

试卷编制的方法与技巧

肖 鸣 政



Shijuanbianzhi
defangfa
yujiqiao

江西教育出版社

试卷编制的方法与技巧

Shijuan bianzhi de fangfa yu jiqiao

● 肖 鸣 政

江西教育出版社

前 言

编制试卷，对于每个教师来说，几乎都是不可避免要做的工作。为了检查自己的教学效果，或者了解学生对教学内容掌握的情况，每个学期教师都必须对他或她的学生进行数次测验或考试。然而，由于种种原因，解放以来试卷编制的理论与方法在我国没有受到足够的重视。师范院校一直没有开设有关这方面的课程。只是党的十一届三中全会以来部分高师的教育专业才对这门课程略有介绍，非教育专业仍然没有开设这门课程，以至有关编制试卷的方法与技巧，广大教师不得不在实践中花费大量的时间去摸索，重复前人的劳动。为此，我们在赣南师范学院部分系科中开设了《教育测量学选讲》课程，使非教育系的学生也能懂得有关试卷编制的理论，能系统地掌握有关试卷编制的方法与技巧。《试卷编制的方法与技巧》一书就是在总结几年来的教学经验的基础上，结合目前我国考试改革的实际需要，为适应广大师生自学而编写的。其目的除了想把它作为师范院校学生课外阅读参考书外，主要是想帮助广大中小学教师系统地学习有关试卷编制的基本方法与技巧，促进校内日常课堂考试的规范化与科学化，适应并积极配合即将在我国普遍推行的标准化考试。具体地说，主要想通过它来帮助广大中小学教师解决以下几个问题：（1）如何设计试卷来满足特定教学目的的要求与特定考试的需要；（2）如何编写能测量学生知识水平、理解能力和应用能力的各种试题；（3）如何通过试卷反馈信息的分析，改进教学和提高编制试卷的质量；（4）如何获得某些编制试卷的专门技巧；（5）如何科学地解释考试

的结果。

全书共分十一章,依次围绕编制试卷的自然顺序而展开。试卷是测试的工具,为此必须首先了解测试的种类、特点及其作用,这是第一章的内容;试卷是用来测试教学效果的,故教师必须对教学目标或学习成就进行分析并明确其结构,这是第二章所讲的;为了保证试卷编制工作展开的高效性,教师必须对试卷的质量特征和整个的编制程序有个全面的把握,这将在三、四两章叙述;如何编写客观型试题来测量学生所具备的知识和能力,将分别在五、六、七三章叙述,其中对于选择题的编写给予了特别完备的介绍;如何组织试题成为试卷,将在第八章叙述;测试后如何分析试卷中的各种反馈信息,改进教学与试卷编制的工作,这将在第九章叙述;论文型试题是我国的传统试题,教师应对它的特点、应用及其评分误差的控制有足够的了解,这将在第十章叙述;最后,测试结果出来了,教师应知道如何解释或处理它才比较科学,这就是第十一章所要讲的。在所有这些章节中,都广泛吸取了国内外经典著作中的精华和近年来广大教师在命题研究和改革中的最新成果。同时力求以通俗易懂的叙述和各种典型例子来说明有关试卷编制的方法和技巧。

本书的编辑出版,《赣南师范学院学报》编辑部做了大量工作,郭贤逢副主任、副编审对书稿作了审阅、修改,肖家亨副教授、南昌22中姚燕平老师以及赣南师范学院数学系、化学系选修《教育测量学选讲》课的学生也提出了许多修改意见,在此一并表示衷心的感谢。

由于水平有限,时间仓促,书中不免存在错误与疏漏,谨请读者批评指正。

编者

1988年10月于赣州

目 录

第一章 绪论	(1)
一、教育测量学在我国发展的概 况	(1)
二、教育测量学中的几个基本概念	(3)
三、测试的种类及其特点	(7)
四、试卷在教学过程中的作用	(14)
第二章 成就的种类及其分析	(19)
一、目标的表述	(19)
二、目标的分析	(23)
三、成就的种类	(29)
第三章 衡量试卷质量的主要指标	(32)
一、效度	(32)
二、信度	(35)
三、其它质量指标	(39)
第四章 试卷编制前后的有关工作	(45)
一、一般校内日常测试的流程	(45)
二、测试的目的及其决定	(46)
三、测试内容的确定及其组织	(49)
四、试卷的应用与分数的评定	(51)
五、测试内容的分析与确定	(54)
六、细目表及试卷蓝图的制定	(57)
七、其它工作	(62)
第五章 测量知识水平常用的试题类型及其编写	(67)
一、填空题与简短答案题及其编写	(72)
二、是非判断题及其编写	(80)

三、选择题及其编写	(83)
四、搭配题及其编写	(114)
五、小结	(120)
第六章 测量理解能力常用的试题类型及其编写 ..	(122)
一、测量对概念理解的试题及其编写	(123)
二、测量对抽象结论或复杂概念理解的试题及其编写 ..	(128)
三、小结	(132)
第七章 测量应用能力的试题类型及其编写	(133)
第八章 试卷的组织过程	(140)
一、试题的选择	(140)
二、试题安排顺序的确定	(141)
三、标准答案的制备	(142)
四、审查试卷并给出指导语	(144)
五、打印前后的工作	(145)
六、答卷的采用	(147)
第九章 试卷的分析与试题的修改	(150)
一、整体分析	(150)
二、试题分析及其修改	(160)
第十章 论文型试题的编写与评分	(176)
一、论文型试题的特征和用途	(177)
二、良好论文型试题的标准	(180)
三、论文型试题的编写	(183)
四、论文型试题编写具体过程的示范	(189)
五、论文型试题的改进与评分	(191)
六、统考中论文型试题评分误差的控制	(208)
第十一章 试卷分数的解释	(223)
一、常用分数的种类及其转换	(223)
二、成绩的评定与分数的解释	(233)

附录.....	(237)
一、试卷样本示例	(237)
二、把一个综合目标分解为行为分成的实例	(256)
三、目标双向细目表	(258)
四、目标库中的目标卡样本	(260)
五、试题编写检核表	(261)
六、试卷分析实例	(267)

第一章 绪 论

一、教育测量学在我国发展的概况

教育测量学是一门以考查和评定学生学业成绩为研究对象的科学。它是本世纪初兴起的一门新学科，主张通过系列化的工作，把试卷的编制建立在合理的难度、区分度、效度和信度的基础上，把评分结果的解释和分析奠定在某种统一的客观标准上。因而，它能比较如实可靠地测量学生的学业水平。实践证明，在教育测量学理论指导下编制的试卷，对于人才的选拔、教学质量的检查、教育管理以及教育理论的科学化起了重大的作用。近年来，随着效标参照测试的出现，以及计算机技术带来的测量手段的现代化——计算机题库、计算机阅卷和计算机统计分析的出现，极大地推动了教育测量学科的发展。教育测量学受到了人们从来没有过的重视，许多国家已把它作为师范学生的必修课。

教育测量学产生于19世纪末，辛亥革命后就引入了我国，当时不少师范学校都相继开设了这门课程，许多学者都从事于这门学科的研究。然而，由于解放初期我国受着苏联的影响，把它斥之为伪科学，因此教育测量学在我国长期以来一直是无人敢于问津的领域，以致许多教育工作者对这门学科一无所知，不懂得还有专门研究测试及其试卷编制这门学科。所以迄今为止，不少人还是沿用过去那种翻书找题、拍脑袋出题的经验做法。其结果是使测试的效度、信度、难度和区分度都得不到保证，评分缺乏客观统一的标准，考试

无法反映学生的真实水平，常常是高分低能。试题有时太难有时太易，有时数量过少，有时又内容太偏。1983年北京师范大学心理学系测试研究小组曾运用试卷分析的理论和方法，对1978~1981年四年的高考试卷进行了分析，发现只有少数几份试卷具有较高的信度，即在0.85以上，其余的都比较低。尤其是政治试卷，1981年的信度仅为0.317；不少试卷的效度竟出现了负值，同时许多试题的难度也都不合要求。例如，增设附加题，一般是想把尖子学生与普通学生的分数距离拉开，以便优中择优。这就要求附加题的难度应该比正题大一点（此处的难度定义为没有通过某题考生的比率）。但1981年语文试卷的附加题的难度仅仅为0.27，比试卷中大多数正题都更容易，这显然就失去了增设附加题的作用。由此，人们不禁要问，既然涉及千家万户、牵动亿万人神经，由专家编制的高考试卷，倘有如此许多的不足，那么日常教学中一般教师所编制的试卷谁知会有多少缺点与错误呢？

在举国上下坚持改革开放的今天，在全国标准化考试即将全面展开的形势下，那种闭门思索、翻书找题的经验命题法早已无法满足于教育与社会的需要了。专家早已指出，目前我国流行的传统命题方法早已不利于考生智能的发展和德、智、体的健康成长，有碍于我国智力的开发和人才的挑选工作，必须尽快改革这种陈旧的测量方法。国家教委锐意改革，1985年已在广东进行了高考标准化的试验，1986年又在山东、辽宁、湖北、广西等省扩大试验，1990年将在全国普遍推行。因此测量学生成绩方法的科学化、现代化是大势所趋，人心所向。党的十一届三中全会以来，我国已在师范教育专业中开设了教育测量学这门课，现在有些师范院校又把

它作为选修课在非教育专业中开设，这对我国高考的标准化，课堂测验及日常教学测试的科学化将有着重大的意义。

二、教育测量学中的几个基本概念

1. **测量**：测量的英文是Measure (ment)，按辞海的注释是：用量具或仪器来测定零件的尺寸、角度、几何形状或表面相互位置的过程。显然这一定义的范围过于狭窄。推广一点，我们可以把测量定义为：测量是按照一系列规则用数字描述事物或对象的数量、质量、程度或频率的过程。如测量高度时，我们是应用数字按某个确定的长度单位和重叠的法则进行度量，得出一个比较结果。测量物体运动的速度是按某个单位时间内物体位移的多少这一法则用数字来描述物体运动的快慢。测量这一概念应用到教育上，则就有所不同了。这里测量指的是把学生的行为与事先设计的标准行为进行系统化比较，然后赋以数值的整个过程。其中系统化与比较程序对教育测量来说是非常重要的。

教育测量，若从它测量的对象来看，可分为广义的教育测量与狭义的教育测量。

广义的教育测量以教育的全部领域为测量对象。它涉及教育的一切方面：教育与社会、教育与经济、教育与政治等。例如，教育体系、教育目标、课程设置、教育内容、教育方法、教育管理、教师活动、师生关系等等。

狭义的教育测量以学生的学习成就或以教育目标为对象，它包括学生德、智、体等各个方面。

最狭义的教育测量仅以学生的认知领域中的成就（目

标)为对象,专门研究如何测量学生在文化科学知识方面的学习成就(目标达到的程度)。

2.评价:评价的英文一般指evaluation,在教育上,指的是运用系统的方法对教育活动或行为的价值或效果作出合理的判断。它不同于一般的评价,教育评价一般是有理论根据的,是运用科学的系统程序,根据教育目标、针对教育效果而进行的;或是以掌握教育计划、衡量教育成就,并以改进教育活动为目的的。

教育评价与教育测量、教育评定、教育评估均有所不同。

(1)教育评价与教育测量有所不同。教育测量通常限于对某人或某单位所持有的教育质量(包括数与质两方面)作出测定,重在事实的获取并以“量化”原则来表示,因而较具有客观性;而教育评价建立在对事实资料的测定上,重在事实的解释、诊断和价值的判断,带有一定的主观性。例如,甲、乙两个人在某学科测试中同样都是80分,那么,在测量上他们是一视同仁的,毫无区别。但在评价上则不然,在同类测试中,也许甲是原来的70分上升到80分而乙则由原来的85分下降为80分。因此评价甲成绩进步而乙成绩下降。由此可见,测量本身侧重于注意量的多少或所处的位置,是客观地、正确地将学生达到的教育目标的程度加以量化。而评价侧重于价值的判断,是对测量结果有多大的教育价值的判断,并以改进教育活动为宗旨。

单个的教育测量难以获得有关教学效果的全部信息。例如,有人指责测验所提供的是关于为学生未学到什么的信息而不是关于他们已学到了什么的信息。而单个的教育评价却必须同时涉及许多方面,它可以考查学习的全部领域,可以

对知识、态度、兴趣、品德、行为、智力、能力倾向、身体发育乃至家庭环境作出全面而综合的评论。显然，它比较适合于现代教育全面性的要求。但是它一不可能代替教育测量，因为评价必须建立在测量的基础之上。二在科学性与技术性上不及测量成熟，带有相当的主观性。

(2) 教育评价与教育评定或教育评估也有所不同。

教育评价侧重于对评价对象的个体特征和差异的评价。而评定 (assessment) 侧重于对不同背景及不同情况下团体的整个教育状况作出评价，更强调目标性。它们的关系是：教育评定是一种特殊的评价。1985年，泰勒在接见来访者曾说明 “assessment” 一词是他1963年提出的，用于指一种特殊的评价，即对不同年龄、背景儿童团体的整个学习状况的评价，以区别人们常用的有关测验和评价的概念，以区别以往的小范围的侧重于学生个体特征的教育评价。

教育评估 (estimate) 通常指对教育的事实、条件与现象的价值予以某种估量。目前在我国被认为是对学校办学方向、办学水平和办学条件进行的一种评价。这就是说，评估只是教育评价某方面的内容。因此，教育评价与教育评估也有所不同。

(3) 评价的目的、原则、种类及其方法

评价对于学生来说，其目的主要在于：为诊断与改善学习进步或落后的情况提供资料；为升学、就业与因材施教提供依据；检查教学效果为教学内容与教学方法改善提供依据；帮助学生正确地认识自己的性向、能力并刺激其学习动机；使教师、家长、以及社会正确地了解学生的学习情况。

评价时，遵循以下四条原则是最为根本的。首先，注意

教育目标的导向性。评价在方向上必须指向于教育目标，整个评价工作必须围绕着教育目标进行，促进教育目标的实现。因此，评价者必须充分把握教育目标的精神实质，把教育目标具体化——成为评价目标。评价即就是考察教育目标在学校教育工作中具体地实现了多少，学生最后在教育目标上达到了什么程度。其次，要注意教育评价的时空性。从广度上来说，评价要考察学习的全部领域，包括知识、理解、态度、兴趣、道德、行为、智力、性向、适应状况、勤惰、身体发育、健康情况乃至家庭环境；从深度上来说，要涉及评价对象的实质，许多教育目标，必须持续地进行观察、测验、评定、抓住对象的本质特征，不要被偶然的表面现象所迷惑；从内容上来说，要密度大，即以尽可能少的指标反映尽可能多的信息。这就要求单项评价与全面评价相结合，定性与定量相结合，从方法到指标使整个评价都具有综合性。第三，要注意评价分析诊断的性质。评价过程中，不但要对以往的教育工作及其成果作出价值评定，而且要着重诊断成绩与不足的原委，分析现状、预测未来。只有这样才能使评价的结论有根据，使评价的指导有针对性。第四，要注意评价的立体性。即评价不能单独由某一方面决定，而应由被评价者本人、同学（同事）和评价者（一般是教师或专家）共同评价。

评价的类型按评价对象来分，主要有以下九种：①学习成绩的评价；②智力评价；③性向评价；④身体评价；⑤人格包括态度、行为、道德、情绪、适应等的评价；⑥环境评价；⑦教师评价；⑧学校评价；⑨课程教材评价。

评价的方法，常用的有6种：从比较的对象来分，有横

对评价法、绝对评价法与个体内差异评价法三种，从评价方式上来分，有数量化评价方法与非数量化的分析评价方法、综合评价方法三种。相对评价法是在被评价对象的集合中，选取一个或若干个作为标准物，然后把各个评价对象与标准物进行比较评价。

绝对评价法是先在被评价对象的集合之外确定一个标准物，这种标准物称为客观标准，然后把评价对象与客观标准进行比较评价。

个体内差异评价法是把被评价集合中的各个元素的过去和现在相互比较，由此进行的评价。

数量化评价法是以数学为工具并将评价结果数量化的一种方法。

分析评价法是预先根据评价的观点，把评价内容分解为几个项目，然后分别评价。

综合评价法是对评价内容进行整体性的评价。

三、测试的种类及其特点

测验 (test(ing)) 是心理学与教育学中常用的一个术语。在教育领域中，它一般指用纸和笔为工具测量学生学习成绩的一种形式。例如，我们的单元测验、期中考试、期末考试等等以试卷为工具来测量学生成绩的形式都是测验。

考试 (examination) 对于学校教育来说，是一种必须执行的制度，是整个教学过程中必要的环节，在我国已有几千年的历史了。实际上就其考核检查与鉴别挑选的本质来说，考试与测验是相同的。因此我国一般只提考试而说不说测

验。然而,两者也有不同之处。考试带有浓厚的政治色彩,这与我国封建社会长期以来,以考试取士的传统有关,更多地是突出了它鉴别挑选的特点,突出了它的权威性、官方性以及由上施于下的特点。而测验则不然,它更带有学术性,更多地是突出它检验反馈的特点。为了综合兼顾我国长期使用考试的习惯与测验的学术性及国际性的说法,我国目前不少人把考试与测验统称为测试。以下我们就不妨把考试、测验与测试视为同义来介绍有关内容,以免因名称不一而发生混淆。

测试,按考生表现的方式来分,有笔试、口试和演示(又称实践操作测试)等三种形式;按测试时考生可用参考资料的限制来分,有开卷与闭卷两种测试形式;按测试编制、实施与评分等整个过程的科学性来分,有标准化测试与非标准化测试两种形式;从编制者不同的角度来分,有专家编制测试与教师自编测试两种类型;按测试目的来分,有准备度测试、安置性测试、形成性测试、诊断性测试与总结性测试五种形式;按解释分数的标准来分,有常模参照性测试与目标参照性测试两种形式;按测试场所可分校内测试与校外测试两类;按测试时间又可分为期中测试、期末测试、学年测试和平时测试等形式;按测试的用途可分为入学测试、毕业测试和学科测试等;按测试的内容范围来分则有单元测试与综合测试等形式。除此之外,还可以有其它的分类形式。

1. 标准化测试:它是国际上较为流行的一测试。1985年在北京召开的标准化测试国际讨论会上,各国专家对“标准化测试”的概念众说不一,一般认为标准化测试是按照系统的科学程序组织,具有统一的标准,并对误差作了严格控制

的考试或测验。具体来说，它包括以下四个方面：(1) 试题及试卷编制的标准化：即整个试卷的编制过程，一般要经过试题的编制、审查、试测、分析、筛选或修订这么几个阶段，最后达到编制计划所要求的各项指标。(2) 施测过程的标准化：要求整个的测试过程均要按照测试手册上有关的规定进行。(3) 评分记分的标准化：要求评分与记分按统一的标准进行。目前一般力求实行机器评分与记分。(4) 分数解释标准化：要求分数的解释直观、科学、可比和可加。标准化测试主要是用来选拔或大规模评定学生成绩的。

2. 教师自编的测试：之所以这样称呼主要是相对于编制过程来说的。与标准化测试相比，它的编制缺少试测和筛选等阶段。主要是凭教师日常的教学经验来判断或保证试卷编制计划中的各项指标要求。它主要是用来帮助教师教学的。按它在教学过程中施行的先后顺序，可分为以下几种：

(1) 准备度测试：这种测试一般是在某一课程或单元教学的开始进行的。主要测量学生在某课程或单元学习上所具备的基础知识与技能。这种测试一般难度低范围小。

(2) 安置性测试：这种测试主要是用来测量学生已达到教学计划中所欲达到目标的程度。测量所得到的结果，一般是用来作为编班分组的主要依据。它一般也是在教学开始时进行的。

(3) 形成性测试：这种测试常在教学期间进行。其目的在于鉴别学生的学习是否成功。换句话说，它要查明学生在哪些内容上有了令人满意的进步，在哪些单元上还需要提供帮助。由此不断地提供反馈信息，使师生均了解自己教或学过程中的不足，达到教与学的最优化。这种测试有以下几个

特点：①它是根据教学目标来编制的。凡是课程中重要的部分都需要测试。②它只注意学生是否完成了指定的学习任务或达到了预定的教学目标，而不太重视与别人比较的相对地位。因此它是属于“学习性的测试”，难度比总结性测试略容易一些。③它一般是由一组在有限的教材范围内的综合性试题所组成的。试题的排列以类别和难度为主要依据。常常是将形式同类的试题排在一起，再将同类形式中的试题由易到难地进行排列。

(4)诊断性测试：它主要是用来确定哪些学生在什么地方有特殊的学习困难，以便进行针对性的补课。它一般也是在教学期间进行的。对于特殊的学习困难的诊断和补救，常采用多种评鉴工具，如智力测验、学习态度测验和兴趣测验等。诊断测试只限于对特殊学习困难的诊断。这种测试的特点较重视学生对特殊试题的反应，或是对整组试题的反应，而学生所取得的总分反而不太重要。而且，它特别注意学生的共同错误，偏重于教学中所欲达到的学习结果(目标)的普遍抽样。

(5)总结性测试：它一般指我们期终或教学结束时的测试。其目的在于决定哪一些学生已掌握了本课程或本单元的教学内容，决定哪个学生的成绩应得何种等第。这种测试的特点是，其每个试题都具有一定的综合性与代表性。每一类试题的比率与整个课程各类学习成就所占的比例相一致。而且难度与范围要比其它各类测试大一些。

3.常模参照测试：这种测试通常以包括考生在内的某个具有代表性的团体为标准来决定每个考生在这一参照团体中的相对地位。考生的成绩常采用由考生在测试上的原始分数