

计算机考试的喜与忧 PDF转换可能丢失图片或格式，建议阅读原文

[https://www.100test.com/kao\\_ti2020/138/2021\\_2022\\_\\_E8\\_AE\\_A1\\_E7\\_AE\\_97\\_E6\\_9C\\_BA\\_E8\\_c98\\_138892.htm](https://www.100test.com/kao_ti2020/138/2021_2022__E8_AE_A1_E7_AE_97_E6_9C_BA_E8_c98_138892.htm) 随着社会信息化对高素质IT人才的呼唤以及人们对提升自身IT水平的渴求，我国的计算机证书考试如雨后春笋般发展起来。面对种类繁多的计算机考试，人们有褒有贬，有喜有忧，有人祝它健康成长，有人急呼立即叫停。我们对这些考试热应如何冷思考呢？计算机证书考试品种繁多、目不暇接。论成绩非常巨大，谈问题十分明显。既充满希望、令人欣慰，又助长浮躁、令人忧虑。我们希望各种计算机考试好自为之，择善从之，兴利除弊，不断适应人才市场的需要。中国是考试大国 我国是考试历史最悠久的国家。自隋朝实行科举制度以来，已有1400年历史。考试成为中国文化博大精深的组成部分，科举弊端积重难返。废除科举近百年了，人们仍用“理科状元”、“文科状元”来赞誉各地高考的优胜者。我国又是考试规模最宏大的国家。例如，2005年高校计划招生510余万，其中本专科生475万，比去年增长8%；硕士生31.6万，增长15%；博士生5.5万，增长2%。如果按50%的录取率估计，参加考试者以千万计，无疑是世界上规模最大的考试。从考试种类上看，可谓名目繁多。有学校考试、社会考试；有学历考试、非学历考试；有选拔性考试，如升学考试、公务员考试；有水平性考试，如等级证书考试、资格认定考试。随着我国社会信息化的发展、人才需求的变化以及教育结构的不断调整，非学历证书考试迎来了广阔的发展空间。其中，计算机考试就属于非学历证书的社会考试。计算机考试应运而生在

全国性的计算机证书考试中，出现最早的是“软件水平考试”；规模最大的是“等级考试”；理念最佳的是“应用证书考试”。

1. 计算机软件考试 1987年软件水平考试开始试点。1991年人事部和信息产业部发布了《中国计算机软件专业技术资格和水平考试暂行规定》，由信息产业部电子教育中心实施。该考试的最大特点是“以考代评”。人事部规定今后对软件人员不再实行资格评审，考试合格者将取得相应技术任职资格。考试有初级程序员、程序员、高级程序员、系统分析员，分别对应技术员、助理工程师、工程师、高级工程师的职称。到2003年底，有100万人参加了考试，其中近20万人取得合格证书。该考试得到社会的认同，为IT人才的培养作出了贡献。但通过率较低，“以考代评”不易兑现，发展得不够快。2003年人事部、信息产业部又发39号文件，把软件考试纳入全国专业技术人员职业资格证书制度，并与日本相应级别的考试互认。从2004年开始，考试分为每年两次。共分5个专业类别：计算机软件、计算机网络、计算机应用、信息系统、信息服务。每个专业有3个层次：高级资格(高级工程师)、中级资格(工程师)、初级资格(助理工程师、技术员)。

2. 计算机等级考试 1994年教育部考试中心推出“全国计算机等级考试”(National Computer Rank Examination)，简称NCRE。它分为一、二、三、四级，一级为低端，四级为高端，但对各级证书的含金量没作任何规定。经过十年的发展，到2004年已累计考生1350万，取得证书者约450万。2004年考生就有310万，比上年增长23%。显然，它成为规模最大、发展最快的计算机考试，对形成庞大的计算机人才培养与考试市场作出了突出的贡献。

3. 计算机应用证书考试 1997年教

教育部考试中心的中英教育测量学术交流中心推出“全国计算机应用技术证书考试”(National Applied Information Technology Certificate Test, 简称NIT)。它借鉴了英国剑桥大学考试委员会举办的剑桥信息技术(Cambridge Information Technology, 简称CIT)的成功经验,采用了系统化的设计、模块化的结构、个性化的教学、规范化的考试,适合各种行业人员岗位培训的需要。NIT考试分为过程式考核、作业设计及综合考试三部分。过程式考核及作业设计在培训过程中完成,最后的上机考试由考试中心统一命题,对考生的独立操作和解决问题能力进行综合测试。从考试理论来分析考试是根据教育学和心理学理论,按照教育测量学方法,通过预制的试题,得到被测试者的反应结果,从而对其教育成就和心理特质进行描述。通常用“分”作为度量单位。

1. 选拔性考试和水平性考试 根据对“分”的解释,考试可分为选拔性考试和水平性考试两类。用考试理论的术语,选拔性考试称为“常模参照考试”;水平性考试称为“标准参照考试”。表1列出了两种考试模式的比较。“常模”代表被测群体实际达到的程度,考大学和考研究生都是常模参照考试,这种考试要求有比较高的区分度。而“标准”则是一组明确界定的目标或能力,它不注重区分度和难度,但是一定要符合考试大纲。从NCRE和NIT的实践看,命题都严格按照考试大纲进行,覆盖面很大,基本达到标准参照考试的要求。而软件水平考试虽然也是标准参照考试,但似乎有常模参照的特点,可能是为控制比较低的通过率的原因。

2. 教育测量理论 教育测量理论有两个:一个是经典测量理论,另一个是项目反应理论。经典测量理论(The Classical Test Theory, CTT)借

用普通物理测量的类似方法，对考生回答的结果与其心理特质存在的误差进行分析，形成区分度、等值、信度、效度等概念，使用相当广泛。它的不足是信度估计不精确、难度量表与能力量表不一致。项目反应理论(The Item Response Theory, IRT)建立在考生潜在能力的基础上，研究考生对试题的反应与其潜在能力之间的关系。优点是通过对模型测试考生能力可以精确估计测量误差，对考生能力的估计不依赖于题目的特殊选择。缺点是假设过于理想，技术复杂，可操作性差。

### 3. 难度与区分度

在经典测量理论中，难度就是考生答题时感到的困难程度，用考生群体在该题目上的通过率或得分率表示，又称为统计难度。它用0~1之间的数值表示。在等级考试中，把题目分为难、中、易三种，每个科目都有明确的难度分布细目表。每次命题后都要估计难度系数，而每次考试后再根据抽样数据统计难度，一般都比较接近0.5。在IRT中，我们把难度分为A、B、C、D四个层次：A为容易题(0.7~1之间)、B为中等偏易题(0.5~0.7之间)、C为中等偏难题(0.3~0.5之间)、D为难题(0~0.3之间)。在命题实践中，还采用了内容难度的概念，使其与考生样本无关，而由题目自身的属性决定。区分度是区分不同水平考生能力程度的指标。不过，预先对它难作估计。

100Test 下载频道开通，各类考试题目直接下载。详细请访问 [www.100test.com](http://www.100test.com)