

第一章

化学教学测量与评价概述

学习指导

(1) 掌握化学教学测量与评价的理论和方法,是适应化学教学管理和考评方法日趋标准化、科学化的需要,也是对新世纪化学教师现代教育科研能力结构的新要求。

(2) 化学教学的测量与评价是中学化学教学活动的一个不可缺少的环节,但它绝不是单纯为了证明教学效果或为了评等级、排名次,其目的在于对化学教学活动所涉及的主要因素及其效果给予价值上的判断,进而调节和改进教学,使教学过程真正成为一个首尾相接的可控制系统。

(3) 教学测量与评价主要包括教师教的质量(主要是授课质量)和学生学的质量(主要是学业成绩)这两大内容。

(4) 教师授课的质量直接决定着教学质量,所以对化学教师教学质量的测评又应以课堂授课质量为主,并以设计科学的评价指标体系作为评价标准。

(5) 由于学生的学业成绩既反映了学生的学业水平,也在一定程度上体现了教师教学的能力和水平,所以学生学的质量测评又应以化学学业成绩为主要内容。

(6) 考试、测验是衡量学生学业成绩的主要方式,学业成绩的测验与考试又分为常模参照性测验(用于衡量学生学业成绩相对水平)和目标参照性测验(用于衡量学生学业成绩的实际水平)两大类。因而学业成绩的考评又有相对评价和绝对评价之分。

引 言

教学是学校里一项最主要的工作,是培养人才、实现教育目的的基本途径。教学过程是由相互连接的各种教学环节组成的一个统一体,其中之一就是对教学质量的测量与评价,以提供信息改进教学和对被检测对象作出某种资格证明,这是教学过程的一个十分重要的环节。教学质量的测量与评价的范围很广,包括了教学活动的各个方面,如教学过程、教学成果的测量与评价。在中学,最大量的是对教师教学水平(包括教材处理、教学设计、教学方法、教学效果等)和学生学业水平的测量与评价。

长期以来,很多学校总是把考试作为衡量教学质量的惟一方式,把一次考试、几张试卷的成绩作为一把尺子来衡量教学质量,甚至以此来决定学生的命运,因而带有很大的局限性和片面性,不能反映出全部教学的质量。现实中存在的片面追求升学率及大学新生中一些人“高分低能”的现象以及教学工作无标准的状况都和缺乏对教学质量进行全面的、科学的检查与评价有关。

实际上,教学是教师教、学生学的统一活动,因此对教学质量的测量评价就应该主要包括教师教的质量(主要是授课质量)和学生学的质量(主要是学业成绩)这两大方面。

本书正是从这两方面入手,分别对化学教师授课质量和化学学业成绩如何实施科学地测量与评价进行研究。又由于这种测量与评价所采用的基本工具大多是各种形式的考试和测验,因此,本书也用较多的篇幅去探讨如何实现考试与测验的标准化、科学化。

第一节 化学教学测量与评价的概念及功能

一、什么是测量

测量(measurement),通常是指人们对客观事物进行某种数量化的测定。如对物体的长度、重量、温度以及时间、空间、运动等物理特性的测量。它是一种直接测量,其操作方法已被人们共同接受并有一些较好的量具来表现这些结果如天平、尺子、温度计等。这种测量具有以下三个基本要素。

(一) 单位

单位(unit)即计算数量的单位。如测量长度以米、厘米等为单位。这种单位的特

点是等值的 如 1cm 与 2cm 之间的间距等于 5cm 与 6cm 之间的间距。

(二) 参照点

参照点 (reference point) 即确定事物数量的计算起点。参照点有两种：一种是绝对的零点 即以“ 恰恰没有一点重量 ”“ 恰恰没有一点长度 ”为计算起点 对其测量只要统一单位就行了 另一种是人定为 的参照点 即相对零点 如以海平面为测量陆地高度的起点，以冰点为测量温度的起点等。最好的参照点是绝对零点，人定的参照点越接近零点越好。

(三) 量表

量表 (scale) 是具有单位和参照点的连续体，是测量的工具，也称量具。

实施物质测量 既要求量具本身的标准化 (即统一单位和参照点) 又要求使用量具方法的标准 (即统一操作程序)，否则就没有准确性和可靠性。

上述均为物质实态的直接测量。那么 对人类的精神现象 (心理现象) 如人的知识、技能、智力、兴趣、气质、性格、品德等精神特性 能否测量呢？

回答是肯定的，其测量所要依据的基本理论和方法就是本书所要依据的基本理论和方法——教育统计学、教学测量学和教育评价学。

二、什么是教育统计学

(一) 统计学

统计学是研究统计原理和方法的科学 属于应用数学领域。从广义上说 统计学是对于事物总结信息的搜集和分析，并以此为根据进行推断的方法和理论。

统计学分为两大类：一类是与研究对象的特征密切结合的各科专门统计学 另一类是数理统计学。

统计学最初从研究人口、土地等国情开始 经过三百多年的发展 至今已经成为一门拥有众多分支和具有方法论特色的科学。

(二) 教育统计学的产生与概念

教育统计学的历史并不长，到现在才 100 年左右。它是统计学和教育科学互相结合、渗透而产生的一门边缘科学 (图 1-1)

可见 教育统计学既是统计学的一个分支 又是教育科学的一个分支 但又与二者不同，所以它的定义为 “ 教育统计学是运用统计学中的一般原理和数理方法研究教育现象的一门应用科学 ”。

例如 著名的世界第一条遗忘曲线——艾宾浩斯遗忘曲线的内容和规律是教育心理学范畴的，而教育统计学在其中的作用则是形象地表达了规律。

（三）应用范围

教育统计学可用于正规学校教育的研究，也可用于业余教育的研究；它可用于儿童身心发展规律的特点的研究，也可用于教育事业发展规划和速度的计算；它可给心理实验结果以统计检验，还可为教育机关对人、财、物的科学管理提供必要的数 据 等等。无论哪一方面的研究 必须在充分占有数据的基础上才有可能科学分析和说明，从而掌握其一般发展趋势，探索出规律性的东西。

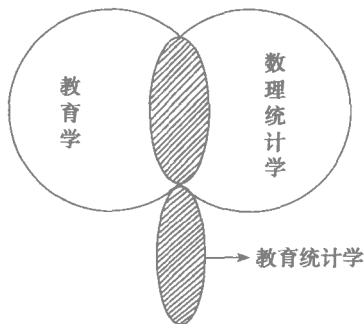


图 1-1 教育统计学是边缘科学

三、什么是教育测量学

（一）教育测量 学的产生与概念

教育测量 (educational measurement) 作为一门科学产生于 20 世纪初西方的客观测验运动，但其源流却可追溯到我国古代的考试制度。早在公元前 840 年的西周时期，我国就初步建立了学校教育制度，实行了教育考评。我国专门以考试为主选拔人才的制度始于隋炀帝大业二年（公元 606 年）开创的科举考试制度，它通过分科考试来选取人才 采用了贴经、墨义、策问、诗赋等方法去测试考生 此可谓开教育测量之先河。欧美各国 18 世纪末、19 世纪初采用的以考试的方法选拔官吏，是从我国学去的。

教育测量虽起源于我国 但是我国古代的考试，一般是试题的数量少 知识的覆盖面窄 不足以代表所测学科知识的内容 而且评分又难以保证客观性 因此 它不是科学的教育测量。

1904 年美国心理学家桑戴克 R.L.Thorndike 编著出版了《心理与社会测量》一书 该书介绍了心理统计方法和编制测量的基本原理。其后他又编制了《书法量表》、《拼音量表》、《作文量表》、《图画量表》等标准测量工具 使教育测量走上了科学化的道路。因此，桑戴克被称为教育测量学的鼻祖。

早期的教育测量主要用于客观化、标准化地考查学生学业。 20 世纪 30 年代起又扩展到考查学生的学习态度、兴趣、品德、性格以及教育工作的各个方面。因此 现代教育测量的定义应为：“教育测量就是对学生们的学习能力、学业成绩、兴趣爱好、思想品德以及教育措施上许多问题的数量化测定”。它主要用于对学生精神特性的测定。

（二）教育测量的特点

教育测量学是把教育统计学方法运用于考试分析而产生的一门学科，它主要研

究教育测验的编制、实施和测验结果的解释。教育测量和物理特性的测量一样，也具有单位、参照点、量表这三个基本要素，但它不像物理特性的测量那样可以直接测定，而是具有以下特点。

1. 教育测量一般是间接测量

这是因为人的知识、智力、才能是人的大脑活动情况，还无法直接测量。我们只能通过学生对测验题目的反应和一些行为表现，运用推理、判断的方法，来间接地测量出他们的知识水平、智力高低和品德好坏。

2. 教育测量的度量单位是相对的，且不等距

物理特性的度量单位一般是绝对的，1m就是1m，无论何地都相同（当然也是有条件的，具有一定的相对性，如在海平面上1kg重的物体，在高山顶上就会不到1kg重）。

在教育测量中，一个学生在一次班级测验中得70分，可能是这个班较高的分数，而又可能是别的班和外校较差的分数。这是因为不同班级考试题目难易不同，学生成绩分布情况不同，因而分数的价值并不相等。同样，一个考分为80分的学生的知识量并不等于在同一测验中得40分的学生知识量的2倍。因此，教育测量的度量单位是相对的。一个学生的学业成绩好坏、智力高低和能力大小等，都是就其所在团体的整个分数序列、行为序列中的地位来说的，其测量的分数单位并非绝对的。同时，分数单位又非等距的，如56分和58分、98分和100分，同样差2分，但性质不同，后2分的取得可能在难度上要大于前者。虽然在教育测量中所使用的单位并非绝对等值，但教育测量必须有单位，则是绝对的。

3. 教育测量的参照点是相对的

教育测量中所用的参照点，可以说都属于人定的参照点。而人的心理现象是没有绝对零点的，即使一个学生考试得了零分，也不能说他的化学知识为零。所以参照点是相对的。另外，教育测量中用的参照点还有一个很大的限制，即从参照点计算的分数，不能以“倍数”的方式解释。如甲生的化学成绩是100分，乙生是50分，我们则不能说甲生的化学知识掌握程度是乙生的2倍。

教育测量应用的参照点有许多种，有以编制测验者推测的零点作为参照点的，有以一个年级或一个年龄段的学生作业的平均分数为参照点的。由于测量应用的参照点不同，所测得的分数也不相同。为了对这些不同的分数进行比较，必须进行分数的转换。这也是不同考试的原始分数非但不能直接相加，也不能直接加以解释的原因。

4. 教育测量要控制的变量很多

如题目编制本身、测试过程、被试者的心境、内部条件等。所以教育测量要达到科学化，首先要实现标准化。

5. 教育测量要制作标准化量表

要测量某种事物，就需要先有一个具有单位和参照点的连续体，然后用这个连续体去测量某种事物，以确定该事物的数量，这个连续体就叫量表。教育测量中所使用的量表多以文字试题、考卷的形式出现，也有以图形、符号、操作要求形式出现的。

教育测量常用的量表有百分量表、常用于测量学生学业成绩的年龄量表、等级量表、T 量表等。

（三）教育测量的种类

教育测量的种类随分类标准不同而有所区别。

1. 以测量的对象来分类

（1）学业成绩测验：这类测验用于测量学生的知识、技能，如数学测验、化学测验等。也有人将此类测验称为教育测验，这是本书主要研究的内容。

（2）智力测验：这类测验用于测量学生的观察、识记、分析、判断、推理等思维活动的能力，如迷津测验、填图测验。

（3）人格测验：这类测验用于测量学生智力以外的心理的各种特性，如性格测验，兴趣、情感测验等。

（4）特殊能力测验：这类测验用于测量学生某种特殊能力的程度，如音乐能力、绘画能力测验等。

2. 以测验结果的参照点来分类

（1）常模参照测验：其目的在于鉴别或比较被试者之间的个别差异，也就是确定每个被试者的成绩在集体中所处的位置，比较所用的参照点就是测验的常模。所谓常模，就是指该考生集体的平均分、标准差和成绩分布状态。某人成绩在常模之上则高于平均水平，在它之下则低于平均水平。入学统考、高考、化学竞赛一般都具有这类常模参照测验的性质。

（2）目标参照测验：其目的在于检验被试者是否完成规定的教学目标（见第二章），以及完成的质量如何。这种测验的参照点则是事先制定好的教学目标或根据它浓缩了的测量目标，如摸底测验、期中期末考试以及不计入高级学校招生考试录取总分的毕业会考等，都属于目标参照性考试。

这两种测验在功能和参照点方面虽然明显不同，但这种划分也不是绝对的，常模参照测验的结果也能反映某一学科部分的（而且是主要部分的）教学目标的完成情况，而目标参照测验一般也能反映出被试者在总体中的位置。近年来也有将这两种类型的测验结合进行的尝试。

四、什么是教育评价学

（一）教育评价学的产生与概念

对教育评价（educational evaluation）我们并不陌生，在教育工作中常用的考核、甄选、检查、总结、评比、奖惩、视导等都含有教育评价的因素，是教育评价的原始的、初级的形式，只是未用其名而已。

在 20 世纪 30 年代才第一次出现“教育评价”的概念，它是在 20 世纪初出现并盛

极于美国的教育测验运动的基础上产生的。最早呼吁人们把注意力从“测验”转向“评价”的是美国教育学家泰勒(R.W.Tyler)。

20世纪30年代初,美国进步主义教育协会在全国挑选了30所中学进行课程改革研究,历时8年(1933~1940),又称“八年研究”。泰勒教授担任了“八年研究”评价委员会主任,为新课程实验设计了一套有利于教育测验和考试的教育评价方法。因此,现代教育评价的观点诞生于20世纪30年代的“八年研究”。从20世纪30年代到50年代末,泰勒的评价思想一直占据着教育评价领域的主导地位。

20世纪50年代中期,泰勒最杰出的学生本杰明·S·布卢姆(B.S.Bloom)发表了《教育目标分类学——认知领域》,成为现代教育测量和评价的重要基石。自此,教育评价思想发展较快。从20世纪70年代至今,教育评价已进入成熟期。我国虽在20世纪80年代才重新将科学教育评价提到议事日程,但发展也较快。

从教育评价产生、发展的过程可见,教育评价概念是一个变化着的概念,其方法也在不断改进之中,但是,其中有一点是始终不变的,即教育评价乃是对教育进行价值判断、为教育决策提供信息的过程。泰勒于1986年在其新著《教育评价概念的变化》中,把教育评价归结为“检查教育思想和计划的过程”。

所以,我们可以这样去阐述教育评价的概念:教育评价是根据一定的教育目标,运用教学手段,对教育效果进行价值判断,为教育决策提供信息的过程。

(二 教育评价的范围和分类

教育评价涉及的范围很广,主要有:学业成就评价、智力能力评价、性格、道德品质评价、身体评价、环境评价、教师评价(主要包括教师资格和教师教学质量的评价)、课程、教材评价、学校评价等等。

在使用各类测验,收集信息资料进行评价时,根据评价的基准不同,教育评价又可分为绝对评价和相对评价。

(三 教育评价的过程

其过程一般可分为三个阶段:第一,具体形成教育目标,即教育目标的确定;第二,设计可以准确测量这些目标的程序,即编制各种测验;第三,综合各种资料和数据并形成价值判断。

(四 测验、测量、考试、评价等术语的相互关系

测验是引起某种行为的工具。通常是指运用某些仪器、试题来引起人们的某种行为,从而测定人们的某种特征,它是进行数量化分析和科学推断的前提和手段。测量比测验的含义要广泛些,测量不仅包括着运用仪器、试题来测定事物的质量与特性,而且还包括着运用调查、观察等方法来测量事物的质量与特性。不过在以往的测量书中,测验与测量常互相通用。

测验与考试也不尽相同。我们平时所说的考试,通常只凭教师自己的经验去出

题评分,带有主观随意性,而测验则是经过较细致的科学分析才编制出题目,在测验的程序和评分方面也有较严格的要求;考试一般用于考核学生的学业成绩,测验不仅用于此,还使用于心理特征的测量。

当然离开教育测量学的观点去谈考试与测验,二者在有些场合没有区别,如期中考试与期中测验。但有些场合则不能混用如“高考”不能说成是“高测”,“心理测验”不能说成“心理考试”。二者在汉字中出现的时间也不同,“考试”一词于东汉时出现,而“测验”一词直到 20 世纪初才在汉语中出现。

一般而言,测验的应用较广,它不含淘汰的意思,其目的在于了解情况,如小测验、时事测验;而考试则用于比较正式的场合,如毕业考试。如考试与测验作为测量与评价的一种手段,两者可以不加区别。

测量与评价既有联系,也有区别。测量是对客观条件的一种“铁面无私”的数量化描述,重在量的测定,有较大的客观性,但若对其结果不进行评价就是无意义的东西。所以测量的结果是评价的重要依据,但又不是惟一的依据。因为评价比测量的范围要大,评价需要质量并重,主客观兼容,依据更多的方法和资料进行。如对学生毕业成绩的评价,除依据测量所得分数外,还要依据学生的平时作业、课堂回答、以往的学习成绩等情况。

在 20 世纪 30~40 年代,测量、评价不分,20 世纪 50 年代以后分开来了,因而有了独立的教育评价学。

五、化学教学测量与评价的概念

教育是指学校中培养人的全部活动,教育测量与评价也就涉及了教育工作的各个方面,而教学只是这全部活动中的一部分。教学,是学校管理中一项最主要的工作,教学是教育这个种概念中的属概念。它是由相互连接的各种教学环节所组成的一个统一体——教—学—管。其中之一就是对教学质量的检查与评定。这是教学工作必不可少的重要组成部分。它可以判别教学的质量如何,进而对教学起到反馈调节作用。这种对教学达到了一定质量的检测与判断过程,就是对教学的测量与评价。

化学教学测量与评价就是根据教学目标,运用可行的测量技术对化学教学活动所涉及的主要因素及其效果给予价值上的判断,从而为调节和控制教学及对被评价对象作出某种资格证明。

由于测量与评价是两个既有区别又有联系的概念,我们也可分而述之。

教学测量的定义:教学测量就是对教师的教学质量和学生的学业成绩作数量化的描述。

教学评价的定义:教学评价是根据教学测量所获取的资料、数据等信息比照教学目标对教学效果作出价值上的判断,它是为教学决策提供依据的过程。

教学测量的目的是为教学评价提供信息,而教学评价的目的则是通过价值判断为教学决策提供依据。因此,教学测量是实施教学评价的工具。但如果在教学测量之

后,不进行教学评价,教学测量就毫无意义。所以我们把教学测量与评价联系在一起去研讨。

要指出的是 合理的、有积极意义的教学评价 又必须建立在可信可靠的教学测量的基础之上,如果教学测量的数据可疑,则据此进行的教学评价也无意义。

例如,在一所学校的初中化学毕业考试中,第一年的优良率为 50% 第二年为 60% 第三年为 70%,于是人们就会习惯地以此声称化学教学质量逐年提高。但当我们把三年的化学试卷放在同一水平的三个班级中去测试,发现第一张试卷的优良率是 50% 第二张是 80% 第三张是 90% 可见三张试卷一次比一次容易 这样的测量本身就没有可比性,当然据此作出的评价也是无意义的。

六、化学教学测量与评价的主要功能

(一) 诊断功能

通过教学测量与评价,可以了解教学的各方面情况,从而判断它的质量和水平、优点和缺点、矛盾和问题,为制定解决问题的策略提供依据。这是其基本功能。

(二) 强化功能

也就是通常说的鼓舞监督作用。通过测评,可以帮助师生及时了解教学活动的进程与效果,不断得到反馈与强化,调整教学方法和学习方式,从而推动教学活动的进行。

(三) 甄选功能

测量与评价也是选拔人才、鉴别能力的有力工具。通过教学测量与评价,能够对学生的学成成就作出鉴定,选拔社会需要的人才;也能够对教师的教学水平作出鉴定,为人事决策提供依据。

(四) 教学功能

测量与评价本身也是一种教学活动。因为在这个活动中,学生在知识、技能、技巧等方面也获得了长进。

可见,教学离不开测评。也可以说,教学测评直接作用于教学活动的各个方面,无论是教学目的、任务 还是教学过程、环节和方法。例如 怎样确定教学目标 教学水平高低的标准是什么?怎样考核教师的教学工作?怎样考核学生的进展与发展?对这些能否作出科学、准确的测评,不仅直接决定了教学效率的高低、培养目标的落实、培养人才是否符合国家需要,还直接影响着师生的积极性。

显然,教学测量与评价是提高化学教学质量的重要手段,是教学工作必不可少的一个基本环节。

教学测量与评价的功能还可以从“三论”特别是从信息论上得到新的、有力的论证。用信息论观点来解释,教学质量的测评,使师生双方不断获得反馈信息,从而使教学过程真正成为一个“首尾相接”、“循环往复”的可控制的系统如图 1-2 所示。

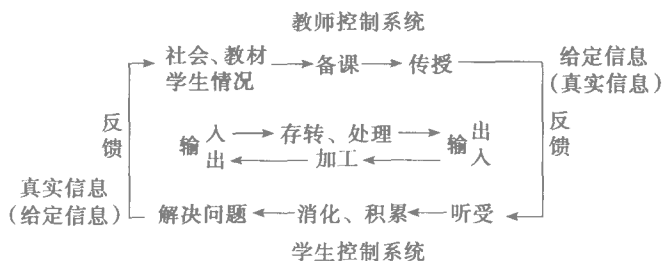


图 1-2 教学测量与评价的功能

第二节 化学教学测量与评价的基本类别

化学教学测量和评价涉及的范围虽然较广,但由于教学是教师教、学生学的统一活动,所以化学教学测量与评价也可以划分为教师教的质量(以授课质量为主)和学生学的质量(以学业成绩为主)这两大基本类别。

一、教师化学教学质量的测量与评价

教学中 教师起着主导作用 教学的方向、内容、方式、方法、进程、结果 都是由教师的教授决定的;学生的学习动机、学习方法以及学习效果,都要受到教师教授的影响,教师授课的质量直接决定着教学质量。因此,对化学教师的教学质量的测量与评价应该以教师授课质量为主,具体方法将在第五章专门研究。

(一) 化学教师授课质量测量与评价的一般类别

课堂教学质量测评的内容很多,一般而言,它包括了教学目标、教学内容、教学结构、教学手段、教学状态、教学效果等六大方面。但是实际的课堂教学质量测评并不都是全面而系统进行的,更多的往往是根据某些具体的目的而进行的专门(分类)测评。可以根据不同的需要,从各种侧面、各种角度展开。因此,也就形成了课堂教学测评的若干种类,如:

- (1) 根据不同的评价主体,可以分为自我测评和他人测评。
- (2) 根据不同的评价目的,可以分为:

- 1)合格课测评。
- 2)优质课测评。
- 3)典范课测评。
- 4)检查课测评。
- (3)根据不同的评价内容，可以分为：
 - 1)新授课测评：基本概念、基本理论、元素化合物知识、有机化学知识。
 - 2)实验课测评：演示实验、指导学生分组实验。
 - 3)复习课测评。
 - 4)习题课测评。
 - 5)讲评课测评。

(二) 化学教师授课质量的测评标准

化学教师授课质量的测评标准与学生学业成绩的测评标准不同，它不像学生测评的认知内容和目标那样明确，难以完全用测验的方法取得量化的结果，往往是通过设计测评的指标体系作为测评标准。如可以设计较为通用的关键指标作为一般标准(如上述六大内容)也可以分门别类地设计具体的测评指标作为分类标准。

二、学生化学学业成绩的测量与评价

学生的学业成绩是反映教学效果的一种有效方法 教师的能力和水平、教学工作实施过程中所作的努力，最终都应从学生的学业成绩反映出来。所以化学教学测量与评价的研究又应以学生化学学业成绩的测量与评价为主要内容。

(一) 化学学业成绩的测量

迄今为止，测验考试仍然是衡量学生学业成绩不可取代的重要方式。所以学业成绩测量也可按不同教学阶段测验考试的特点和要求，分为不同类型的测验或考试。

(表 1-1)

表 1-1 学业成绩测验 考试 的类别与特征

测验类型	测验目的	测验目标	测验特征
诊断性测验(一般在教学之初进行)	1. 了解学生是否具有学习新课程所需的知识能力 2. 了解学生对未学的课程目标所能掌握的程度(即潜力)	1. 学习新课程所必需的知识和能力 2. 新课程教学目标的代表性样本	根据每次测验的功能确定它应属于目标参照还是常模参照测验
形成性测验(一般在教学进程中进行)	1. 为学生和教师提供学习进步情况的反馈信息 2. 诊断学习困难的原因	1. 如有可能，包括所有的单元教学目标或最必要的目标 2. 以学生中存在的各种可能的错误为依据	测验一般不难(难度系数较大，属于目标参照测验)
总结性测验(一般在教学结束阶段进行)	对学生的学习和水平达到何种程度作全面检查	选择课程目标的代表性样本	主要属于目标参照测验

（二）学业成绩的相对评价和绝对评价

由上可见，各种类型的学业成绩测验又都可概括到两大类测验之中。所以通常所言的学业成绩测验也可简单地划分为常模参照测验与目标参照测验两大类。

教学评价与教学测量是对应的，有一类测验就有一类评价。因此，教学评价若按教学进程划分可以分为三类：诊断性评价、形成性评价、总结性评价。若按分数解释的参照点划分，又可对应为常模参照评价和目标参照评价。

上述各类评价还可根据各自的特点统一归为两大类：相对评价和绝对评价。具体内容见第三章。

第三节 化学教学测量与评价的常用方法和步骤

一、常用方法

（一）考试、测验法

此法主要为评价学生学业成绩和能力所用，可分为标准化考试（测验）与非标准化考试（测验）两种。这种方法可在较短时间内同时收集许多学生学习状况的反馈信息，便于作出定量分析，且易于掌握，所以得到了沿用。其局限性在于：在限定的时间内能够考查的内容和范围很窄，学生学习课程的兴趣、能力、意志力、创造性等状况用考试、测验的方法较难测出。此外，非标准化的考试、测验（如教师自编测验）还不可避免地带有主观性和片面性。

（二）观察法

观察法是在教育的自然场景下了解观察对象的方法。此法与测验法截然不同，被观察者像往常一样地活动和学习，不会产生或感到任何的压迫感。所以此法收集的资料始终是被观察对象的常态表现，都是自然、真实的。它用于测评被观察者的能力、行为、性格、态度、纪律等不易为考试测验所获得的内容，也可用于了解学生的化学实验操作技能。

这种方法适用范围广，条件的宽容度大，不用特殊设备和工具。但其短处是：第一，由于观察者个人的主观偏见，或观察中有先入为主的倾向，易将主观印象带入观察记录中，作出有违事实的判断；第二，重要的、有价值的资料的获取，若只限于在某个环境中的一次观察是无法实现的，一般来说，只有在各种各样的场景中多次反复观

察才可能获得。由此可见，此法是很费财力和精力的。

(三) 问卷调查法

问卷调查法是以书面提出问题的方式搜集资料的一种研究方法。研究者将所要研究的问题编制成问题表格，以邮寄方式、当面作答或追踪访问方式填答，从而了解被调查者对某一现象或问题的看法和意见，所以又称问题表格法。问卷调查区别于调查表法，它调查范围较窄，偏重于意见、态度或看法，并往往以个人或一群人为对象。

问卷调查法的优点在于方便实用、省时、花钱少。由于可以不署名，在某种情况下结论比较客观；能搜集大样本信息资料，收效大；便于整理归类，能做量的统计处理，使调查结果具有一定代表性。该调查方法的局限性在于：如果问卷中的问题不明确或题量过大，或被调查者不合作，都会影响结论的代表性。应用范围较广，搜集的资料往往是表面的，还不能深入了解深层次的内心世界的真实情况；若部分调查对象不作回答，难以知道不回答的原因，也会影响问卷的效度。问卷法的运用，关键在于编制问卷、选择被调查者和结果分析。

1. 问卷类型及题型

(1) 结构型问卷：结构型问卷也称为封闭式问卷 (closed form questionnaire)，是把问题的答案事先加以限制，只允许在问卷所限制的范围内进行挑选。

结构型问卷包括以下题型：

1) 是否式：把问题可能的答案列出两极情况，从中择一，“是”与“否”；“同意”与“不同意”。

2) 选择式：从多种答案中挑选最适宜的一个或几个答案，然后做上记号。

3) 评判式：每个问题后列有许多答案，要求被调查者依其重要性评定等级或顺序，所以评判式也叫排列式、编序式，是用数字表示几种答案的排列顺序。

4) 画记式：按同意或不同意，在答案上分别画“√”或“×”。这是一种核对表形式。在核对表的细目中，被调查者通过选择一个提供选择的答案来回答，与选择式、评判式不同之处在于，答案在连续统计上并不代表分点，而是称名类型。

(2) 非结构型问卷：非结构型问卷也称开放式问卷 (open form questionnaire)，问卷由自由作答的问题组成，是非固定应答题。

这类问卷提出问题不列可能答案，由被调查者自由陈述。就题型分析，可以是填空式的，也可以是回答式的。

非结构型问卷往往用于以下情况：一是较深层次的问题研究。被调查者不受研究者和题目答案选择范围已界定的限制，按各自对问题的理解回答。这种问卷能如实地反映出被调查者的态度、特征、对有关情况的了解程度以及所持看法的依据等。因此可用于探讨那些只能进行描述性分析的较复杂问题，以及获得有关人士对某些问题的看法。二是在研究初期，对所研究的问题或研究对象的有关情况还不十分清楚的情况下，采用开放式问卷，来帮助研究人员设计封闭式问卷。一般做法是：在小

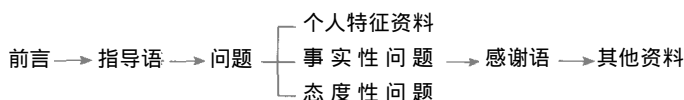
范围内进行问卷调查,并对搜集的资料进行归纳分析。在掌握相当的资料后,再采用结构型问卷进行较大规模的调查和进行定量分析。因此,在一定意义上,开放式问卷调查正是封闭式问卷调查的基础。

这种问卷调查搜集到的材料丰富、具体,往往能得到许多意想不到的、很有价值的资料。但由于答案不集中、材料分散,难于对答案进行横向比较,所以不易进行统计处理。

(3) 综合型问卷 综合型问卷 (comprehensive form questionnaire) 形式一般以封闭型问题为主,根据需要加若干开放性问题。也就是说,将研究者比较清楚、有把握的问题作为封闭性问题提出,而对那些调查者尚不十分明了的问题作为开放性问题放入,但数量不能过多。经调查,在积累一定材料的基础上,问卷中的某些开放性问题就有可能转为封闭性问题,这也是问题设计时常常使用的技巧。

2. 问卷的组成

问卷设计质量的高低是问卷法实施成功与否的决定性因素,因此有必要对问卷的结构做一个充分的了解。问卷一般由如下几个部分组成:



3. 问卷的编制与实施程序

问卷的设计过程,是研究者根据调查研究的目的和需要编写问题和形成问卷的过程。编制程序包括下列步骤:

- (1)明确研究目的,根据研究目的和假设范围搜集所需资料,并确定调查对象。
 - (2)列出问卷调查所需研究问题的纲要,确定所要搜集的信息和问卷类型。
 - (3)围绕主题草拟问题,列出标题和各部分具体项目。
 - (4)征求有关人员、专家的意见,修订项目。
 - (5)试测,从总体样本中抽取 30~50 人作为试测样本,以检查问卷表述的方式、项目、内容能否被受试者所理解,并求出信度、效度。
 - (6)再修订。根据试测结果,对项目内容、排列方式加以改进,然后打印。
- 至此,问卷的编制工作完成,可以按计划发放问卷,进行正式调查。

问卷法的基本实施程序可以概括如图 1-3。

4. 设计问卷问题的基本要求

问卷问题的设计是问卷编制的关键,应考虑以下几点:

(1)问题和范围:是用于小范围的典型调查还是大范围的统计调查;是了解人们思想态度方面的意向性问题,还是主要了解过程方面的事实材料。

(2)问题的内容:所列项目对研究目的是否具有较好的覆盖面,答案要能较全面反映要研究问题的主要方面,且不交叉、重叠。

(3)问题的数量 问卷的长度要适度,一般以 30~40 分钟的作答时间为宜。问题

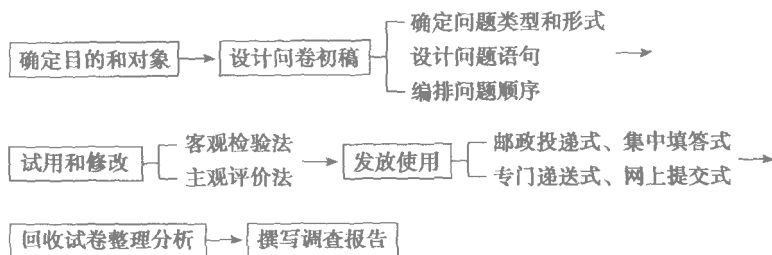


图 1-3 问卷法的基本实施程序

太多 被调查者容易产生厌倦情绪 导致敷衍塞责或不予回答 若问题太少 又不能得到有关研究的基本事实材料以致影响研究结论。因此, 可问可不问的问题要删除。而一些较复杂的甚至要查资料才能回答的问题要尽量避免列入。

(4) 问题的文字表达: 一般要求文字表达准确 简明扼要 容易回答。在结构上, 一个问题只宜含一个疑问, 不应包含两种以上内容的提问; 在语言上, 一般不用假设或推测语, 切忌因繁杂和意义含混而引起误解或无从回答。

(5) 问题的排列顺序 调查表首 要说明为什么要进行调查 要解除被调查者的顾虑。关于被调查者的个人特征资料应放在问卷的前面部分; 能引起兴趣的问题、简单的问题放在前面, 而容易引起紧张的、牵涉个人问题的或复杂的问题可放在后面。同时要按内容或性质, 把以同类方式回答的问题编排在一起。封闭式问卷中, 划分水平程度的答案, 由低到高, 或由高到低, 要随机排列, 以免产生定势而不认真作答。另外, 问题的长短要适度, 并尽可能在选择答案中分出等级。一般为三至七级, 多为五级 如“很差、比较差、一般、比较好、最好”。

5. 问卷的发放回收

(1) 确定发放的形式 问卷发放有不同的形式 且各有利弊。无论哪种形式 在卷首均应说明调查目的及填答要求。邮寄问卷时, 应附回件邮资, 收到填好的问卷后要写封感谢函。

(2) 问卷回收率的计算: 一般而言, 回收率如果仅 30% 左右, 资料只能作参考; 50% 以上 可以采纳建议 当回收率达到 70% 以上时, 方可作为研究结论的依据。因此 回收率一般不应少于 70%。

如果有效问卷的回收率不足 70% 要再发一封信及一份问卷 follow-up)。

二、基本步骤

(一) 确定评价的依据

即确定教学目标。教学目标是师生通过教学活动应取得的预期教学成果, 它是用可观察的师生行为来描述的。如在某一单元教学后, 学生应该掌握哪些知识内容,

应该会做什么。只有将教学目标用表达得当的、可测量的语言确定下来，才能实施测评。目标确定了，测评的目的、方向也就随之确定了。因此，教学目标乃是教学测评的依据。进行教学测评的第一步工作就是要制定教学目标。

（二）选择和设计测量工具、实施测量

即设计可以准确测量教学目标的程序。如用调查法、考试法或制定测评指标体系等。要注意所用的评价工具应当具有高效度、高信度，能包含问题的全部，评分客观 容易实施 记录方便 解释利用具体 经济实用。

（三）作出价值判断

即综合各种资料和数据与教学目标比较并形成价值判断。

三、注意事项

- （1）要注意相对评价与绝对评价相结合，注重绝对评价。
- （2）要注意定性评价与定量评价相结合，注重定量评价。
- （3）要注意形成性评价与总结性评价相结合，注重形成性评价。

教学测量与评价是一项十分复杂和艰巨的工作，需要懂得教育统计、教育测量、教育评价等学科的基本原理和方法才能做好。应当提倡定性评价与定量评价相结合，注重定量评价。凡能量化的应以量化，凡不能量化的也不要硬去量化。为了更准确地运用教育测量与评价的基本原理，对如下七个问题的反复思考是很有必要的。有人将其归纳为‘6W-H’思考方法 或曰‘七个何’评价方法。

Why——为什么评 即评价的目的（为何评）

What——评价什么 即评价的对象（评何事）

Who——谁评价 即评价者和评价组织（何人评）

Whom——评价谁？即教学情况所属者（评何人）。

Where——在什么地方评 即评价的环境（何地评）

When——什么时候评 即评价的时间和阶段（何时评）

How——怎样评 即评价的标准、方案、方法、指标和如何实施 以及怎样作出判断结论 并提出建议（如何评）