

学科试题编制原理与技术丛书

高凌飏 丛书主编

历史

试题编制原理与技术

黄牧航 周朝阳

著

本书介绍考试的基本理论和编制试题的基本要求与技术，是广大教师 and 高等院校学生提高自己教育专业水平的基本学习材料，也可供有兴趣的家长在了解考试命题时参考。

SPM

南方出版传媒

全国优秀出版社
全国百佳图书出版单位



广

学科试题编制原理与技术丛书

高凌飏 丛书主编

历史

试题编制原理与技术

LISHISHITIBIANZHINYUANLIYUJISHU

黄牧航 周朝阳 著

SPM

南方出版传媒

全国优秀出版社
全国百佳图书出版单位



广东教育出版社

图书在版编目 (CIP) 数据

历史试题编制原理与技术 / 黄牧航, 周朝阳著.
—广州: 广东教育出版社, 2015.9
(学科试题编制原理与技术丛书 / 高凌飏主编)
ISBN 978-7-5548-0818-4

I. ①历… II. ①黄… ②周… III. ①中学
历史课—试题—编制 IV. ①G633.512

中国版本图书馆CIP数据核字 (2015) 第189120号

策 划: 李朝明 李 智
责任编辑: 熊全慧 严洪超
责任技编: 姚健燕
装帧设计: 何 维

广东教育出版社出版发行
(广州市环市东路472号12-15楼)

邮政编码: 510075

网址: <http://www.gjs.cn>

广东新华发行集团股份有限公司经销
佛山市浩文彩色印刷有限公司印刷
(佛山市南海区狮山科技工业园A区)

787毫米×1092毫米 16开本 20.75印张 415 000字

2015年9月第1版 2015年9月第1次印刷

ISBN 978-7-5548-0818-4

定价: 52.20元

质量监督电话: 020-87613102 邮箱: gjs-quality@gdpg.com.cn

购书咨询电话: 020-87615809

前言

在人类历史上，自从有了教育，就离不开考试。无论是教育者，还是受教育者都需要知道施教与受教的效果，以利于进一步的教与学，考试便自然而然地产生了。任何考试都是一把双刃剑，有其利也有其弊，不可能是十全十美的。以高考为例，赞扬的说，“文革”后恢复高考是“挽救了我们的民族和国家”；批评的说，高考是“青少年素质发展的最大障碍”，“高考不倒、中国不好”。事实上，大规模公开考试是社会分工的产物，既对今天的教育以及社会有着不可替代的促进作用，又对教育工作和青少年的发展产生了重大的负面影响。只有到了马克思所预言的共产主义的高级阶段，脑体差别以及由社会分工不同带来的人的社会地位和生活状况的差别消失了，考试的功利性才可能消失。即使到那时候，选拔还是存在，考试还可能是选拔的一种重要手段。作为教育系统内部、以检查诊断教育效果和问题、促进学生的学习和发展为目的的考试或考评就更加不会消亡。不过到那时可以说考试是真正服务于学生的学习，真正以促进人的发展为目标。考试的演化和进步一定是渐进式的，在可以预见的将来，以选拔为主要目标的各种考试，包括带功利性的大规模公开考试和促进学习为主要目标的日常教学考评都将长期存在。以选拔为主要目标的考试不仅有选拔功能，也对学生的发展有一定的促进作用；以检查促进学生学习为主要目标的考评也会受功利的影响而对学生的的发展产生负面的效应。面对这样的现状，我们必须、也只能不断加强对考试本身的研究，包括考试的理论和技术的研

究，不断提高考试和考评的质量水平，让考试和考评的正面效应达到最大，负面效应缩为最小。

无论是选拔性的考试还是日常的考评，其核心过程都是向应试者（学习者）提出特定的问题要求应试者加以解答或解决，通过应试者解答或解决的表现来判断他达到的水平和存在的问题。因此可以说，命题是考试考评最为核心的一环。题型定得好加上题目命得好，可以考出应试者的理论水平和实践能力。即使像科举这类考试也是如此，如早期科举中的策论，经常由皇帝或高官根据实际统治的需要而提出问题让考生对策，可以选拔出有真才实学的人才。若题型定得不好加上题目出得也不好，如要求举子死记“四书”“五经”的帖经，只能导向死读书而不求甚解，更不会运用。为防止猜题，帖经的题目越出越偏，最后都没办法出题了，只好取消。古代如此，今天也是这样。20世纪80年代刚恢复高考时曾一度有不少试题就是导向死记硬背的，随着考生应考准备越来越细，这类考试题目也就越来越偏，再也无法继续。历史的经验说明，要发挥好考试的正面导向作用，就必须认真研究考试理论，认真研究如何命制好试题。

研究、学习和掌握考试的理论不仅是社会和教育事业的需要，也是每一位教师和学生自身发展的需要。从教师的角度看，当一名称职的教师，第一，要有高的思想境界和好的道德操守，做到“行为世范”；第二，要热爱学生，热爱教育事业，全心全意地投身于教书育人的工作之中；第三，要有厚重的专业知识，包括某一学科以及教育教学的专门知识，做到“学为人师”；第四，要有实践教育教学工作的能力。考试作为教育教学工作中不可或缺的一环，是每位教师在实际工作中都必须做的。考试有专门的理论体系和实践要求，只有具备了相关的专业知识和必要的实践能力，才能做好这一工作。

具体来说，首先，教师需要在教学的过程中，通过正式的和非正式的考试，检查学生的学习效果和学习过程中存在的问题，并以此为镜反观自己教学的成果和存在的问题。教师是这一类考试的组织者、实施者和结果利用者，必须了解这类考试的性质特点，懂得如何编制试题，如何根据学生的应答情况分析鉴定学习的效果和存在的问题，如何在后续的教学中利用考试的结果。其次，学校、教育行政部门会定期组织各种终结性的水平考试以检查教学工作的绩效，判断教学质量水平的高低。在学校教育的每个阶段结束之后，社会都将组织公开的考试来选拔学生，确定人才的流向。作为教师必须知道这两种终结性考试，包括水平考试和选拔考试的性

质特点，了解其目标和要求，知道如何应对这两类考试，才能帮助学生做好应考的准备并取得好成绩，同时用考试结果的分析来改进自己的教学。优秀的教师还有可能直接参加这两类考试的命题和阅卷，则更需要了解考试命题的理论和方法。一句话，教师需要具备关于考试的理论知识和设计、组织、实施考试的能力，这是教师专业素养水平的一个重要成分。命题则是其中最为关键的一环。

了解考试命题的过程和相关知识也是帮助学生提高学习水平，促进个人全面发展的需要。首先，考试是学生向教师、向家长、向社会汇报自己学习成绩的重要手段。无论从哪个角度看，取得好成绩都是学生的本分。要做到这一点，最根本的是要扎扎实实地学好功课，同时也要了解考试的要求和做法，了解考试中各类试题的特点和解答方法技巧，才能做好应试的充分准备。学生学习的目的当然不是为了汇报成绩，而是为了自身的发展。自身的发展包括身体、心理、道德、思想以及学业等方面，其中学业的发展就离不开考试。通过考试，可以知道自己的学习进展情况和达到的水平，发现学习中存在的问题，及时地加以弥补。通过考试，对学习的反思有了具体的内容，才可能找到自己学习方法上成功的经验和存在的问题，进而形成一套对自己来说行之有效的学习方法，掌握学习的规律。

考试随着教育的产生而产生，也将随着教育的长期存在而一直存在下去。考试过程的核心环节是命题，掌握考试命题的理论和实际技巧是衡量教师专业水平高低的重要方面，了解考试命题的过程和特点也是学生提高成绩和学会学习所必须的能力。正因如此，我们邀请中小学各学科的考试命题专家，编写了这套书，介绍考试的基本理论和编制试题的基本要求和技巧，供中小学教师和大学师范专业的学生作为发展提高自己的教育专业水平的基本学习材料，也供有一定基础和兴趣的高中学生作为了解考试命题的指导书。

本套书分学科编写，共有九册。尽管各册的具体安排和章目略有不同，但基本框架是一致的。第一章“考试与命题”简要地介绍了考试的一些基本理念和知识，介绍了作为考试参照标准的教学目标及其分类体系，以及分析解释考试结果和鉴定考试质量的概念和统计指标。希望通过这些简单的介绍，使读者对考试命题相关的理论领域有一个总体的、超越学科的认识。第一章还介绍了考试工具的设计，即试卷和试题编制上的一些超越学科的共同问题。这些内容虽然超越具体的学科，但都是具有经验性和实用性的，也是通常教育评价或考试理论著作所没有覆盖的，对具体编制考试的试卷和试题有实际的指导意义。

从第二章开始，各章的内容按照学科的知识内容和能力特点展开阐释，大致包括以下几个方面：

第一，各学科考试的沿革（重点在新中国成立后）和试题特点：考试方式与试题形式的变化；教学目标和对试题的要求；知识类型与试题功能、特点；各类考试试卷和试题的编制原则理念，以及考试命题中常见的问题与困惑。

第二，形成性测验和表现性考试中的命题技术：形成性测验的目的和特点；形成性测验常用的题型；形成性测验命题须注意的问题；形成性测验的命题技巧；形成性测验试卷的组合；答案和评分要求。表现性试题的命题要点和注意问题；表现性试题的评分方法，表现性试题的适应范围。

第三，阶段性终结考试的命题：阶段性终结考试（期中、期末考试）的目的和性质；阶段性终结考试命题需注意的问题；阶段性终结考试的命题技巧；阶段性终结考试试卷的组合；答案和评分要求。

第四，大规模公开考试的命题特点：大规模公开考试（中考、高考）的性质与特点；大规模公开考试对试题的要求；大规模公开考试的常见试题类型；大规模公开考试中典型试题的命题思路分析。

第五，命题质量鉴定与反馈：试题质量的鉴定；对命题的问题的检讨和矫正；对学生进行考试的反馈和试题分析时需要关注的问题。

第六，国内外各类考试名题鉴赏：列举各类优秀试题案例并对案例进行分析。

本套书的写作既有理论又不囿于理论，用具体的试题以及对具体试题的分析来阐述作者的观点，对教师来说是易读易懂的，既有理论的提升，又有学科实际技能的提升。本套书还对一些中外优秀试题的实例进行了介绍和点评，相信这些对拓宽读者的眼界会有所帮助。

高凌飏

2015年7月

目 录

第一章 考试与命题	1
第一节 考试的基本理念和知识	1
第二节 试卷和试题	9
第二章 历史科考试的沿革	15
第一节 中国古代的历史科考试	15
第二节 清末和民国时期的历史科考试	19
第三节 新中国成立后的历史科考试	25
第三章 中学历史科表现性考试的命题技术	34
第一节 中学历史科表现性评价的基本内容	34
第二节 培养处理数据能力的历史科表现性评价命题	40
第三节 进行形式灵活多样的历史科表现性评价命题	72
第四章 中学历史科阶段性终结考试的评价体系及 命题技术	99
第一节 中学历史科阶段性终结考试的评价体系	99
第二节 中学历史科阶段性终结考试的选择题命题技术	104
第三节 中学历史科阶段性终结考试的主观题命题技术	117
第五章 中学历史科大规模考试的命题特点和趋势	148
第一节 中学历史科大规模考试的命题特点	148

第二节	中学历史科大规模考试的命题趋势	179
第六章	历史教师选题组题的基本思路和方法	211
第一节	历史教师选题的基本思路和方法	212
第二节	历史教师组题的基本思路和方法	228
第七章	中学历史科的命题质量鉴定与反馈	257
第一节	中学历史科选择题命题质量鉴定与反馈	257
第二节	中学历史科主观题命题质量鉴定与反馈	266
第三节	中学历史科试题难度控制的基本方法	281
第八章	国外中学历史科试题评述	299
第一节	英国历史科学案设计评述	299
第二节	美国高中AP课程历史考试及试题评述	309
参考文献	323

第一章

考试与命题^①

第一节 考试的基本理念和知识

一、考试^②的概念

在现代教育评价体系中，考试只是一种收集评价证据的过程、手段或方法，指的是根据一定的教学目标，让被试者在规定的时间内，按照规定的方式和要求，解决预设的问题，并对被试者的表现和解答的结果进行评定以确认被试者的表现水平，据此对被试者进行评价。

考试有许多不同的类型：在教学过程中实施的通常称为形成性考试，在教学告一段落时实施的通常称为终结性考试。如果从考试结果的利用出发进行分类，可以将考试分为诊断性、水平性和选拔性的考试。从考试的问题—应答形式出发，可以将考试分为开卷考试和闭卷考试，纸笔考试、口试或操作考试，应答式考

① 本章作者为高凌飏。

② 与考试相近的另一术语是测验，从本质上看，考试和测验没有实质性的差别，习惯上考试的规模相对大一些，测验的规模小一些，或者把学段终结时进行的测验称为考试，在日常教学过程中的测验保留称为测验。但也不是绝对的。本书不区分测验和考试，只用考试一个术语。

试或表现式的考试，等等。

不同类型的考试有不同的特点和用途。闭卷考试是最为常见的一种方式，考生必须在规定的时间内回答所有的试题，一般不允许带任何参考材料。这就要求考生必须记住考试范围内重要的知识，能够熟练地运用各种相关的技巧以回答考试的问题。这样的考试有利于考查学生对知识的了解和掌握。开卷考试一般也有时间的限制，但可以携带必要的参考材料以备查阅。考生可以不必记住相关知识的细节而把精力集中在如何运用知识解决或解答问题上，这种方式更有利于考查学生的能力。

纸笔考试通常以文字符号作为媒介，有时也加插必要的表格和图形作为辅助，这样的考试要求考生有一定的阅读和表达能力。听力考试和口试则以有声的语言作为媒介，主要用于语言类的考试，重点检测考生对语言的聆听和表达能力。最常见的操作考试是理科的实验操作考试，以实际动手为主，考查的主要是动手操作的能力。

应答式的考试根据试题的要求对试题做出回答，一般是通过计算或推理或论证得出某个符合题目要求的结论，结论可能只有唯一的一个正确答案，也可能有多个正确答案，也可能完全开放以至于没有所谓正确的答案，关键在于得出答案的过程和逻辑是否合理。表现式的考试要求考生按照试题的要求进行表演或解决问题的活动，一般不会有唯一正确的答案或表现方式，表现的过程和结果一样重要。

在过去的一个世纪中，教育评价的概念发生了不小的变化。然而不管如何变化，作为评价方法和工具的考试却始终在发挥其应有的作用。当然，考试的技术本身也随时代的进步和评价概念的变化而不断变化，然而其本质性的东西却依然如故。

二、教学目标及其分类

考试需要有确定的目标，其参考的基准就是教学目标。由于方法上的局限，任何考试都只能考查到部分的学习效果，因此考试目标往往只是教学目标中的一个部分。

教学目标描述通过教学活动所期望得到的学习效果，包括认知的、方法技能的以及情感态度与价值观方面的。教学目标很多，如果不加以梳理，就会像一团乱麻一样毫无头绪，既不利于教与学，也不利于考查评价。如何梳理？最好就是依照分类学的原理，通过分析目标的本质特点，把特点相同或相近的目标归为一类，再通

过不同类别之间的对比找出每一类的本质特征，这样既可以使目标的体系变得清楚明确，也可以加深对目标涵义的理解。这就是进行教学目标分类的目的。

（一）布卢姆的分类及安德森的修正

首先对教学目标进行分类的是布卢姆（Bloom）和他的团队。布卢姆假设：在教学过程中，较复杂的认知行为中含有较简单的认知行为，简单认知行为的组合构成较复杂的认知行为，复杂程度提升到一定程度之后就发生质变，成为与简单认知行为有本质区别的新的、高级的认知行为。基于这样的假设，布卢姆提出了逐步递进的台阶型分类框架应用于认知领域的教学目标：台阶的最底层为了解（知识），逐级上升到理解、应用、分析、综合、评鉴等层次（见图1-1-1）。了解的层次包括各种各类的知识：具体的知识（术语、具体事实），方法的知识（惯例、趋势和顺序、分类和类别、准则、方法论），普遍原理和概念的知识（原理和概念、理论和结构）。理解包括了转化、解释和推断。应用指对知识的运用。分析包括要素分析、关系分析和组织原理分析。综合指的是能够进行独特交流，制订计划或操作，并进一步推导出抽象关系。评鉴包括依据内在证据和外部证据进行判断两种情况。

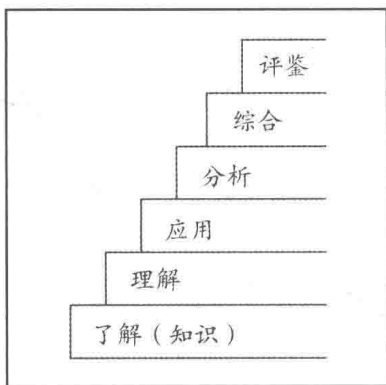


图1-1-1 布卢姆的目标分类体系

布卢姆的体系有许多优点：第一，阶梯形递进的框架模型易于理解和操作。第二，从认知、情感意志、动作技能三个领域去确认学习成果的水平，具备了多元智力的雏形。第三，摆脱了学科和知识类型的束缚，为不同地区和学科间的教学交流搭建了一个共同的平台，为大规模考试提供了共同的参照体系。布卢姆分类体系的提出恰好是在二战之后各国急于提高教学质量、监控教学效果的时候，布卢姆的分类学正好满足了这种需求，各国的大规模考试都以此为框架来构建考试的目标，它成为20世纪教育领域内影响最为广泛的理论之一。

然而，布卢姆教学目标分类也存在相当明显的问题。第一，布卢姆的分类体系在指导思想上是理论性的，用以对目标进行描述的语言却是经验性的，导致了理论与实际操作上的矛盾，层次逻辑混乱。第二，布卢姆以思维的复杂程度的线性积累作为分类的线索，与真实的情况不符。第三，在布卢姆的分类体系中知识是一个没有深度的平台，只有思维操作的不断复杂化才构成了教学目标层次的不断提升，这

种做法人为地将知识内容与过程割裂开来。实际上，思维操作与知识内容是无法分开的。第四，布卢姆强调对教学目标和学习行为进行分解。但是评价不是一个分解行为的过程，相反需要对教学目标全面地加以考虑，需要对各种信息进行综合。分解的结果导致评价的简单化和低层次化，即只测量了低水平的思维技能而忽视了高层次的思维技能。

为了克服上述问题，布卢姆的学生安德森（Anderson）修正了布卢姆的分类体系。安德森不再把知识当成一个没有深度的平台，而是展开为一个维度，另一个维度是认知操作。知识维度包括了事实性、概念性、程序性和反省认识这四类知识，认知过程维度包括了记忆、理解、运用、分析、评价、创造等六个层次，每一个层次的认知操作都与每一类知识相互作用，构成了一个二维的目标分类体系（见图 1-1-2）。

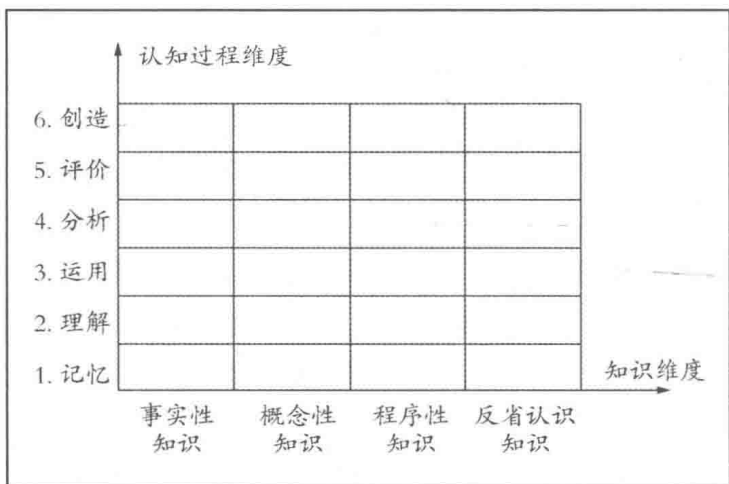


图 1-1-2 安德森的目标分类体系

安德森的修订只是使布卢姆的分类体系的框架结构变得更清晰合理一些，但框架本身赖以成立的假设依然是布卢姆的认知水平层次随认知复杂性增加而台阶式地发展的假定，因而布卢姆体系所存在的问题依然存在。为了克服布卢姆的分类体系所存在的问题，研究者提出了许多新的分类框架。其中值得注意的有马扎诺（Marzano）的新分类法和彼格斯（Biggs）的 SOLO 分类法。

（二）马扎诺的分类

马扎诺提出一个新的模型来解释人的心智行为。他认为人的心智行为是由三个系统控制的：首先是自我系统，决定是否投入到某项新的任务中去。其次是元认知系统，决定投入新任务所要达到的目标和方法。再次是认知系统，具体实施学习或解决问题的新任务。而在整个行为过程中，这三个系统不断地与知识储存进行互动，所有的心智行为都是这三个系统与知识储存之间互动的结果。

在此基础上，马扎诺进一步确定了六种不同的思维加工水平：信息提取、理

解、分析、知识应用、元认知系统、自我系统。他同时将知识分解为信息、心智程序和心理动作程序。和安德森一样，马扎诺假定每一层次的思维操作都与各个领域的知识进行互动，构成了不同水平的认知操作，如表1-1-1所示。

表1-1-1 马扎诺的分类体系

自我系统		动机、情绪反应、效能、重要性							
元认知系统		准确度、清晰度、过程、目标							
认知系统	知识应用	调查、实验、问题解决、决策							
	分析	认定、概括、差错分析、分类、匹配							
	理解	象征、整合							
	信息提取	执行、回忆、再认							
			细节	构想	技能	过程	技能	过程	
			信息		心智程序		心理动作程序		

马扎诺用知识与认知操作之间的相互作用模型，取代布卢姆的从简单到复杂台阶式递进的假设，显得更加合理。特别是他将动机意向和元认知的问题与知识有机地结合起来的做法，更加符合教育的实际。马扎诺是教育实践家，他的分类法更倾向于实用，因而越来越受教师的欢迎，特别是在北美及周边海域的岛国。但是马扎诺的分类法依然没有完全离开布卢姆的台阶模型，区分不同层次的界线不够清晰。

（三）SOLO分类

SOLO的英文原文是Structure of the Observed Learning Outcome，意为“可观察的学习成果结构”。SOLO分类即依据“可观察的学习成果结构”进行分类的理论和方法。SOLO理论创立者彼格斯认为，作为考试评价，重要的不是学习行为而是学习结果，是学生解答问题时表现出来的、可以观察到的认知反应水平。找出学生表现出来的、可以观察到的认知反应水平的结构，就可据以对学生的表现进行分类。

彼格斯认为，决定学生的认知反应水平的因素有两个：一是反应的功能方式；二是在某种功能方式下的反应结构的复杂性。基于这两点，彼格斯将皮亚杰的发展阶段论迁移到知识学习上，进一步假设：不论是儿童还是成年人，当他们在学新

的知识时, 认知的发展是有阶段的; 不同的认知发展阶段之间, 存在着认知水平上的跃迁, 这种跃迁不是简单的量的积累, 而是一种质的变化。彼格斯提出, 可以从能力、思维操作、结论的一致性与收敛方式、应答结构等四个方面区分学生的回答水平, 并进一步具体地归纳出学生思维水平的五个层次, 如表1-1-2所示。

表1-1-2 基于认知发展阶段的SOLO应答层次描述

SOLO层次	描 述
前结构 (Pre-structural)	这是一种低于目标方式的反应, 学习者对问题的回答是混乱的, 要么是拒绝回答问题, 要么是同义反复, 要么就瞎说一气, 回答根本没有一致性的感觉, 甚至连问题是什么都没有弄清楚就收敛(指获得答案)了
单点结构 (Uni-structural)	学习者只能联系单个素材解决问题, 因此没有一致性的感觉, 只接触到一点就立刻跳到结论上去
多点结构 (Multi-structural)	学习者能联系多个有限的、孤立的素材解决问题, 虽然想做到一致性, 但由于基本上只注意孤立的素材而使收敛太快, 解答不完整
关联结构 (Relational)	学习者利用问题线索、相关素材及素材的相互关系解决问题, 并能在设定的情景或已经历的经验范围内利用相关知识进行概括, 在设定的系统中没有不一致的问题, 但因只在一个路径上收敛, 在系统外可能会出现不一致
抽象扩展结构 (Extended Abstract)	学习者利用问题线索、相关素材、素材的相互关系及假设解决问题, 能对未经历的情景进行概括, 解决了不一致性的问题, 认为不必使结论收敛, 即结论开放, 容许多个在逻辑上相容的解答

SOLO 分类法的进步在于它具有严格的理论形态, 比较合理地解释了功能方式、学习周期和阶段等概念, 体现了现代心理学研究的结果。SOLO分类法将评价的目标从学习者的行为转移到学习行为的结果, 从学习结果在结构上的复杂程度出发来评价学生的学习质量, 评价目标的界定清楚、明确, 从而提高评价的效度和信度。SOLO分类法不再采用简单的线性发展模型, 而是采用螺旋式上升的水平层次, 更符合人类的认知规律, 更好地解释人在学习不同任务时的不同表现。正因如此, SOLO分类法正在发挥越来越大的影响, 国际上有重大影响的PISA测试也利用SOLO作为建立评价框架的主要理论依据。当然, SOLO分类法仍然是基于知识认知建立评价的标准, 而不是全面评量学生的发展。对于情感态度的领域的评价, SOLO理论仍然没有突破。尽管有这些不足, 依据SOLO分类法评价学生的学习成果, 更能反映学习的质量水平, 而不仅仅是知识的数量积累, 这与当前强调学生素

质发展的诉求是一致的，值得广大教育研究者深入研究，以期在评价实践中取得更大的进展。

三、考试成绩的描述和分析

考试的结果给出了一组分数，以此代表学生的水平或成绩。如何科学地描述、分析分数的情况，从中得出有意义的结论，这就是教育统计学的任务之一。教育统计学是一门专门的学问，需要专门的学习。本书不打算系统地介绍统计知识，只是从教师的角度，简单地讨论一些需要把握的统计量。

（一）几个简单的统计量

1. 集中量数：描述一组数据的集中趋势的统计量，其中最常见有算术平均数、中数和众数。算术平均数是一组数据的数值总和除以数据的个数所得的商，适合用来代表一组数据的总体情况。中数也称中位数，是位于一组数据中间位置的那个数的数值。在考试的情况下，中数反映中间状态的学生的学业成绩。众数指一组数据中出现次数最多的数值，反映一组数据中最大的群体的情况。

2. 差异量数：差异量数描述一组数据的离中趋势，也即分散程度。最常用的差异量数为方差和标准差。方差是一组数据中的每一个数与平均数的差值的平方和的平均值。标准差则为方差的平方根。每一个数据与平均值的差的平方，反映的是这个数与平均值点的距离，再加以平均，反映的就是平均距离。方差和标准差都反映了一组数据与平均值点的平均距离。距离越大，显然数据的离散程度就越大。

3. 相关系数：不同的事物之间总会有某种程度的关系，如因果关系、共变关系或相关关系。有相关关系的两种事物同时都在发生变化，但两者之间既没有因果关系，也不是受第三因素掌控的共变关系，它们之间的关系是复杂的。

反映这种相关关系的系数称相关系数。相关系数的计算原理比较复杂，不同的计算方法可以得出不同的相关系数，反映不同的相关关系。相关数值有正负之分，正相关说明两种事物的变化方向是相同的，负相关则相反。相关系数的绝对值大小在0~1之间。数值为0表示两者完全没有相关，数值为1说明两者完全一致。

一般情况下，因为计算过程中残差的积累，导致任何两组足够大的数据都会显示某种程度的相关，所以，判断相关的显著性就非常重要。只有显著的（足够大的）相关才是有意义的，否则没有意义。相关是否显著（有意义）的判断标准是根据经验界定的，判断相关是否显著的统计量为 p ，也称为阈值。通常用于判断的阈

值有 $p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ 。

（二）效度和信度

效度和信度是描述测量或考试的质量的重要指标。效度指的是测量工具的有效程度，即测量工具是否真的测到所要测的东西。效度是一个非常重要的概念。在考试的情况下，效度反映考试结果的解释在多大程度上是有意义和有用的。在考试中常用的效度有三种，即内容效度、效标效度和结构效度。内容效度通常使用质的检验方法，邀请专家对考试的目的、标准、目标、内容结构、解答试题的思维过程、应试策略等方面进行讨论，确定试卷的有效性。效标效度是用某一权威的考试作为效标，计算被检验的考试和作为效标的考试结果之间的相关，以此作为被检验的考试的效度指标。采用这种方法时应注意，作为效标的考试和被检验的考试必须目的相同、性质特点相同，才能进行这样的检验。实践中常见一些人以高考作为高中平时测验的效标，这种做法是不恰当的。结构效度的检验方法有两种，简单的方法是用双向细目表来检验试卷内考试内容、能力要求、题型题量等各方面的分布的合理性。严格的量化方法是用验证性因素分析确定试卷实测数据与预设的结构模型之间的吻合度，后一种方法相对比较复杂，在学校内由教师主持的考试中不常用。

信度指一次测量的可信程度，反映了测量可能存在的误差。在考试的情况下，通常使用试卷的内部一致性、对半相关或者与某个作为效标的考试的相关，作为考查信度的指标。信度检验是最常用的、最重要的统计检验。从统计方法上看，信度指标的计算都是针对特定的一组数据而言的，数据变了，计算出来的指标就不相同。因此，任何一份考试卷第一次使用时的信度是好的，不能保证第二次使用的信度一定就好，必须重新检验。

（三）难度和区分度

难度反映考试或试题相对被试者的难易程度。通常用通过率表示难度，通过率越高、难度越小。

区分度反映考试或试题鉴别被试者的水平差异的本领或精确程度。考试中常用高低分组的平均分差值、试题与试卷的相关、试题得分情况分布曲线的斜率作为区分度的指标。

难度和区分度是一对关系密切的量，在考试分数分布呈正态分布的情况下，难度居中时区分度最高，难度过高或过低都会导致区分度下降。从整体上看，难度值中等的试卷和试题，有较好的区分度。但是一份试卷也不能全部由中等难度的试