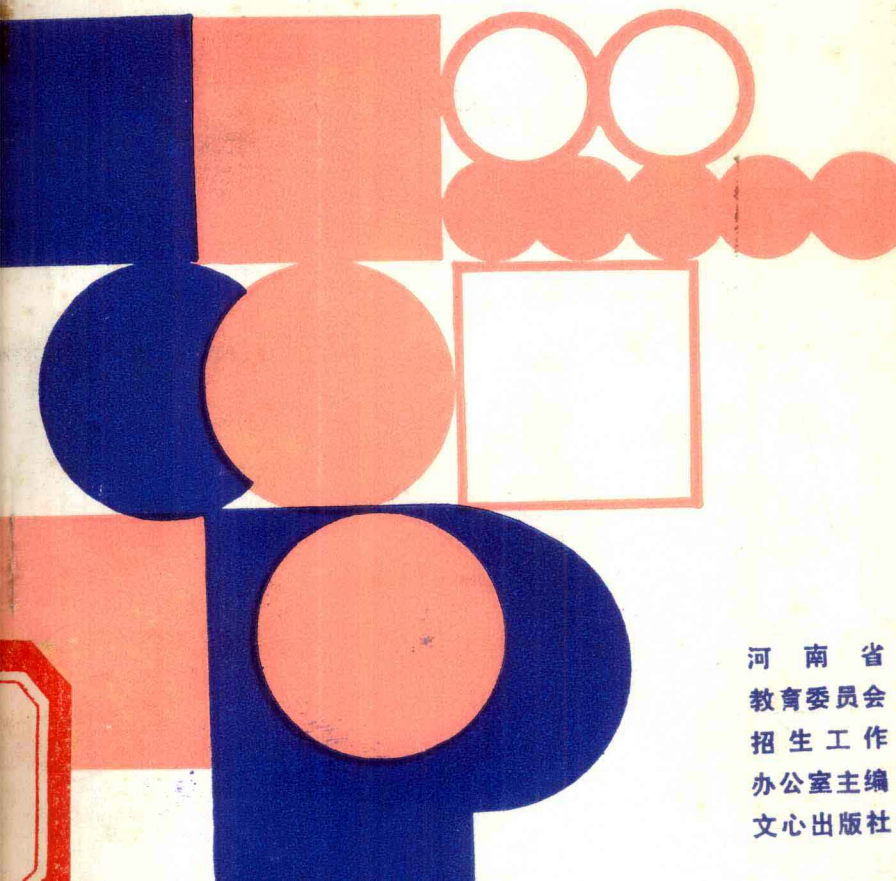


KAOSHI
YU
BIAOZHUN
HUA

考试与标准化



河南省
教育委员会
招生工作
办公室主编
文心出版社

考试与标准化

河南省教育委员会 主编
招生工作委员会

文心出版社

208352

外语教材

主编 李 强
副主编 王 强

考试与标准化

河南省教育委员会 主编
招生工作委员会

文 心 出 版 社 出 版
郑 州 解 东 印 刷 厂 印 刷
河 南 省 新 华 书 店 发 行

787×1092毫米 32开本 5 印张 100千字

1987年3月第1版 1987年7月第1次印刷

印数 1—8,500册

统一书号7423 103 定价1.00元

前 言

什么叫考试？在学校教学过程中，有哪些不同类型的考试，它们各起什么作用？如何组织这些考试，怎样处理和解释考试分数，并根据这些分数所传递的信息作出有益于改进教学活动的正确决策？

什么叫标准化考试，它与传统考试有什么区别？怎样组织标准化考试？中小学教学如何适应标准化考试？如何通过考试来评价教师和学生的教学质量？

怎样编写各种类型的试题，怎样努力使日常教学过程中的考试标准化？

.....

考试对教学活动有着一定的调节功能，这就是人们常说的“指挥棒”作用。如何正确发挥考试的“指挥棒”作用，从而使之更有益于学校教学，这就是我们编写这本小册子的初衷。

为此，我们走访了一些学校，耳闻目睹了种种对考试模糊的、不全面的认识和做法。这就越发使我们感到：端正对考试的认识，科学地组织各种考试对于提高学校教学质量是多么重要！

这是一本面向各级各类学校教师和教育行政干部的业务

进修书，因此我们力求使之紧密结合教学实际，对一些专门性的统计和测量原理，在保证科学性的前提下力求简明、通俗、易懂。

如果广大读者合卷之余，能对考试有一完整认识，并对考试研究有一积极的态度，努力使日常考试标准化，从而有益于教学质量的提高，那么，我们将感到极大的欣慰，因为编写本书的目的达到了。

本书由河南教育学院任建胜同志执笔编写，北京师范大学心理学系冯伯麟老师、广东省教育科学研究所张敏强老师提出了宝贵意见，张敏强老师还撰写了其中第八章和第五章的部分内容。黄委会科研所杜幼玲同志为本书绘制了大量插图。

河南省教育委员会招生工作委员会

五、分数合成的标准化.....	(63)
第六章 自编测验的标准化.....	(70)
一、测验设计.....	(70)
二、试题编写.....	(76)
三、评分.....	(87)
四、分数解释.....	(90)
第七章 中小学教学与标准化考试.....	(101)
一、中小学教学与“双向细目表”.....	(101)
二、学会使用考试.....	(105)
第八章 利用电子计算机技术, 促进考试现代化.....	(110)
附录 I 教育统计学基础知识.....	(127)
一、特征量数.....	(127)
二、相关系数.....	(134)
三、正态分布.....	(142)
附录 II	(148)
表 1、正态曲线分布表.....	(148)
表 2、积差相关系数(γ)显著性临界值.....	(153)
附录 III 广东省1986年物理科标准化考试命题双向细目表.....	(154)

第一章 对考试的正确认识

一、什么叫考试

考试，作为学校教育的重要组成部分在我国由来已久。自唐代建立科举制以来，人们一直习惯沿用传统的考试方法，而对考试理论的研究并不重视，以至于人们对“考试”这一经常接触的概念不能正确地理解，甚至曲解。

什么叫考试呢？我们将考试理解为：对照教育目标，对受教育者由于学校教育而得到的发展所进行的测量和评价。即考试包含着测量和评价两层含义。

所谓测量，指的是客观、正确地把学生达到教学目标的程度数量化（即考试的评分）。在此，测验（考卷）是测量的工具；评价是以测量结果的价值为课题，即对考试分数的解释、分析，进而作出判断的过程。在这里，解释分数的参照系是评价的工具。

显然，测量是考试的手段，评价是考试的目的。客观地评价是以正确的测量为前提的，两者互为依存，不可偏颇。然而，在日常教学的考试中，人们常满足于评分，并简单地根据分数多少给学生“排队”，而忽视了正确地解释分数中包含的有益于改进教学的信息。

二、考试在教学活动中的功能

现代教育理论证明：教学效果的反馈是提高教学质量的关键。在教学过程中，根据不同决策的需要而实施与之相应的各种考试，有利于教学信息的反馈。

（一） 考试与决策的种类

1、教学决策与形成性考试

在教学过程中，教与学的双方都希望了解教学活动的效果，以便作出调整今后教学方向和重点的决策。适应于教学决策的考试叫形成性考试，它的主要目的是在教学过程中通过测量和评价使师生双方都了解以往的教学状况。因此形成性考试中最重要的一点是使师生明确下一步的教学对策。教学中的“单元考试”应属于形成性考试。

对于教师来讲，形成性考试的意义不仅在于对学生学习质量的评价，更为重要的意义在于检查教师自身的教学效果，得到反馈信息，从而改进和调整今后的教学。

所谓学习质量，实质上是学生由于教学活动而引起的各种发展的概括，它包括学生的知识、能力、技能以及价值观念、情感等方面的发展和实现。尽管影响学生学习质量的因素是复杂的，然而最为重要的因素是教师的教学工作。因而教师教学质量的好坏直接体现于学生的学习质量上。通过考试，教师对自己以前的教学工作作出评价（如教学预期目标的实现程度，不能达到预期目标的原因，学生学习中的薄弱

环节)，得到反馈信息，从而调整今后的教学活动，扬长避短，提高教学质量。

形成性考试应该在某一阶段的教学告一段落，另一阶段的教学即将开始时进行。形成性考试的考卷，应该在本阶段教学内容范围内取样构成。由于内容范围相对较小，考卷应尽量包括所有的教学重点、难点和必须掌握的内容。

2、补救决策和诊断性考试

“长善救失”，“因材施教”。早在二千多年以前我国的思想家们就提出了这一精辟的教育理论。《学记》中写道“教也者，长善而救其失也”。其中的“救失”就是现代教学中的补救决策。

“失之何处”，“救以何方”。诊断性考试针对前者，补救决策则以后者为目标，两者相辅相成。

诊断性考试之所以不同于形成性考试是因为形成性考试面向全体学生，而诊断性考试只面向个体学生；形成性考试以某一阶段的教学内容为考试范围，目的在于得到今后如何教的信息；诊断性考试以以往教学某一薄弱环节为考试范围，目的在于发现造成这一薄弱环节的原因。因此，可以把诊断性考试理解为是形成考试的具体形式之一。例如，我们通过形成性考试发现甲、乙两个学生都没能正确地求证一道有关三角、平面几何题，那么能否据此认为他们两人在以往的学习中都没有很好地掌握这方面的内容呢？我们希望更确切地知道他们究竟是由于没有掌握三角知识，还是由于没有掌握平面几何知识而造成不会做该道题，就可以分别对这两位

学生施以平面几何诊断性考试和三角诊断性考试，以了解是在哪些内容上产生了阻碍。结果表明，甲生基本上掌握了平面几何、三角方面的知识，只是综合应用能力不够，乙生因为平面几何、三角方面的知识都没有完全掌握，从而出现了答题错误。

诊断性考试是一种针对性很强的考试，使我们能够对学生的技能、能力和知识进行精确分析，“诊断”学习困难的症结所在，为补救决策提供信息。

教师针对学生在以往学习中某一方面的困难和缺陷（诸如知识、技能、能力等），采取相应的补救措施，叫“补救决策”。补救决策应在诊断考试所提供的信息的基础上作出。

3、课程决策和调查性考试

教师在教学伊始，通过调查性考试，了解他将要教的学生已掌握的知识，所具备的能力以及存在的缺陷等情况，从而对要教的课程做出调整的决策叫“课程决策”。

课程决策主要服务于教师，它使教师对所教课程的增补、删减、顺序以及今后教学的重点和可能出现的难点心中有数。

以制订课程决策为目的的考试叫“调查性考试”。调查性考试的试卷内容，不一定限于某一教材内容的范围，还可包括与该课程有关的学科。

4、评价决策与总结性考试

评价决策指经过一段教学之后，教师对学生的学习情况

作出全面的评价，并据此对学生的学习质量作出判断，提出成绩报告单。

服务于评价决策的考试叫“总结性考试”。它的主要目的是对学生的学习做出全面的总结，所以它应在某一教学过程结束时进行。其考试内容应包括这一教学过程的主要内容及对能力要求。期末考试即属于总结性考试。

5、选拔决策与预测性考试

选拔决策是指高一级学校对是否录取某一学生继续升学所做的决策。服务于选拔决策的考试叫“预测性考试”。它对学生是否具备在高一级学校顺利地完学习的能力进行预测。尽管预测性考试也对学生在低一级学校的学习进行测量、评价，但它的主要指向是考生未来可能的发展，通过考试将考生在上述两方面达到的程度进行排列，以利于选拔。

在教学的不同阶段，教师要做出不同的教育决策。为此必须在教学过程的不同阶段实施目的不同的考试，从而获得有利于各种决策的不同的信息。

对于不同的教育决策，应使用与之相应的不同形式的考试，这样才能获得有利于不同的教育决策的反馈信息。对任何一位准备进行考试的教师来说，必须首先回答以下问题：

本次考试的目的和内容是什么？

明确了“为什么考”和“考什么”，才能编写考题。理

由很清楚，目的不同，决定了考试的内容、方法以及对考试结果的解释也不同。如果不明确考试的目的，就不可能编写出有效的考卷，也不可能根据考试结果得到有益于改进教学的信息，这种考试就没有必要了。

（二）考试的功能

考试在某种意义上评价着教育目的的实现程度，判定着教学效果的好坏，客观地向教与学双方提供反馈信息。

对于教师，考试的意义在于通过对学生学习质量的测量和评价来检查教师的教学效果，从而改进和调节今后的“教”。

对于学生，考试是对其学习质量的评价尺度，同时也是学生自我评价的基本指标。一般地说，学生通过考试得到好的评价，亦即受到正反馈时，学习能给他带来好的、安定的情绪体验，使他的自尊心和自信心增强，从而使学生的求知欲得到强化。与此相反，经常得到考试的否定评价，即过分的负反馈，可能引起学生过份的自我防御和回避，从而抑制了学生的求知欲望。

由于考试结果客观上对学生的学习质量进行了评价，为得到好的评价，学生必然追求好的考试分数，无论在学习方法、思维方法或其它方面都尽可能地适应考试，因而考试在客观上又一定程度地制约着“学”。

考试为教学提供反馈信息，从而成为既是指导、控制教学过程的手段，也是促进学生的学习兴趣，进一步发展学生能力的具有广泛功能的手段。对于教学活动来说，考试具有

调节功能、组织功能、学习功能和发展功能。

应该强调的是，考试的上述功能必须以测量的客观性、正确性为前提。否则，考试的反馈信息功能将成为教学过程中的一种异己力量。

根据心理学的效果律可知道，任何行为如果产生不愉快的后果，人们就可能避免这种行为，或根绝这种行为。这里，愉快的后果的刺激物为阳性强化刺激物，不愉快的后果的刺激物为阴性强化刺激物。在教学过程中，对学生学习行为最直接的强化刺激物可以说是考试分数。由于传统考试只侧重于测量学生对知识的识记，而不易于测量学生的能力，加之识记大量的知识较其它学习方法更易于适应这种考试，所以尽管教学双方可能意识到学习兴趣和基本能力发展的重要性，但在传统考试面前也不得不采用大量传授和大量识记的方法，致使各种练习泛滥成灾。这里体现了考试分数的阴性强化作用——“两害相权，择其轻者从之”。值得注意的另一个事实是，学生的自我认识 and 自我评价，最初是基于别人的评价之上的，正确、客观地对学生的进步进行评价，是培养学生自信心的必要条件，这正是推动他们去追求成功和创造的精神动力。遗憾的是，传统考试通常不能发现学生的潜力，评价学生的进步，不能全面、客观地测量和评价学生的学习，从而使学生不能很好地认识自己、评价自己、发现自己，也使教师无法有意识地引导、培养学生，使全体学生在原有的基础上向着各自的发展方向发展。

第二章 优良考试的质量要求

当我们希望通过一次考试取得信息来帮助作出决策时，我们总要自问：“这次考试要测量和评价什么？即考试的目的的是什么？”“这次考试目的的实现程度怎样？”“考试分数可靠吗？怎样解释、评价这些分数？”等等。

这些问题正是评价考试质量的基本指标，考试的效度、信度和解释分数的参照系。

一、效 度

当我们要测量某一物体的长度时，便使用钢尺，因为对测量物体的长度来说，钢尺是十分有效的工具。但是，假如你使用钢尺去测量物体的重量，就成了一件荒唐事。因为对于物体的重量这一测量目的来说，有效的测量工具是磅称或天平。

生活常识告诉我们，任何测量工具都是相对一定的测量目的才有效，因此不论进行任何测量，在测量之前都必须明确测量目的，并根据测量目的选择最有效的工具，从而保证测量的有效性。否则，我们可能干出类似拿钢尺去测量重量的荒唐事。

教育测量也不例外，我们首先必须明确这次考试的目的

是什么，是测量学生掌握某一方面知识的情况，或某一方面能力的发展状况。与一般测量相比，教育测量相对更为间接，教育测量的对象和工具难以明确，其测量结果也难以数量化。

例如，我们希望测量学生思维能力的发展情况，但我们无法将学生的思维能力直接地置于某一量尺上进行测量，而只能借助于根据我们对思维能力的理解而编制的考卷来间接地测量思维能力。这时我们所使用的工具——考卷是不明确的，我们不知道这份试卷的每一道题是否都在测量思维能力，也无法保证学生在回答每一道题时都只依靠了他的思维能力而没有借助其它能力。

“这样的考试是否也犯了拿钢尺测量重量的错误呢？”

教育测量的间接性使我们无法直接回答这个问题，但对教育测量来说，确切地回答这一问题又至关重要。因此，对任何一个考试，人们首先要明确回答这次考试能否考出我们所希望考的东西，考试对考试目的是否有效。

效度指的正是考试对于考试目的测量的有效性程度。它是一个优良考试首先必须具备的指标。

效度总是针对一定的考试目的而言的。一次对于选拔学生十分有效的考试，并不能有效地诊断出学生在掌握某一方面的知识时所存在的困难，而一次十分有效的调查性考试也不可能十分准确地评价学生在某一阶段的学习状况。没有一种效度对于任何考试都适用。由于考试目的的不同，效度大

效有以下几种：

(一) 内容效度

内容效度指考试对预定要考内容测量的有效性，它用考试内容与预定要考内容之间的一致性程度表示。

预定要考的内容主要是教学大纲、教科书中规定的教学内容，它包括要求学生掌握的基础知识及应具备的一般能力。一个考试要具备较好的内容效度，关键在于考试内容必须能反映预定要考的内容。所以这种考试对考卷的基本要求是试题的取样必须有代表性（能够代表预定要考的内容）。

由于这种考试主要涉及学生是否掌握了某一范围的教学内容，或是否达到某些规定要求达到的教学目标，因而称之为内容参照测验或标准参照测验，日常教学中的考试多属此类。

在设计一个内容参照测验（标准参照测验）时，首先必须明确考试的范围和这一范围所包含的内容，然后确定各部分内容的比重（如表1），从而保证考试内容所具有的代表性；其次是将考试内容与教学内容和规定的教学目标进行分析对照，看看前者在多大程度上体现了后者，如果教学内容和规定的教学目标基本上能体现在考试中，那么就可以推断这个考试具有较好的内容效度。