高校中青年教师教学基本功指南

（下册）

◎ 段文美 张轩萍 邓 蕊 著

主审 卢 莉

红旗出版社

高校中青年教师教学基本功指南

下 册

卢 莉 主审

段文美 张轩萍 邓 蕊 著

红旗出版社

图书在版编目(CIP)数据

高校中青年教师教学基本功指南/[illegible]

第六章　高等医学院校学生质量评价

评价是一项科学性、专业性很强的工作，所以评价方法应根据评价对象加以选择，尤其要注意受教育者的认识规律和年龄特征，以期取得良好的评价效果。教育评价的主要对象是学生，学生学习质量的检查与评定是整个教育过程的有机组成部分。它是以教育目标为依据，通过测验、考试手段，系统获取学生个体发展和学习效果信息的过程，学业成绩的评定是在检查的基础上对学生个体发展和学习效果的价值判断，以衡量学生达到教育目标的程度。

第一节　学生质量评价概述

一、学生质量评价的意义

从系统论观点分析，学业成绩的检查与评定，对学校对社会都有实际意义，它是教学结果的内外信息的反馈，它的具体意义表现在以下几个方面。

（一）检测　通过考试可以检查评定学生完成学业的数量与质量，对学生个体掌握知识、技能的深广度及熟练程度，了解学生自学能力和分析问题、解决问题的能力，根据客观评定的成绩，作为升留级、学位授予、毕业证书及毕业分配的重要依据。

（二）导向　通过考试对教与学的双方，均有导向作用。对教师的教学工作，考试也是检查，教师可以了解教学效果，

总结教学经验，通过考试结果的各种信息，改进教学工作，对学生可以促进智力发展，提高学习水平，通过考试使所学知识系统化，也是加深理解和巩固提高的过程。

（三）激励　考试本身带有一定的教育性，考试可以帮助学生自我认识，使之了解自己的学习状况，认识自己在学习目的、学习态度、学习方法、意志品质、兴趣爱好等方面的长处与短处，从而明确方向，增强进取心。

（四）选拔鉴定　考试是国家落实培养目标选拔专门人才的措施，使培养合格的专门人才得以保证，考试评价的结果也成为选拔、鉴定、使用人才的重要依据。

二、学生质量评价的标准

教学过程中对学生学业成绩的测验，由于目的要求的不同，有两种不同的测验标准。一种为目标参照测验，另一种为常模参照测验。

（一）目标参照测验

指用于衡量学生实际水平的测验。“目标”是指某一固定的标准，如教学大纲、课程目标，作为判断考生在多大程度达到预定课程目标的水平。目标参照测验通常采用绝对评分方法记分，如百分制，100 分为满分，60 分及格，但及格线可以浮动，有任意性，美国曾有以 75 分为及格线的做法。

（二）常模参照测验

指用于衡量学生相对水平的测验。“常模”是指某一特定的学生集体在某项测验中实际具有的学业成就，即标准化样本在测验中的平均成绩，以此作为判断考生在该集体中学业水平的相对位置。常模参照测验通常采用相对评分方法记分，如 A、B、C、D、E，或优、良、中、及格、不及格，此种测验无法反映学生达到教学目标的

程度。

一次良好的考试，有时两种测验标准同时可以予以解释和应用，如某学生考儿科学，试题为100道，该生答对了80道，以目标参照测验解释该生答对80道，如以百分制记分实得80分；该生成绩超过全年级90%的学生，这种解释即为常模参照性解释。

三、学生质量评价的原则

从测量评价的方法学来说，在实施测量评价时应满足有效、普遍和合适三方面的基本要求，为达到上述基本要求，哈伯德博士（Hubard）等人就设计测量评价方法与工具，提出了以下指导原则：

（一）可靠性

可靠性又称信度（Reality），是指某种测试方法和工具所提供的结果在不同条件下所具有的可重复性，一次可靠的测试，不论主考人是否相同，但采取同样的方法重复进行或分次进行，其结果均能保持一致。信度也是统计学上的概念，可以用可靠性系数或信度系数予以表示。

（二）有效性

有效性，又称效度（Validity），是指一种评价测试方法在多大程度上能对评价对象进行有效的测量。有效性与可靠性有联系又有区别，信度是保证效度的必要条件，效度依赖信度，有效的结果必然是可靠的，但是信度不依赖效度，可靠的不等于有效的，因此任何一种评价测量，在确保可靠性的前提下，一定要做效度分析。效度分析应注意以下三种有效性：

1. **内容有效性**：是指被测量的事物或行为与本次评价目标是否紧密相关，如以师资队伍、教学设施、教学环境为指标来评价学校办学条件和规模，则相关性很高，同样指标评价学校管理水平，

则相关性不高，因指标中未包含管理水平的基本因素，如管理队伍结构、素质、决策调控能力及管理质量等。

2. **方法有效性**：是指某一种测试方法是否适用于对特定的对象进行有效测量，如考试方法采用多选题、是非题测量知识的记忆、理解和应用是有效的，但用以测量非认知领域的技能水平或态度则是无效的。

3. **预测有效性**：是指某一次评价结果，能在多大程度上预测另一种情况，如基础阶段学习质量的评价，可以推测专业学习阶段质量、临床技能考试质量、毕业后临床工作能力等。

有效性也是统计学上的概念，检测有效性以考试成绩为例，可以用本次考试得分与另一次同样内容的权威性考试得分，进行相关分析，相关性越高，有效性越大。

（三）综合性

学生质量的测量与评价，无论对个人或集体都应注意综合性原则，例如对医学生的学业成绩是通过掌握知识、技能和表现的行为来反映的，对医学学生的评价是通过德、智、体综合测评反映的。

（四）复合性

每一种教学测试方法与工具均有各自的测量范围，目前尚无一种可有效测量所有能力的工具，因此尽可能采取复合测量技术。

（五）客观性

一种测试工具所取得的结果，如何正确地、客观地赋值，使评价方法与评价标准取得一致，如考试评分，应尽可能控制可变因素，排除主观随意性。

（六）区别性

学生质量测量评价的结果，应有区分度，没有区分度的检测是无效的。为了取得较好的、客观的区分度，必须制订合理的、科

学的指标体系。如以组织考试为例，应注意命题的质量，要有覆盖面，不出偏题或难度过大的题，也不出过于容易的题目。区别性与难度也是一个统计学上的概念，可通过对试题的分析，求取区别指数、难度指数。

表 6-1　目标参照测验和常模参照测验对比表

要　点	目标参照测验	常模参照测验
目　的	判断考生达到课程目标的程度	区别考生学业成绩的差异
依　据	以课程目标为依据	以一学生集体实际水平的常模
试题样本	包括教学内容较窄，考题较多	包括教学内容较多，考题较少
测量方法	选择型试题少用	选择型试题为主
试题质量	按课程目标命题，即使偏易也应保留	难度适中，跨度大，区分度较好
评分方法	绝对评分	相对评分
信度、效度	一般认为不适应统计方法计算	适用传统的统计方法计算

四、考试方式介绍

通常，考试方法可区分为口试、实践考查、笔试三大类。这些考试方法都是按照考试目的设计的，例如笔试中的多选题，其重点是考察基本知识掌握的程度；简答题的重点是考查应用知识的能力；短论是考查分析问题的能力。随着教育目标测量方法的增多，还有可能设计出更有意义的方法。

由于各种考试方法是根据不同目的而设计的，因此，在教学过程中不可能应用一种方法来全面测量学生的知识和技能，每一种考试方法有其优点也有一些缺点。现将各种考试方法的优缺点介绍如下。

（一）口试

这种考试的优点是教师直接同考生单独接触，方式灵活，教

师有机会提出补充问题，学生便于作出个人论点的系统阐述。对学生的思辨能力和语言表达能力也可作出直观性检验，可由一名主考与若干名辅考联合进行评议。这种考试方式的缺点在于缺乏统一的标准，成绩缺乏客观性，并且不能重演，花费时间过多、进度慢，且培养一批高素质主考人员难度大。这些问题都限制了口试的运用。

（二）实践考试

考场一般设在实验室、医学现场或特定的模拟环境。这种方式的优点是，主考人可当场观察和审核学生操作技能的各个方面，直接了解学生分析问题和解决问题的能力，可以观测和考核学生的看法及对一般情况的反应性，考核学生在紧急情况下，鉴别轻重缓急的能力，以及处理各种材料的能力，是能力考试的有效形式。其缺点是，无论在实验室，还是在现场，考查的标准很难达到一致，耗费时间更多，也不便于大批学生同时进行考试。

国外正在推广应用的模拟考试，通常是：先通过病人提供一段简短病史，或先看一段事先选择好的录像，然后由学生按照所提问题的程序，一步一步地进行临床过程的处理。最后根据每道程序所采用的行动，由主考人员作出客观的评定。这种考试过程只要设计程序合理，可以完全实行自动化处理，解决耗时过多的弊端。

（三）笔试

笔试是当前国内外最普遍采用的考试形式。它可在规定的时限内保证学生完成最大数量的答案，通过考试内容还可主动调节学生的学习方式。

笔试的试题形式分为两类：一类是开放式试题，如写作题、论述题、问答题、翻译题、演算题等等。这类试题可由考生提出各种各样的答案，靠阅卷员的判断来决定答对的程度。由于判分过程中难免掺杂有主观因素，故这类试题也常称为主观性试题。

另一类是封闭式试题，如是非题、改错题、填空题、配伍题、多选题等等，这类试题只有一个答案是对的，这个答案可以由考生提供，也可以由试卷提出几个选择。由考生挑选一个正确的答案。由于答案对错分明，评分时不会掺杂有主观因素，谁来评卷，分数都一样，所以这类试题就常称为客观性试题。

第二节　标准化考试评价方法

随着经济建设和科学技术的发展，标准化的工作已受到世界各国的重视，而且发展速度异常之快。考试的历史是随着人类社会的发展而发展的。它已经经历了由不规范到规范，由面试到笔试，由手工劳动到电脑控制，由主观性考试到客观性考试的演变过程，标准化考试也属于标准化工作的范畴。作为一种新的考试方法，标准化考试在国际上已有几十年的历史了，但仍在发展之中。为了适应教育改革和教育评价的需要，我们应对标准化考试进行认真的研究并逐步加以推广。

一、标准化考试的意义、特点和功能

（一）标准化考试的意义

标准化考试属于客观性考试。它是根据现代考试理论，运用现代统计手段，严格按照科学程序设计与实施，并且有统一标准的考试。标准化考试，就其形式而言，同其他客观考试和问卷法相比，似乎没有什么特殊之处；但在内容及其程序的设计上又区别于教师自行编制的测验。标准考试是由专家对优秀考试的诸种条件的研究而亲自制作的，经过科学手续制成标准，即称常模。只要把考试后的结果同这一标准对比分析，便可判断被试者的程度。这个标准具有代表性，因而应用范围很广，小至地区，大至全国，

对任何一个学生的程度都可以测定，有标准化的意义，所以称做标准考试。一般把这种考试又称做正式考试，而把教师自制的测验称做非正式考试。

一般来说，教师自制的测验只能在本班或本校之内，就学生的知识和能力进行测定和比较，但不能用于多数学校或全国，所以自然无法同较大范围内的学生水平相比较。而标准考试，对任何学校或任何个人都可以利用，都可以比较，这是因为它有一个可比的标准——常模。这正是标准考试之生命力所在。要了解每个学生或每所学校的成绩在地区或全国的地位，就可以利用标准考试。

(二)标准化考试的特点

1. 考试水平具有代表性

标准考试是一种标准化的考试。所谓标准化，是指考试的每一个环节都有一定的质量标准要求，即从试题的编制、考试的实施到评分、记分、分数的合成及解释都要标准化。

标准考试有可能对照的标准，即常模。所谓常模，是指个体之间进行比较的标准。某一学科的常模，就是指在该科考试中某团体的平均水平，常以平均数和标准差来表示。由于各种考试的目的、性质、难度不同，其考试原始分数的价值也不同。所以，在评价的过程中一个孤立的分数，如果没有确定的常模做标准进行比较，是无法判断其高低是否的。在标准考试中，常模就是比较分数、解释分数的依据。评价时，将评价对象与常模比较，从而判断其在团体中的地位。这种评价，主要是通过横向比较来判断优劣。而要判断评价对象达到目标的程度，则要以既定的目标为标准。

2. 考试对象不受限制

学校里的学业成绩考试的对象是学生，考试内容以教学大纲和教材为准。而标准考试，则是一种水平考试，因此，不仅不受考试内容的限制，而且不受考试对象的限制。标准考试着眼于一定

的水平，虽然也能反映某些教学大纲和教材内容的要求，但却不考虑参加考试者所学内容是否相同以及学了些什么。

标准考试的内容和范围，由于不局限于某一套教学大纲和教材，适应性很强，所以同一国家不同地区的考生或者不同国家的考生，只要具备考试所要求的水平，均可应试。例如，美国的高等学校入学水平考试(SAT)，每年应试者近200万；美国为外国留学设计的英语水平考试(TOEFL)，其对象很广泛，不论哪个国家和地区的考生，只要通过这种考试，均可取得留学美国的资格。可见，标准考试无论使用者是否相同，都可以测得同样的效果。

3. 考试分数稳定，可靠性强

由于标准考试是以科学程序建立起来的常模为标准，所以分数的合成和解释都有科学的依据，每次考试的分数都保持稳定，不会因考试的地点、时间及工作人员的不同而发生变化。也就是说，每次考试的分数都有稳定性，不存在分数升值或贬值的问题。

标准考试兴起于20世纪30年代，走过近70年的历程，实践证明其稳定性是比较强的。例如，美国的TOEFL考试，虽然历经几十年，但考生分数的平均值却长期稳定在550分左右。我国的EPT考试，多次考试分数的平均值都稳定在90分左右。这表明，标准考试的分数可靠性比较强，考生的分数能反映其真实的水平；考生所获得的分数与他的真实水平的差距越小，考试分数就越可靠。实际上，这说明标准考试的信度高。但传统考试却做不到这一点，往往是多次考试难易不等，分数上下波动很大，考试结果也因分数不等值而无法进行比较。

4. 考试题量大，有效性强

标准考试以客观性试题为主，形式多样，诸如选择题、填充题、改错题、是非题、判断题等。每个问题只有一种正答形式，按正答数量多少得分。无论是机器评分还是人工评分，其结果都是一

样的。这种客观性试题的评分办法，可以排除各种人为因素的干扰，确保考试结果的有效性。

标准考试有的还采取客观性试题与主观性试题相结合的形式，但不管采取何种形式，试题量都是很大的，覆盖面也宽。例如，美国医学院的基础课考试，试题量多达 800～1000 道，平均每门课 200 道题左右；GRE 能力倾向考试试题为 214 道，规定考试在 180 分钟之内答完；美国的 SAT 考试，题量通常为 220～240 道，限定时间为 180 分钟，平均每小时答题 75～80 道。我国的 EPT 考试，题量为 160 道，考试时间为 160 分钟。

题量大，内容多，范围广，就有条件测量考生的知识和能力，达到预期的目的。例如，美国的 SAT 考试，数学部分的试题，不仅能考核对数学基础知识和基本概念的理解，面且能考核综合运算能力、定量分析能力、推理判断能力以及解决实际问题的能力；语言部分的试题，既能考核基本词汇和语法，又能考核阅读理解、分析、运用、归纳、推理及书面表达等能力。这表明，标准考试的试题能代表所要测量的知识和能力，考试结果能准确地测量预定的目标程度，这就叫做效度高。

5. 考试程序科学化，质量能得到控制

标准考试的考试设计和实施的全过程，均按照系统的科学程序进行，严格规定考试的手续和时间，采用先进的统计方法，并且要求全体人员按照这一规定去做，最大限度地减少误差，使考试质量得到有效的控制。例如，设计阶段须控制的有考试目的、考试大纲、编题方案、编题及审题、预测、试卷、评定标准、考试说明书等；实施阶段须控制的有实施方案、印制试卷、施测、阅卷、分数转换与统计分析、考试分析报告、考试结果处理等。

为了更好地实行质量控制，许多国家还成立了考试机构，并拥有一批高素质的专家队伍，包括考试专家、学科专家及专业技术人员等。例如，美国教育服务公司有专职考试工作人员 2700 余

名，其中各类专家近 700 名、博士和硕士 400 余名；而英国的 UCLES 考试委员会，下设国内外各种考试的学科委员会 20 多个，各类专职工作人员 250 余名，每逢考试还需要雇用工作人员 4880 余名。

（三）标准化考试的功能

1. 能推进教育质量的全面管理

标准考试的核心，是其标准化。没有标准化，标准考试也就无从谈起，或者说是不能存在。所谓标准，从广义上看，是衡量事物的客观准则，它存在于社会生活的各个领域，其中也有质量管理领域。标准化已在企业的科学管理中发挥了巨大的作用。据统计，1981 年底国际标准化组织制定的国际标准有 4580 个，到 1984 年底增至 5692 个，增加了 24%；每年制定的标准数由 1981 年的 521 个增至 1984 年的 603 个，增长 16%。截至 1997 年底，我国颁布的国家标准总数已达 18359 个，其中强制标准 2000 多个，推荐性标准 16000 多个。

对教育质量实行标准化的管理也是人心所向、大势所趋的事。教育体制改革的根本目的是提高民族素质，多出人才，出好人才。为达到这个目的，不仅需要改革旧的教育思想、教学内容和教学方法，而且需要改革考试方法，使考试方法逐步走向现代化、标准化。这是实行标准化管理的需要。只有采取检验人才质量的标准，才能做到准确量才、人尽其才、人尽其用。实行标准考试，可以促进教育质量的科学管理。标准化与质量管理的关系极为密切，标准化是质量科学管理的基础，没有标准化，就不能更好地进行质量管理活动；没有质量管理，标准化的实现也就没有可靠的保证。

实行标准考试的过程，就是一个标准化的质量管理过程。例如，标准考试要求命题标准化、答案标准化、施测标准化、评分标准化、统计标准化、解释标准化，等等。由于标准考试一般规模都比较大，考生数量众多，需要统计的数据庞大，要想把各个环节的

质量都控制在标准化的范围之内，就必须借助电子计算技术，用机器取代人工操作。20世纪50年代，用计算机统计考分，但数据还是人工输入，所以速度慢，且有差错。60年代，考试统计技术更新，用光学扫描器取代人工输入，速度快，准确度高，效率倍增，光学扫描器每小时可阅卷600000份。70年代以后，电子计算技术在考试方面的使用范围逐步扩大，已不局限于考试的最后阶段，而是扩展到考试的全过程，除了考试的设计和实施这两个阶段以外，还用于考生的资格审查、考试结果分析、网上录取、试题储存管理等方面。这种现代化的考试手段的使用，有力地促进教育质量的全面管理，推动了教育标准化的进程。

2. 能促进评价手段的科学化

教育评价手段的科学化程度，是衡量一个国家的教育现代化的重要标准之一。现在，许多发达国家都用标准考试这个现代化的测量手段来评价本国的教育质量，还利用这一手段进行国家之间的学力比较。标准考试是得到国际公认的评价手段。例如IEA先后在1964年和1981年进行过两次国际数学教育调查，并进行了国际学力比较；去美国留学的学生，不管其英语水平受到过何种评估，均不足为据，要判断其英语水平的程度，就必须经过美国的托福考试才能认可。托福考试是标准考试，是得到100多个国家公认的，所以凡去美国学习的留学生，均需参加托福考试。

3. 能促进国际间的文化交流

在国与国之间互派留学生时，或因某种原因需要在别国的高等学校读书时，往往存在不承认对方国家学力水准的问题。究其原因，可以列出许多条，但主要的还是没有经过公认的标准考试。你说你的水平高，我说我的水平高，采取不承认主义，不利于国际间的文化交流。如果各国能使用公认的学力标准，那么这个问题就比较容易解决了。在国际上，推行公认的标准考试，是解决此类问题的好办法，而且有助于国际间的文化交流。

为了推行教育标准化的工作，促进国际交流，联合国教科文组织于1976年制定了《国际教育标准分类》。这个教育标准分类，是联合国教科文组织对许多国家的教育情况做了大量的调查后制定的一个教育统计标准。该分类把学前教育到研究生教育分成八个教育层次，每个层次下设若干学科领域，各学科领域又分设若干课程计划组，这样就形成了教育层次、学科领域和课程计划组三级结构。这个分类标准共有八个教育层次，大约100个学科领域和500个课程计划组。它不仅使各会员国在国内和国际间收集、整理和提供教育统计资料时有了一个国际通用的可行工具，而且有利于国际间的教育比较。

总之，推行教育标准化的工作，特别是推行标准考试，不仅有利于教育质量的全面管理，而且有利于评价手段的科学化，还有利于国际间的文化教育交流。因此，我国应有组织地、有计划地开展标准考试的理论和方法的研究，同时应继续试行和推广标准考试，以适应教育改革的需要，适应经济和社会发展的需要。

二、标准化考试的原则

任何一种考试方法，包括标准化考试，在设计的时候都要遵循以下四项主要原则：

1. **客观性**。一种客观的考试方法，由任何主考人对同一组学生进行分等时，大体上应能达到一致的水平。

2. **正确性**。考试可以被认为是一种可以任意使用的测量工具，不应受其他因素的干扰。例如测量学生分析问题和解决问题的能力时，它不应涉及按记忆做出回答的内容。

3. **可靠性**。这里所说的可靠性，就是指考试成绩的可信程度，它是一个统计学上的概念。因此从这一意义上来说，同一种考试方法在不同条件下对同一组学生进行考试时均应产生相同的成

绩；或者在基本相同的条件下，应能取得相同的成绩；或者同一范围的试题被分成两次进行考试时，也可产生同一水平的成绩。

4. **相关性**。即考试的选题标准与测量方法的一致性，也就是说，对必须评定的知识和技能，能否用某一种考试方法测量出来。

其他值得注意的因素有：

“平衡性”，即有关每一教育目标的试题数目与教育目标测量方法理想分量的比例彼此应当协调。

“同一性”，即考题与教学内容之间应当协调。

“区别性”，即测量方法的每一部分，在特殊情况下也能区别出好的和较差的学生。

“效率”，即单位时间内，用这种测量方法能保证学生完成最大数量的答案。

“时间”，由于允许用以考试的时间有限，如果在考试中引进无关因素，如猜测、冒险或碰巧，那么，这种考试方法是不可靠的。

实际上，我们在设计考试和命题时往往会背离这些原则，命题时往往抓不住重点，有些题目措辞表达非常含糊（多选法尤为多见），这样，学生用以理解考题所花的时间也许比回答这一考题的时间还要多，甚至强迫学生按照主考人的偏见和过时的概念来回答问题。这是需要尽力避免的。

三、标准化考试的制定

（一）考试大纲标准化

考试大纲和考试结果是提供反馈信息的主要手段之一。考试大纲可以保证考试标准的稳定性。就以题型的题量而论，如果每次考试采用的都不一样，要想控制试卷的难易度几乎不可能。再其次，对考生来说，如果在考试之前对怎样考法毫无所知，也容易产生心理负担，特别是增加临场的心理负担，还会分散考生全面复习功课的精力。因此，考试大纲具有“安民告示”的作用，促进学