

# 开发有效、 可信、公平的考试



一些高风险考试的目的是为了筛选出那些具有必须的知识、技能和能力的考生，也有些考试将应试者的表现放在统一的标准下比较优劣。无论目的如何，考试都必须有效、可信和公平<sup>(1)</sup>。

## 效度、信度和公平性

### 效度

教育和心理测试标准<sup>(1)</sup>将“效度”描述为“开发和评估考试时应考虑的最基本因素”。简而言之，考试的效度与以下两个问题密切相关：

1. 考试是否考核了其想要测试的内容？
2. 对考试分数的解释是否合理和适当？

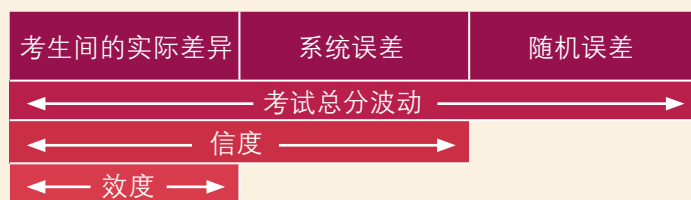
评估考题开发的过程，分析考题间关联性及其是否与考试方案相符等均可用来证明考试的效度。考试效度也可以通过比较考试分数与外部相关变量来进行评估。例如，某大学入学考试成绩可与学生表现进行比较，从而获取相关的考试效度信息。

### 信度

信度指考试的一致性和随机误差的大小。考试之所以有用是基于应试者的知识水平具有稳定性的假

设。然而，一定程度的考分误差是难以避免的。应试者的表现受许多因素影响，如焦虑或考试当日的状态等。尽管这些因素在应试者的控制能力之外，但考试主办方和组织者有义务标准化那些可能导致考试分数非正常波动的可控性因素，如考试环境和考试本身。

应试者应完全凭借知识层面的理解能力来答题，除此以外任何影响考生选择正确答案的因素均被视为误差。我们应该了解两种类型的误差。系统误差反映应试者之间的差异，与考试目的无关。如，定量推理试题要求考生有较强的口语表达能力，就会引入一个影响考试成绩的变量（语言能力），其与考试目的（评估定量推理能力）无关。随机误差是临时、偶然因素对考试结果的影响，如应试者看错题或论述题评分错误。下表反映了分数差异、效度和信度之间的关系。



# 开发有效、可信、公平的考试

## 公平性

教育和心理测试标准<sup>(1)</sup>指出，“公平一词被用于许多地方，没有一个具体的技术含义。”该标准列出了公平一词就考试有关的四种用法：A) 没有倾向性；B) 公平地执行考试流程、计算和使用考试成绩；C) 考试结果的等值化，能力相同的应试者应得到相同的考试成绩，无论他们属于哪个群体（如种族或民族）；D) 获得学习资料的机会均等。

试题开发过程中包括考题倾向性和敏感度审核，并可利用统计方法识别出那些对某些考生群体有倾向性的考题。试题表现差异（DIF）分析可以在控制考生能力差异影响的前提下，甄别出在不同考生群体之间表现不同的试题（如性别、民族、社会人口状况、年龄）。也就是说，答对考题的概率在不同群体之间有所不同，即使是具有同等能力的应试者也是如此。差异较大的试题需要经过更严格的内容审核，以调查试题可能存在的倾向性，并决定是否继续在题库中留用。应该强调的是，试题功能差异并不一定等同于倾向性，一道试题在不同群体间的差异可能有其他合理的解释。

标准的通过分数线设置、组卷和等值化对于建立公平的考试流程和评分方法非常重要。这些概念将在下文中讨论。

## 通过分数线设置

对于要求判定及格/不及格的考试，必须设置一个通过分数线，通常有三种方法：

1. 整体法 - 一个固定的百分比分数（如60%）
2. 定额参照法 - 一个固定的及格率（例如，排名前60%的考生）
3. 标准参照法 - 按照必须达到的能力水平来设定绝对及格分数

在这三种方法中，通常不会使用整体法，定额法只在特殊情况下使用（如有限的参加培训课程的名额），而标准参照法是我们推荐的方法。特别是对执照或认证类高风险考试来说，标准参照法更是业内广泛认可的首选方法。Pearson VUE的心理测量师正是使用标准参照法帮助客户设置理想的通过分数线。

## 组卷

Pearson VUE在计算机化考试组卷方面具有丰富的经验。我们的内容开发人员和心理测量师与客户一同选择适合的组卷模式，包括固定问卷、线性动态或自适应考试。固定问卷是指生成多个包含一定数量考题的且在统计意义上等值的试卷版本，线性动态考试是根据定义好的内容和统计特征从题库中随机抽取考题的组卷方法。

## 名词解释

### • 经典测试理论（CTT）

一个基于任何考试都是由两个独立部分组成之假设的考试开发与评估框架。这两个部分为：1) 考生的实际水平，2) 随机误差。根据特定考生群体中正确作答人数的比例，以及单题得分和考试总分的相关性（试题-考试相关性）对试题进行分析，并根据应试者答对的题目数量为其打分。

### • 计算机化自适应考试（CAT）

是一种计算机化考试，它根据应试者回答前面试题的表现从题库中挑选后面的试题。根据试题反应理论（IRT），这种测试的目的是为每个应试者挑选难度适中的试题。应试者答题表现好，后面的考题难度就会增加；反之则会降低。自适应考试的长度可以固定或不固定。

### • 结构无关差异

“考试分数受与其预定结构无关过程影响的程度。”（美国教育研究协会、美国心理学协会和美国国家教育测评委员会（1999）教育和心理测试标准，第10页）

在自适应考试中，计算机根据考生前面答题反映出的能力水平为其抽取后面的考题，从而更准确地评估考生的能力。选择组卷方法需要考虑到题库的规模和质量、考生数量和考试发送的频率。

## 等值化和量表化

通过等值化对及格分数进行调整，以使不同版本的试卷同等地反映出不同考生的能力，即某一应试者不会得到比其它应试者更容易或者更难的试卷。

Pearson VUE通常利用试题反应理论（IRT）将两个或两个以上问卷版本中的试题放在同一标尺上。IRT运用统计模型将所有试题参数放入同一标准量表中，从而建立起IRT标准题库。这样组卷和设置通过分数线时会考虑到不同问卷版本间的轻微难度差异，且在抽取考题时，未经测试和量表等值的考题不会出现在试卷中。如采用固定试卷，则会预先生成一定数量的经等值化的问卷版本，考试时为考生随机选取一个版本。如采用LOFT考试，则在考生开始考试时即时生成试卷。

## 考试功能

在计算机化考试环境下，除了确定考试心理测量统计特性之外，考试主办方还需要考虑哪些计算机化特征可以帮助达到考试目的。计算机化考试可以实现的多项功能也意味着，需要有技术性的考试参数，作为对考试方案或考试蓝本（考试内容大纲）的补充。

技术参数包含以下试题细节及其呈现方式：

- 试题预览
- 图表和图片的要求（例如，尺寸和格式）
- 与试题一起显示的辅助信息（如附件、说明、计算器）及其显示格式
- 是否允许考题/页面回看

实施计算机化考试时，应确保其有助于更好地评估应试者的能力，而不会造成其它方面的干扰或影响。

通过与Pearson VUE合作，并按照心理测量学最佳实践严谨地开发考试并充分利用计算机化考试的功能，考试主办方可以创建有效、可信和公平的考试。

## 参考资料

1：美国教育研究协会、美国心理学协会和国家教育测评委员会（1999）；教育和心理测试标准；华盛顿特区：美国教育研究协会。

2：国际测试委员会（2005）；国际计算机化和互联网考试指南，网址：<http://www.intestcom.org/Downloads/ITC%20Guidelines%20on%20Computer%20-20version%202005%20approved.pdf>（2012年1月8日检索）。

de Klerk, G. 经典测试理论（CTT）。M Born, C.D. Foxcroft & R. Butter (Eds.) 测评在线阅读，国际测试委员会，<http://www.intestcom.org/Publications/ORTA.php>（2011年12月5日检索）。

### • 结构表现不足

“考试未能抓住结构重点的程度。”（美国教育研究协会、美国心理学协会和美国国家教育测评委员会（1999）教育和心理测试标准，第10页）

### • 等值

用统计方法调整不同试卷版本的考生分数使其使用同一量表的过程。

### • 试题反应理论（IRT）

分析应试者对一组试题表现的统计模型，它的基本假设是，应试者正确答题的概率取决于应试者的“能力”和一到三个试题的特征。考题的三个特征用“参数”来表示。

## 美洲

全球总部

明尼阿波利斯市, 明尼苏达州

+01 888 627 7357

[pvamericassales@pearson.com](mailto:pamericassales@pearson.com)

[www.pearsonvue.com](http://www.pearsonvue.com)

费城, 宾夕法尼亚州

+01 610 617 9300

[pvamericassales@pearson.com](mailto:pamericassales@pearson.com)

[www.pearsonvue.com](http://www.pearsonvue.com)

芝加哥, 伊利诺斯州

+01 888 627 7357

[pvamericassales@pearson.com](mailto:pamericassales@pearson.com)

[www.pearsonvue.com](http://www.pearsonvue.com)

## 欧洲、中东和非洲

伦敦, 英国

+44 0 207 775 6737

[vuemarketing@pearson.com](mailto:vuemarketing@pearson.com)

[www.pearsonvue.co.uk](http://www.pearsonvue.co.uk)

曼彻斯特, 英国

+44 0 161 855 7000

[vuemarketing@pearson.com](mailto:vuemarketing@pearson.com)

[www.pearsonvue.co.uk](http://www.pearsonvue.co.uk)

迪拜, 阿联酋

+971 44 535300

[vuemarketing@pearson.com](mailto:vuemarketing@pearson.com)

[www.pearsonvue.ae](http://www.pearsonvue.ae)

## 亚太区

北京, 中国

+86 10 5989 2600

[pvchinasales@pearson.com](mailto:pvchinasales@pearson.com)

[www.pearsonvue.com.cn](http://www.pearsonvue.com.cn)

德里, 印度

+91 120 4001600

[pvindiasales@pearson.com](mailto:pvindiasales@pearson.com)

[www.pearsonvue.com](http://www.pearsonvue.com)

墨尔本, 澳大利亚

+61 3 9811 2400

[pvseasiasales@pearson.com](mailto:pvseasiasales@pearson.com)

[www.pearsonvue.com](http://www.pearsonvue.com)

东京, 日本

+81 3 6891 0500

[pvjsales@pearson.com](mailto:pvjsales@pearson.com)

[www.pearsonvue.com/japan](http://www.pearsonvue.com/japan)

致力于开发出  
有效、可信和公平的考试

如需了解更多信息, 请访问: [www.pearsonvue.com](http://www.pearsonvue.com)

培生教育有限公司及其分支机构2014版权所有, 保留所有权利