稍多。

国家教委考试中心负责人指出：广东省的试验表明，“我国高考完全可以实行标准化，但我国高考标准化不应照搬美国的具体形式（如全部用选择题）；我国应当也完全可以在总结传统考试经验的基础上吸收外国考试的长处，创造适合中国国情的标准化考试（如主客观题型结合）。广东的试验还表明，实现标准化以后，高考可以成为科学、合理、稳定的考试。这是科举以来，我国考试形式上的重大改革和突破。它有利于克服由于片面追求升学率而造成的某些弊病；有利于克服押题、死记硬背等现象；有利于中学生打好基础、培养能力、发展智力。”由此可见，高考的标准化将是今后相当长时期的高考改革的方向。

三 化学标准化试题库

随着国内标准化考试研究和实践的深入发展，标准化考试的一些程序和方法也开始逐步应用于更加广泛的化学教育测评领域之中。近年来，国内已经建立起了一些化学标准化试题库，并逐步将其应用于常规化学教学之中。

题库是指具有必要参数的、质量符合测试要求的项目集合。题库中的项目一般都经过试测、统计分析、精心选择，项目参数表明项目的性质，如难度、区分度、题型、学习水平等。利用题库命题，可以在一定程度上避免主观因素的影响，保证试题的科学性、客观性，同时可以大大减轻教师的工作负担，提高工作效率；还为客观、公正地评价教师的教学质量提供了一种比较有效的手段。现以北京景山学校和中国长城计算机公司联合研制开发的 CBE 化学题库系统为例，介绍化学标准化题库的基本结构和功能。

CBE 系统共有项目近 2 万个。这些项目是由专业化学教师根据现行中学化学教学大纲和高等学校招生考试大纲的要求，参考了北京、上海、天津、福建等全国 20 多个省市的教学经验和资料，以统编中学化学教材的章为单位，以滚雪球的方式精心设计的。其中大部分项目已在不同地区、不同类型的学校进行了试测，确定了相应的参数。这些参数的种类及其意义见表 6-3。

曾灼先等：1987 年广东省化学高考试卷分析，化学教育，1988 年第 3 期。

杨学为：标准化考试的理论与实践，广东高等教育出版社 1988 年版。

表 6 - 3CBE 题库系统参数的种类及其意义

类型	参数	意义	类型	参数	意义
章	1	卤素	认知水平	1	识记
	2	摩尔		2	理解
	3	硫、硫酸		3	应用
	4	碱金属		4	分析综合
	5	物质结构、元素周期律	难度	1	容易
	6	氮和磷		2	较容易
	7	化学反应速度和化学平衡		3	比较难
	8	电解质溶液		4	难
	9	硅、胶体	题型	1	填空题
	10	镁、铝		2	判断题
	11	铁		3	选择题
	12	烃		4	计算题
	13	烃的衍生物		5	问答题
	14	糖类、蛋白质		6	写方程式
知识块	1	基本概念	综合度	1	包含知识因素 1—2个
	2	基本概念		2	包含知识因素 2—3个
	3	元素化合物		3	包含知识因素 3—4个
	4	有机化学		4	包含知识因素 4个以上
	5	化学计算			
	6	化学实验			

从上述参数可以看出该题库的基本结构。一套参数确定之后，一组等值的（测量目标相同的）项目就随之确定了。

利用该系统命题，只要给出考试的范围、难度、认知水平、题型以及题的数量，即输入必要的参数，题库即可以高效、自动地生成试题（答案同时生成，另行打印），并显示、打印出字迹清晰、卷面整齐、规范、图文合一的试卷。出同一水平的试题可以同时提供几份复本。中学化学教学测量需要的几类试题都可以提供。该系统还可以对项目进行修订、补充（有 2 万个项目的扩充余量）。化学标准化题库的建立和投入使用，表明国外标准化考试的理论、方法在我国化学教育领域已经比较广泛而深入地得到了应用，化学教育测量和评价手段已经开始逐步现代化。

第七章 情感领域化学教育目标及测评

我们在第二章讨论化学教育目标的内容分类时，已经说明：情感领域的化学教育目标是化学教育目标系统的重要组成部分。随着人类社会的进步，人们的教育观、人才观也发生了变革，包括各种非智力因素在内的情感领域的教育效果越来越引起人们的重视。在化学教学中，不仅要促进学生掌握知识、发展能力，还要对他们进行辩证唯物主义观点教育、爱国主义教育，注意培养和激发学生的学习动机、发展学习兴趣、养成科学态度和正确的学习方法等。这就是说，不仅要帮助学生达到认知领域的教育目标，还要达到情感领域的教育目标。

由于情感和认知是两个不同的目标领域，情感领域的化学教育目标从内容、分类、教学方式及其测量和评价方法上都有其自身的特点，需要认真加以研究。本章讨论情感领域化学教育目标的具体内容、学习水平分类以及该领域目标测评的方法技术特点。

第一节情感领域的化学教育目标

一 重要意义

(1)情感领域化学教育目标作为化学教育目标系统的重要组成部分,是化学教学大纲对化学教学中德育教育要求的具体化,是学生全面、和谐发展所必不可少的素质要求,而不是什么可有可无的东西。按照情感领域化学教育目标的要求,必须防止那种“只管教书,不问育人”的错误倾向。

(2)情感领域化学教育目标的实现直接影响认知和实验技能领域化学教育目标的实现。这主要表现在:情感领域目标对认识活动具有“定向”作用。如果学生对化学有着浓厚的兴趣,就会使他的注意力转向化学课程及其他有关的活动;情感领域目标对认识活动的动力作用。学生对化学的兴趣和爱好,能够引发和维持化学学习动机;在化学学习遇到障碍或阻力时,学生已有的意志品质和坚定信念可以支持学生克服困难、排除阻力。

(3)在化学教育活动中,以情感领域目标和认知及实验技能目标为共同指导,把情感教育和传授知识、培养能力结合起来,不仅能够保证化学教育活动的正确方向,而且还可以使之生动、活泼,提高教学效率和教学质量。

二 具体内容

由于目前对情感领域本身还没有一个公认和一致明确界定,用“兴趣”、“态度”、“欣赏”、“价值观”、“情绪”、“意向”或“倾向”等术语所表达的内容相当空泛。对情感领域化学教育目标的具体内容也是看法不一。有人从情感和意志两个方面制订了11个具体的情感领域目标。

我们认为,按照情感领域的一般意义和化学教育的具体特点,可将情感领域化学教育目标的具体内容分为以下3个部分:

1.辩证唯物主义观点与科学方法论

物质观;

运动观;

对立统一观;

质量互变观;

内因与外因——宏观化学现象与微观化学本质的关系、物质的组成、结构、性质、制法、用途之间的关系;

共性和个性、个别与一般——典型元素与元素族的关系;

分析与综合方法;

归纳与演绎方法;

实验方法;

抽象与概括方法。

2.价值观念

爱祖国、爱社会主义;

爱科学、欣赏化学中的美;

了解化学在社会生产、生活中的作用;

对化学学习的兴趣。

3.科学态度

实事求是；
理论联系实际；
坚韧不拔的学习意志；
勇于探索的创新精神；
以化学实验为基础。

上述三类情感领域化学教育目标各有其不同特点。辩证唯物主义观点与科学方法论目标主要与学生的认识有关，但又不同于认知领域目标中对具体化学知识的认识，而是更高层次上对事物的一般规律的认识，即哲学认识；这种认识对形成世界观和方法论直接相关，因而将这些目标区别于认知领域而划归为情感领域可能更合适些。价值观念目标主要与学生在化学学习中的情感体验有关，学生对化学学习的兴趣实际上是对化学学科的价值认识及学习过程中好奇心的满足为基础的。科学态度目标则与学生的态度、意志等个性倾向有关。

三 学习水平分类

克拉斯沃尔和布卢姆等人曾于 1964 年发表了情感领域的教育目标分类。与认知领域类似，他们也想寻找到一个“连续体”，使情感领域教育目标的分类也具有连续性、累积性和层次性等科学分类的性质。他们把这一连续体描述成一个内化过程，情感成分从单纯觉察开始，经过具有一定的动力阶段，最后达到对个人行为的控制。这一过程的 5 个基本阶段就是情感领域目标分类的 5 个层次：接受（留心）；反应；估价；组织；性格化。像认知领域的目标分类一样，每个大的层次都可以再进一步分为不同的亚层次（共 13 个亚层次）。在最低的层次上，学生只是接受某一现象，表现为从察觉、愿意接受到有选择的注意；在下一层次上，开始有情感地对现象做出反应，表现为默许、履行到满意。接着，是把行为和感受观念化，并组成一种结构。当这种结构成为一种人生观时，便达到了最高层次，成为人的个性的一部分。对这一分类体系的详细了解，还请参阅克拉斯沃尔和布卢姆等人编著的《教育目标分类学·第二分册：情感领域》。

我们认为，克拉斯沃尔和布卢姆等人提出的分类体系为情感领域化学教育目标的学习水平分类提供了很好的思路。但这一分类的较高层次并不能仅仅依靠化学教育就能达到；加上目前化学教学研究与实践水平的限制，还不能像认知领域那样划分出较多或更为细致的水平层次。由于情感领域化学教育目标的三方面内容各有其不同特点，需要对这三方面目标分别做出不同的学习水平分类。我们提出的分类体系如表 7-1 所示。

表 7-1 情感领域化学教育目标的学习水平分类

内容分类	学习水平分类
辩证唯物主义观点与科学方法论	1. 接受：对目标内容接受或注意的程度 (1) 觉察：意识到目标内容的存在 (2) 愿意接受：对目标内容有接受的意向，不是回避或拒绝 (3) 关注：在学习过程中控制注意力，使之指向目标内容，即对目标内容的选择性注意
	2. 意识：对目标内容有比较清楚的认识 (1) 归纳：从目标内容的不同表现形式中意识到其共同特征 (2) 概括：对目标内容有明确的知觉、正确的理解
	3. 反应：运用目标内容对有关事物和现象做出正确的反应 (1) 鉴别：运用目标内容对有关事物或现象做出合理的分类 (2) 分析：运用目标内容对有关事物或现象进行深入分析，得出一般性的结论
价值观	1. 接受：对目标内容接受或注意的程度 (1) 觉察：意识到目标内容的存在 (2) 愿意接受：对目标内容有接受的意向，不是回避或拒绝
	2. 体验：对目标内容产生情感体验的程度 (1) 兴趣：对目标内容产生积极而不是消极的情感体验 (2) 乐趣：对目标内容产生快乐的情感体验
	3. 追求：有意识地、主动地获取与目标内容有关的情感体验 (1) 志趣：为获取与目标内容有关的情感体验做出努力 (2) 信念：对目标内容已经确信无疑，并对它们产生了深刻的情感，形成了信念

内容分类	学习水平分类
科学态度	1. 接受：对目标内容接受或注意的程度 (1) 觉察：意识到目标内容的存在 (2) 愿意接受：对目标内容有接受的意向，不是回避或拒绝
	2. 体会：有意识地对目标内容进行实践，体会目标内容的意义和作用 (1) 理解：理解目标内容的意义，明确目标内容对学习行为的基本要求 (2) 实践：有意识地按照目标内容要求自己，初步达到目标内容对学习行为的要求
	3. 个性化：自觉按照目标内容要求自己，已经形成一种行为习惯 (1) 自觉：主动按照目标内容要求自己、自我克服偏离目标要求的错误倾向 (2) 习惯：已将目标要求内化为个性的一部分，表现为自动、稳定的行为习惯

在具体制订情感领域的化学教育目标时，可以与制订认知领域的目标一样，先确定情感目标的具体内容，然后再根据其内容分类，确定适当的学习水平。

由于目前比较缺乏关于情感领域化学教育目标研究和实践的资料，上述关于情感领域化学教育目标的内容及学习水平分类的认识都是相当粗浅的，还有待于进一步研究、深化。

第二节 情感领域化学教育目标测评

一 测评的意义与特点

1. 测评的意义

(1) 促进情感领域化学教育目标的实现。

利用测评的定向和激励功能，针对情感领域化学教育目标进行测评活动，是促进该领域目标实现的有效手段。过去相当长的时期里，化学教育实际中不重视对学生非智力因素的培养的现象，与我们缺乏有效的情感领域目标测评手段有很大关系。在高度重视学生全面发展的今天，将情感领域化学教育目标作为测评的基本对象，有利于我们更加准确地把握学生情感发展的现状，有利于化学教学要求的全面落实。

(2) 诊断学生在化学学习中存在的问题。

在学生的学习活动中，非智力因素的重要作用已被越来越多的化学教育工作者所认识。作为情感领域化学教育目标所涉及的需要、兴趣、动机、情感、意志、性格等非智力因素，虽然不直接参加对化学知识的认知过程，但他们作为学习的动力系统，却制约着学习的积极性。学生的学习成就，实际上是学生的智力因素与非智力因素相互作用的产物。因此，当学生的化学学习发生困难时，并不一定是由于学习的知识基础或能力水平等方面的原因，而有可能是非智力因素方面的原因。这就需要通过情感领域化学教育目标的测评，准确诊断学生学习困难的真正原因所在，从而对症下药地找到解决困难的有效途径。

2. 测评的特点

(1) 测评结果的变异性。

对认知领域目标，主试可将测量目标转化为学生的作业，要求学生尽量完成作业，主试可以根据学生在完成作业时的最高表现，按一定的客观标准进行评价；而对情感领域目标的测评，却不是要求学生给出最高表现，一般只是让学生做主观报告以呈现其内心世界。在不同的测评情景、不同的测评目的等条件下，学生会以不同的方式作出反应。学生很可能会推测哪些反应对自己有利而哪些反应会对自己不利，因而并不据实报告自己的真实反应，而是受取悦心理的支配。学生报告喜欢化学，可能并不是因为他真的喜欢化学，而是希望以此来取悦于化学教师。这样就会造成假象，使测评结果有很大的变异性，影响测评的信度和效度。

(2) 没有确定的测评标准。

对认知领域目标的测评，学生反应的对或错、好或差都可以有确定的评价标准；而对情感领域目标的测评，却很难有确定的评价标准。学生在兴趣、态度、价值观等方面差异是其个性特征的表现，与学生的个性发展特点有关，因而是因人而异的，确实没有一个收敛的评价标准。例如，我们不能简单地规定，喜欢化学就好，而不喜欢化学就不好；即使我们做出这样的硬性规定，只能强化学生的取悦心理，给出虚假反应，同样不能得到可靠的测评结果。

(3) 测评目标不易明确界定。

目前，情感领域的化学教育目标本身还未能像认知领域目标那样比较明确、具体，各个不同目标之间的区别与联系又比较复杂。这就给情感领域目标测试工具的设计带来了困难，测试项目与测试目标之间难以对应，测试结果的计分、解释也不能像认知领域测量那样明确、具体。

(4)目标本身的缓慢实现。

认知领域化学教育目标可以在较短时期内实现，如学生对化学知识的记忆、理解、应用可以通过某一教学阶段即可习得，即使某些比较复杂的能力，也可以在一学期或一学年的教学过程之后获得。这样，学生在知识、能力方面的行为变化就可以通过测评而表现出来，以作为我们评价学生学业成就的依据。但是，达到情感领域的化学教育目标就可能比较缓慢，因而，学生在这一方面的变化可能要经过相当长的时间才能表现在测评结果中。

(5)对个人隐私的尊重。

学生的化学学习成就、化学知识与能力水平是可以人所共知的。学生也愿意通过化学考试、竞赛等方式表现自己的知识能力水平，并争取获得奖励。但是，学生常常不愿意被要求表达他在情感领域到达目标的情况。一个人的兴趣、态度、价值观等往往被认为是个人的私事，这种私事一般是受到尊重的。除了个人愿意，其个性特征不可能受到别人的仔细审查。对个人隐私的尊重，是情感领域目标测评应当高度重视的因素。

根据上述对情感领域化学教育目标测评特点的分析，我们认为，目前在进行情感领域测评时，要把注意力放在为化学教学过程提供有效信息上，不要过多地涉及评价；要以学生集体为测评对象，一般不去涉及学生个人。尽可能消除学生接受测评时的思想顾虑，使他们给出真实的反应。

二 测评的问卷调查法

1. 问卷调查法的长处和可能存在的不足

问卷调查法是情感领域化学教育目标测评常用的一种方法。这种方法是由测评者根据测评目的设计出问卷或“量表”，要求学生按照问卷所提出的问题，提供有关自己的行为、观点、看法等情况的报告，作出相应的回答，从而获得有关情感目标发展的信息。类似于心理测量中的自陈量表法。问卷的内容一般是那些只有学生自己了解得最清楚的个体素质和别人难以观察到的那些以个体意识倾向及个性心理特征为基础的各种行为。

与谈话法、观察法等其他测评方法相比，问卷调查法的长处在于：可以节约时间，能够在短时间内调查到大量的学生，且可以涉及到大量的、内容广泛的测评目标；调查资料比较集中，相对便于统计、分析和整理；简便易行。

问卷调查法可能存在的不足是：所提出问题可能被学生误解，问题稍有含糊，就得不到所需要的回答，即使问题明确，学生还容易产生误会；由于情感目标一般比较复杂，不是几个字或几句话就能提供完整的资料；学生可能不严肃、不认真地回答问题，不露真情、随便填写，甚至捏造事实、弄虚作假，而回答是否真实可靠又往往无法证实；发出的问卷有时不能全部收回，容易影响答案的代表性等等。

2. 调查问卷的设计要点

(1)要明确调查目的，严格限制调查内容范围，不可滥用。调查应限于具体事实和个人反应两个方面，并分清事实和意见。

(2)要通过问卷的指导语和其他方法，使学生了解调查的意义和作用，并采取积极合作的态度，实事求是地回答问题；要能够和学生建立起彼此信赖的关系。一般情况下，不要求学生写出姓名。

(3)提问的内容和方式要尽可能简单，便于学生回答。难答的质问要尽量

避免，学生多半不愿意回答的问题不要问；不要问那些涉及到许多材料的问题，避免冗长的文字答案；问题不宜太多，可问可不问的问题尽量不问。一般不直接提问，而是提出几个反映主题的间接问题，由此推测学生的反应。

(4)慎重使用提问的用语。不能有提醒、暗示学生回答意向的语句；避免由于不恰当的提问而对学生产生刺激、引起学生的某种强烈情绪或出现某些不必要的联想。在涉及学生本人情况时，提出问题不要有明显的褒贬色彩。

(5)在问卷中适当设置检查测验有效性的项目。例如，对同一类问题以两种不同的方式提问，看学生在两个问题上的答案是否一致，以此来检验学生是否认真作答；或是给出明显不合理的选择答案，若有学生选择这样的答案，即指示该生未能认真回答问卷。

(6)可采用强迫性选择技术，即给出的问题答案都是学生所希望的，或都是学生所不愿意的，但要求学生必须通过比较从中挑选出最合适的答案。

(7)问卷拟定好之后，要与了解学生情况的、有经验的教师共同检测问卷，可能时也要进行预测，发现问题及时修改。

3. 调查问卷常用的问题形式

(1)问答式。

又称自由记述式。提出开放性问题，让学生自由给出答案。

例 1 你喜欢学习化学吗？为什么？

(2)选择式。

提出问题并列几种答案，让学生选择其中一种或几种作为答案。

例 2 我喜欢化学或促使我学化学的原因是：

- A. 化学比较容易学
- B. 化学实验有趣
- C. 化学与实际生活联系紧密
- D. 化学是一门必修课，我必须及格才行
- E. 我以后要从事有关化学方面的研究工作
- F. 化学老师讲课生动有趣
- G. 我的其他看法

(3)填空式。

提出问题，让学生给出简短答案。

例 3 请你估计一下自己的学习成绩（是全班前 10 名、前 20 名、前 30 名或 30 名以后）：

我的各科平均成绩在班里属于_____；

我的化学成绩在班里属于_____。

(4)是非式。

也称为二分质问法。提出问题或给出一句叙述，让学生做出肯定或否定回答。

例 4 对下列问题，合乎你的情形时在“是”字上打“√”；不合乎时在“否”字上打“×”。

- | | | |
|-----------------|---|---|
| 我上化学课时，常常精神不集中。 | 是 | 否 |
| 我很讨厌化学实验。 | 是 | 否 |
| 你喜欢自己动手做实验吗？ | 是 | 否 |
| 化学作业是不是留得太多了？ | 是 | 否 |

……

(5)排序式。

提出问题，列出几种答案，让学生按自己的认识把它们排顺序。

例 5 你在每次上化学课时，是一种什么样的心情？在下列的几种答案中，按照你自己的情况把它们排出顺序。你认为最主要的就在答案前的（ ）中填上 1，其次比较主要的就填 2，照此类推。

- () 不能逃课
- () 喜欢听老师讲化学知识
- () 喜欢看老师做化学实验
- () 没有什么想法
- () 要应付考试

.....

(6)对应比较式。

提出问题，列出一对相关的答案，让学生通过比较确定其中之一为适合自己的答案。

例 6 你学习化学的目的是什么？请你从下列每一对关于学习化学的目的叙述中，挑选其中比较适合于你的画上 。

- ①—
 - A. 为了考上大学
 - B. 为了将来从事化学事业
- ②—
 - A. 为了学到今后有很多用处的化学知识
 - B. 为了获得比较好的化学考试成绩

.....

(7)评定量表式。

提出问题，或给出叙述，然后列出一组答案，并给答案分出等级或赋予分数构成量表，让学生根据自己的看法选择其中的答案，根据答案即可折算出等级或分数，最后可以累积每一问题的得分以说明结果。

例 7 请你仔细阅读下面的句子，并在最能表达你意愿的字母上画一个“ ”。每个字母所表示的意义分别为：A—很同意；B—同意；C—说不清；D—不同意；E—很不同意。

在化学上，我下了比其他学科更多的功夫。ABCDE
一翻开化学书我就犯困。

A B C D E

我认为化学知识在实际生活中应用广泛。

A B C D E

我喜欢自己动手做实验。

A B C D E

.....

三 测评的其他方法

1. 观察法

学生在情感领域的发展水平，总是要表现在外部行为、习惯上。因而可用行为观察法来测评学生的情感领域化学教育目标的发展水平。这种方法适用的测量目标广，运用的机会多，也比较容易为广大化学教育工作者所掌握。它可以随时用来检核学生的行为，可以在不影响正常教育、教学活动的情况下进行。学生的化学学习兴趣、科学态度等都可以用观察法来测评。在用观察法作为情感领域化学教育目标的测评工具时，要对观察过程进行科学的控

制，注意做到以下几点：观察要有目的、有计划，明确观察的内容和重点；观察者要有客观的态度，不能“先入为主”；观察要深入、透彻，系统连贯；观察者要掌握观察对象——学生的基本情况和具备有关学生心理发展、变化的知识；观察要按统一的标准及时记录；观察结果要做出合理的整理、分析和评价。

2. 谈话法

谈话法和问卷法一样也是通过问题来探测学生的化学学习兴趣、观点、科学态度和其他内心活动，但不是用纸笔测验，而是通过口头交谈进行的。其优点在于：比较有人情味。在谈话中可以逐步打消学生的思想顾虑，待进入比较和谐、彼此信任的状态时，再提出问题，可增加回答的可靠性；谈话时可直接观察学生的外在行为表现，了解学生对测评的态度是积极的还是消极的，确定答案的可信程度；当学生对问题不够明确时，可以当场解释，不至于产生误解。谈话法的缺点在于：耗费时间较多，无法作大面积调查；对谈话结果的分析易受谈话者个人偏见的影响。

为了保证谈话法能作为情感领域化学教育目标测评的有效方法，需要达到以下基本要求：谈话要有准备，要明确谈话的目的，准备好交谈的问题及提问的顺序，了解接受谈话学生的基本情况，对谈话技巧作一定的演习等；

注意选好谈话的时间、地点，尽量创造自然、和谐的谈话气氛，使谈话对象无拘无束，畅所欲言；提出的问题要避免对谈话对象产生刺激，问题应是谈话对象能够回答的，谈话者所用的语言也是谈话对象易懂的；要注意避免以主观好恶影响谈话对象的态度，不要提供暗示，不要催促回答；谈话时一般不作记录，谈话后要及时回忆、准确记录，谈话对象的谈话内容和态度要分开记录；对谈话结果的分析，要以有关理论为指导，从谈话的实际出发，防止主观臆断。

3. 个案研究法

由于情感领域化学教育目标测评涉及的因素较多，我们有时很难对大样本的学生都实施测评；在这种情况下，我们可以选择一些有代表性的、少量的学生样本，作为研究对象，通过对这些少量学生进行较长时间的、比较全面的观察、调查及其他研究手段，收集他们在情感领域化学教育目标方面发展、变化的尽可能多的信息，以此作为推测大样本学生在情感领域发展、变化的依据。这就是个案研究法。被选择出的少量学生样本就是所谓个案。

个案研究法的优点是具有连续性、典型性和真实性。但常因研究时间长，受客观条件变化的影响较大，不易坚持到底。使用这一方法时要注意对样本的行为变化随时记录和跟踪记录，不能随意取舍研究信息。

可用于情感领域化学教育目标测评的方法还有投射法、轶事记录法等，投射法可看作调查问卷中的特殊问题或问题情景；轶事记录法可看作个案研究中的方法之一。因而对它们不再作专门论述。

第三节学生学习化学的兴趣和态度的问卷调查与研究

在情感领域化学教育目标测评活动中，进行得较多的是关于学生的化学学习兴趣与态度的调查。这是因为，学生化学学习兴趣与态度对学生掌握化学知识及培养能力有着比较直接和明显的影响。一般用于测评学生化学学习兴趣与态度的方法是设计调查问卷，就学生有关的行为表现编制出各种问题，让学生对这些问题做出反应，然后统计学生的各种反应的比例，作为分析、得出结论的依据。

我们曾就学生的化学学习兴趣 and 态度问题做了一些初步的调查和研究。1992 年春夏，我们用自己设计的“化学学习兴趣与态度量表”对北京市 5 所中学、11 个教学班的 400 余名学生做了调查，并统计、分析了调查结果，得到了一些有意义的结论。现将本次研究的过程及结果简要报告如下。

一 研究的目的是调查量表的设计

本研究的主要目的是：探索以初步定量的方法描述学生化学学习兴趣 and 态度状况的可能性，检验所设计的“中学生化学学习兴趣与态度调查量表”对学生化学学习兴趣 and 态度测评的有效性。

本量表以 20 个评定式问题为主，设定每个问题的每种答案的分数，累积学生在这 20 个问题上的总得分，作为学生化学学习兴趣与态度状况的表征。在这 20 道题中，有 10 道题是“肯定式”的叙述，另 10 道题是“否定式”叙述，分别代表学生化学学习兴趣与态度的两种指向；每道题后的 5 个备选答案，分别是对每个叙述的肯定或否定的程度，以测量学生的化学学习兴趣与态度指向的强度。此外，还有 4 道选择式问题、3 道填空式问题和 1 道问答题，其设计目的是进一步搜集有关信息，检验前 20 个问题的有效性。

该量表全文如下：

同学们：

你们好！为了了解当前广大同学对化学学习的兴趣和看法，发现目前教学中存在的问题，以便教师改进教学，提高同学们对化学学习的兴趣，使同学们既愉快又有效地掌握化学知识，我们设计这份问卷。恳切地希望同学们给予支持和合作。

我们将保证调查结果不对各位同学产生任何不利影响。本问卷不要求你写上姓名。希望同学们能认真、如实地进行回答。

谢谢你的合作！

一、请你仔细阅读下面的句子，并在最能表达你意愿的字母上画一个“ ”。每个字母所表示的意义分别为：A—很同意；B—同意；C—说不清；D—不同意；E—很不同意。

1. 我上化学课时，常常会精神不集中。

A B C D E

2. 我很讨厌化学实验。

A B C D E

3. 在化学上，我下了比其他学科更多的功夫。

A B C D E

4. 一翻开化学书我就犯困。
A B C D E
5. 我认为化学知识在实际生活中应用广泛。
A B C D E
6. 我喜欢自己动手做实验。
A B C D E
7. 对老师的课堂提问，我能积极思考作答。
A B C D E
8. 反正高考不考化学，学了化学也没有多大用途。
A B C D E
9. 课堂有化学演示实验会提高我的听课兴趣。
A B C D E
10. 有空闲时间时，我会找一些有关化学知识的文章来阅读。
A B C D E
11. 化学作业留得太多了。
A B C D E
12. 化学理论概念太抽象了。
A B C D E
13. 化学元素化合物知识有些杂乱无章。
A B C D E
14. 我每次都能很认真地完成化学作业。
A B C D E
15. 化学越学越没劲。
A B C D E
16. 化学不容易学好。
A B C D E
17. 化学比其他科目更吸引人。
A B C D E
18. 化学知识与物理、数学一样，是有规律可循、有条理的。
A B C D E
19. 要学好化学，主要是靠死记硬背。
A B C D E
20. 化学学得越深越有趣。
A B C D E

二、选择题下面两道选择题，请你选择一个答案。在所选答案前打 。

1. 你喜欢化学吗？
A. 很喜欢 B. 比较喜欢
C. 一般 D. 不太喜欢
E. 很不喜欢
2. 你对化学有兴趣吗？
A. 一直有兴趣 B. 开始有兴趣，现在没兴趣了
C. 一般 D. 开始没兴趣，现在有兴趣了
E. 一直没兴趣

下面两道题，选择的答案可以不止一个，如果你觉得给出的答案不够

完全，就把你的看法写下来。

3. 我喜欢化学或促使我学化学的原因是：

- A. 化学比较容易学
- B. 化学实验有趣
- C. 化学与实际生活联系紧密
- D. 化学是一门必修课，我必须及格才行
- E. 我以后要从事有关化学方面的研究工作
- F. 化学老师讲课生动有趣
- G. 我的其他看法

4. 我不喜欢化学或促使我不学化学的原因是：

- A. 成绩不好，努力后也没多大的提高
- B. 学习负担太重，各门功课压得我无力对化学增添兴趣
- C. 老师讲课不生动
- D. 我参加高考时不准备考化学
- E. 据说化学药品对身体有害
- F. 化学学习太枯燥
- G. 我的其他看法

三、请你列出你最喜欢的 3 门课和最不喜欢的 2 门课的名称。

最喜欢的课

第一	第二	第三

最不喜欢的课

第一	第二

四、填空

1. 你的性别_____。

2. 请你估计一下自己的学习成绩（是全班前 10 名、前 20 名、前 30 名或 30 名以后）：

我的各科平均成绩在班里属于_____；

我的化学成绩在班里属于_____。

五、如果你愿意，请回答下面这道题：

你希望老师怎样上化学课？

二 调查结果的分析与讨论

1. 调查对象

本研究共收回有效问卷 382 份。样本分布如下表所示。

表 7-2 化学学习兴趣与态度调查对象分布

教学班代号	给出有效问卷数	所属年级	所属学校
B11	41	高一	市重点 1
B21	40	高二	市重点 1
S21	46	高二	市重点 2
S32	48	初三	市重点 2
D11	45	高一	区重点 1
D21	41	高二	区重点 1
Y31	38	初三	一般校 1
Y16	41	高一	一般校 1
H14	42	高一	一般校 2

2. 调查结果的统计方法

由于我们的研究重点是探索所设计的测量量表对学生化学学习兴趣和态度进行定量化描述的可行性和有效性,所以我们将第一大题的 20 个项目分别赋值,并作为统计的重点。对“肯定式叙述”的 10 道题(3, 5, 6, 7, 9, 10, 14, 17, 18, 20)的 5 种备选答案 A, B, C, D, E 分别赋值 5, 4, 3, 2, 1;对“否定式叙述”的 10 道题(1, 2, 4, 8, 11, 12, 13, 15, 16, 19)的 5 种备选答案 A, B, C, D, E 分别赋值 1, 2, 3, 4, 5。如果在某道题上学生未选答案,则视为其选择“C(说不清)”,计为 3 分。这样,学生在每道题上都有一个分数,累计学生在 20 道题上的全部得分,作为本量表的总分。不难看出,该量表的满分为 100 分;当学生的得分大于 60 分时,表示该生对化学学习的兴趣与态度是肯定的,分数越高,其肯定的程度越大;当学生的得分低于 40 分时,表示该生的化学学习兴趣与态度是否定的,分数越低,否定的程度越大;当学生的分数介于 40—60 之间时,表示该生对化学学习的兴趣和态度没有明确指向。

我们利用电子计算机做了以下各项统计工作:

- (1)分班、分学校、分年级统计量表分数、分数分布、平均分、标准差;
- (2)总体 382 个样本的量表分数、分数分布、平均分、标准差;
- (3)分组和总体样本在各个计分项目的得分率、各备选答案的选择率;
- (4)分组和总体样本在非计分项目上的主要答题情况;
- (5)总体样本的计分项目得分与量表总分的相关(按积差相关计算)。

限于篇幅,这里只给出与结论有关的一些数据(见表 7-3),其他统计结果略去。

3. 结果分析与讨论

(1)量表分数表明学生化学学习兴趣与态度的一般状况。

图 7-1 化学学习兴趣与态度调查总体分数分布图

分组量表分数的平均分分布在 66.15—73.56 之间,标准差分布在 7.87—10.18 之间,表明来源不同的样本组平均分与标准差的差异并不十分显著。这一结论不难理解。尽管不同的学校在师资、生源、教学设备与条件等方面存在明显差异,这些差异可能会给学生在认知领域化学教育目标的发展方面带来显著影响,但却不一定会对学生化学学习兴趣与态度的发展带来决

定性的影响。

从总体上看，学生对化学学习的兴趣和态度基本上是积极的而不是消极的（总体平均分=70.34 > 60），但这种积极的程度并不很高。

总体量表分数分布接近正态，表明学生在化学学习兴趣与态度方面的行为发展是不平衡的，接近于在知识、能力等认知领域的发展状况。我们认为，量表所得出的这一结论是可以接受的；量表能够将这一现象揭示出来，很值得引起注意。

(2)得分率较高的项目。

总体样本平均得分大于4.0的项目有4个，即项目2, 6, 8和9。其平均得分分别为4.36, 4.16, 4.12, 4.32。

项目2, 6, 9都是关于化学实验的，反映出学生喜欢化学实验(项目2)、喜欢动手做实验(项目6)、演示实验能够提高学生的化学学习兴趣(项目9)。这一结果再次证明，化学实验是提高学生化学学习兴趣的有力手段。统计结果还表明，高年级学生偏爱课堂演示实验，说明他们已经意识到演示实验在认识化学现象及其本质方面的重要作用；而低年级学生更偏爱亲自动手做实验，表明他们对化学实验的直接兴趣还很强。

项目8是关于化学学习目的的问题。学生在该项目得分较高，说明学生已经能够认识到化学学习的更广泛的意义，并不因为高考不考化学就不去学它。在我国正在进行考试制度改革、某些高考方案中可能不设化学考试科目的今天，学生能不为考试而认真学习化学，是有重要意义的。

(3)得分率较低的项目。

总体样本平均得分低于3.0的项目也有4个，即项目3, 10, 16, 17。其平均得分分别为2.64, 2.68, 2.81, 2.95。

学生对项目3, 10, 17的反应都说明，在当前中学生的课业负担较重的情况下，学生一般都不会给化学以特殊的注意，他们不得不在中学的各个必修课程上平均使用力量，不能偏科。在项目17上，选“说不清”的学生多达46%。这可能是学生的化学学习积极性趋中而并不表现出特别强的主要原因。

学生对项目16的反应，与其所在的学校和年级有关。认为化学不容易学好的倾向，一般学校比重点中学更明显，低年级学生比高年级学生更明显。可见，学习条件及学习结果必然要影响学生对学科难度的认识。

(4)非计分项目的回答情况。

非计分项目的设计目的是为了验证计分项目测量结果的可靠性。在回答“你喜欢化学吗？”的问题时，选“比较喜欢”的有167人，占43.72%，选“一般”的有126人，占32.98%；在回答“喜欢化学的原因”时，选“化学实验有趣”的有188人，占49.21%。这些结果，与计分项目的测量结果是一致的。

在回答“不喜欢化学的原因”时，有211人作答，选B, C, D, E, F, G的都有，其中选C(“老师讲课不生动”)的有91人，占答题人数的43.12%。可见，喜欢化学的原因是比较一致的，而不喜欢化学的原因却是因人而异，而教师又是其中的重要影响因素。

(5)项目与总分的相关情况。

20个计分项目与量表分数都呈现较高相关。见表7-4。

表 7 - 4 计分项目得分与测验总分的积差相亲值

项目序号	1	2	3	4	5	6	7	8	9	10
相关系数	.559	.494	.284	.587	.436	.312	.539	.566	.227	.644

项目序号	11	12	13	14	15	16	17	18	19	20
相关系数	.332	.425	.487	.431	.637	.390	.581	.496	.438	.480

绝大多数项目得分与总分相关达到了显著性水平，说明学生在项目上的反应与其量表分数之间存在着较好的一致性。

根据上述分析，我们认为，中学生化学学习兴趣与态度量表作为情感领域化学教育目标测评的工具，所测出的关于学生化学学习兴趣与态度的状况是比较符合实际的，因而可以认为该量表是比较有效的。

第八章 化学实验教学目标及测评

化学实验是化学教学内容的重要组成部分。以实验为基础是化学教学的基本特征之一。通过实验教学，不仅可以培养学生的化学实验技能，还可以为学生形成化学概念和理论提供丰富的感性材料，有助于培养学生的观察能力和思维能力、严谨的科学态度和对学生进行科学方法训练、辩证唯物主义教育等。

但在化学教育实际中，化学实验教学常常未能得到应有的重视，其作用也未能得到充分发挥。造成这一状况的原因是多方面的，而缺乏明确、具体的化学实验教学目标及相应的测量与评价手段，则是其中的重要因素之一。

本章着重讨论化学实验教学目标明确化、具体化的方式和化学实验教学目标测评的方法与技术。

第一节 化学实验教学目标

化学实验教学目标是以为化学实验教学为主要手段而使学生产生的教学大纲所要求的行为变化。化学实验教学目标与认知领域及情感领域化学教育目标的主要区别在于：在目标的内容上，主要涉及要求学生掌握的化学实验技能，在相当大的程度上是与学生的外部行为表现相联系的，而不是像认知领域和情感领域目标那样，主要与学生的内部行为相联系；在实现目标的过程中，围绕化学实验而组织的教学活动是不可缺少的条件。教师的演示实验、学生实验等是实现目标的主要手段。

根据化学实验教学目标的特征，培养学生的化学实验技能是化学实验教学目标的主要内容。我们认为，学生的化学实验技能不仅包括化学实验操作技能，还要包括实验观察能力。

一 实验观察能力的培养

1. 明确实验观察能力培养目标的意义

观察能力是指进行有意识的、有计划的、持久的知觉活动的的能力。实验观察能力是指能够观察化学实验现象（通过必要的变革和可控制的现象），排除各种偶然、次要因素的干扰，把握住研究对象的本质的能力。在化学教学中，学生的实验观察能力的培养显然主要与化学实验相联系，学生的外在行为表现为“眼动”。

观察能力是化学教学大纲规定的要培养学生具备的四种能力之一，因而是化学教育目标的基本内容之一。我们认为，通过化学实验教学培养学生的观察能力，既是结合化学学科的教学特点培养学生观察能力的主要而有效的途径，又是培养学生的化学实验操作技能和进行化学学习的基本条件。我们将实验观察能力的培养作为化学教学中观察能力培养的主要内容，可以突出化学学科教学的特点；而将实验观察能力与实验操作能力区别开，并使之相对独立出来，有利于进一步引起对观察能力培养的重视。我们相信，明确地将实验观察能力的培养列为化学实验教学目标的基本内容之一，有利于大纲中观察能力培养要求的进一步落实，也有利于化学实验教学质量的提高。

2. 实验观察能力培养的基本内容

培养学生的实验观察能力，要包括对下列观察能力的基本品质的培养：

(1) 观察的目的性。

明确观察的对象、条件和要求。即做好观察前的准备，掌握观察的计划、步骤。比如能够明确实验所要解决的问题，是探索性实验，还是验证性实验；知道先看什么，后看什么等。无明确目的的观察，只能是一般的感知活动而不是真正的观察。

(2) 观察的条理性。

观察时能够遵循合理的顺序和系统步骤。比如，按照时间的先后顺序或按照位置的前后顺序、远近顺序等进行观察，观察活动应有条理，而不是杂乱无章。

(3) 观察的理解性。

在观察过程中，能够积极开动脑筋，使头脑中的认知结构积极影响观察，以理解观察对象，使观察更为深刻、全面。即俗话说：“会看的看门道，不会看的看热闹”。

(4)观察的敏锐性。

一方面能够对所观察到的现象迅速地做出反应而不迟钝；另一方面又善于发现那些容易被忽视或在其他较强刺激掩盖下较弱的但又是很重要的实验现象。

(5)观察的持久性。

在整个观察过程中，能够自始至终保持高度的、同等的注意力，而不是在实验开始时由于新奇的刺激而高度注意，随后却慢慢地失去耐心，不认真观察。

3. 实验观察能力培养目标的制订

上述实验观察能力的五种品质，是实验观察能力培养的基本内容。各个品质之间不存在水平高低，而是有一定的相对独立性。但学生在形成这些品质的过程中，也会表现出一定的差异，这就使我们有可能根据学生在形成观察能力的不同品质时表现出的水平差异，像认知领域的目标分类那样，来制定实验观察能力基本品质的学习水平分类体系。不过，目前这方面的研究刚刚开始，我们还不能对每种品质的不同学习水平进行详尽的描述，但我们至少可以对其做出粗略的划分。例如，可将观察的目的性划分为3个学习水平，暂时分别称为1级、2级和3级。3种水平的意义粗略界定如表8-1。

表 8-1 “观察的目的性”三种学习水平的意义

学习水平	意 义
1 级	知道观察的目的，但比较抽象而不具体；不明确观察的重点与步骤
2 级	基本明确观察的目的，知道观察的重点是什么，但还没有观察的具体方案
3 级	明确观察的目的，观察之前有较详细的观察计划，拟定的观察步骤正确、合理

观察能力的其他品质也可以作类似的界定。随着我们对化学实验观察能力测评实践的发展，逐步掌握大量的学生在有关方面差异表现的资料，就可以对各种品质的学习水平做出更加清晰、合理的界定与说明。

在具体制定学生的实验观察能力培养目标时，首先根据实验活动的具体内容，确定要培养的观察能力的基本品质的内容，然后依据教学计划的总体安排和学生的具体情况，确定能够达到的适当学习水平。请看表8-2所列示例。

表 8-2 观察能力培养目标——“氧气的实验室制法”

实验活动内容	观察能力培养内容	学习水平
氧气的实验室制法	观察的条理性	2 级

二 实验操作技能培养

1. 化学实验操作技能的学习水平分类

(1)国外动作技能领域教育目标分类学理论。

动作技能领域是教育目标分类学理论的三大领域之一。相对于认知领域和情感领域而言，动作技能领域未能引起人们的足够重视。布卢姆和克拉斯沃尔等人于1956年、1965年分别提出了认知领域及情感领域教育目标分类学理论之后，并未能继续进行动作技能领域的目标分类。这一工作是后来由美国的哈罗和辛普森(E. J. Simpson)各自完成的。

哈罗和辛普森提出的分类学理论的基本框架与认知领域及情感领域相

同，所不同的是分类的具体内容。

哈罗提出的分类学理论旨在帮助教育工作者和课程编制者澄清有关儿童的动作经验，与儿童的生长和发展关系密切，主要是为学龄前教育所用的。该分类模式共有 6 大层次、20 个亚层次。这 6 大层次分别是：

- 反射动作；
- 基础性的基本动作；
- 知觉能力；
- 体能；
- 技巧动作；
- 有意沟通。

辛普森提出的分类学理论主要是为从事职业技术教育领域的教育工作者提出的，因而它更多地涉及职业技术工作所需要的能力和技能。该分类模式共有 7 大层次、10 个亚层次，第一大层次的第一个亚层次还进一步划分了 6 个细目。这 7 大层次分别是：

- 知觉；
- 定势；
- 指导下的反应；
- 机制；
- 复杂的外显反应；
- 适应；
- 创作。

对哈罗和辛普森分类理论的进一步了解，可参阅《教育目标分类学·第三分册：动作技能领域》。

(2) 化学实验操作技能的学习水平分类。

国内一些化学教育工作者已经对化学实验操作技能的学习水平分类提出了多种不同的看法，例如：

- 初步学会、学会、熟练__；
- 重复模仿、技能、技巧、设计__；
- 见识、学会、技能、技巧、设计；
- 知觉、模仿、初步学会、学会、创新；
- 知识、模仿、熟练、应用、创新；
- 观察、准备、模仿、表现、熟练、创造；
- 见识、模仿、学会、熟练、技巧、创造。

上述各种分类的理论框架是基本相同的，都是按照布卢姆等人的教育目标分类学理论所提供的思路设计的，与辛普森提出的分类体系相当接近；上述分类的区别在于：要不要有以认知为基础的“知觉”、“知识”、“见识”等层次？有的虽然也列出了“知觉”层次，但又明确指出它不属于测量目标；可以划分为几个层次？与此紧密联系的问题是，最高的水平层次是

上海市化学化工学会：化学教学，1988 年第 3 期。

耿茜：初中化学实验教学目标的理论与实践（北京师大硕士研究生论文），1990 年。

侯文光：教育测量与教学评价，明天出版社（济南）1991 年版，第 97 页。

欧阳钟仁：现代启发教学法，台湾版。

江苏省教育委员会教研室：初中化学教学目标实施手册，第 2 次修订本，1989 年，第 7 页。

什么？要不要有“创新”、“创造”等层次？

我们的观点是：在化学操作技能领域，还是不涉及那些认知领域的内容或那些以认知领域内容为基础的目标为好。这样可以明确划分不同领域的测评范围，避免重复。由于技能领域目标的测评比认知领域目标测评更困难一些，因而凡是那些能够用纸笔测量的内容，如实验原理、对实验仪器的名称、用途、性能及使用注意事项的认识等，一般不再出现在操作技能领域内，以提高测评的效率；就目前中学化学教学的实际状况和大纲的具体要求看，水平分类的层次还不宜太多，水平层次过高，可能会脱离中学的教学实际，因而还是不出现“创新”之类的水平层次为好。现将我们提出的化学实验操作技能的学习水平分类列在表 8-3。

表 8-3 化学实验操作技能的学习水平分类

分类层次及名称	意义
模 仿	能够重复教师的实验操作；或在教师的指导下完成实验操作
初步学会	能够独立完成实验操作，但还不够熟练，表现在操作的规范化和完成操作所需要的时间上都有明显的可改进之处
熟 练	能够独立地按照正确的实验操作方法和步骤，迅速完成操作，并表现出一定的操作技巧
设 计	能够将已经掌握的实验原理知识和操作技能“迁移”到新的问题情景中，根据实验的目的和要求，独立设计实验的步骤和方法，并正确地完成实验，得到合理的实验结果

2. 化学实验操作技能培养目标的制订

(1) 实验操作技能的内容。

实验操作技能的内容大体上可以分为两个部分：一是实验仪器的使用技能，如“酒精灯的使用”、“量筒的使用”等；二是实验操作技能，如“固体药品的使用”、“加热”等。每一部分都含有许多与具体的教学过程相联系的具体的实验项目。

(2) 实验操作技能学习水平的确定。

制定实验操作技能培养目标，就是确定上述各个具体的实验项目的学习水平。在确定实验项目的学习水平时，一般需要考虑到以下因素：教学大纲的具体要求；学生未来升学或就业的需要；在后续课中出现的频率；实验教学的具体条件。

(3) 实验操作技能培养目标的序列。

与认知领域目标系统一样，实验技能领域目标也有一个目标出现的前后顺序问题。从理论上说，实验技能的发展顺序应该是与认知、情感领域的发展顺序有所区别的，实验技能培养目标要有自身独立的体系和结构。国内中学化学统编教材一直没有将实验技能培养序列独立出来，而是服从于化学知识结构的需要，结果造成对初中学生的实验技能水平的要求反而比对高中学生的要求高这样一种不尽合理的情况。化学实验技能的发展序列与化学认知水平的发展序列的矛盾仍然是目前化学课程编制需要研究解决的重要问题之一。目前，只能按照课程中化学知识教学的序列和要求安排实验技能培养目标的序列了。

表 8-4 和表 8-5 给出了化学实验技能培养目标的示例。

表 8-4 化学实验技能培养目标

——实验仪器的使用（示例）

	模仿	学会	熟练	设计
量筒				
胶头滴管				
坩埚钳				
...	

EMBED Word.Document.6 \s

第二节 化学实验教学测评

化学实验教学测评是实现实验教学基本保证和重要条件。我们知道，黑板画实验不能代替演示实验，这是因为学生实验观察能力的培养、引导学生进行从宏观实验现象过渡到微观化学本质的思维训练，都离不开活生生的演示实验本身和教师对实验所作的深入分析、讲解和启发；书面考试不能代替操作考核与观察能力测评，是因为实验操作技能及实验观察能力只有通过实际操作和实际观察才能真正显示出来。学生书面考试得高分，只能说明学生知道应该如何去做，但这并不等于学生一定真的会做。实验教学过程和结果（目标到达度）都是实验教学测评的基本内容，所以化学实验教学测评要把过程测评与结果测评结合起来。过程测评着重考察教师的教学过程，包括教师的演示实验技能、引导学生观察并培养观察能力的技能以及利用实验进行认知、情感目标教学的技能，这些内容将在第十章中讨论。本节将着重讨论结果测评——对学生化学实验教学目标到达程度的测评。

一 化学实验观察能力的测评

实验观察能力测评的关键，是要能够设计出适当的观察情景使学生的实验观察能力及其差异表现出来。目前，尚未发现这方面的研究报告。我们认为，可以设计专门的观察能力探测题，即选择适当的化学实验，由教师进行演示操作，让学生独立观察并记录观察结果；根据学生提供的观察报告来评价学生的实验观察能力。实验观察能力的五种品质及其重点测评内容如表 8-6 所示。

表 8-6 实验观察能力的五种品质及其重点测评内容

观察品质	重点测评内容
观察的目的性	观察的目的、内容是否明确；能否指出观察的重点
观察的条理性	观察的前后顺序是否合理
观察的理解性	重点实验现象记录是否准确；观察过程中的“即时想法”是否深刻；对实验现象的解释是否合理
观察的敏锐性	对实验现象的记录是否全面、细致，特别是对那些“转瞬即逝”的现象能否准确记录
观察的持久性	对实验的注意力是否前后保持一致（观察的其他几种品质的表现是否随时间变化）

可以针对某一项或某几项重点测评内容，选择相应的实验，分项设计出实验的具体步骤及相应的对学生的观察要求，并确定各项的权重。然后根据学生的观察报告，按其达到观察要求的程度分项评分并相加，得出学生的测评成绩。

例 1 观察能力探测题示例。

项目名称：中和滴定与溶液的导电性

实验装置与内容：将 1mol/L 的氢氧化钡溶液 50mL 注入 250mL 的烧杯中，以碳棒为电极接通直流电源和小灯泡，形成通路。用盛有 1mol/L 硫酸溶液的滴定管滴定。

观察指导：认真观察滴定的全过程，并将所观察到的实验现象和对现象

所做的解释尽可能全面地记录到观察报告中。实验结束后10分钟内将报告交给老师。教师的演示实验操作步骤与评分标准列于表8-7。

表8-7 演示实验操作步骤与评分标准

序号	操作内容	观察到的现象	解释	评分标准
1	安好装置,接通电源	灯泡发光,亮度较强	氢氧化钡是电解质,其溶液可导电	现象与解释各1分
2	向烧杯中逐滴加入硫酸,至45mL时暂停滴定。滴定中用玻璃棒不停搅动	烧杯中有白色沉淀生成并逐渐增多,灯泡逐渐变暗	氢氧化钡和硫酸发生中和反应,溶液中自由移动的离子减少	现象3分,解释2分
3	继续滴加硫酸至50mL过半滴	灯泡由暗至完全不发光,继续有沉淀生成	滴定至终点,溶液中的离子已全部变成沉淀和难电离的水	现象3分,解释2分
4	继续滴加硫酸至70mL	灯泡又开始发光,其亮度逐渐增大	硫酸也是强电解质,其浓度逐渐加大时,溶液导电能力增大	现象2分,解释1分

操作时可同时口头说明操作内容;

每一步骤之间停顿1—2分钟,以便学生记录和解释所观察到的现象;实验结束后10分钟让学生整理观察报告。

在设计观察能力探测时要注意:内容新颖。被观察的实验最好是学生在过去未曾看过,但又是在现有的知识基础上能够理解或解释的,以避免练习效应的干扰;探测目的明确,所要考察的观察品质都能在测题中得到体现;演示操作阶段明确,各个阶段时间(包括学生记录实验现象的时间)分配合理;对演示操作各个步骤的要求明确、具体、全面,保证操作的正确、规范;选择的实验简单、易行、成功率高。

二 化学实验操作技能的测评

实验操作技能表现为学生完成化学实验的实际的、专门的操作技术水平。在测评时也需要把过程测评和结果测评结合起来,即既要考察学生的实验操作过程,又要考察学生的实验结果。操作过程包括学生实际动手完成实验的全部行为;实验结果不仅是实验产物的质量和数量,还包括学生对实验的观察记录、结果分析及得出的结论等。实验的过程和结果之间虽然经常呈现正相关,但两者相分离的情况也并不少见;忽视其中任何一方面,都可能影响测评的全面性。过程测评内容较全面,所得到的信息比较直接、准确,但需要对学生逐个观察、记录和评价,要耗费的人力、物力、时间较多;结果测评则省时省力,但有时不够准确、客观。因而,当过程测评和结果测评相结合时,既可以保证测评的全面性,又可以提高测评工作的效率。

1. 实验操作技能的过程测评

实验操作技能的过程测评,一般是先设计与测评内容有关的实验,要求学生指定的时间内完成;通过直接观察、记录学生进行实验操作的全过程,评定其操作水平。为使评定客观、准确、可靠,一般需要对被选实验进行任务分析,即确定完成实验所必需的基本操作步骤和程序,再根据测评目标及

操作步骤的重要性程度，确定每一步骤的分数权重；在此基础上，制定出实验操作技能测评量表（示例可见本章

第三节)。

在评定学生进行实验操作的质量水平时，可采用以下指标：熟练性，包括操作的协调能力；准确性，包括操作动作的规范性；应变性或敏捷性，包括对意外或突发实验现象的处理；实验速度，即完成实验的时间长短。

为了避免学生认知领域发展水平的差异影响其实验操作技能测评的成绩，可以提前公开测试题，教师帮助学生充分理解实验原理，并让学生掌握评定标准；鼓励学生提前进行自我练习、自我训练。

为了缩短测评时间，提高测评效率，可以训练学生做评定委员，帮助学生明确测评标准，掌握评定方法，并通过在正式考核之前的试评，进一步提高学生的评定水平。

提前公布测试题和训练学生当评定者不仅有利于测评工作的顺利展开，同时也可以发挥出测评的教学功能，有利于促进学生实验操作技能水平的提高。

2. 实验操作技能的结果测评

实验操作技能的结果测评，一般是要求学生在规定的时间内完成某一个或几个化学实验，但测评者并不直接观察学生的操作过程，而是只根据学生的实验结果对学生的实验操作水平做出评定。一般常用的实验有物质的制备、物质的鉴定或鉴别等。根据实验结果评定学生操作水平的指标有：实验结果的数量、质量或正确程度；完成实验所需要的时间；对实验过程（操作步骤、实验现象等）记录或描述的准确程度；对实验结果的分析 and 解释的合理程度。

由于结果测评是对学生实验操作技能的一种间接测评，因而学生给出的实验结果有可能并不反应其真实的实验技能水平。为了降低这种错误出现的可能性，使测评结果更加准确、可靠和公平合理，就需要精心选择测试题和严格控制测评过程。测试题要选择那些对学生而言是新的实验内容，或是演示实验、学生实验的变式，以避免练习效应的影响和防止学生预先知道或猜测出实验问题的答案。为了防止测评过程中学生之间的相互影响，可以设计一组测评目标相同而实验内容各异的等值复本；让学生分组抽取测试题，使实验桌相近的学生用不同的测试题进行测试。

由于影响结果测评的因素更多一些，所以常用等级评定（相对评价）的方法给出评定结果。

第三节 中学化学实验操作技能测评研究

化学实验技能测评一般都是由测评者个人对受试的实验操作过程或结果的质量及水平做出评定，这种评定方法难免会带有测评者个人的主观因素的影响。能否找到一种与评定者个人主观因素无关的、更为客观的测评标准呢？我们设想，学生完成实验的时间可能是评定学生实验操作水平的一个比较客观的指标。为了检验这一设想是否成立，我们进行了下面的研究。

一 研究目的和测评量表的设计

本研究的主要目的是：探索能否以学生完成实验的时间作为评定学生实验操作水平的一个比较客观的指标。

本研究采用过程测评的方法。测试题要求学生完成两个实验，评定者观察受测学生的实验过程，即时判断受试操作是否正确，并记录结果；同时记录受试完成实验所用时间。

本研究所设计的实验考核试题及实验考察记录卡如下：

高一化学实验操作考核试题

1. 用 98% 的浓硫酸(密度为 $1.84\text{g}/\text{cm}^3$)配 50mL 0.01mol/L 硫酸溶液(提示：所需浓硫酸为 2.5mL)。

2. 实验室需要配制一些氯化钠溶液，但氯化钠晶体中混入了少量硫酸钠和碳酸氢铵。请按照下列实验步骤配制氯化钠溶液。

- (1) 用托盘天平称取含杂质的氯化钠晶体 1.8g；
- (2) 用 20mL 左右的蒸馏水溶解氯化钠制成溶液；
- (3) 向溶液中滴加过量的氯化钡溶液；
- (4) 向溶液中滴加过量的氢氧化钠溶液；
- (5) 过滤，弃去滤渣，取滤液；
- (6) 将滤液倾入大试管中，加热至无氨气放出(用石蕊试纸检验)。

实验考察记录卡见表 8-8。

表 8-8 实验考察记录卡

试题	实验操作	正误	所用时间(分)
1	(1) 量取浓硫酸		
	(2) 稀释(注酸入水)、搅动		
	(3) 转移稀酸入容量瓶(酸液冷却、玻璃棒引流)		
	(4) 洗涤烧杯，转移酸液(控制用水量)		
	(5) 定容(滴管滴加、读取、倒转、摇匀)		
2	(1) 称取氯化钠样品		
	(2) 量取蒸馏水、配制深液(读取、用玻璃棒搅动)		
	(3) 滴加氯化钡溶液		
	(4) 滴加氢氧化钠溶液		
	(5) 过滤(过滤器制作、过滤操作)		
	(6) 滤液转移并加热驱氨(用试纸检验)		

说明：1. 此卡所列项目按照实验所需的基本操作步骤编制。

2. 主考人要认真观察考生的每一步操作并逐项记录：操作正确的打“ ”；操作错误或未做的打“ × ”。
3. 使用秒表准确记录考生所用的实验时间（分）。

二 测试过程

样本来自北京市四所中学（含市重点、区重点和一般中学）的高一年级，共 120 人，分为 5 组，每组 24 人。测试时，要求被测学生完成考核试题上的两个实验；同时有 24 位学生任监考人。监考人的任务是：观察被测学生的实验过程，按照实验考察记录卡逐项记录观察结果，同时记录完成实验的时间。

实验考核题提前发给被测学生，让他们事先预习，并可在教师的帮助下，解决实验中涉及的有关实验原理等知识方面的问题。

实验所需要的药品、仪器等全部事先备好。

三 测试结果分析与讨论

学生完成实验的时间分布于 16—34 分之间，各个时间段的人数分布如表 8-9。

表 8-9 按学生完成实验时间分组统计结果

时间段 (分)	16—18	18—20	20—22	22—24	24—26	26—28	28—30	30—32	32—34
人数 (人)	5	7	12	14	16	18	27	6	2

学生完成实验所用的平均时间为 26 分 44 秒，标准差为 5 分 11 秒。

现将学生在各个时间段的人数分布按照完成实验时间由长到短的方向作直方图如图 8-1。图 8-1 学生完成化学实验时间分布直方图从图 8-1 不难看出，学生在完成实验所需时间上的人数分布接近正态。我们认为，这一结果是十分有意义的，它反映了学生的化学实验操作技能水平的实际状况，说明学生完成化学实验操作的时间能够表征学生的化学实验操作的技能水平，可以作为评定学生化学实验操作技能水平的有效指标之一，为我们找到合理评定学生实验操作技能水平的方法提供了实践基础。

从上述认识出发，我们提出，可以根据学生进行化学实验操作的正确程度和完成实验操作所需要的时间两个方面的指标，综合评定学生的实验操作成绩 X 。即按下式计算：

$$X = a \cdot X_c + b \cdot X_t \quad (\text{式 8-1})$$

式中， X_c 为学生的操作成绩， X_t 为学生的时间成绩， a ， b 为两种成绩的权重系数。

学生的操作成绩 X_c 可以将实验的每个具体步骤赋予一定的分数值，然后按照实验考察记录卡观察学生的实际操作过程直接评分，或是按照学生完成正确操作的百分率而评分；学生的时间成绩 X_t 可根据学生完成实验操作所需要的实际时间，先将其转化为标准时间，然后再转化为时间成绩。即按下式计算：

$$X_i = \frac{T_i - T_{\text{平均}}}{S_i} \cdot c + d \quad (\text{式8-2})$$

式中, T_i 为学生 i 完成实验所用的实际时间, $T_{\text{平均}}$ 和 S_i 分别为被测学生组完成实验的平均时间和标准差, c, d 为常数, 其作用是将分数变到常规范围之内。

式 8-1 和式 8-2 中的常数 a, b, c, d 可根据测试的具体目的和被测学生的具体情况而设定。例如, 当我们强调实验操作的正确性、规范性时, a 值要适当大一些; 而当我们注重实验的熟练程度时, b 值则可规定得高一些。式 8-2 是参照标准分的转化公式而得, 其中 c 值一般取 10, d 值可根据 T_i 的分布范围, 使 X_i 的值接近并小于 100 即可。

例如, 某学生的操作成绩为 85 分, 完成实验的时间为 29 分 10 秒, 即 1750 秒。 $T_{\text{平均}}=26$ 分 44 秒=1604 秒, $S_i=5$ 分 11 秒=311 秒。设 $c=10, d=75$, 则

$$X_i = \frac{1750 - 1604}{311} \times 10 + 75 = 79.7$$

再设 $a=0.6, b=0.4$, 则

$$X=0.6 \times 85 + 0.4 \times 79.7=82.9$$

即该生实验操作技能水平测试的最后成绩为 82.9。

试验研究的结果表明, 按照式 8-1 和式 8-2 所得出的评定结果与教师平时观察结论有较好的一致性。

我们认为, 学生完成化学实验的时间与其实验操作的熟练程度直接相关, 因而可以反映出学生的实验操作的技能、技巧水平; 对学生完成实验时间的测量比较客观、准确且易于实现。所以, 学生完成实验的时间能够作为评价学生化学实验操作技能水平的有效指标之一。

第九章 学生解题思维过程的测量与评价

化学教育要培养学生的能力，已经成为当今化学教育工作者的共识。培养学生运用知识解决化学问题的能力，是化学教育中培养学生能力的主要内容。化学教育测量与评价也必须为培养学生解决问题能力服务。

一般的教学测量与评价，都是通过测验了解学生掌握知识的状况和能力发展水平，根据学生对测验问题的反应，间接地推测学生解决化学问题的水平及其可能存在的问题；而对学生解决化学问题的思维过程的测量与评价，是直接以学生的解题过程为测评对象，了解影响学生解题的各种因素，从而给教师提供更直接、更准确的信息，以便选择或设计更有针对性的教学措施，更有效地培养学生解决化学问题的能力。因而对学生解决化学问题的思维过程的测量与评价，应当作为化学教育测量与评价的重要研究内容。

本章将介绍有关问题解决的心理学研究现状，汇报我们对学生解决化学问题的思维过程研究的主要方法及初步成果。

第一节 有关问题解决的心理学研究概述

在科学技术迅猛发展的现代社会，需要人们解决的新问题层出不穷，对培养受教育者的解决问题能力的要求也越来越高。因此对解决问题的研究也就显得更为重要，这一课题也必然会引起许多心理学家和教育学家的关注。他们用不同的方法从不同的角度研究这一课题，已经获得了许多成果。了解他们的研究方法及其成果对我们研究学生解决化学问题的思维过程无疑是有益的。

一 尝试-错误理论

提出这一理论的代表人物是美国著名的学习心理学家桑代克。他也是最早用实验方法研究学习和解决问题过程的学者。他以动物为实验对象，把处于饥饿状态的猫关在一个特制的木笼中，笼外放有食物。猫面临的问题是如何逃出笼外取得食物。它可能只需要转动一个扣头，或拉动一根绳子，或压下一个门闩，门就可以打开。开始，猫到处乱碰乱抓，偶然碰到开关，逃出笼外取得食物，但把这些逃出笼外取得食物的猫再放回到笼内，猫解决问题的错误和尝试的次数就逐渐减少，解决问题的时间也相应减少。最后，只要把猫一放回笼子，它就会立刻打开开关逃出笼外。从这类实验桑代克得出结论：解决问题的过程就是尝试-错误的过程，其进程为：错误反应逐渐减少，正确反应逐渐增多；各种正确的反应组合后，产生一系列成功反应的动作。在解决问题的过程中，并不存在什么神秘的思维和推理。他将这一结论推广到人类解决问题的过程。尝试-错误理论没有看到思维、推理在问题解决中的作用；实际上，并不是事事都需要千百次的尝试错误。

二 顿悟学说

德国心理学家苛勒（W.Köhler）是这一学说的代表人物。他也进行了一系列的动物实验。其中最典型的是猩猩取香蕉实验。将猩猩关在笼内，笼外放着水果，但因距离太远，猩猩靠前肢无法取得。观察猩猩怎样利用现有条件，找出解决问题的方法（如拉动一端缚着水果的绳子、把两根短竹竿接起来够着水果、先用短竹竿够取长竹竿再用长竹竿够水果等等）。上述方法，有的猩猩能够马上发现，有的一两年也发现不了。这里，猩猩无法进行“尝试-错误”。苛勒认为，猩猩是通过“顿悟”来发现解决问题的方法的。问题的解决是由于理解了部分与整体的关系，是整个情境的新综合，组成新的整体，而不是已知动作的重组。盲目的“尝试-错误”并不是解决问题的主要方法。顿悟学说重视整体的作用，强调对材料的组织，但又忽视量变，轻视经验的积累。

三 五阶段理论

这一理论是美国实用主义教育学家杜威（J.Dewey）提出的。解决问题的5个阶段是：

(1)问题的意识与提出。

要解决问题，首先要发现和提出问题。问题的解决者必须意识到问题的存在，并进行初步怀疑、推测。在日常教学中帮助学生发现和提出问题，增加对问题的敏锐程度，是十分必要的。

(2)问题的确定。

明确问题的空间，认识到它与其他问题之间的各种关系，确定问题解决所要达到的目标。这种对问题空间的分析常常是解决问题的关键。

(3)形成假说。

在分析问题空间的基础上，拟定各种解决问题的可行方案，这便是形成假说的阶段。为了提高方案的可行性，在考虑到各种各样的可能性时，不仅要越周密越好，同时还要考虑到测试或验证这些假说的具体途径。

(4)假说的检验。

由于解决问题通常都可能有几种不同的途径，所以可能的解决问题的方案或假说也就不止一种。一方面，要凭借以往的经验从中挑选出可能性最大的一种或几种假说来予以验证，另一方面还要借助于一定的技术在实际中证实假说。

(5)最佳途径的选择。

经过检验，找出解决某一问题的最佳途径，即得到最好效果的方法。最后，还要把解决某一问题的经验加以归纳，结合到认识结构之中，用以解决同类的或新的问题。

杜威提出的这一理论对于我们分析解决问题的过程有很重要的参考价值。但这五个步骤并非是固定不变的，各步骤之间也常有一定的交叉重叠。

四 现代认知心理学的研究

用计算机模拟问题解决过程，是当代认知心理学研究这一课题的主要方法。美国的尼温(A. Newell)和赛蒙(H.A. Simon)等人用专门编制的计算机解题程序，结合口述记录的方法来探讨人们如何利用记忆来解决问题的操作过程和策略。当代认知心理学综合了尝试-错误理论与顿悟学说的观点，认为人们解决问题一般经过三个阶段：一是了解问题空间，找到问题的表征形式；二是在记忆中搜索有关知识，形成解决问题的动作状态；三是不断进行反馈式评价，衡量操作过程与目的状态和始发状态的距离。借助于计算机，认知心理学对问题解决过程作了更为细致、具体的描述。

问题空间是由始发状态、操作状态和目的状态组成。所谓始发状态就是问题没有得到解决前的开始状态；目的状态就是问题解决后的状态，是努力的目标与结局；从问题的始发状态向目的状态过渡便是操作状态。不同的问题有不同的问题空间，即始发状态和目的状态是否明确、操作状态中解决问题的途径的复杂情况不同，都会构成不同的问题空间。了解问题空间，对问题本身进行明确的表征，是解决问题的关键所在。

问题解决的基本操作过程以手段-目的分析方法为基础。这一方法包括5个主要步骤：确定目标；审视现状与目标之间的差距；寻找缩短差距的方法；建立接近问题目标的子目标；通过操作，实现从子目标向总目标的过渡。解决问题过程中的每一次操作，都是一个“条件-行动”的组合，即一种包含了识别过程和动作过程的组合。每当识别到所需要的条件得到满足时，就会激起某种动作；当几组条件同时得到满足时，就需要选择一种最佳的行为后果。人们解决问题的一般途径是把总目的分成许多小的步骤，通过寻找新的信息，满足达到各种目的状态的条件，通过不断认识和操作来实现总的目标。

在探索解决问题的途径时所采用的基本策略有两种，一种是正向探索，

即从问题的起点开始，采用一些直接的方法进行操作，直到问题的解决；另一种是逆向探索，即着眼于问题解决的最终目标，以及达到这个目标的必要条件和先行步骤，然后一步步地退到原来问题所提出的起点。在这种逆向探索中，更依赖于问题解决者以往的经验。

通过分析专家和生手在解决问题时的策略发现，专家由于掌握了丰富的专门知识和实践经验，形成一种灵活的认知模型，对新问题的解决有较好的应变和判断能力，能顺利地打开节点和通过问题空间，使问题得到较好的解决。这就充分说明了知识在解决问题时的作用。

第二节 学生解题思维过程的测评研究

有关问题解决的心理研究，揭示了人类解决问题的一般过程及其某些基本特征。但由于化学问题在内容上的特殊性，一般的心理学研究的结论并不一定适用于学生解决化学问题的特殊过程。为了探寻测评学生解决化学问题能力的有效方法和手段，我们在已有的心理学研究成果的指导下，通过设计化学试题，调查了解学生解决这些试题的基本过程，经统计分析，归纳出学生解决这些化学试题的一般过程及其主要特点，得出了一些有意义的结论。

一 调查问卷的设计

考虑到中学化学教学的实际状况，我们以与研究目的无关的影响因素较少的高一学生为主要研究对象；根据处于高一学年第二学期的学生一般学习水平和知识基础，选编了3道化学问题；同时，为了调查的方便和有效，在对部分中学教师、学生和高师化学专业本科四年级学生进行初步调查的基础上，分析并给出了每道问题的“一般思路”或“一种思路”，并用“框图”的形式，将我们所设计的这种思路较为详细地表示出来。以3道化学试题及其解题的一般思路构成了调查问卷的基本内容。

中学生化学解题思维过程调查问卷

学校_____ 年级_____ 姓名_____ 性别_____

亲爱的同学：

解化学题是化学学习的重要组成部分，为了探索化学解题的最佳方法和解化学题的一般规律，请你来做几道题，看能否从中发现你解化学题的特点、长处和不足，以利于今后更好地学习化学。殷切地希望得到你的合作与支持。做题时，请你做到以下几点：

(1)请尽可能快地完成试题，并记录每道题所耗费的时间。我把时间写在黑板上，请你在开始做题时抬头看一下黑板，记下起始时间，解完这道题后，再记下完成时间。

(2)请尽量写下自己的全部解题思路，类似于平时解题的打草稿过程，请写清具体步骤。

(3)这不是测验，调查的主要是你的解题思路，请不要有任何顾虑，做对做错都没有关系，只请你能够说明解题思路即可。

谢谢你的合作！

现在请开始做题，并计时。

第1题：答题时间：起始_____，完成_____。

0.5mol A 原子跟足量的 HCl 溶液反应，在标准状况下可产生氢气 11.2L 和 0.5mol ACl_n ， A^{n+} 可被氧化成 $\text{A}^{(n+1)+}$ 。 $\text{A}^{(n+1)+}$ 的 M 电子层中有 13 个电子。

请回答：

(1)元素 A 的核外电子排布是怎样的？A 是什么元素？

(2)A 在周期表中的位置如何？

第2题：答题时间：起始_____，完成_____。

某二价金属 M 含有结晶水的盐 A，溶解后加入 AgNO_3 得白色沉淀。A 的水溶液中加入 NaOH 不生成沉淀，向此碱性溶液中通入 CO_2 则生成白色沉淀 1.97g，此沉淀加入稀 HCl 全部溶解时产生 224mL CO_2 （标况）。向此盐溶液中加入稀 H_2SO_4 生成不溶于稀 HNO_3 的白色沉淀 2.33g。1 克 A 加热到 125 时失去 0.147g 水生成无水盐。问：

(1) 根据 A 溶解后加入 AgNO_3 溶液产生白色沉淀，该沉淀不溶于稀 HNO_3 ，推断 A 是哪一类盐？

(2) 根据上述数据推算出金属 M 的原子量。

(3) 求 A 的化学式。

第 3 题：答题时间：起始_____，完成_____。

现有两种氧的单质 O_2 和 O_m 的混合气体，从其中取出体积相同的两份。往一份中加松节油， O_m 完全被吸收，结果气体体积减少 $V\text{mL}$ ；加热另一份使其中的 O_m 全部分解成 O_2 ，则气体体积增加 $\frac{1}{2}V\text{mL}$ （以上气体体积均在同温同压下测定）。根据上述数据，推断 m 的数值。

解题思路调查问卷

学校_____ 年级_____ 姓名_____ 性别_____

亲爱的同学：

你已经按时完成了试题，非常感谢你的合作！

下面列出了每道题的一种解题思路，不一定是最佳思路，是帮助你回忆自己的解题思路用的。请回答：

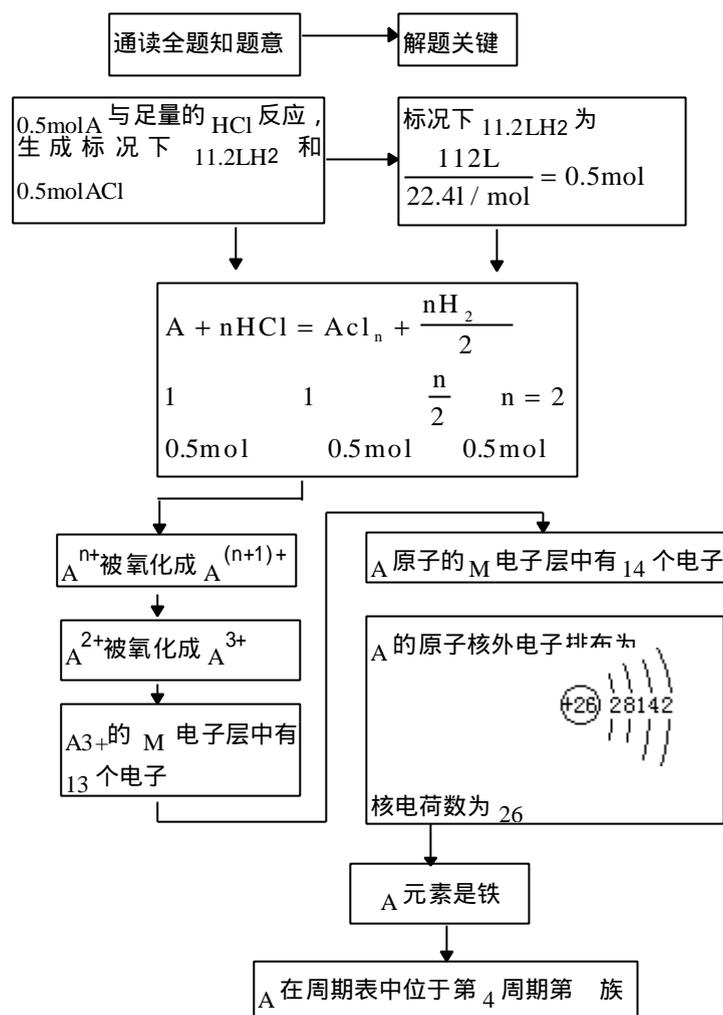
(1) 你的思路和列出的思路是否相同？_____（相同、不同）如不相同，请将不同处在图上标出或说明。

(2) 在解题过程中耗时最多的那个步骤前标上“ ”。

(3) 造成解题困难的主要原因。

再次感谢你的热情协助！

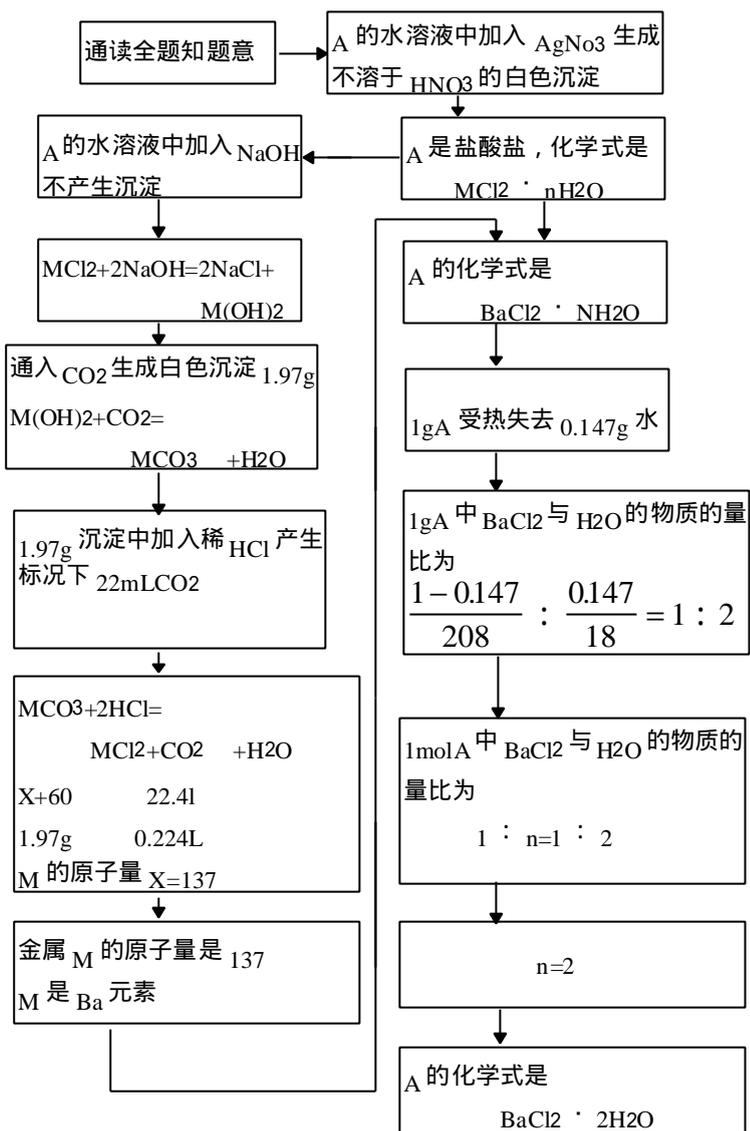
(1) 思路：



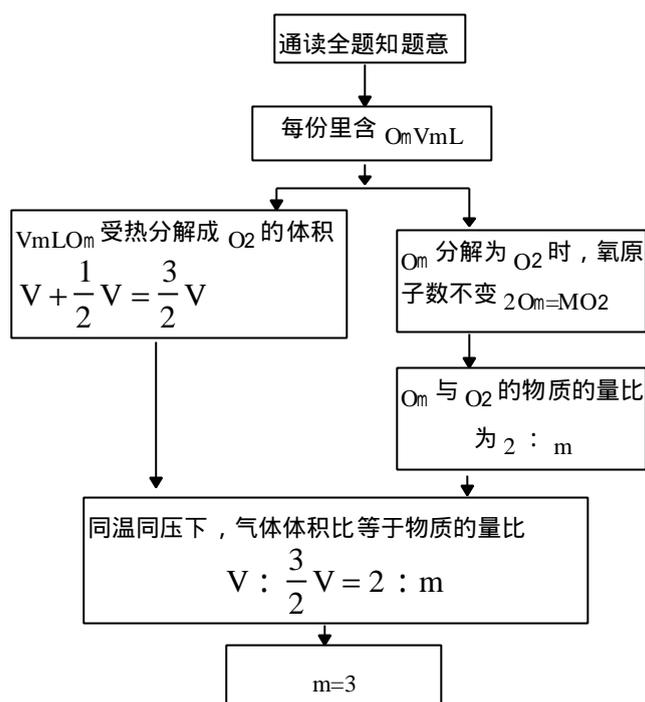
(2) 耗时最多的步骤 _____

(3) 解题困难的主要原因 _____

第 2 题



- (1)思路：
- (2)耗时最多的步骤_____
- (3)解题困难的主要原因_____
- 第 3 题
- (1)思路：



(2) 耗时最多的步骤_____

(3) 解题困难的主要原因_____

二 调查结果

利用一节课的时间, 对选定的被测学生进行集体测试。测前先向学生说明测试的目的、方法, 努力消除学生对测试的顾虑并争取他们重视和配合这项工作, 同时, 力求给学生提供一个与平时学习状况一致的解题环境, 以避免无关因素对测试结果的影响。调查问卷的第一部分: 试卷分发完毕后强调解题的时间要求, 即鼓励学生尽快完成试题。由主试规定开始解题的时间, 并每隔一分钟在黑板上做一次记录, 以便于学生记录完成每道试题的起始时间。当 90% 的学生完成试题后, 要求学生停止做题, 并发放调查问卷的第二部分——思路问卷, 让学生对照已给的思路, 回忆自己的解题过程, 标明与已给思路不一致的地方, 并指出解题过程中耗时最多的那个步骤, 即该学生在该题中所遇到的瓶颈。在对所收回的问卷做出初步分析之后, 挑选少量思路比较典型或代表性较好的学生进行座谈或个别交谈, 进一步了解或确认学生的解题过程及其特点, 以提高调查结果的可信性。

本研究原计划对北京市三种水平的学校——市重点、区重点和一般中学都进行抽样调查, 但在实际实施调查后发现, 一般中学的学生对调查不能积极配合, 调查结果明显不可靠; 同时考虑到本研究的性质, 最后确定以重点中学的学生为调查对象。限于研究条件, 最后选定的被测学生样本共 122 人, 回收有效试卷 109 份。调查结果统计如下:

(1) 解题所需时间 (分) 及其人数分布 (直方图)。

解答正确的学生共 54 人, 占 49.5%。

第2题：

时间(分)	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
人数(人)	1	1	4	9	12	17	23	16	12	8	3	2	1

解答正确的学生共 65 人，占 59.6%。

第3题：

时间(分)	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11
人数(人)	10	13	15	16	17	9	7	4	4	2

解答正确的学生共 59 人，占 54.1%。有 12 名学生未答此题。

(2)学生在解题过程中的典型行为表现及解题思路(根据学生答卷统计、归纳)。

第1题：

序号	学生的解题行为	出现次数
1	原子结构示意图不正确	28
2	核外电子排布较慢	19
3	考虑 M 电子层电子数为多少耗时最多	17
4	周期表运用不熟，搞错 A 在表中的位置	10
5	完全错误，思路混乱	7
6	无素化合价与元素所在周期表中的族数没有搞清	5
7	审题花费时间较长	5
8	列比例式求算 n 值是解题困难的主要原因	4
9	一种思路：逆推法，先猜测 A 是 Fe，后证明	4
10	书写化学方程式最困难	2

解题思路 1：已知条件 n 值 电子排布 核电荷数为 26 Fe 在周期表中的位置。

解题思路 2：猜测 A 可能是 Fe 逆推、验证。

思路 1 使用较普遍。瓶颈主要存在于对过渡金属元素的原子结构不熟悉，不能熟练运用原子结构的有关知识。

第2题：

序号	学生的解题行为	出现次数
1	求算 M 的原子量及计算所含结晶水个数这两步耗时最多	56
2	先据题中的数字特点猜测 M 为 Ba，再进行计算加以验证	12
3	基本符合所给思路	11
4	审题不清，下手困难	6
5	无困难	4
6	寻找计算 Ba 的原子量的方法耗时较长	3
7	书写化学方程式耗时较长	3
8	对题中所列各物质的性质不熟	3
9	把 A 的化学式 $ACl_2 \cdot nH_2O$ 过于复杂化，导致求不出 n 值	2
10	思想先入为主，看完题后已知答案，表达整理不清（学生语）	2

解题思路 1：与 Cl^- 共存 可能是 K^+ ， Na^+ ， Ca^{2+} ， Ba^{2+} 与 H_2SO_4 生成不溶于酸的白色沉淀 Ba^{2+} M 的原子量 A 的化学式。

解题思路 2：同问卷思路。

解题思路 3：猜测 M 约为 137 验算证明 M 为 Ba A 的化学式。

本题的瓶颈为： 计算量大； 题目较长、审题困难。

上表中第 1，2，5 三种行为为三种典型的解题思路。本题“瓶颈”不明显。未能解出此题的主要原因为不能正确写出 O_m 分解为 O_2 的化学方程式。

三 调查结果讨论

(1) 学生解题所需时间是表征学生解题能力的重要指标。

从学生在 3 道试题上所需解题时间的人数分布看，都接近正态分布。这一结果是很有意义的。在班级教学的情况下，学生按学习成绩高低的人数分布一般也是正态。这并非是偶然的巧合，而可以看做是对同一事物——学生的化学学业成就的两种不同的表征方法。但是与用考试分数代表的学生学习成绩相比，学生解题时间可能更直接、更可靠地表征了学生的化学学业成就。有经验的化学教师都认为，化学基础知识越扎实，应用知识解决问题的能力越强的学生，解决化学问题所需的时间就越短。这种“经验式”的认识在本研究的结果中得到了比较满意的证实。通过统计分析和问卷后的座谈及个别交流，我们发现，学生的化学学习成绩和其所需的解题时间之间存在着较明显的正相关，即化学成绩好的学生，所需解题时间短；反之，化学学习成绩差的学生，所需的解题时间就长。

学生在解题时间上表现出的差异，还为我们更有效地发现学生间的差异提供了一个可用的方法。我们知道，一般的考试或测验，都是根据学生能否正确回答问题来区别学生的，但是对于同样能够答对问题的学生而言，这种测验或考试是无法再做进一步区分的。然而却可以利用他们在解题时间上可能存在的差异对他们做出进一步的区分。

可见，学生的解题时间可以作为表征学生解题能力的重要指标，而且还可能成为测评学生化学学业成就的更为细致、具体的指标。

(2)学生的化学基础知识是影响学生解决化学问题能力的重要因素。

研究发现，学生解决化学问题的思路选择和所需时间与学生对相关化学知识的掌握水平密切相关。学生在解题过程中出现的瓶颈问题，在相当大的程度上是由于未能较好地掌握相关的化学知识造成的。如第1题，有67.9%的学生所显示的瓶颈是由于未能在应用水平上掌握原子结构和元素周期律等基础知识。基础好的学生能够正确地选择解题思路，所需要的解题时间较短；而基础薄弱的学生则表现为在选择解题思路时比较困难、犹豫不决，较盲目地进行“尝试-错误”，显然所需要的解题时间也较多。调查结果清楚地表明：化学基础知识的掌握是解决化学问题的基本前提，是提高解题能力的先决条件。无知必定无能，而且对化学知识的掌握决不仅仅意味着对知识的死记硬背，如果不能将所学知识融会贯通，不能把握灵活运用知识的条件和方法，同样不能解决化学问题。

由于学生解决化学问题的过程能够更为直接和具体地反映出学生对相关知识的掌握情况（如对知识记忆的准确程度、灵活运用知识的水平等），因而考察学生的解决化学问题的过程可以作为进行化学教育测量与评价的一种十分有意义的信息源。

(3)学生的审题能力和计算能力对学生解题过程的影响。

审题是指通过阅读试题明确解决问题的方向和要求，把握试题所给出的已知条件，并初步判断问题目标与已知条件之间的关系或联系，为选择或确定解题思路提供依据。调查表明，有10%以上的学生表现为审题困难，即在开始阅读试题与动手解题之间耗费了大量时间。如第2题，由于题目本身内容较多，有的学生读完一遍之后感到不知所云，只好再读第二遍、甚至第三遍；还有的学生不能把握试题所提供的隐含信息，如第3题，有的学生未能注意到“往其中一份中加入松节油， O_m 全部被吸收，气体体积减少了 VmL ”这一隐含信息，不能判断出“ O_m 的体积就是 VmL ”这一已知条件，因而不能确定出正确的解题思路。

我们认为，审题能力是与观察能力、分析综合能力、逻辑思维能力等一般能力密切相关的一种综合能力。因而学生审题能力对解题过程的影响是十分复杂的。为了培养和提高学生的审题能力，除了要注意培养与审题能力有关的一般能力之外，还要在平时教学中加强审题训练，有针对性地解决学生在审题过程中出现的典型问题。

计算能力是指学生在通过化学计算而解决问题时所表现出的能力。当化学问题与数学运算有关时，计算能力常常成为一些学生能否解题的关键。调查结果显示，有51%的学生在做第2道题时，计算步骤耗时最长。提高学生的化学计算能力，也是提高学生解决化学问题能力的重要方面。但问题的另一方面也需要引起注意：不能把化学测量变成数学测量。即在设计化学计算类的试题时，要注意避免过难、过繁的数学运算，试题的内容应紧扣化学基本概念、基本理论和重要的元素化合物知识，计算过程应是化学基础知识的运用过程；没有必要在化学测量中让学生花费过多的时间和精力去进行数学计算的练习。

(4)试题质量对测试结果的影响。

从调查结果看，试题2在学生的解题时间上得到了比较满意的分布状态；学生在该题的行为表现与学生的平时化学学习成绩之间相关程度也是最明显的。因而可以认为该题相对质量较高。可见，试题质量会对学生解题过

程的测评结果产生不可忽视的重要影响。本研究中所用的试题 1 和试题 3 的质量还有待于进一步提高。在本研究的实施过程中，我们着重观察、了解学生在解决化学问题的过程中所存在的问题及其原因；但同时发现，有些学生在解题过程中表现出一定的思维的独创性、新颖性和敏捷性等特点，这是十分可贵的，也是很值得引起注意和加以研究的。因限于研究的条件未能就此展开工作，这是本研究的一个缺憾。

第十章 化学课堂教学评价

一般意义上的化学课堂教学评价是指：以一节化学课的教学为研究对象，以化学课堂教学的目标和化学教学原则为依据，对化学课堂教学的过程和效果给予价值判断的过程。

化学课堂教学评价的主要被评价者是化学教师。根据评价主体的不同，可以将化学课堂教学评价分为教师的自我评价和他人评价。教师的自我评价是教师在整个课堂教学过程中，有目的地、比较系统地搜集反馈信息，及时调整自己的教学计划、措施、方法、手段等教学行为，从而实现对课堂教学的有效控制，帮助学生顺利达到教学目标。自我评价是教师的基本教学技能之一，有关的方法、原则、依据等，在一般化学教学论著作或化学教学法教材中都有较详尽的论述。本章将着重讨论以“他人”——除课堂教学活动的主体教师和学生之外的其他评价者——为评价主体的、以化学教师在课堂上的教学活动表现及其所产生的学生反应或教学效果为主要评价内容的课堂教学评价活动。

第一节 化学课堂教学评价的主要功能

以听课为主要形式的课堂教学检查，在我国几乎已经形成一种传统的教学管理方式。课堂教学评价可以看做是对传统的课堂教学检查的改进，与听课相比，它可以实现更多的功能。课堂教学评价的功能主要表现在以下三个方面：

一 全面落实化学教学目标

课堂教学是学校实现其培养目标的主要途径。传统的教学检查和一般以学生达到教学目标的程度为主要内容的教学评价活动，都以学生的学习结果为评价的基本依据，通过检查学生的学习结果的质量和其中存在的问题，来间接地评价课堂教学的质量。但是这种结果评价还不能完全代替过程评价，即学生的学习结果并不能全面地反映课堂教学的全部质量状况，化学教学目标中的很多重要的内容还不能直接反映到学生的学习结果中。例如，结合化学知识教学对学生进行政治思想教育和心理品质的培养；按照化学学科的特点，注意培养学生的化学实验能力、进行科学方法教育、养成科学态度等，都是化学教学的基本目标要求，但这些内容目前还很难完全体现在对学生的学习结果测量中。正是由于过去只偏重于结果评价，导致部分教师忽视那些不能直接、客观测量的化学教学目标的重要组成部分，从而出现了“片面追求升学率”、“只教书不育人”、“黑板画实验、课下背实验、考试写实验”等严重违反教育、教学规律的错误行为。

通过课堂教学评价，可以直接考察教师是否将认知、情感、动作技能三大领域的各种教育目标都落实于教学过程之中，从而保证化学教育目标的全面落实，纠正或防止那些只顾认知目标而忽视情感、技能领域目标的错误做法。

二 全面评价教学水平，激发教学积极性

评价教师的教学水平，也必须将过程评价和结果评价结合起来。只看学生的学习成绩好坏来决定教师的教学水平，是比较片面的。因为影响学生学习成绩的因素是多方面的，而不是仅仅取决于教师的教学水平。教得好不一定就能使学生学得好或考得好；而学生的考分高也并不能充分证明一定是教师教得好。如果我们不能认识到教师教学水平与学生考试成绩之间相背离的可能性，不管学生来源、学校教学的实际条件等差异，单凭学生的考试成绩评价教师的教学水平，就可能严重挫伤部分教师的教学积极性。通过课堂教学评价，让教师的学科基础知识、教育思想和观点、教学方法与艺术、教学态度等在课堂教学中充分表现出来，并将其作为评价教师教学水平的重要依据，就能使所得到的评价结论比较客观、全面、公正。同时，课堂教学评价活动的展开，容易扩大教学评价活动的影响，有利于形成一种重视教学、重视教师的良好风气，从而有利于提高广大教师深入开展教学研究、努力提高教学水平的积极性。

三 帮助教师改进教学方法，提高教学水平

课堂教学可以集中表现教师的教学水平，并反映出教师在基础知识、教学观点、教学方法、教学态度等方面的不足甚至缺陷。虽然学生的测验结果

也可以反映出教师在教学中存在的某些问题，但结论多来自于一种间接地分析或推测；通过课堂教学评价，则可以比较直接、客观地发现教师教学中存在的问题，从而可以更直接和有针对性地提出改进教学的途径和方法，提高教师的教学水平。

开展课堂教学评价活动的最终目的是为了改进课堂教学。但是，在不同情况下评价的具体目的也会有所不同；而评价目的又决定了评价的程序和方法。因此在进行课堂教学评价活动时，同样需要首先明确评价目的，根据目的的要求，确定评价过程的侧重点，采用与之相适应的评价方法和手段。

第二节 总结性的化学课堂教学评价——评优课

总结性的课堂教学评价是通过评价活动对教师的课堂教学水平进行分等鉴定，为人事决策如评选优秀教师、为教师提职晋级等提供依据。评优课是目前开展得最为广泛的一种总结性的课堂教学评价。

一 评优课指标体系的建立

1. 建立评优课指标体系时应注意的问题

建立一个好的评优课指标体系，是保证评课质量的基础。指标体系的内容及其权重分配，反映了指标体系的设计者对化学课堂教学本质或规律的认识和化学课堂教学质量观。好的指标体系，要有比较坚实的理论基础和实践依据。目前，还没有一个能被广泛接受或公认的评优课指标体系出现，对此仍在不断探索之中。这里，只提出几个在建立化学评优课的指标体系时需要重视的问题。

(1) 要做到全面、具体与切实可行相结合。

评优课的指标体系会对教师产生较强的导向作用。指标体系中的重点内容，必然会是教师注意的重点并作为努力的方向；而指标体系中的非重点内容或所忽视的教学要求，也会使教师忽视或有意识回避。因此，指标体系的片面性容易导致不良后果。为了使评价结果客观、可靠，还要求每一项指标都尽可能具体、细致，以便于评价者做出价值判断。如果指标过于简单、笼统，则不可避免地会给评课过程带来较多的主观性。但并不是说，指标越多、越细就越好。实际上，指标过多、过细，会使其中许多指标形同虚设，因为要评价者在很短的时间内对过多的指标都逐一做出价值判断是相当困难的。所以，指标太多，就会增加评价过程的随机性，还会徒劳地增大评课的工作量，降低评课工作的质量和效益。

因此，在建立评课指标体系时，要注意根据评课人的实际水平及评课条件，既切实可行又全面、具体。要做到指标体系详略得当，就必须找到那些能够反映化学课堂规律的、最为重要的评价指标体系既能够充分保证化学的质量标准，又不至于使指标体系过于琐碎、庞杂，虽然全面，却难以应用。

(2) 要反映化学课型与教学内容的特点。

评优课的指标体系一般都要涉及教学目标的确定、教学内容的选择、教学方法、演示实验、教学态度、教学效果等方面。然而教师教学水平的发挥，常常和化学课的类型、内容有关，而在实际评优课中，又很难做到让所有被评教师上同一类型或同一内容的课。因此，需要在全面要求的指标内容的基础上，根据不同的课型或教学内容，相应确定有关指标的权重，以制订出与被评课的课型或教学内容相应的评价指标体系，这样可能会使评优课更为合理、公平些。例如，对于理论知识课，那些要求教师引导学生积极进行由宏观到微观的思维活动等方面的指标应赋予较大的权重；而对元素化合物知识课，那些要求教师做好演示实验并注意引导学生观察实验、培养观察能力并利用实验掌握知识等方面的指标也要赋予较大的权重；对化学演示实验较多和演示实验较少或几乎没有的两种课，关于教师做化学演示实验方面的指标显然要有不同的权重或对评价的内容加以适当增减。

(3) 要对课堂教学的实际效果予以适当的评价。

从理论上说，教师教学水平的高低，最终应反映到学生的实际学习效果

上。而对教师课堂教学实际效果的评价，却不是一件简单的事情。在进行评优课的实际中，常常根据当场对学生进行的课后测验成绩来评价教师的课堂教学效果。应当说，这种评价方式带有一定程度的客观性和合理性；在课后测验内容合理的前提下，学生的测验成绩能够作为评价教师的教学水平的重要依据。但是也应该看到，学生的课后测验成绩并不仅仅受到课堂教学过程的影响，还与学生原有的基础知识、智力水平差异、情感因素、练习效应、家庭文化背景等多方面因素有关；而且课堂教学效果往往表现出一定程度的滞后性，即课堂教学所引起学生的某些行为变化，并不能立刻就表现出来，因而也难以反映在课后测验上。所以对课堂教学实际效果的评价，既要重视课后测验的结果，又不要给学生的测验成绩赋予过大的权重；还要注意学生在课堂教学进行过程中的实际反应，如课堂学习气氛、学生集中注意力的程度等，综合各个方面的信息，适当评价课堂教学的实际效果。

2. 化学课堂教学评价指标体系的建立

(1) 应用调查统计法建立指标体系。

好的课堂教学评价指标体系的建立，要求设计者对化学课堂教学本质或规律有正确、深刻的认识，有正确的化学课堂教学质量观。为了防止指标体系设计过程中的主观性和片面性，可以通过比较广泛地征求那些有较丰富的实践经验和一定理论水平的化学教育专家、学者对化学课堂教学评价指标体系的意见，应用统计分析方法，综合所征求的各种意见，在此基础上形成指标体系。具体方法如下：

设计调查问卷。问卷要列出通过初步调查所收集到的有关课堂教学评价的各种指标，其数目一般应多于准备实际采用的指标数。请被调查者通过比较各个指标的具体内容，确定其重要性程度，即列出各个指标的等级。

计算各个指标的权数，即利用排序指数公式对调查结果进行统计处理。

设采用含 4 个指标的调查表对 24 位化学教育专家进行调查，所得结果如表 10-1 所示。

计算公式为：

$$W_i = \frac{\sum a_j \cdot n_i}{N \cdot \sum a_j} \quad (\text{式}10-1)$$

式中， W_i 为平均权数； a_j 为调查权数，共有 4 个等级，其权数分别为 4，3，2，1， $\sum a_j = 4 + 3 + 2 + 1 = 10$ ； n_i 为对指标 i 所赋予不同权重的人数； N 为被调查的总人数。

表 10-1 评优课指标体系调查结果（示例）

	特别重要	很重要	比较重要	可有可无	平均权数
	4	3	2	1	
指标 1	9	8	5	2	0.300
指标 2	3	5	5	11	0.200
指标 3	8	6	3	7	0.263
指标 4	4	5	11	4	0.237

将表 10-1 中的各项数据代入式 10-1，则可得到 4 项指标的平均权数。

$$W_1 = \frac{4 \times 9 + 3 \times 8 + 2 \times 5 + 1 \times 2}{24 \times 10} = 0.300$$

$$W_2 = \frac{4 \times 3 + 3 \times 5 + 2 \times 5 + 1 \times 11}{24 \times 10} = 0.200$$

$$W_3 = \frac{4 \times 8 + 3 \times 6 + 2 \times 3 + 1 \times 7}{24 \times 10} = 0.263$$

$$W_4 = \frac{4 \times 4 + 3 \times 5 + 2 \times 11 + 1 \times 4}{24 \times 10} = 0.237$$

确定指标，并对指标权重作归一化处理。可根据统计计算所得的平均权数值的大小顺序，按照设计要求选择平均权数值较大的指标构成指标体系，并经过归一化处理后，确定各个指标的权数。例如，若根据表 10-1 的调查结果和设计要求，选择三项指标构成指标体系，则应选取指标 1, 3, 4 而淘汰指标 2。指标权数的归一化公式为：

$$W_i = \frac{W_{i_0}}{\sum W_{i_0}} \quad (\text{式10-2})$$

式中， W_i 为归一化之后指标 i 的权数； W_{i_0} 为归一化之前指标 i 的权数。

按照式 10-2 计算指标 1, 3, 4 的权数：

$$W_1 = \frac{0.300}{0.300 + 0.263 + 0.237} = 0.375$$

$$W_2 = \frac{0.263}{0.300 + 0.263 + 0.237} = 0.329$$

$$W_3 = \frac{0.237}{0.300 + 0.263 + 0.237} = 0.296$$

调查统计法使用简单，意义明显，可广泛应用于化学课堂教学评价以及其他教学评价指标体系的建立过程之中。

(2) 用于化学课堂教学评价的参考指标体系。

表 10-2 给出了化学课堂教学评价的一般指标体系，供设计具体的化学课堂教学评价指标体系时参考。该体系按理论知识课和元素化合物知识课两种类型给出了不同的权数分配。这里的课型划分的主要依据是演示实验在课堂教学中的作用或分量。

表 10 - 2 化学课堂教学评价的参考指标体系

项目	评价内容	理论课	元素化合物课
教学目标 和	教学目标合理, 内容选择适当	5	7
	讲授正确, 思路清晰、合理	8	10
教学内容	传授知识和培养能力相结合	7	10
	开展思想教育, 理论联系实际	5	8
教学方法 和	突出重点, 突破难点, 抓住关键	10	12
	主导教学, 重视启发, 勇于创新	5	8
教学效果	时间分配合理, 环节连贯紧凑	5	8
	教与学和谐一致, 气氛活跃, 学生听课效果好	10	12
演示实验 和	准备充分, 装置美观, 操作正确	8	0
	时间控制适当, 现象鲜明	7	0
指导观察	指导观察方法, 培养观察能力	5	0
	利用实验结果, 紧密结合教学	5	0
教学素质 和	语言简练、流畅、逻辑性强	7	9
	板书简要、工整、清晰、美观	6	8
教学态度	仪态大方, 轻松自然, 富有感染力	3	3
	因材施教, 照顾全面, 善于处理意外情况	4	5

关于化学课堂教学评价的指标体系以及评价实施的具体方法, 可参见《化学教育》杂志 1992 年第 1 期。

二 实施评优课时的有关问题讨论

1. 怎样综合不同评价者的评价结果

在化学课堂教学评价实践中, 评价结论需要综合不同评价者的评价结果而得到。评优课的评价者常常是由被评教师、被教的学生和评价专家组(一般由教研人员、教学管理人员及有经验的教师组成)三部分人构成。那么应当怎样综合教师自我评价、学生评价和专家评价的意见而得出评优课的评价结论呢?

我们认为, 评优课作为一种总结性评价, 往往与教师的切身利益直接相关, 因而必须要有一定的权威性; 为此, 在综合不同评价者的评价结果时, 宜以专家组评价意见为准, 教师自我评价和学生评价只作为专家组评价的重要依据或供专家讨论评价结论时的参考资料。这是因为: 教师的自我评价经常容易受到教师的个性特点、情绪因素及被评者心理的较强影响, 使评价结果失真。在评价实践中不难发现这种现象: 那些自尊心、自信心较强的教师往往对自己评价偏高; 而自信心不足、自我意识不强、性格较懦弱的教师又往往对自己评价过低; 评价结果对教师的切身利益影响越大, 这种失真的程度越大, 产生失真的可能性也越多。学生评价更容易受到教师、环境条件、从众心理及其他心理因素的影响。由于认识水平的局限, 学生对评优课指标体系的理解、对教师课堂表现的观察和价值判断能力与评优课对评价者的要求之间存在着相当大的距离, 因而很难保证学生评价的客观性。学生评价常常表现为对教师的评价偏高, 绝大多数学生都不愿给教师打低分; 当个

别学生对教师的评价较低时，往往又因为存在着与评优课本身无关的其他原因。这种现象在中学生中较普遍；年级越低，学生评价结论的客观性往往越差。

为了提高评优课的工作效率，还可以考虑直接以专家组作为评优课的唯一的评价者。在实施评价时，专家组成员要注意从教师和学生两方面广泛收集评价信息；当专家组内意见不一致时，不要将不同的评价意见简单地加权平均，而要通过讨论，基本取得共识，在此基础上再下结论，从而提高评价结论的科学性和可靠性。专家组成员的选拔和素质的提高对保证评优课的科学性水平显然也是有重要意义的。

2. 怎样看待教师的临场发挥失常

个别被评教师平时教学水平较高、教学效果较好，而在评优课时却临场发挥失常而只得到较低的分，这时，应当如何给出评价结论？这是评优课实践中经常遇到的一个问题。

我们认为，评优课要以被评教师的临场教学水平的发挥作为评价的唯一依据。无论被评教师的平时教学水平如何，被评课的教学水平高，就应该获得较高评价，反之，只能得到较低评价。这就如同对体育比赛的裁判，决定比赛成绩的只能是运动员在赛场的临场表现而与运动员的平时训练成绩无关。实际上，评价专家组成员一般都难以对所有的评价对象的平时教学水平有比较全面、深入的了解；如果将被评教师的领导或同事的介绍作为评价的依据之一，显然这种依据难免带有介绍者主观因素的影响。因此被评教师的平时教学水平是不宜作为评价依据之一的。即使被评教师平时教学水平确实较高，而在评优课时由于某种偶然因素的影响导致发挥失常，则仍可认为被评教师的教学中存在着值得改进之处，因为真正教学水平较高的优秀教师应当具备较强的控制偶然因素、保证教学质量的能力，这正是一般教师需要努力培养的教学能力之一。

所以评优课以被评教师的临场教学水平的发挥作为评价的唯一依据，不仅可以使评优课的结果更加公平、合理，还能够促进教师加强自己控制影响教学的各种因素、保证教学质量的能力。当然，与之相应的是评优课的经常化和制度化，给教师以更多的参与竞争的机会，克服那种碰运气的侥幸心理，从客观上减少导致教师临场发挥失常的因素。

第三节 形成性的化学课堂教学评价——检查课

一 检查课的意义

检查课是指教学管理人员平时对教师课堂教学质量实施检查的一种形式。如督学、校长、教导主任、教研组长等人员去听某位教师的一节课，就是检查课最为常见的一种形式。检查课的听课过程，实际上就是对教师教学质量做出评价的过程。与评优课相比，检查课的特点表现在：

1. 以促进教学目标的全面落实为主要目的

检查课的主要目的并不是要对教师进行分等鉴定，而是在于检查教师是否全面落实了课堂教学的各个目标。通过检查课，可以了解教师对教学计划的执行情况，了解教师为落实教学目标采取了哪些措施或进行了哪些尝试，了解教师在课堂教学过程中有哪些成功之处和存在哪些缺陷或不足，进而督促和帮助教师改进教学工作，或采取某些必要的教育行政管理措施，以保证教学目标的全面落实。正是由于检查课具备这种直接有利于教学目标实现的功能，我们可以称之为形成性的课堂教学评价。

2. 以教学目标为评价标准的绝对评价

评优课是以参评教师的相互比较评出优秀或先进教师，因而是以相对评价为主的；而检查课却不需要对被评教师进行分等鉴定，因而也不需要将被评教师作相互比较；为了实现检查课的目的，必须以教学目标作为检查的唯一依据。由于教学目标可以看做是一种绝对标准，因而检查课就可以认为是一种绝对评价。检查课的这种性质，使其不会带有评优课的那种较明显的竞争性，因而也不会对被评教师产生过多的心理压力，相对比较容易实施，所得结果客观程度也更大些。

3. 重视实施检查课的意义

我们在讨论化学教育目标的类别时已经谈到，化学教育目标中存在着一类体验目标，如科学态度的培养、爱国主义教育等方面的目标，需要教师在课堂教学中经常创造出合适的情境和气氛，让学生不断地感受、长期地体验，在不知不觉中逐步达到教育目标的要求。这类目标是化学教育目标中不可缺少的重要组成部分，它们的实现更多地要依靠化学教学的过程，因而被称之为过程目标；但由于它们难以直接体现于对教学效果的最后测验之中，所以过程目标在实际化学教学中往往被忽视，这也是片面追求升学率、俯首听命于高考指挥棒的典型表现之一。如果我们重视实施检查课，以过程目标直接作为评价的重要内容，就可以作为一种有力措施，促使教师重视化学实验在实现过程目标中的重要意义，在传授知识的同时，培养能力，并进行科学态度和方法、热爱社会主义祖国的教育，以保证过程目标的实现，从而有利于培养社会主义建设所需要的合格人才。

二 化学检查课的实施要点

1. 化学检查课需要注重的评价内容

检查课不需要像评优课那样必须有较全面的评价指标体系，而只需要侧重于检查教师是否具有全面、正确、合理的教学目标，特别是是否具有与化学知识目标相应的过程目标；侧重检查教师是否能够围绕目标有效地组织教学过程，特别是能否在教学中体现出过程目标的要求。我们认为，根据当前化学教学实际中存在的主要问题，在检查课的实施过程中，要注重以下两个

方面的评价内容：

(1)对化学实验教学的评价。

要注重检查教师能否遵循以实验为基础的化学教学规律，想方设法让学生多亲自动手做实验，引导学生观察实验，让学生亲身体验通过化学实验探索规律的发现或验证活动，在理解化学概念、掌握化学知识的同时，受到科学态度和方法的熏陶、感染和教育，并培养解决问题的能力；着重改变和防止那种不重视化学实验如以口头描述实验代替实际动手操作实验的错误做法，以及那种只是简单地做一些验证式实验而不注重开发化学实验的智力价值和教育价值的简单化做法。

(2)对引导学生进行思维训练教学的评价。

要注重检查教师能否引导学生从宏观的化学事实、现象去思考、探索和发现事物的微观化学本质，能否引导学生进行由表及里、由简单到复杂、以简驭繁、由宏观到微观的思维过渡，在教给学生以化学知识的同时，有效地培养他们的思维能力；着重改变和防止那种只管学生知、不管学生想的填鸭式教学。

2.实施化学检查课的其他要点

检查课作为日常教学管理工作的一部分，应当经常化和制度化。在实施检查课之前，要制订出具体的计划和方案；要根据各个教学阶段的不同特点，确定检查的重点内容，并制定出比较简单、易行、重点突出的检查指标；检查的对象应是全体任课教师；在时间和检查对象的具体安排上要采用随机方式，不需要事先通知被检查者，以便获取更客观的教学反馈信息。

在实施检查课的过程之中，不需要像评优课那样随时做出价值判断、给被检查者评分或给出等级评定，而是要根据检查指标，准确而全面地做出听课记录。这就要求检查者不能单纯地带着耳朵听，而是要事先了解课堂教学内容，准确把握课堂教学目标，全面而有重点地观察课堂教学的全过程，进而全面评价教师的课堂教学活动。

在实施检查课之后，要结合被检查教师的具体教学特点，评价课堂教学的成功与不足，明确指出被评教师在课堂教学过程中表现出的长处与缺点，使其明确今后改进教学的努力方向。检查结果要记入教师个人的业务档案，使之成为教师提职、晋升的依据之一。

第四节 诊断性的化学课堂教学评价——研究课

一 研究课的意义

研究课是指评价者帮助教师诊断课堂教学中存在的问题，并提出和帮助教师实施改进课堂教学的措施、计划、方案，以提高教师的课堂教学水平。与评优课和检查课相比，研究课的特点表现在：

(1) 研究课以教师在课堂教学中可能存在的主要问题为评价重点。

研究课并不需要对课堂教学过程进行全面的观察和评价，而仅仅侧重于教师在课堂教学中可能存在的问题；研究课并不要求指出教师在课堂教学中表现出所有长处和不足，而只是侧重于发现教师在课堂教学中的主要缺陷并分析判断其存在的原因。这就像医生给病人看病一样，并不需要对病人进行常规体检而评价其身体的健康状况，而是要诊断病人主要疾病及其致病的原因。正因为研究课的这一特点，我们可以将其称为诊断性的课堂教学评价，或称之为课堂教学的临床诊断。

(2) 研究课以提高教师的教学水平为直接目的。

研究课的目的不是为了对教师区别优劣或是对教师进行检查、督促，而在于帮助教师发现课堂教学中的问题、寻找解决问题的途径和方法，最后改进课堂教学质量、提高教学水平。研究课的这一特点，可以有效地消除教师对评价活动的疑虑，使其有可能受到教师的欢迎；教师不是像评优课和检查课那样只是比较被动地接受评价，而可能会更加积极主动地参与到评价过程之中。研究课以帮助教师改进教学为直接目的。教师对研究课持积极态度，使研究课有可能成为改进教学的更直接而有效的手段。

(3) 研究课是当前提高教师的教学水平和开展化学课堂教学研究的重要手段之一。

课堂教学是学校教育的基本形式；课堂教学质量是学校教育质量的基本保证，因此提高课堂教学质量是改进学校教育活动的途径。教师作为课堂教学活动的主导者，其教学水平直接影响课堂教学活动的质量。实践证明，仅仅依靠教师自身发现课堂教学活动中的问题并判断其原因是比较困难的。很多教师难以明确自己教学活动中的问题所在；更多的教师即使了解了问题所在却苦于没有解决问题的有效途径和方法。目前已逐步展开和推广的评优课和检查课对提高课堂教学质量已经产生了积极的推动作用。但是，评优课和检查课更多地只是提供了改进课堂教学的动力机制，也能够发现课堂教学中存在的某些问题，但却未能提供改进教学的途径和方法。从这个意义上说，研究课是评优课和检查课的必要补充。研究课不仅帮助教师发现问题，而且还要帮助教师解决问题，借以提高教师的教学水平，因而它是提高教师教学水平的重要手段。

研究课虽然是以被评教师个人的课堂教学行为为研究对象，但通过对大量教师的课堂教学行为的研究，必然能够发现其中带有普遍意义的一般课堂教学问题，这就为我们发现和掌握课堂教学的一般规律奠定了基础。在化学教育研究中，通过研究课，既可以发现化学课堂教学实际中存在的普遍问题，也可以将已经产生或引进的化学教育、教学理论应用于实际教学之中，积累丰富的第一手资料，使总结或概括教学经验、引进国外先进的教学理论和方法在研究课中得以检验和发展。这正是当前化学教育研究所急需的研究方法与手段，它必然会推动化学教育研究的进一步发展。

二 化学研究课的实施要点

1. 实施研究课的基本程序

(1) 课前准备。

要了解被评教师的有关想法，如对自己课堂教学的哪些方面不满意、希望得到什么样的指导和帮助等；熟悉被评教师的课堂教学方案；与被评教师共同讨论和制订课堂观察计划，确定课堂观察的重点行为及其观察方式。

(2) 课堂观察。

研究课的主要任务是观察和评价课堂教学活动哪些是有效的、哪些是无效的甚至是有害的，评价的依据主要是教师的课堂教学行为与学生对这些行为的反应。因此在实施研究课时，要根据观察计划，尽可能准确地记录教师的行为方式与学生的行为反应。由于课堂教学中的人际交往和信息交流相当复杂，在一堂课中观察的重点只能是 1—2 个。需要记录的观察信息是大量的，可以根据研究的重点、评价者的经验、研究的具体条件采用适当的记录方式，如采用录音或录像的方法，也可以用一种特制的专门表格，在单位时间里记录下教师的某一种行为及其学生的反应，如每隔 90 秒钟记录一次教师的行为（讲述、板书、实验等）和学生对教师行为的关注程度等。

(3) 课后分析与评价。

先对观察记录进行分析处理，对研究重点所涉及的教师行为及其学生反应进行归类，以此作为分析、评价的基本依据；然后与被评教师共同讨论，比较教师的实际教学行为与理想教学行为之间的差距，分析其原因。在对教师的课堂教学行为进行分析、评价时，不要套用某种固定不变的课堂教学模式和评价模式，而要注意结合教师自身的素质、个性等方面的特点，找到符合被评教师特点的问题的原因，做出准确诊断。在此基础上，制订出改进课堂教学的计划。这一改进计划又可作为制订下一次研究课方案的依据，从而保持课堂教学研究的连续性。最终实现对实际课堂教学的改进而获得研究课的实际效果。

2. 实施化学研究课的其他要点

(1) 研究重点的选择。

确定研究重点主要是根据被评教师的实际需要。一般而言，在化学课堂教学中，教师的提问方式、对学生的有效课堂学习行为的强化技术、板书的设计技巧、演示实验的操作与运用、化学教学语言艺术等，都可以作为研究课的重点。在化学课堂教学中，教师对演示实验和化学语言的运用，尤其要予以重视。一些教师特别是青年教师，往往不善于运用化学演示实验帮助学生理解和掌握知识、培养能力、激发兴趣；不善于以较严密的逻辑性去组织、展开生动而形象的化学教学语言，以引导学生实现从宏观到微观的思维过渡。可以运用研究课的手段，有针对性地解决类似的问题，切实提高教师的化学教学水平。

(2) 对评价者的要求。

相对于评优课和检查课而言，研究课对评价者有着更高的要求。首先，评价者应当具备丰富的化学教学经验和一定的化学教学理论水准，这样才有可能帮助教师准确发现课堂教学中存在的问题和提出切实有效的解决问题的途径与方法。从这个角度而言，与其说他是化学课堂教学的评价者，不如说是化学教学的研究者更合适些。其次，评价者不仅具备了必要的知识、经验

和能力，而且确实以帮助教师提高教学水平为唯一目的，并采取了相应的工作方式和态度，这样才可能与被评教师建立起良好的人际关系，获得他们的信任、尊敬与合作，保证研究课的顺利展开和取得积极的效果。通常，研究课的评价者只需要 1—2 人即可。

