

中学教育
科学丛书

评价学业成就的方法

李玉芝 赵裕春 编著



光明日报出版社

评价学业成就的方法

李玉芝 赵裕春 著

光明日报出版社

《中学教育科学丛书》顾问、编委名单

(按姓氏笔划为序)

顾问： 王 铁 江山野 吴 畏

张 健 张承先 钱伟长

顾明远

编委： 田东平 李镜流 阎金铎

陈孝彬 姚维斗 高文元

崔相录

责任编辑：李镜流 崔相录

评价学业成就的方法

李玉芝 赵裕春 著

光明日报出版社出版

(北京永安路106号)

新华书店北京发行所发行 北京四季青印刷厂印刷

787×1092毫米 32开本 4.625 印张 100 千字

1987年7月第一版 1987年7月第一次印刷

1—23,000册

统一书号：7263·062 定价：1.10元

为《中学教育科学丛书》序

茅 以 升

近年来，随着我国现代化建设事业的发展，越来越多的同志深切认识到教育的重要。诚如中共中央关于教育体制改革的决定所指出的：“今后事情成败的一个重要关键在于人才，而要解决人才问题，就必须使教育事业在经济发展的基础上有一个大的发展。”教育是上层建筑，它的存在和进步，要依靠一定的经济条件。而从长远来看，教育又必须走在经济的前头，为经济起飞准备大批优秀的人才。我相信，在党中央的领导下，我国的教育事业必将有一个较大的发展和更加美好的前景。作为一个老教育工作者和科技工作者，我感到由衷的喜悦，并愿聊尽绵薄，为发展祖国的教育事业，继续贡献力量。

现在，光明日报出版社要出版一套《中学教育科学丛书》，我认为这是一件很有意义的工作。

中学教育是整个教育系列中关键的一环，具有承上启下的作用。它要使亿万青少年学到基本的科学文化知识和技能，并为更高一级的职业技术教育和高等教育输送合格的学生。因此，发展教育，开发智力，提高全民族的科学文化水平，必须从中小学抓起。“百年树人”。中等教育的水平和质量，是关系全局、举足轻重的大事，万万忽视不得。

由于十年浩劫和长期“左”的思想的影响，我国教育从

DAM25/06

总体上来看是落后了。这不仅表现在教材体系、教学设备、师资队伍和学生素质等方面，也表现在教学方法特别是教育科学研究方面。教育科学是一门综合性的学科，既有严密精确的一面，又有复杂和多样的特点。中学教育还是基础教育，以传授各门学科的基本知识为主，但同样有一个教育理论、教学方法问题，需要我们努力探索、研究和创新。我以为，当务之急是必须尽快改变那种层层加码，把学生当作教材的奴隶、作业的苦力的状况；改变老师讲、学生听(记)、满堂灌的“填鸭”式教学方法；采用启发式的教学方法，调动学生学习的主动性和积极性，培养学生具有初步的创造性思维能力。不但使他们牢固掌握基本的理论、概念、原理，而且能够灵活运用；不仅知其然，还要知其所以然。当然，这说起来容易，真要做到并不那么简单。这就有待于教育科学研究的开展，还需要教师们的努力。

教育科学的研究，不但是专业科技人员的事，也是广大教师的事。教育科学也有实验室，这个实验室就是课堂。作为塑造人类灵魂的工程师，教师不应当满足于当“教书匠”，而应该成为教育家，成为培养人才的艺术家的，造就一代又一代现代化建设人才。

我希望，《中学教育科学丛书》将对我国广大的中学教师加强理论修养、改进教学方法有所帮助；在提高中学教育质量，发展教育事业中起到一定的作用。

一九八五年十月

目 录

第一章 学习与学业成就的评价	(1)
一、什么是学习.....	(1)
二、教育评价与学业成就评价.....	(4)
三、教育评价的由来和发展.....	(8)
第二章 评价学生学业成就的主要方法——测验法	(14)
一、测验是评价学业成就的主要方法.....	(14)
二、测验的类别.....	(15)
三、常模参照测验和标准参照测验.....	(17)
四、标准化学业成就测验.....	(20)
五、教师自编测验.....	(22)
六、良好测验的基本条件.....	(24)
第三章 教师自编测验与教学目标分析	(28)
一、分析掌握教学目标是教师自编测验的首要前提.....	(28)
二、终极教学目标、中程教学目标与具体教学目标.....	(29)
三、具体教学目标对教学工作的重要意义和当前在这方面存在的问题.....	(31)
四、拟定具体教学目标应遵循的原则.....	(32)
五、教学目标分类法介绍.....	(34)
第四章 教师自编测验的试题形式	(43)
一、论文式测验的意义、优缺点和使用时应注意之	

事项·····	(43)
二、客观式测验·····	(43)
三、问题情境测验·····	(61)
第五章 教师编制测验的类别和方法 ·····	(64)
一、教学活动的流程与评价·····	(64)
二、诊断性测验的作用和类别·····	(66)
三、形成性测验的作用、编制方法与实施·····	(73)
四、总结性测验的特点和编制方法·····	(77)
第六章 评价学生学业成就的其它方法 ·····	(95)
一、视察法的意义、种类、记录方法和注意事 项·····	(95)
二、评定法的意义、种类及注意事项·····	(101)
三、问卷法的意义、种类及注意事项·····	(104)
四、谈话法的意义和应用谈话法应注意的事项····· ·····	(110)
五、人物推定法的意义、应用及注意事项·····	(111)
第七章 测试结果的解释和应用 ·····	(116)
一、相对评价的意义和数量表示方法·····	(116)
二、绝对评价的意义、数量表示和解释方法·····	(127)
三、个人内差异评价的意义和运用方式·····	(133)
四、向学生解释测验结果和向家长报告学生的成绩 ·····	(136)

第一章 学习与学业成就的评价

什么是学习——教育评价与学业成就评价——教育评价的由来与发展

一、什么是学习

学习虽是随处可见和我们每个人经常从事的活动，但要对它作一番概括性的描述，或下一个简短而全面的定义却并不容易。因为，学习包括的内涵极广，学习的情境、层次也复杂多样，一般说来，我们对学习这一术语的使用有两种情况：一种是狭义的学习，是特指学校中的有目的有计划的知识技能的学习；另一种是广义的学习，是指日常生活中无特定目的的、附带的、甚至是偶然的学习。广义的学习虽不是人类社会的主要学习形式，但却是大量存在的，婴儿从呱呱坠地开始，逐步学会走路、说话、穿衣、吃饭，直到懂得生活是怎么回事，世界是怎样的，以及必须做些什么才能维持自己的根本生存等等，所有这些经验，并不都是专门从学校中，从老师处学来的，其中很大的比例是从与父母、亲戚、兄弟姐妹、同学同事的交往接触过程中，从劳动、游戏以及其它种种活动中，非正式地、漫不经心地学会的。学校里的学习也少不了要以这种学习为基础进行。

心理学中关于学习概念的含义则更加广泛。它不仅包括人在学校中的和在日常生活中的学习，而且包括动物的学习。大猫教小猫捕鼠，大燕带小燕学飞，动物园里的工作人员训练猩猩、大象、海狮耍把戏等等都是学习。而且，不仅这些显著可见的，比较高级的动物是如此，就连肉眼不易看到的，简单细小的原生物草履虫，据科学家们的实验室观察证明，经过练习也能减少它们在毛细管中旋转的时间，这显然是由经验引起的行为变化。从这样的广义来说，学习就是人及动物在生活过程中所发生的行为变化。这种变化不是由于先天的本能倾向所产生，也不是由于机体的自然成熟或学习者的暂时生理状态(如疲劳、饥饿、药物反应、机体损伤等)所引起。这些暂时状态虽然有时也能引起个体行为的变化，但它们不属于学习的范畴。另外，偶然出现而又迅速消失的行为变化，譬如，在足球场上偶而踢进一球，这也不是学习。心理学家们一般把学习看作是后天凭藉经验产生的比较持久的行为变化。

人类的学习，不管是在日常生活中发生的还是在有目的、有计划的教育过程中发生的，自然都是凭藉经验产生的。但是，后者与前者不同。后者是教师按照教育目标，通过对学习者进行一定的训练而出现的，在行为上比较持久的变化。这些变化有认知方面的，情感方面的，也有动作习惯方面的。

关于这些变化是怎样产生的，它的实质是什么的问题，有种种不同的看法，归纳起来有两种：

一种是联结主义的学习观。桑代克是联结主义的早期代表。他通过对动物和人类的实验研究指出，学习的实质在于形成一定的联结，这种联结是通过尝试错误逐渐建立起来

的：俄国的巴甫洛夫，虽然没有宣称自己的学习观是联结主义的，但他认为学习就是通过一个刺激代替另一个刺激建立种种条件反射的过程，即暂时神经联系的形成；斯金纳也认为，一切行为都是由反射构成的。但行为有两类，一是应答性行为，一是操作性行为。应答性行为是由刺激引起的反应，如巴甫洛夫的经典条件反射——铃声(刺激)代替食物引起唾液分泌(反应)。操作性行为，是由操作结果引起的反应，如老鼠在斯金纳设计的实验箱中学会了按压操纵杆以取得食物。在这里，按压操纵杆的操作成为取得食物的手段或工具。所以，这种条件反射叫做操作条件反射(或工具条件反射)。因此，斯金纳也分学习为两类：一是反射性的，即通过巴甫洛夫所描绘的经典条件反射实现的学习；二是操作性的，操作性的学习是受操作结果的影响而实现的条件反射。斯金纳认为，操作性的学习更能代表实际生活中人的学习情况。他把人的行为和学习主要看做是操作性的。但他也指出不应当忽视反射性的学习，不过它不是主要的。

另一种学习观是认知论的学习观。布鲁纳是美国当代认知心理学的主要代表人物。他认为，学习不在于被动地形成刺激——反应的联结，而在于主动地形成认知结构。这和早期的格式塔心理学的观点基本上是一致的。格式塔心理学认为，当个人为追求他的目的而看出利用环境因素(包括他自己的身体结构)的新方式时，便产生了顿悟，学习就是获得新的顿悟或者新的意义。但他们对顿悟产生原因的解释带有唯心主义的神秘色彩。现代认知心理学家在抨击联结主义者把学习看作是刺激与反应之间的直接的机械的条件的条件反射理论时指出，在刺激与反映间应有一个认知过程(中介过程)。学习就是通过认知获得意义，从而形成认知结构的过

程。所以，他们十分重视已有的知识经验的作用(即原有的认知结构的作用)，认为只有当新学习的材料与原有认知结构联系起来，新旧知识发生相互作用时，新学习的材料才能在学习者的头脑中获得意义。从而也就消除了格式塔心理学的神秘主义色彩。

对联结主义者来说，行为是指身体的活动，而且他们更注意的是外显的、可观察得到的行为。认知心理学家也使用行为一词，但它主要是指心理行为，所以在他们看来，学习是人在知识、技能、态度、价值观或信仰上的持续不断的改变。这种行为，不是直接可以观察得到的，但是可以从可观察到的行为举止中推断出来。

二、教育评价与学业成就评价

教师的任务在使学生经过一定时期的学习之后，知识更丰富，技能更发达，并形成良好的学习态度、习惯和思想品格等，也就是使学生在知识、技能、态度和品格等方面朝向预定的目标变化。学生是否达到了预定的目标，有多少进步，以及和其他学生相比成绩如何？是通过教育评价做出判定的。教育评价是对照从教育目标导出的评价标准，对学生的学习成绩给予价值判断的系统过程。

学业成就是学生学习成果的重要方面，它包括知识、技能，也包括兴趣、态度、习惯等诸多方面的成就，这是较广义的理解，一般主要是指知识、技能方面的成就。学生头脑中的认知结构(即知识结构在头脑中的反映)的变化不能直接测得，但可利用测验间接加以测量。测验是根据教学目标选定的作业题目，通过学生对这类作业题目的反映，可以判断学生是否学得了知识、技能和具有理解、应用、分析综合等

能力。

学业成就测验是测量学生学业成就的主要工具。利用这个工具对学生的学业成就加以衡量以得的数量资料（测验分数），是对学生成绩的客观描述，就是一种测量。测量后的评价，是指对他得到的成绩（分数）所做的价值判断。譬如，某生的测验成绩是45分。单就这个分数说，仅是他的成绩的量化的代表，并不说明它有什么样的价值意义。用他这个分数与其他学生的成绩相比，是好还是差，属于优、良、中、及格、差哪一个级别；或者和一个确定的标准相比（如某一单元教学目标），说明他是否达到了目标、相差多少、有什么问题；或者与他过去的成绩相比，是进步了还是退步了，都是对他的测验成绩所作的价值判断，即对学生学业成就的评价。可见，测验、测量与评价虽然意义相近，常被混用，但实际上它们是有区别的。评价不仅要对学生学业成就测量结果进行好坏、优劣、高低的判断，还要了解有关的其它方面的情况，如学生的能力、努力程度、学习态度、某些个性特点、学习条件、健康状况等。以便找出取得这一结果的原因，做出教育上的决策和给予必要的帮助。所以，评价还意味着具有诊断和治疗的功能。

学业成就评价在教与学的活动中具有主要的地位和作用。

首先，学业成就评价可为教师提供必要的反馈信息借以调整和改进教学，提高教学水平。

教学是师生双方面的活动。学业成就评价的结果，反映着学生学习的成果，也反映着教师教学的效果与效率。教师从中可以看出自己拟定的教学计划是否体现了教学目标？在教材方面有什么问题？各种概念的阐述是否正确清晰？教学

方法是否妥当？以及采用的某些技术或资料的效验如何等等。这些由评价得到的反馈信息就为教师完善教学指导计划，改进教学方法，以至修订教学目标提供了依据。

其次，学业成就评价在为教师提供反馈信息的同时，也为学生的学习提供反馈信息，有促进学习的作用。

事实证明，通过评价证明了的成功的学习，会给学生带来满意与愉快的学习体验，这种体验能增强学生的学习信心，提高他们的学习兴趣，强化他们的学习动机。而被评价证明了失败的的学习所引起的不愉快体验，也会使学生为了避免这种不良后果的再度发生，在以后的学习中更加努力。评价的本身也是一种主要的学习经验。通过评价肯定了正确的意义和联想，校正了缺点与错误，澄清了混乱的概念，分出了学习材料中已经掌握和尚未掌握的部分。这样，在以后的学习中学生就会把精力集中于更重要的部分。另外，学生还会根据经常获得的外部评价经验，逐步学会独立地进行自我评价。为了促进这一目标的实现，教师应尽量让学生参与对自己测验结果的解释。这样做，除了有锻炼学生自我评价能力的好处之外，还有两条现实意义：一是可以增加学生自我了解和对测验结果自我接受的程度，避免由测验带来的消极影响，如因成绩不良产生不安和自卑，或因获得了好成绩而骄傲自满等；二是教师可用由学生提供的非测验资料补充说明测验的结果。

第三、学校和家长也可从学业成就评价中得到有用的信息，这样在学校家长的共同关怀与配合下，通过师生的共同努力，就会更好地达到既定的目标。

第四、学业成就评价有助于教育科学的研究。教育科学的研究与其它科学研究一样，需要以客观的测量搜集可信的

资料。比如分组实验时等组的建立,标准的确定,定期效果的检查和最后结果的比较,都少不了要以学业成就评价所得资料为依据。同时,各级各类学校课程与教材的修订与改善,除须征询教师、教育学专家、心理学专家、课程专家以及家长和有关社会人士的意见外,更多的还是要以学生学业成就的评价结果为依据。学业成就评价的结果可以解答或部分地解答:课程是否有所偏倚?教材是否充分?教材的编排是否合理?难度是否适中?各级各类学校间的课程是否上下衔接左右关联?以及有关教法的规定如理论与实践的比例,讲授与自学的安排等是否恰当等问题。

学业成就评价的重要作用虽如上述,但其应用不是孤立的。为了提高教学效果和更好地达成教学目标,仅仅依靠从学业成就得到的反馈信息,去调整教学过程和改进教学方法还是不够的。教师在确定教学目标、制定教学指导计划之初,就需要取得有关学生智力、相应的特殊能力、过去的学力、性格特征、学习的态度兴趣、课程教材以及学校家庭环境等方面的情报资料。也就是说,教师也必须通过对这些因素的评价,掌握与学生当前学习有关的全部信息,才有可能制定出最优化的教学指导计划。同样,为了使学业成就评价在教学活动中能够真正起到诊断和治疗的作用,也必须参照上面所说的那些有关方面的材料,对学生的测验成绩作综合全面的解释才行。还有,为了检查和说明教学效果,有时对教学指导计划和教学方法也需进行评价。所以,教育评价是一个更为广泛的概念,上述各项都是教育评价的对象。但在学校教育活动中,教育评价主要是指对学习成果——学业和思想品格两方面的价值判断的系统过程。其中最经常,而且大量运用的是对学业成就的评价。它在学校教育活动中占据

中心位置。其它方面的评价往往是围绕着学业成就评价进行的。因为学业成就评价虽然是以学生的学业为评价对象，但并非就学业论学业，它必须结合有关方面的情报资料对学生的学习成绩作综合性的全面分析和解释，并在此基础上调整改进教学，才能更好地完成教学任务。为此，教师也就必须对其它与学生当前学习效果有关的一切因素作出相应的评价。足见学业成就评价和教育评价的其它各方面也是互相不可脱离的。

三、教育评价的由来和发展

任何时代的教育都要考查其教育效果。考查的方法随着时代的进展，也不断有所变化与改进。从教育发展看，最初无论中外考查教育效果都是采用口试。口试只能一对一的进行，效率既低，又难取得一致的标准，后来随着教育的发展和学生人数的增加遂有笔试的出现。我国起自隋唐时代的科举考试，实是世界上最早应用笔试的先例。在欧洲直至1702年英国的剑桥大学才开始使用笔试。美国则更晚，1845年笔试才应用于普通学校。笔试可用同一题目同时测量许多学生，较之口试效率有了很大的提高。但是，当时的笔试都是论文体的考试，这种考试试题数目少，难以考查到全部的教学内容，而且评分全凭教师个人的认识和好恶，存在着主观与笼统的弊病。因此，自本世纪初在美国就有教育测验运动的兴起。

教育测验运动虽发端于美国，但德国的冯特于1879年在莱比锡大学设立心理实验室，用实验和测量的方法研究人的心理；英国的高尔顿1882年在伦敦成立人类学实验室研究个别差异，以及皮尔逊为这种研究所设计的统计方法等，对教

育测验运动都有很大的影响。美国的卡特尔，曾就学于冯特，后来又回到英国向高尔顿学习人类学和向皮尔逊学习统计学。回国后便设立心理实验室，并于1890年发表文章主张心理测验。及至卡特尔的学生桑代克时，教育测验运动才蓬勃发展起来。1904年桑代克发表了被认为是测验学史上具有划时代意义的巨著《精神与社会测验学导论》，提出“凡是存在的东西都有数量，凡有数量的东西都可测量的著名观点。在这种思想的指导下，他领导学生编制标准测验及从事教育测验工作。此后，在不到廿年的时间里，美国相继发表了数以千计的心理测验和教育测验，把教育测验运动推到了最高峰，并普及于世界各地。

那时，保证测验客观化的重要手段是使测验标准化，即测验的内容与形式对所有被试者都是一样的，施测的条件也完全相同，有统一的施测程序和指示语，有标准的时限和记分方法等。此外，标准化测验还以同一团体在某一测验上的成绩水平作为参照标准，使个人的成绩与这个标准(常模)相比较，以了解个人在这个测验的次数分配上所处的相对位置。这样就可以较为客观地比较学生之间的差异。测验的标准化在一定程度上保证了测验的客观性，并能测出学生之间的差异。但这类测验偏重于测量记忆性的知识，不能测出学业成就的全部内容，更难测得全部教育效果。当然，也就无法综合各方面的材料对学生的学业成就进行关联性的解释。

1930年前后，由于美国经济大恐慌而引起的教育危机，促使一些教育家开始对教育的根本问题进行探讨，教育观点也随之有所改变。从主智的教育转而主张全人格的教育。当时的进步主义教育联盟就曾以发展人的多方面才能为教育目标，首先对课程的内容进行了长达八年之久(1933—1940)的

实验研究。与此同时，为了检查比较新旧两教育形式的效果他们推举了泰洛等人组织起评价委员会。该委员会一方面批判当时教育测验中存在的问题，一方面以教育评价的思想研究考查教育成效的方法，并设计了一套教育评价法。从而使美国的教育测验运动于四十年代初转向了教育评价。

从教育测验运动转向教育评价，并不意味着反对用客观的方法收集有关的数量资料，这些资料依然是教育评价的依据。教育评价不同于过去的教育测验的主要特点在于：

1. 教育评价是实现教育目标的一种考查手段，它是按照教育目标，利用一定的测量工具，对学生到达目标的程度进行的价值判断。因此，教育评价与教育教学目标有着密切的联系。在评价过程中，教育、教学目标或评价目标越清楚、越具体，教育评价的效果也就越大。

2. 教育评价具有分析、诊断的性质。它着眼的不是学生的得分或名次，而是学生在测验中反映出来的问题。通过对这些问题的诊断和分析做出进一步的指导或补救的决策。因此，教育评价有一定的连续性，要在整个的教育教学过程中，不断地收集资料，以便根据学生的发展状况给以帮助。

3. 教育评价力图通过学生的全部资料，对测验结果做综合性的全面分析。为收集学生的各方面的情报资料，在评价方法上不限于纸笔测验，还广泛利用观察法、评定法、问卷法等其它评价技术。

4. 教育评价注重学生的主动性，鼓励学生参与学习(教学)目标的选择和测验结果的解释。目的在于促进学生自我评价，以增进自我了解，和自觉调整下一步的学习。

从上述教育评价的特点可以看出，教育目标是教育评价的核心问题。美国大学考试委员们，在这方面做了不少研究，