

GMAT

美国 MBA 报考成功方略

(第1册)

主编 吴建国

审校 *David Nightingale*

- GMAT-CAT 考试须知
- 语文能力考试指导
- 语文能力分项练习
- 语文能力模拟考试
- MBA 项目和奖学金
- 578 所美国商学院简介

北京航空航天大学出版社

431
W830
1

GMAT

美国 MBA 报考成功方略

(第 1 册)

主 编 吴建国

副主编 陈 宏 张明生

编 委 张万安 欧阳旭东 徐志长 张新颖

顾 问 余名叔

审 校 [美] *David Nightingale*

北京航空航天大学出版社

图书在版编目(CIP)数据

GMAT 美国 MBA 报考成功方略(第1册)/吴建国主编.-北京:北京航空航天大学出版社,1999.9

ISBN 7-81012-908-2

I. G… II. 吴… III. 英语-研究生-入学考试-自学参考资料 IV. H31

中国版本图书馆 CIP 数据核字(1999)第 24460 号

GMAT

美国 MBA 报考成功方略(第1册)

主 编: 吴建国

责任编辑: 梁德润

责任校对: 陈 坤

北京航空航天大学出版社出版发行

北京市学院路 37 号(100083) 发行部电话(010)82317024

<http://www.buaapress.cn.net>

E-mail: pressell@publica.bj.cninfo.net

各地书店经销

河北省涿州市新华印刷厂印装

*

开本: 787×1092 1/16 印张: 24 字数: 614 千字

1999 年 8 月第 1 版 1999 年 8 月第 1 次印刷

印数: 1-4000 册

ISBN 7-81012-908-2/H·103 定价: 36.00 元

前沿报告(代前言)

21 世纪语言测试的新视点—计算机适应性考试 (CAT)

正值世纪之交,美国“教育考试服务中心”(Educational Testing Service,简称 ETS)宣布,自 1999 年起,ETS 所主办的全球性教育考试项目 TOEFL (Test of English as a Foreign Language),GRE (Graduate Record Examination),GMAT (Graduate Management Admission Test) 将告别“纸与笔”(Paper & Pencil,简称 P&P)的考试形式,代之以“计算机适应性考试”(Computerized Adaptive Test,简称 CAT)。

“一石激起千层浪”。ETS 的这一举动立即引起了全世界考生以及语言测试专家的关注。人们纷纷猜测:这只不速之“猫”(CAT)究竟是什么?为什么要用 CAT 取代 P&P?

其实,CAT 并非 ETS 的独家创造。早在 20 世纪 70 年代,一些语言测试专家就注意到了大规模的 P&P 考试所暴露出来的问题。例如,为了准确测试不同应试者的实际水平,试卷中必须安排不同难度级别的试题,但总数以多少为宜,则一直未能在技术上得到很好的解决。全体考生面对同一份试卷,水平高和水平低的考生都不得不耐心地完成不适合自己的能力的试题。数小时下来,不仅对考生造成了生理和心理上的压力,而且使测试结果的准确性受到影响。

面对这些问题,于是有人设想,假如让高水平的人回答高难度的试题而让低水平的人回答低难度的试题,岂不是可以省时省力,各得其所吗?这便是“适应性”思想的开端。美国语言测试专家 Frederic Lord 博士则是这一思想的开创者之一。他经过多年的实验,建立起一个“适应性考试”的理论构架,为后来的 CAT 发展奠定了基础。“美国军事研究院”是 CAT 研究的积极推行者,他们为美国陆、海、空三军研制的“军队服务职业智能考试”是对 CAT 的一次具体尝试。这个考试所积累的数据为后来 CAT 的进一步发展提供了宝贵的经验。但由于当时计算机技术尚不发达,而且成本高,CAT 研究一度进展缓慢。

进入 20 世纪 80 年代,一大批语言测试专家和心理学家携手加入 CAT 的研究行列。研究者们从题库模型、系统设计、试题信度与可比性诸方面对 CAT 进行了全方位的研究与实验,取得了突破性的进展。随着 CAT 由理性思考阶段进入实验阶段,其可行性与优势也渐渐显现出来。

计算机技术的发展为 CAT 的发展起了推波助澜的作用。20 世纪 70 年代时,计算机进入语言测试领域,扮演阅卷和处理成绩报告的角色;80 年代时,计算机进一步在大规模考试管理和题库建设方面发挥了重要作用;进入 90 年代,计算机技术突飞猛进,特别是多媒体技术的发展,使得计算机大踏步进入语言教学和职业培训领域,并使得对学习者实施个别指导成为可能。利用计算机测试个人能力与水平的工作也迅速提到了时间表上,一种充满生机的新的语言测试形式—CAT 诞生了。ETS 研究管理部副总裁一语道出了语言测试专家们的欣喜之情:“在走完了由襁褓到青年的路程之后,我们现在终于有了一个有着强大技术支持的新的测试工具—CAT。”

在被问及为何要以 CAT 取代 P&P 时,ETS 的专家们列数了 CAT 的十大优势:

1. 考试的安全性得以提高。与 P&P 考试相比,CAT 的试题储存在试题库之中,考生

只能在考试时接触到其中的一部分试题。即使经过考试记住了几道题,但题库中的试题数量之多,考生再次考 CAT 时遇到同样试题的机会极小。因此,P&P 考试时经常发生的试卷丢失,试题泄露等安全问题一举得以解决。

2. 考生可以自主控制答题速度。在规定的总时间范围内,考生可以根据自己的情况选择答题速度,不再受到 P&P 考试各部分所规定时间的限制。考生可从容不迫地完成某一部分的试题,然后进入下一部分试题。同时计算机自动记录下考生在每道试题上所花的时间,留作对考生在所测技能熟练程度方面的一个考查。

3. 考生得以充分发挥自己的水平。CAT 在对考生的现有水平作出初步判断后,选送给考生的试题使考生既不感到难以驾驭而失去信心,又使考生感受到挑战而跃跃欲试。其次,由于试题最大限度地涵盖测试目标所规定的点和面,考生得以全面展现自己的水平而不会出现漏测现象。

4. 节省考生的时间与精力。CAT 与 P&P 考试的最大区别是,CAT 一次测试的题目数量大大少于 P&P 考试。且于试题的难度与考生的实际水平相近,因此水平高的考生不必回答低级难度的试题,水平低的考生也不必为回答高级难度的试题而绞尽脑汁或猜测,各得其所,省时省力。

5. 减少考试过程中的操作失误。在 P&P 考试中,考生在答题纸上标示答案时,经常会出现漏标或错行现象。改动答案时,或由于橡皮使用不当而污损甚至弄破答题纸,或由于答案所涂的范围不够大,造成错判、漏判而影响成绩。CAT 考试中,考生移动鼠标在计算机屏幕上答题,考生的每个回答计算机都要求确认;未回答的问题,计算机不会让考生通过;如有改动(CAT 考试中对于首次选择的答案,在按“确认”键之前有一次重新更改的机会),计算机会自动删除旧答案,记录新答案。整个过程准确无误。

6. 当场获得考试结果。P&P 考试一般要待考试结束后两个月甚至更长的时间才能得知考试成绩,而 CAT 考试则可在考试结束后的几分钟内当场打印出成绩来。这不仅节省了考试管理方面大量的财力和人力,也使考生对自己的现有水平有了即时的了解。

7. 考生自由安排考试时间。实行 CAT 考试后,考生参加考试的时间变得十分灵活。从理论上讲,一年 365 天,一天 24 小时均可安排考试。考生可以选择在适合自己的时间报名参加考试,并选择将理想的成绩报告给所申请的学校。考生不再受到 P&P 考试次数和固定报名时间的限制,也不必担心由于某次考试未能注册而失去考试的机会。

8. 促进并加速题库建设。CAT 考试对题库有很高的要求,题库内的试题要不断更新和扩大。计算机可将预测用的试题在 CAT 考试过程中输送给考生去做而不会干扰考生的正常答题。预测的结果用作试题分析,将符合标准的试题送入题库,保证题库“吐故纳新”,题源不断。

9. 使试题设计多样化,测试全面化与精确化成为可能。由于先进的计算机技术可以同时处理文字、声音和图像信息,CAT 的测试形式不限于多项选择式,正确答案也不局限一个。听力、阅读、写作、拼写,甚至记忆能力均可以在一次测试中完成。

10. 为电子测试手段的不断改进提供了一个技术平台。鉴于 CAT 是建立在计算机技术基础之上的一种标准化考试形式,随着计算机技术的不断发展,CAT 将得以不断更新完善,而电子测试手段的普及又会推动电子技术在教学与培训领域内的应用,从而加快学习速度,促进学习效果。

那么,以电子作为测试手段的 CAT 是如何实现标准化考试的目标的呢?换言之,CAT

的理论依据是什么？

CAT 的理论依据是“试题反应理论”(Item Response Theory, 简称 IRT)。所谓 IRT, 实际上是一个数学模型, 用以描述构成某一试题的各种特定参数之间的关系以及考生在解答某道题时的表现与这些参数之间的关系。通俗地说, IRT 是一种计算试题难度的数学公式, 它包含三个重要参数: 水平系数、试题难度及试题区分度。考生在答题时与试题的这些参数之间产生交互反应, 这些反应再经过某种统计后便得到了考试的结果。IRT 是标准化考试的“灵魂”, 它确保在不同的考试条件下, 如场地、时间、试卷形式等, 所获得的测试结果相同。作为标准化考试的一种形式, CAT 必须达到三个目标: 1. 区分试题的特征, 2. 确定试题挑选的有效规则, 3. 使用同一度量标准。而这些目标都可以通过 IRT 来实现。

为了确保上述目标的实现, CAT 需要一个庞大的、高质量的试题库 (Item Pool)。一个理想的 CAT 试题库必须具备三个特征: 1. 最高区分度; 2. 均等难度分布; 3. 最小猜题可能性。要实现这三个目标, 必须做好题库建设中每一个环节的工作, 如确定考试范围、测试项目、试题数量、试题类型、难度分布等。更为重要的环节则是鉴定、挑选合格试题, 通过 IRT 模型对每道试题的三个参数分别进行测定并对试题进行校准。鉴于 CAT 用于测试考生水平的试题数目大大少于 P&P 试题, 且因人而异, 因此对考生来说, CAT 的每一道题都至关重要, 任何试题的任何缺陷都会给考生的实际水平的评定造成双倍的损失。除“主题库”外, CAT 还建有若干个“子题库”, 以支持补充“主题库”。“主题库”与“子题库”在各项设计指标上必须协调一致。由此可见, 题库是 CAT 的核心。

既然 CAT 是依赖计算机进行测试的一种系统, 它必然包括两大部分: 硬件和软件。CAT 对计算机硬件的技术要求主要有三个: 1. 控制试题输送速度, 两题之间只能有 1 秒的间隔; 2. 控制记录答题时间的速度, 要求达到 0.017 秒; 3. 屏幕分辨率要求达到 64×480 的标准, 其他硬件设备, 如键盘、鼠标、打印机则是常规配制。如果考试对声音和图像有要求的话, 则硬件中还必须包括声卡等。

支持 CAT 的软件必须具备六个方面的功能:

1. 重新启动功能。在考试的过程中, 一旦考生出现误操作, 计算机应能重新启动并迅速回到考生当前正在回答的试题。换言之, “考生工作站”(Examinee Testing Station, 简称 ET) 的运行程序应是一个开放系统, 能以多种形式进入或重新调整, 并能将变化的情况随时传递给“监控中心”(Testing Administrator Center, 简称 TA)。

2. 过程监控功能。考试的全过程通过“TA”实施监控, 了解每个“ET”站的进展情况, 如目前正在进行的试题及其水平等级; 及时对考生提出的“求援”(Help) 信号作出反应; 随时对考生的答题反应进行水平测定并挑选下一道试题; 决定考试开始与停止的时间等。

3. 随机记录功能。自动记录下考生对每道试题正确或错误的选择以及答题时间, 记录已经回答和暂未回答的试题并在适当时间提醒考生回答这道试题, 记录每道试题的水平参数及在考试结束后考生的考试成绩等。

4. 安全保密功能。对试题加密以防止“盗题者”。一旦发现有人盗题, “TA”应能及时收到警报并自动改变正在考试的主题库, 将盗题者引入备用的子题库, 并记录下盗题者的身份证号码, 以便考试后予以处理。

5. 求援应急功能。在考试过程中, 考生一旦出现操作上的问题, 或对考试指示不理解, 按 Help 键后便能即刻得到指示。Help 功能要设计下拉菜单, 以便考生针对自己的问题去选择所需帮助。

6. 操作辅导功能。编写供考生在考前进行的硬件使用与操作说明和必要的辅导性练习,以使考生熟悉操作要领和程序。作文考试要求用键盘输入的,则还要设计键盘使用说明和主要编辑方法。

其次,“ET”站与“TA”站之间还有一个连接形式问题。通常采用互联网形式,这种方法不仅能够节省大量计算机记忆容量,还便于“TA”对全局实施监控。互联网软件还需具备单机运行功能,一旦联网硬件或软件出了问题,考试可以立即换到另一台单机上进行而不致中断。

CAT 是如何工作的呢?让我们来观察一个模拟的 CAT 考试过程。

考生甲和乙来到某 CAT 考试中心,被分别安排在“ET-A”和“ET-B”座位上。

①考试开始,“监控中心”启动。设定甲、乙两人目前状况:

$$D(\text{甲})=0, \quad L(\text{甲})=0, \quad H(\text{甲})=0, \quad R(\text{甲})=0$$

$$D(\text{乙})=0, \quad L(\text{乙})=0, \quad H(\text{乙})=0, \quad R(\text{乙})=0$$

(其中, D 代表试题难度, L 为试题数目, H 为试题参数, R 为答对试题数目)。

②“监控中心”向题库发出指令:选送两道同等难度的试题到“ET-A”,“ET-B”工作站。

③指令完成。“监控中心”记录下第一道试题的难度 D。甲、乙开始回答第一道试题。

④“监控中心”分别记录下甲、乙对该题的反应,并对两人的反应作出正误判断:甲回答错误;乙回答正确。

⑤“监控中心”改写甲、乙目前状况:

$$D(\text{甲})=D, \quad L(\text{甲})=L+1, \quad H(\text{甲})=H-D, \quad R(\text{甲})=R-1$$

$$D(\text{乙})=D, \quad L(\text{乙})=L+1, \quad H(\text{乙})=H+D, \quad R(\text{乙})=R+1$$

⑥“监控中心”向题库发出第二个指令:挑选下一道试题,选送 $D=D-2/L$ 试题送给“ET-A”;选送 $D=D+2/L$ 试题给“ET-B”。甲、乙继续答题(此刻起,甲乙俩人的试题难度开始不同)。

⑦“监控中心”重复③,④,⑤,并向题库发出建立在⑤结果上的下一个指令。在进行若干道试题之后,

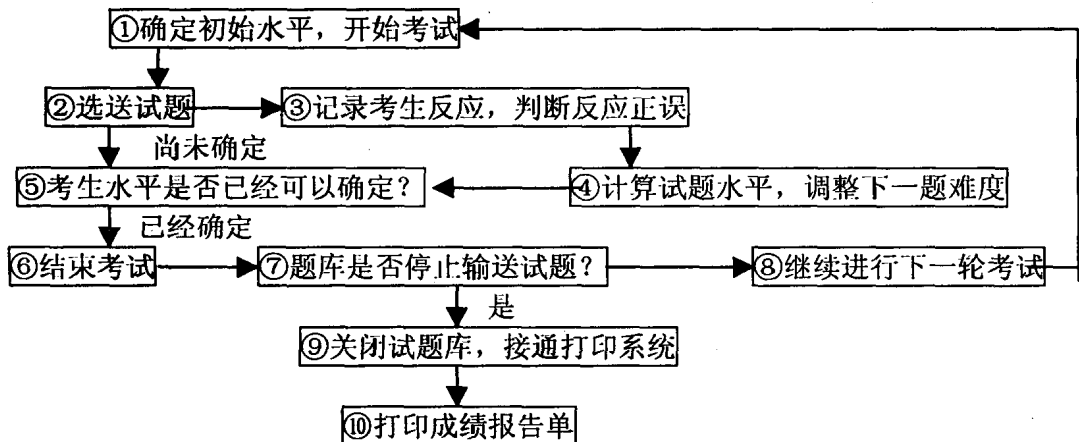
⑧“监控中心”根据甲、乙两人的答对题数以及答对试题的参数计算出甲乙目前水平系数。

⑨“监控中心”作出初步判断:甲的水平已经测定,指示题库停止向“ET-A”工作站输送试题;乙仍需进一步数据,指示题库继续向“ET-B”工作站输送下一道试题,难度为“D+目前状况”。考生甲结束考试;考生乙继续答题,直到乙的水平被确定,才停止向“ET-B”工作站输送试题。

⑩“监控中心”向打印系统发出指令:分别打印出甲、乙的水平报告。甲、乙获得各自的成绩报告单,并离开 CAT 考试中心。考试结束。

归结起来, CAT 的工作逻辑很简单:挑选确定考生目前水平的第一道试题→记录考生反应并测算初始水平→选送下一道试题→记录考生反应,修正初始水平→继续选送试题,持续记录考生反应→累计计算反应结果→确定考生目前水平→结束工作。

下面的流程图表示了整个考试过程。



CAT 考试的前期准备是一项复杂而细致的工作。

首先，即第一个问题是，选送给考生的第一道试题的难度水平如何确定？因为考生还没有开始答题，CAT 无从知道该考生的水平，因此也就缺乏判断的依据。解决这一问题的办法是，采集该考生的某些特定的辅助数据，如年龄、专业、学历。考察一组同类或相近人员以前的考试情况，以他们的考试成绩的均值作为估计该考生的初始水平的依据，因为该考生的实际水平标准误差值不会超出均值的 ± 10 。由第一道试题引发的另一个问题是，它会不会影响对考生最终水平的测定？因为如果第一步走错了，以后步步都有可能走错。为“防患于未然”，须进行大量复杂的模型设计、取样、演示和验证工作。

第二个问题是，考试过程中应该向考生输送什么样的试题？除了在难度上接近初始水平外，试题内容所涉及的知识面的问题如何均衡？这一问题不解决，则考生知识水平的多维性无法测试，结果当然也就无效了。目前的解决办法是，事先就设计好一张“信息表”（Info Table），这张表详细标明每道试题的难度水平及所涵盖知识面的量化数据。考试过程中 CAT 依据这张表来确定选送什么样的试题。

第三个问题是，何时停止考试？是依据某个事先确定的目标是否达到？还是事先确定的试题数目已经完成？或是事先确定的考试时间已经用完？方法之一是，固定试题数目。其好处是便于操作，且每道试题的利用率可以得到准确测定。此方法适用于以 IRT 模型检验整个考试的标准误差。方法之二是，对考生实行不等长度考试。这样做可以确保测试的准确性接近试题设计者试图达到的目标。方法之三是，全体考生进行同一水平标准考试。这种方法可对每个考生的成绩以同等标准误差来测算，同时也便于今后对试题进行标准误差方面的统计分析。方法之四是，规定考试时间。它适于用速度考试之目的，计算考生在规定的时间内答对的题数并以此测试某一速度下所达到的准确率。

最后一个问题，也是最重要的问题是：如何确定考生的最终水平？通常的做法是将所答试题的水平参数累加后求出平均值。

由 P&P 考试改为 CAT 考试，业内人士最关心的问题之一便是 CAT 的可靠性（Reliability）。“可靠性”系指考试的测试结果是否准确无误。评估一次考试是否可靠有三个基本标准：1. 鉴定测试误差产生的主要因素以及由此导致的误差程度。2. 证明在特定情况下两次考试的成绩达到预期的可靠程度。3. 对包含试题、考试形式、评分标准、考试

管理及其他测试因素在内的各次考试结果所得出的结论相同。

为了确保达到上述标准,首先必须知道影响 CAT 可靠性的主要因素有哪些。ETS 测试专家认为主要是两个:一是 CAT 的试题库,二是 CAT 的系统设计。他指出,在 CAT 题库建设的整个过程中,从试题编写、鉴定、抽样、预测到正式使用,处处都存在威胁 CAT 可靠性的因素,而消除这些因素的根本途径是保证试题的高精确度。试题的精确度是由试题的“标准测试误差”(Standard Error of Measurement)来标记的,其数值越低,试题的精确度就越高。理想的标准误差值在 0.1~0.3 之间。系统设计是影响 CAT 可靠性的第二大因素,它表现在两个方面:一是用于试题挑选的算法程序,二是试题曝光的控制程序。

围绕这几个方面的问题,测试专家们对 CAT 的可靠性进行了大量的、大规模的模拟实验。他们将同一批受试者由 P&P 考试得到的水平值与 CAT 考试的预测水平值进行对比,测定 CAT 的可靠性为 $P=0.9$ 。在另一项测试中,两组水平不同的考生同时回答来自 CAT 题库的两组题数相同但难度不同的试题,然后根据考生的答题反应计算标准误差和考生水平,其结果为与第一项实验结果完全相同。这两项实验证明,CAT 的题库与系统设计是可靠的。

一个具有高度可靠性的考试必须具备两种功能:一是指示受试者目前的能力水平,二是预测同一受试者在未来的一段时间内不同环境下的能力水平。CAT 所具备第一种功能已经得到验证,而第二种功能则需在 CAT 实施运行相当长的一段时间后才能对它加以评定。

评估 CAT 考试质量的另一重要标准是“有效性”(Validity),即由考试分数所得到的各种结论或推断是恰当的、有意义的和有效的。衡量 CAT 的有效性包含三个方面的内容:1. 与测试内容相关的有效性。2. 与测试总体构架相关的有效性。3. 与评估标准相关的有效性。鉴于 CAT 使用计算机作为测试的媒体,在与测试总体构架相关的有效性中还存在一个“模式效应”问题,即同样的试题在屏幕上阅读与在卷面上阅读是否会产生不同的效果。

鉴定 CAT 的测试内容是否有效,关键在于确认它的试题是否代表了所测试领域内的水平,如知识、认知能力、技能等。考生的考试成绩是其知识水平的表现,而非其他不相干的知识或能力。考试过程中,CAT 在向考生选送试题时,必须注意把握好试题的内容和水平两个方面的平衡。

所谓“总体构架”的有效性,系指某次测试的意图通过试题的总体设计得以实现。鉴定 CAT 的“总体构架有效性”,可将 CAT 的考试结果与 P&P 进行比较。一是比较 CAT 试题的变量与协变量矩阵与 P&P 的对等关系;二是比较两者在试题变量与协变量矩阵方面的差异。如果两者的矩阵相关性高,则表明 CAT 的总体构架在设计上与 P&P 的构架是相同的。换言之,CAT 的总体构架设计是有效的。

如果将 CAT 用于选拔目的,其预测的有效性就显得十分重要。也就是说,CAT 考试的结果能否达到预测分数高的考生未来成功的机会大于分数低的考生的效果。另一个问题是测试的标准是否有效?检验测试标准的有效性的方法是拿考生当年的考试成绩与现在的工作业绩作一线性的量化对比分析,以此考察当年的考试成绩是否准确地预测到今日的成就。其实,预测并非易事,它牵涉到一系列的问题,例如,考试为什么目的而设?考试成绩能够预测什么能力?学业成绩与工作业绩的评定标准是什么?如果出现业绩表现与当年的考试成绩所预测的结果相反怎么处理?总之,预测评定标准的效度远不及考试本身的效度高。以美国的 LSAT (Law School Admission Test) 考试为例,有人在数十所美国法学院作了一次调查,对学生的第一年的学业成绩与他们的 LAST 的考试成绩进行了预测效度评估,发现各学院间所测得的效度系数相差很大。

又鉴于 CAT 考试是在计算机上进行,其间牵涉到一系列其他问题,例如,考生必须逐题回答,不能在前面一道题没有回答的情况下进入下一道题;答案一经确认后,便不能再更改;考生答题不受时间的限制等。所有这些都是影响 CAT 效度的因素。由于 CAT 尚未大规模运作,所得的信息十分之少,要证明“模式效应”对于考试结果的影响尚需时日。

曾几何时,人们所熟悉的 P&P 考试及传统的“考试成绩=答对题数之和”的计分方法将被 CAT 取代。人们被告知,同一批参考人员所回答的问题将各不相同,每个人所能够回答的可能只有计算机选送试题总数的一半,答对试题的数量不会影响最终成绩的评定。于是,人们不由地要问:CAT 的成绩是怎样计算出来的?它与 P&P 考试成绩有可比性吗?CAT 专家们回答说:只要你的 P&P 成绩是依据 IRT(试题反应理论)计算的,那么你的 CAT 成绩与你的 P&P 成绩就是可比的。也就是说,只要是建立在同一测试理论基础上的考试,虽然形式不同,但所表现的信息曲线是相同的,因此它们之间就存在着直接的可比性。

美国“教育考试服务中心”(ETS)于 1998 年 5 月在 TOEFL 这个项目上进行了一次 CAT 与 P&P 的可比性实验。他们邀请曾于 1997 年 11 月和 1998 年 3 月参加过 TOEFL 的 P&P 考试的 6,500 名考生前往他们设在美国和加拿大的 CAT 考点,参加 TOEFL-CAT 考试。这些考生被告知,ETS 将为他们付出的考试时间提供经济补偿,如果他们的 CAT 成绩好于 P&P 成绩,将承认他们所取得的 CAT 成绩。之后,ETS 公布一张“P&P 和 CAT 成绩协调表”(TOEFL Concordance Table)。这张表由“总分”和“单项分”两个部分组成,每个部分又有“分数对照”和“分数段对照”两个小表。在这些表中,CAT 的成绩与 P&P 的成绩逐一对应。不同的是,CAT 所使用的数量级大大小于 P&P 的数量级。例如,CAT 的 270 分对应于 P&P 的 637 分;CAT 的 190~203 这个分数段相对于 P&P 的 520~537 这个分数段。这张“协调表”试图向人们证明,CAT 成绩与 P&P 成绩是可比的。

这一结论是建立在大量细微而复杂的技术运作的基础之上的。首先是确定 CAT 的评分系统。CAT 的评分系统包含两个概念:分数量(Scores)和分数值(Scales)。其中“分数量”又包含三个概念:试题水平分(Proficiency),题库分(Item Pool Score)和子题库分(Item Sub-pool Score)。考生答对的每一道题均含有这三个分数,答对题目总数的三个分数量之和,经转换成对应的分数值后,便是考生的 CAT 考试成绩。例如,某考生的 TOEFL 听力考了 600 分。这个 600 分应该叫做 600 分值(Scales)。换句话说,该考生的实际答对题数 44 道,这 44 道题的分数量总值为 600 分。分数量转换成分数值之后,分数才有了意义:它标识出考生的实际水平,并使这个水平与其他水平有了比较的机会。

在“TOEFL 成绩协调表”上,P&P 的分数值范围是 310~677,而 CAT 是 40~300,数量级比 P&P 小一倍多,而其底线值则更小。这个数值范围的确定考虑了四个方面的因素:①足够的宽度,避免出现分数值超出底线或溢出顶线;②相对紧凑,使每个分数段都能包括在内;③中心段的位置,使平均分数位于分数值范围的中部;④分值单位,能体现测试的精度。

其次,CAT 与 P&P 的可比性还取决于四个方面的均衡:①测试内容均衡,例如均为测试语法知识。②测试对象均衡,例如受试者水平相同,分数评定的条件分布相同。③受试人数的均衡,因为对比是建立在一对一的基础之上的。④对称均衡,既可以 $a=b$,也可以 $b=a$ 。

经过近 20 年的反复实验与改进,CAT 系统已达到了 P&P 考试的信度和效度,现在已到了“万事俱备,只欠东风”的时候了。随着 21 世纪的临近,CAT 在全世界范围内的广泛

应用，一大批新的测试成果将公诸于世。但 CAT 面临的挑战却远未接近尾声。

其一，在实际运作方面，CAT 还存在着一些有待进一步解决的问题。例如，虽然 CAT 对考生的答题时间未加限制，但 CAT 系统需以某种方式提醒考生注意整个考试所剩的时间以及考试临近结束时的再次提醒。又如，对于未答完的试题作何处理？是按答错题处理，还是忽略不计？再如，考生出现反常现象时——开始时的中等难度试题全答对了，而后面的试题（包括初级难度试题）绝大部分都答错了——是否按作弊处理（该考生有可能事先已经知道开始时的试题）？这是来自考试自身的挑战。

其二，从理论上讲，由同一试题库出题考试，考生之间的 CAT 成绩具有统计学上的可比性。然而在实践中，由于取样失当，模型选用不当，也可能出现对比上的失准。由此可能引发法律纠纷。这是来自社会的挑战。

其三，由于 CAT 的独特模式——随“机”应试，各得其所，水平高的考生一路向“上”，遥遥领先，而水平低的考生则会顺流而“下”，拉开他们与高分者之间的水平差距。低分者可能会因此抱怨 CAT 考试“不公平”，因为他们没有给予挑战难题的机会而无法全面展示他们的实际水平。CAT 考试实际上剥夺了低分者与高难度试题的相遇机会。这是来自考生的挑战。

其四，在选拔人才时，用人单位将在很长一短时间内对 CAT 成绩心存疑虑：这个成绩能够全面反映不同申请者在学识、才华、潜能与经验等方面的差异吗？一次 CAT 考试所使用的试题能否最大限度地包含并反映出受试者在知识、经验、才能诸方面的因素。这是对试题库的挑战。

其五，虽说 CAT 输送给考生的是最适合考生当前水平的试题，那么 CAT 系统又如何把握在一次考试中输送给考生的某个水平的试题数量正好能测试并显示出诸多背景各不相同的考生在学识、经验、才华与潜能方面的实际情况呢？这是对测试模型的挑战。

有挑战才会有突破，CAT 由襁褓走向成熟，每迈出一大步都是在向自身的不足与局限发出挑战而获得向前发展的更大机遇的。从 20 世纪 80 年代初设计的，用于美国陆、海、空三军选拔军事人才的“军队服务职业能力 CAT 考试”到如今，CAT 已大大扩展了自己的疆域。心理学家利用 CAT 研究个体认识能力的差异，社会学家利用 CAT 探讨个体的态度与社会的相互关系，政治家则利用 CAT 测试人民对政府的信心，等等。语言测试学家在这个时候推出 CAT 取代传统的 P&P；显然是适应社会发展潮流之举。可以预言，在 21 世纪这个舞台上，随着人类社会的不断进步，科学技术的迅猛发展，CAT 一定会演出一幕又一幕生动而精彩的“活报剧”来。

参考书目

1. *Computer Assisted Adaptive Testing*, Amirzafari, Jamileh (1987), Princeton, N.J.
2. *Computer Adaptive versus Pencil and Paper Test*, Bergstrom, Betty (1992), Ph.D. Thesis, University of Chicago.
3. *Let Computers Use the Past to Predict the Future: Using Machine-Based Retrospective Correlation Data for Prospective Aptitude Assessment*, Borchardt, F.L.&.E.B. Page (1994), 1994 Language Aptitude Invitational Symposium, Arlington, VA.
4. *The Promise and Threat of Computerized Adaptive Assessment of Reading Comprehension*, Canale, M. (1986), *Technology and Language Testing* 7, I.C.W. Stansfield, ed. Washington, D.C. TESOL. 29-45
5. *Item Equivalence from Paper-and-Pencil to Computer Adaptive Testing*, Chae, Sunhee (1995), Ph.D. Thesis, Chicago: University of Chicago, Dept. of Education.
6. *Using Computerized Tests to Measure New Dimensions of Abilities: An Exploratory Study*, Cory, C.H. and B.Rim Land (1977), *Applied Psychological Measurement* 1, 1.101-110.
7. *The Design and Development of 'Catalyst': a Computer Assisted System for Adaptive Testing*, Crick, Joe E (1979), UM1 #79-27,911. Ed. D. Thesis, Harvard Graduate School of Education
8. *The ACTFL Computerized Adaptive Test of Foreign Language Reading Proficiency*, Dandonoli, Patricia (1989), *Modern Technology in Foreign Language Education: Application and Projects*, edited by F. Smith. Lincolnwood IL: National Textbook Company.
9. *Computer Assisted Language Learning and Testing: Issues and Practice*, Dunkel, Patricia, ed. (1991), New York, NY: Newbury House
10. *Case Studies in Computer Adaptive Test Design through Simulation*, Eignor, Daniel R. (1993), Research Report; RR-93-56. Princeton, N.J.: Educational Testing Service
11. *Construct Validity of Computer-Based Tests*, Green, Bert F. (1998) May 6, 1998, Howard Wainer and Henry I. Braun, eds. Hillsdale, N.J. Lawrence Erlbaum
12. *Comparability and Validity of Computerized Adaptive Testing with the MMPI-2 Using a Clinical Sample*, Handel, Richard Walter (1997), KSU Masters Theses, Dept. of Psychology, Kent State University
13. *Validating an Item Bank in a Computer Assisted or Adaptive Test: Using Item Response Theory for the Process of Validating CATS*, Henning, G. (1991), *Computer Assisted Language Learning and Testing: Research Issues and Practice*, edited by P.Dunkel, New York, NY: Newbury House.
14. *Computer Adaptive Testing of Listening and Reading Comprehension: The Brigham Young University Approach*, Madsen, H.S. (1991), *Computer Assisted Language Learning and Testing: Research and Practice*, P. Dunkel. ed. New York, NY, Newbury House
15. *Computerized Adaptive Testing: A Primer*, Wainer, Howard (1990), Educational Testing Service, Hillsdale, N.J. L. Erlbaum Associates
16. *On Item Response Theory and Computerized Adaptive Testing*, 1993 *Journal of College Admissions* 28, 4, 9-16.

目录

CONTENTS

前沿报告 (代前言)

21 世纪语言测试的新视点——计算机适应性考试 (CAT)

第一部分 GMAT 机考与考前准备 1-20

第一章 GMAT 机考问答 1-3

第二章 GMAT-CAT 考前须知 4-20

第一节 CAT 测试项目、题量与时间分布 4

第二节 CAT 考试报名与填表注意事项 4

第三节 分数查询、成绩报告与重考 5

第四节 CAT 考试工具及使用说明 5

附表一 CAT 屏幕设计示意图 7

附表二 GMAT-CAT 考试报名表 18-19

附表三 额外成绩报告中请表 20

第二部分 GMAT-CAT “语文能力”考试辅导 21-211

第一章 GMAT-CAT “语文能力”考试 21-44

第一节 句子订正 21

第二节 阅读理解 24

第三节 阅读理解题型与答题思路 27

第四节 逻辑推理 31

第二章 GMAT-CAT “语文能力”练习与解答 45-130

第一节 “句子订正”100 题 45

第二节 “阅读理解”100 题 72

第三节 “逻辑推理”50 题 110

第三章 GMAT-CAT “语文能力”综合模拟考试 131-211

第一节 “语文能力”综合模拟考试之一 131

第二节 “语文能力”综合模拟考试之二 147

第三节 “语文能力”综合模拟考试之三 163

第四节 “语文能力”综合模拟考试之二 179

第五节 “语文能力”综合模拟考试之三 195

第三部分 美国商学院 MBA 项目申请 213-312

第一章 美国商学院 MBA 项目简介 213-222

概述 213

第一节 录取标准 213

第二节 录取程序 215

第三节	申请材料与注意事项	216
第四节	命题短文与回答技巧	219
第五节	面试与面试技巧	220
第六节	申请时机与注意事项	221
附表一	美国杜克大学 FUQUA 商学院入学申请表	223
附表二	美国哥伦比亚大学商学院推荐信格式	234
第二章	MBA 项目选择	240-246
第一节	MBA 项目选择的考虑因素	240
第二节	MBA 课程设置	241
第三节	MBA 的就读方式	245
第三章	MBA 项目奖学金申请	247-253
第一节	支出项目与费用	247
第二节	学资筹措与奖学金	248
第三节	奖学金申请的几个技术问题	250
第四节	美国奖学金机构与联系地址及有关网址	252
第四章	100 所美国商学院简介与联系地址	254-294
第一节	美国“五星级”商学院名录与简介	254
第二节	美国“四星级”商学院名录与简介	258
第三节	美国“三星级”商学院名录与简介	268
第四节	美国“二星级”商学院名录与简介	278
第五节	美国“一星级”商学院名录与简介	288
第五章	520 所美国和加拿大大学网址	295-312
第一节	478 所美国大学网址	295
第二节	42 所加拿大大学网址	310
附录	参考答案、错误点评、解答与注释	313-370
附录一	第二部分第二章 “语文能力” 练习参考答案	313
附录二	第二部分第二章 “句子订正” 100 题错误点评	314
附录三	第二部分第二章 “阅读理解” 100 题解答	332
附录四	第二部分第二章 “逻辑推理” 50 题注释	355
附录五	第二部分第三章 “语文能力” 综合模拟考试参考答案	370

第一部分 GMAT 机考与考前准备

GMAT 是 Graduate Management Admission Test 的缩写,是由美国“教育考试中心”(Education Testing Service, 简称 ETS)设计的、美国商学院 MBA (Master of Business Administration)的入学考试项目。

1997 年 10 月,一直以“纸笔”(Paper & Pencil,简称 P&P)形式考试的 GMAT 改为计算机考试。1998 年 4 月,中国境内的 GMAT 也开始了机考。一时间,考生对这一新的考试形式不甚了了,产生了许多疑问。本篇旨在为今后参加 GMAT 的考生答疑解惑,指点迷津,以便尽快适应新的考试形式。

第一章 GMAT-CAT 问答

问题之一: GMAT 报名表上的 CAT 是什么?

答: CAT 是 Computer Adaptive Test 的缩写,即计算机适应性考试,简称“机考”。

问题之二: 什么叫“适应性考试”?

答: 所谓“适应性考试”,是指由计算机根据考生的水平,选送最接近这一水平的试题供考生回答,并据此测定考生的现时水平。

问题之三: 考试刚开始时,计算机是如何知道考生的水平的?

答: 考试刚开始时,计算机选送给考生的是一些中等难度的试题。如果考生全部答对了,计算机便确定考生的现时水平达到了中等水平;如果考生答错了,计算机便知道考生的现时水平低于中等水平。

问题之四: 计算机是如何确定考生的最终水平的?

答: 在第一轮中等难度水平试题考完之后,计算机根据考生的答题情况,初步确定考生的现时水平,然后向考生选送高于或低于考生现时水平的第二轮试题。计算机在得到考生第二轮试题回答的数据后,重新确定考生的现时水平,并调整第三轮试题的难度水平,继续输送给考生回答。这样几经调整后,当考生的答题正确率稳定在某轮水平难度的试题上时,便是考生的最终水平。

问题之五: 计算机是如何确定试题的难度水平的?

答: 试题的难度水平并不是由计算机在考试过程中即时测定的。实际上,试题的难度水平是由专家测定的。考试前,测试专家在设计试题时就对每一道试题的难度水平进行了测定,然后做上标签,存入题库。计算机只是将标有某个水平的试题从题库中辨认并挑选出来而已。当然,题库建设的本身离不开计算机的大力协助。

问题之六：机考中会不会出现考了容易题目的人占了便宜，考了难题的人吃亏的现象？

答：不会出现这种现象。因为是适应性考试，容易的试题难度水平低，考到这部分题的人相对的水平也低。而做难题的考生，其最终测定的水平一定高于做容易题的人。换句话说，大家各得其所，谈不上吃亏或沾光。两名同时参加 CAT 考试的考生，即使两人答对的题目数完全一样，他们的成绩则可能完全不同。

问题之七：GMAT-CAT 的试题会不会比 GMAT-P&P 难？

答：不会。因 GMAT-CAT 和 GMAT-P&P 只是考试手段不同，考试的性质并没有改变。传统的 GMAT 试题印刷在纸上，而现在的 GMAT 试题是在计算机屏幕上进行答题。

问题之八：为什么考过 GMAT-CAT 的人都说难？

答：说这种话的人有两种情况：一种是既参加过 P&P 又参加过 CAT 考试的人；另一种是第一次参加 CAT 考试的人。第一种情况的人说难，是因为他们在 CAT 考试中没有做到在 P&P 考试中的某些难度较低的试题；而第二种情况的人说难，是因为他们在 CAT 考试中根本就没有机会接触到难度低的试题。说起来是一回事：CAT 的试题是最接近他们实际水平的试题，而这些试题都不是轻易可以答对的，心理上自然就觉得“难”了。可以说，只要参加 CAT 考试，就再也做不到低于自己现在水平难度的试题了。这是 CAT 考试的一个显著特点。

问题之九：不懂计算机会不会影响考试成绩？

答：不会。第一，计算机考试只是利用计算机的一些工具，如键盘、鼠标来完成考试答题，并不涉及计算机知识。第二，即使在考试过程中出现误操作，事先设计好的计算机程序会提醒考生，并指示考生如何返回。只要按屏幕上给出的指令，一步步地操作，就不会出现误操作。所以不懂计算机并不会影响考试成绩。

问题之十：在 CAT 考试中如果出现误操作，耽误了时间，会不会因答不完题而扣分？

答：考试中出现误操作确实会耽误一些时间，但与答题并无直接联系。因为 CAT 考试并没有像 P&P 考试那样，规定在一定的时间内答完一定数量的试题。CAT 考试只考虑达到某一最终水平所花费的答题时间，并不计算回答每一道题所花的时间。且时间的长短并不会对成绩产生决定性的作用。为了避免在考试中出现误操作，建议在考前熟悉 CAT 考试的工具及考试的形式，以及考试的指令。顺便说一下，考试中一旦出现误操作，不要慌张，更不要手忙脚乱。注意屏幕上的提示，按指令操作，很快就会返回到正常程序中去。

问题之十一：如果考试过程中计算机自身出现了故障怎么办？耽误的时间怎么算？

答：这种情况一般不会发生。万一计算机出了故障，如死机，不要自行处理，应立即报告监考人员或由技术保证人员负责恢复，或转到备用工作站继续考试。所耽误的时间由监考人员做出记录，并于考试结束后报告 ETS，不会影响成绩。

问题之十二：听说考 CAT 时心理压力大，考试成绩大大低于平时成绩，有这回事吗？

答：这个问题问得好！关于心理压力，一是凡是考试就会有，二是人人都会有，不过程度不

同。至于 CAT 考试所带来的心理压力往往来自于对这一新的考试形式了解不够或缺乏感性认识。只要考前多多练习,熟悉 CAT 考试程序,这种压力便会大大减轻的。至于考试成绩,确实有人的 CAT 成绩低于平时成绩。这里面有一部分是 CAT 考试的形式造成的,如屏幕阅读不习惯,注意力难以集中,眼睛容易疲劳,答题后无法返回去改变答案等。但只要考生心理状态稳定,发挥正常,以上这些不利因素就会得到控制,从而阻止成绩的大幅波动。

问题之十三: CAT 考试中能否返回前面的试题去查对,并改动答案?

答:可以返回到前面的试题,但却不能改动已经被确认过的答案。在 CAT 考试中,对于首次选择的答案有一次重新变动的机会。只要不掀动“确认”(Confirm)键,就可以改动原答案。一旦选择了其他答案,原答案便自动消除。专家建议,对于没有把握的答案暂不作确认。先选出一个答案,然后进入下一题,待完成其他试题后再返回并改变原答案。切记,凡已经确认过的答案都不能再改变!

问题之十四: 实行 CAT 考试后,作文考试是否一定要用键盘输入?

答:是的。ETS 要求 GMAT 的作文考试必须用键盘输入。CAT 写作时,计算机会提供一些简单的编辑功能,但不提供查找拼写错误的工具。习惯使用键盘写作并依赖计算机查拼写错误的人要特别注意这一点,拼写错误是要影响作文成绩的。

问题之十五: CAT 考试的评分标准是否与 P&P 考试不同?两者有无可比性?

答:鉴于 CAT 考试的评分依据是试题的难度系数,而非答对试题的总数,评分系统不同于 P&P 考试。成绩的表现形式也不同,为分段等级制。为了让考生和学校看懂这个成绩,成绩报告单上给出了各段级别与原 P&P 评分系统的对应系数。ETS 保证,这两者之间具有很高的可比性。

问题之十六: 考完 CAT 后能否立即得到考试成绩?

答:从理论上讲,除了作文成绩外,其他两个部分的考试成绩可以立即得到,因为计算机对考试的全过程进行了跟踪记录。从实际上讲,考生得到的只是由计算机打出的一份报告单,告知答题的总数,答对试题的数量,答错试题的数量,答题的总时间以及大概的水平等级等信息。而最终的成绩仍要在 ETS 进行计算,并在考试结束大约 3~4 周后才能得到正式的成绩报告单。

问题之十七: 如果第一次 CAT 考试的成绩不理想,能否紧接着马上再考一次?

答:按照 ETS 的规定,第二次 CAT 考试与第一次 CAT 考试之间必须间隔 4 周。这一规定是出于保护试题库的考虑,也是为了保证考试公平无欺。

问题之十八: 能否在 CAT 和 P&P 之间自选一种考试?

答:不能。美国自 1997 年 10 月起,全世界除少数不发达国家与地区外,自 1998 年 4 月起全部实行 GMAT-CAT 考试。考生无选择余地。