

美国国防采办测试和评估体系的发展脉络、 基本框架及特点

杨培宇, 刘戟锋

(国防科学技术大学 人文与社会科学学院, 长沙 410073)

摘 要: 美国通过建立制度化的国防采办测试和评估体系在军事技术与军队战斗力之间搭建一条坚实稳固的桥梁, 保证了其军队长期处于技术领先地位。通过文献分析, 梳理美国国防采办测试和评估体系发展脉络; 探寻了美国测试和评估的基本类型、组织结构和实施步骤; 总结提炼了其特点, 并为我国军事技术测试和评估体系的构建和完善提供了一定有益的借鉴。

关键词: 国防采办; 测试和评估; 技术评估

中图分类号: N09; F204 **文献标识码:** A **文章编号:** 1674-4969(2018)01-0068-13

美国是世界上公认的军事高技术领域领跑者, 也是率先建立起体制化、规范化技术评估体系的国家之一。自 20 世纪 50 年代以来, 在美国军方的主导下, 美国逐步建立并发展了独具特色的国防采办测试和评估 (Test and Evaluation, T&E) 体系, 帮助军队作战力量在军事技术与战斗力之间搭建一条坚实稳固的桥梁, 成为其确保“技术优势”的一道有力保障。2017 年 2 月, 美国国防部重新修订并发布了第 5000.02 号国防部指令——《国防采办体系的运行》, 对测试与评估做出了新的解释和调整, 并再一次强调了测试和评估在美国国防采办体系中的重要地位^[1]。

1 美国国防测试与评估的发展脉络

1.1 测试和评估的松散组织阶段 (1940 年前)

从 19 世纪末到 20 世纪初期, 美国社会处在一种自由竞争主义生长的发展阶段, 社会制度和

社会氛围相对宽松^[2]。这一时期, 美国国内的科学技术研究活动主要受到民间资本支持, 联邦政府则通过扮演“守夜人”的角色, 对科技研究领域进行适度的引导和有限的支持^[3]。在这样的背景下, 美国国内技术评判的标准主要是资本市场反馈, 能否取得经济效益成为了评判技术优劣的核心指标。

尽管联邦政府缺乏对科学技术研究的整体规划和宏观管理, 但是在一些为数不多由政府主导的研究项目中, 测试和评估的理念开始显露身形。1908 年美国军方同莱特兄弟展开的合作项目便被认为是测试和评估在军事技术发展史上的首次亮相^[4]。当时, 美国军方要求莱特兄弟研制一款军用飞机, 需要达到以下指标:

能够搭乘两名总重 350 磅 (约 158 千克) 的男子;
运送足够飞行 125 英里的燃料;

收稿日期: 2017-09-19; 修回日期: 2017-11-23

基金项目: 中国科协重大调研课题 (2016ZCYJ09)“战略高技术创新体制机制研究”

作者简介: 杨培宇 (1993-), 男, 硕士研究生, 主要研究方向为科学技术与社会、军事技术哲学。

刘戟锋 (1957-2018), 男, 教授、博士生导师, 少将军衔, 主要研究方向为军事技术哲学。

在无风状态下保持 40 英里的飞行时速。为确保所研制飞机具备上述能力,研制合同还特别要求莱特兄弟对飞机的性能进行测试。

完成合同规定测试后,飞机被运送至德克萨斯州,交由弗洛伊斯(Benjamin D. Foulois)上尉接收。为了确保飞机能够满足美国军方要求,弗洛伊斯上尉提出要在作战条件下对飞机进行性能测试。测试中,一些合同指标外的问题被弗洛伊斯上尉发现。例如,由于没有安装座位,导致飞机在着陆时经常发生人机分离的情况。作为修正,弗洛伊斯上尉提出为飞机安装驾驶座位,并通过武装带对飞行员进行固定^[5]。从那时起,初具架构的安全椅和安全带便成为军用飞机的一部分,而这一过程被视为是最早的作战测试和评估。

在这一时期,美国政府的科技管理工作自由而松散,行政力量很少介入科研活动,国家主体主要通过立法手段间接参与国内的科技活动。美国国内研究活动的主导权依旧掌握在民间力量手中,但是在第二次工业革命的推动下,越来越多涉及科学技术的法案被摆上国会议员的案头。与此同时,受到“进步主义(Progressivism)”思潮的影响,一些国会议员开始强调“独立立法机构获得知识和知情权的重要性”^[6]。于是在 1914 年,国会图书馆内部成立了研究服务部,为国会议员提供研究分析服务,其中就包括科学研究领域的分析和评估服务^[7]。

随着第一次世界大战的爆发,美国通过向交战双方兜售武器装备大发横财。与此同时,以飞机、坦克为代表的机械化装备成为战场新的主导力量,军事技术呈现出日趋复杂的发展态势,对武器装备性能参数的定量评估提出了新的需求。于是,一些美国研究机构和武器制造商开始建设专门从事武器装备测试和评估的试验场地,初步实现了武器装备测试和评估体系的规范化,但这一时期测试和评估的目的还是服务于装备研制过程,并未独立形成专业化程度较高的方法和技术体系。

在这段时期,美国国内初步形成了测试和评估的思想,但归根结底是一种经验性的总结,在美国国内技术评估还没有形成统一的规范和制度。

1.2 形成军事技术研究、开发、测试和评估协同管理体系(1940 年~1970 年)

第二次世界大战爆发后,军事技术在现代战场上释放的巨大能量,让美国政府不得不将建立国家科技管理体系纳入议事日程。1940 年 6 月,经时任美国总统罗斯福批准,万尼瓦尔·布什牵头组建了国防研究委员会(National Defense Research Committee, NDRC),负责协调并监督军用技术的研究、开发和生产环节。一年之后,国防研究委员会的行政职能被收回,罗斯福政府成立了战时统管全美研究活动的行政管理机构——科学研究发展办公室。在其主导下,美国完成了包括“曼哈顿计划”、雷达预警系统、精确弹药等在内的一批战略高技术研究项目。

在战略高技术不断取得突破性进展的同时,测试和评估的思想也逐渐被美国军方所重视,其开始斥资建立大型综合实验基地。如 1943 年组建的尤马测试场、白沙导弹试验场以及 1945 年成立的中国湖武器试验中心^[8]。1945 年 7 月,为了应对中途岛战役中日本神风特攻队所发动的自杀式袭击,美国海军组建了海军大西洋舰队复合特遣队,旨在依托海军装备体系对战术战法进行开发和评估。随后,其职能被拓展到技术测试和评估领域^[9],成为美国军方首个行政性质的专业测试评估机构。

二战结束后,世界局势动荡复杂,罗斯福政府战时建立的国防体系已经难以适应新的时代要求。1947 年,时任美国总统杜鲁门签署《国家安全法案》,在原有战争部和海军部的基础上组建了美国国防部。为了获取技术和决策分析支持,首任国防部长福里斯特尔(James Forrestal)于 1948 年组建了武器系统评估小组(Weapons Systems Evaluation Group, WSEG),旨在采用先进的科学分析和研究技术,从公正、客观的角度为国防部

和参谋长联席会议提供武器系统评估报告^[10], 这也一度被视作“国家最具挑战性的安全问题”^[11]。然而由于人手不足, 在实际运作中, 该小组难以满足美国军方旺盛的评估需求。于是在 1955 年, 麻省理工学院在国防部和参谋长联席会议的支持下牵头组建了国防分析研究所 (Institute for Defense Analyses, IDA), 以协助武器系统评估小组完成国防采办中的测试和评估。

1957 年, 苏联成功发射人造卫星和洲际弹道导弹的消息给美国朝野带来巨大的震动。而此时, 美国和苏联之间的意识形态矛盾不断升温, 世界格局演变为以美苏为中心的两极对峙局面。在这样的历史背景下, 应用于军事领域的战略高技术被视为能够帮助美国夺取世界霸权的重要工具手段。全美范围内形成了扩大军事开支的共识, 一度占据联邦研发经费总量的半数之多。

为了进一步控制技术成本、提高技术产出、树立技术领先地位, 应对来自苏联的“技术挑战”, 1958 年, 时任美国总统艾森豪威尔接受胡佛委员会 (The first Hoover Commission) 的建议, 开展了新一轮的国防部重组, 重点对国防部技术领域职能机构进行了整合调整。基于新签署的《国防部重组法案》, 国防部设置了级别仅次于各军种部长的国防研发和工程主任 (Director of Defense Research and Engineering, DDR&E) 一职^[12], 负责协助国防部长处理军事技术研发与工程领域事务, 同时监督高级项目研究局 (Advanced Research Projects Agency, ARPA) 以及导弹研究办公室等机构的项目研制^[13], 具体职责涉及基础研究、应用研究、测试和评估以及设计与工程四个方面^[14]。此外, 根据艾森豪威尔的设想, 国防研发和工程主任还应被视为全美“科学技术领域的风向标”人物^[15]。

围绕着国防研发和工程主任在美国军事技术领域所处的核心地位, 国防部对其内部机构的业务职能范畴进行了进一步整合, 将测试和评估提升到了与研究 and 开发 (Research and Development,

R&D) 同等的地位, 纳入国防研发和工程主任管理职能之内, 形成了国家层面的军事技术研究、开发、测试和评估 (Research Development Test and Evaluation, RDT&E) 协同管理体系。在这一管理体系框架下, 国防研发和工程主任是总揽全局的政策制定者和管理实施者, 接受来自国防科学委员会 (Defense Science Board, DSB) 的咨询建议, 并由 6 名副主任分工协助管理业务工作。其中评估和管理副主任具体负责测试和评估工作的组织实施。此外, 在新的协调管理体系内, 武器系统评估小组的人员规模、职能范围也得到进一步扩大, 这支由各军种人员、国防分析研究所研究人员以及文职人员所组成的评估队伍, 需要对当前和未来的武器装备系统及其对军事战略、编制体制、作战模式的影响进行独立、全面、客观的分析和评估, 并对美军武器装备系统的相对效能和成本投入进行评估^[16]。

重组过程中, 美国在军事技术管理理论上也取得了重大突破。在新的管理体系框架下, 美国军事技术研发项目还被细分为基础研究、探索开发、样机试制、工程开发、作战系统开发、管理支持保障等 6 个阶段。形成了美国国防采办生命周期管理体系的雏形。

总的来说, 这一阶段美国加强了对技术评估的重视, 初步形成了一套具有可操作性的技术评估规范, 但是这一阶段所开展的技术评估还是针对武器系统性能的测试, 而且还未形成独立的技术评估管理机构, 技术评估与技术开发之间的行政管理界限还较为模糊。

1.3 形成独立的测试和评估体系 (1970 年至今)

20 世纪 60 年代中期, 美国在同苏联的核竞赛和太空竞赛中逐渐占据优势。越南战争的爆发, 使美国国会更多的注意力集中在了胶着的越南战场上。因此, 国会并没有对当时国防采办体系以及测试和评估活动投入过多的关注。

事实上, 在前一阶段针对武器装备系统的测试和评估中, 所分析的问题大多数集中在武器装备系统的实际需求和武器装备系统投入现役所需的成本上, 而关于武器装备系统能否达到预定指标和是否具备生产条件等重要问题却鲜有人关注^[17]。之后, 美国国防科学委员会 (Defense Science Board) 在一份日后的回顾性报告也指出当时测试和评估所存在的一个重大缺陷^[18]: 测试和评估机构在组织测试和评估中缺乏与开发人员的交互反馈, 导致测试和评估结果无法及时纠正开发过程存在的潜在技术风险。同时, 由于测试和评估机构与技术开发管理机构、采办机构隶属于同一套管理系统, 也导致测试和评估过程中时常会受到一些主观因素的影响。

60 年代末, 越南战事急转直下, 美国经济进入滞胀阶段, 加之尼克松 (Richard M. Nixon) 就任美国总统后对美国政治、经济和军事政策的大规模调整, 国会加大了对国防预算开支的审查力度。在这样的背景下, 原有 RDT&E 协同管理体系所存在的体制机制问题逐渐显露出来。直到 1970 年, 美国总统国防事务蓝带小组向尼克松提交报告, 直截了当地指出了当时测试和评估体系存在的重大问题^[19]:

“不幸的是, 现行的测试和评估体系几乎不可能获得能够直接适用于决策或有益于分析的结论。一般来说, 这些测试数据没有任何意义, 因为它们通常来自于存在设计缺陷或在未充分控制的条件下进行的测试。”

报告还建议, 为了获得尽可能客观全面的测试评估数据, 需要成立独立于开发人员和项目赞助方的测试和评估机构。

这一报告引起了国会和尼克松的高度重视。依据报告内容, 在国防部副部长帕卡德 (David Packard) 的主持下, 国防部紧急采取了一系列政策措施, 对测试和评估体系进行了系统的重组。期间, 帕卡德亲自向相关业务管理部门发送了一

系列备忘录, 要求实现测试和评估机构的独立运营, 并且强调该机构需具备直接向军种部长和最高军职长官负责的职能。这一期间, 他还牵头成立了国防系统采办审查委员会 (Defense Systems Acquisition Review Council, DSARC), 负责统管国防部测试和评估工作, 并向总统和国防部长提供关于国防采办事务的建议。

随后, 国防部在 1971 年和 1973 年分别发布了第 5000.1 号国防部指令《重大系统采办 (Major Systems Acquisition)》和第 5000.3 号国防部指令《测试和评估 (Test and Evaluation)》, 以正式指令的形式为测试和评估提供了具体的政策指导, 并首次从国防部层面明确了测试和评估的定义和分类^[20], 即开发测试和评估、和作战测试和评估。根据指令要求, 各军种也成立或改建了相应的独立测试评估机构, 包括陆军作战测试和评估局 (Operational Test and Evaluation Agency, OTEA)、空军作战测试和评估中心 (Air Force Operational Test and Evaluation Center, AFOTEC) 以及海军由大西洋舰队复合特遣队发展而来的作战测试和评估部队 (Operational Test and Evaluation Force, OPTEVFOR) 等, 负责军种采办过程中的作战测试和评估活动的组织管理^[21]。

1972 年, 尼克松签署《技术评估法案》, 基于这一法案, 国会成立了技术评估办公室 (Office of Technology Assessment, OTA), 为国会所面临科学技术问题提供客观和权威的分析援助, 作为国会审议研究领域预算、法案的重要依据^[22]。

到了 1983 年, 根据总统国防事务蓝带小组提议, 国会直接在国防部办公室 (OSD) 内设立了作战测试和评估主任办公室, 取代国防系统采办审查委员会成为新的测试和评估工作的主管机构。作战测试和评估主任直接向国会负责, 这也标志着国防采办中的测试和评估进入到了美国国家宏观战略决策视野。1988 年, 为了进一步规范测试和评估的组织程序, 国防系统学院发布了《测

试评估管理手册(Test And Evaluation Management Guide)》,对测试和评估中的细节问题进行了进一步的明确。

为了解决测试和评估中暴露出来的专业人才缺乏问题,1991年,国防部发布第5000.57号国防部指令,在贝尔沃堡组建国防采办学院(Defense Acquisition University, DAU),对测试和评估展开系统研究,并负责测试和评估人员的培养和《测试评估管理手册》的修订。

至此,美国在国防采办过程中初步形成了独立于研发管理的测试和评估体系,相较于原先的协同管理体系,新的测试和评估体系机构设置更加合理,职权边界更为清晰,方法过程更加科学。

2 美国国防采办测试与评估体系的基本框架

测试与评估在美国国防采办体系和计划中的根本目的是:找出需要减少或消除的技术风险的部分。在开发的早期阶段进行测试与评估来表明概念化方法的可行性、将设计风险最小化、确定设计方案、比较和分析权衡,以及评估作战效能和作战适用性。随着系统的设计和开发,测试的重点逐渐从发展测试与评估(其主要涉及工程设计目标的实现)向作战测试与评估(其重点是作战效能、作战适用性和可支持性的问题)转移。

测试与评估做出的主要贡献是识别和报告可能对系统的性能、可用性或可支持性产生不利影响的缺陷。

2.1 美国国防采办体系中的里程碑设置

国防采办是通过投资为军队获取新型或改进的武器装备、信息系统和服务能力的活动^[23],是保障国家安全战略的部署和武装力量的发展的重要手段。自20世纪70年代发布“5000”系列国防采办指令以来,在国防部的行政框架下,经过不断的修订调整,美国已经形成了一套涉及需求确定、技术研发、试制生产、部署保障诸多环节的渐进式国防采办体系^[24]。测试和评估则是国防采办体系下一类重要的组织方法和手段,依托于整个国防采办体系展开。

在现行的国防采办体系中,美国通过划分阶段、设置里程碑将采办活动覆盖到了装备建设的全系统全寿命周期,并根据项目方向、要求不同将国防采办细分为硬件集中型(Hardware Intensive)、软件集中型(Software Intensive)等6型采办项目类别。尽管各型采办类别在细节上有所不同,但是在总体上采办过程是趋于一致的,即将国防采办分为5个阶段,通过里程碑、决策点的设置实现对国防采办过程的总体管控,其采办模型如图1所示。

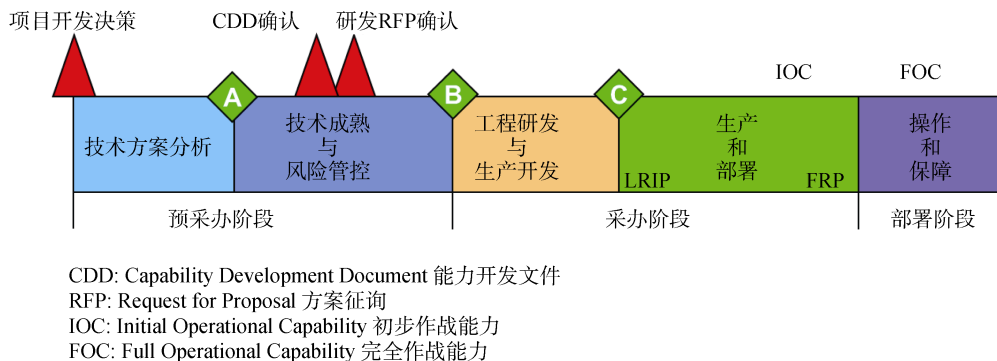


图1 国防采办体系里程碑设置

在整个采办过程中,里程碑的设置是国防部对采办项目组织进度控制的重要手段。所有采办

项目必须通过里程碑的评审决策,才可以进入下一个采办阶段,不同阶段有专门的里程碑决策机

构(Milestone Decision Agency ,MDA)负责评估^[25]。从一个阶段到下一个阶段的进展,不是按照计划的日程表,而是通过实际解决关键的问题和实现测试和评估验证的预设阈值。对于一个新的采办项目而言,里程碑决策机构有权决定其进入采办过程的任一节点,前提是满足该节点前里程碑评判“进入准则”,这样一种设置大大降低了一些技术相对成熟项目进入最终采办环节的时间成本。在里程碑 A,决策机构需要判定技术成熟度,并由测试和评估机构初步制定测试和评估计划和策略;在里程碑 B,决策机构则会评判技术是否可用于工程研发,同时测试和评估机构会发布一份基于顶层需求设计的指导性文件——《测试和评估总体规划(Test and Evaluation Master Plan , TEMP)》,以对项目的关键技术参数(Critical Technical Parameters , CTP)、性能特征(performance characteristics , MOE)和关键操作问题(Critical Operational Issues , COI)以及测试和评估的总体目标、责任和时间表做出解释和规定;而在里程碑 C,决策机构则会判断项目是否可以进入生产和部署阶段,可以批准项目进入小规模生产(LRIP)或全速率生产阶段(FRP),测试和评估机构则会根据项目采办进展进一步更新《测试和评估总体规划》。

在里程碑的判定过程中,测试和评估的结果发挥着举足轻重的作用。事实上,里程碑决策机构对于测试和评估结果的期望并非是有利于采办项目的,因为他们更希望测试和评估在采办过程中所做出的贡献是能够及早发现并判定项目潜在的技术缺陷和不足。一方面,对于整个采办过程而言,一份系统的、客观的测试和评估报告能够为改进项目缺陷、提升性能提供具有借鉴意义的参考和警示,如果没有测试和评估结果作为决策依据的坚实基础,决策机构将无法做出合理公正判断;另一方面,在项目采办的后期对系统缺陷进行修正,会使得项目修正成本从 10%增加到

30%,如果严格计划并执行测试和评估,可以极大地减少项目采办过程中的各方面成本。图 2 展示了系统的生命周期成本(Life Circle Cost ,LCC)以及决策对项目支出的影响。

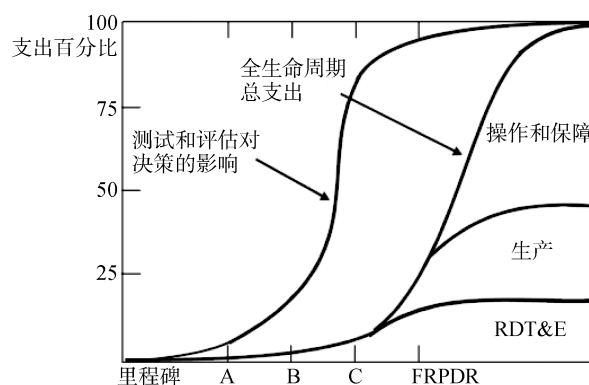


图 2 测试和评估决策对成本的影响

2.2 美国国防采办测试与评估的类型

测试和评估的根本目的是为决策者提供必要的信息,确认并验证系统的设计指标,评估技术性能参数的程度,确定系统是否有利于提升军队战斗力。在开发的早期阶段,为了证明概念方法的可行性、评估设计风险、验证设计方案,比较和分析权衡以及判断系统是否满足作战需求,测试和评估的主要形式是开发测试和评估。随着系统的不断开发,测试和评估的中心逐渐转变为系统对于现实作战需求的满足程度,也就是系统作战效能和作战适用性,这一阶段的测试和评估主要采取作战测试和评估。虽然从时序上看,开发测试和评估和作战测试和评估活动的组织是相互衔接的,但是二者在武器系统的发展过程中并不是严格的连续组织环节。在一些特定的环境下,综合组织开发测试和评估与作战测试和评估有助于节省人力物力成本。

2.2.1 开发测试与评估

国防项目采办是一个涉及到设计、建模、测试、识别缺陷、修复、重新测试的迭代过程,理想状态下,随着采办进程的不断开展,待采办项目系统的各项指标性能将无限趋近于预期设计指

标。在这过程中, 开发测试和评估 (Development Test and Evaluation, DT&E) 发挥了重要作用。

开发测试和评估用于支持项目系统的开发和大规模生产, 是使系统逐渐实现技术成熟的测试环节, 其目的方面是验证设计的合理性, 并使之风险最小化, 另一方面则是督促设计生产机构完成预定设计指标, 并为后续作战测试和评估提供一定的数据基础。开发测试和评估通常需要大量的仪器和测量工具, 由模拟环境中的工程师、技术人员或军方操作、维护、测试完成。

在实际组织开展中, 开发和评估的测试活动包括了任何适用于验证技术进展情况的工程类型测试, 相关人员会根据项目进展组织系统组件测试、硬件/软件集成测试和生产资格测试。其主要内容包括寿命周期测试、设计极限测试以及可靠性、可用性和可维护性测试。其中寿命周期测试主要评估系统长期暴露于预期环境下的性能可靠性, 这些测试用于确保系统不会由于金属疲劳、部件老化或长期暴露于环境而导致系统失效; 设计限制测试 (DLT) 是将系统暴露于极限环境 (如低温、扬尘) 下的测试, 以确保系统在身处极限环境时依旧能够提供可靠的性能。

开发测试和评估保障了工程设计和开发过程的顺利开展, 一般由政府管理机构进行规划和监督, 由承包商负责具体组织。具体来说, 承包商通过对项目开展开发测试和评估来降低潜在技术风险, 验证设计指标, 并确保产品能够被政府和军方所接收。此外, 开发测试和评估的结果也用于在系统投入使用时对系统的军事效用进行预估。政府管理机构也可以在作战测试和评估之前进行技术合规性测试, 通过开发测试评估, 可以确认系统是否达成预期技术指标, 并且已准备好进行下一步现场测试。

2.2.2 作战测试和评估

作战测试和评估是在所采办系统完成初步设计、产出样品时组织的测试和评估活动, 其关注

的重点是所采办系统是否满足未来美国军队的作战需求, 是直接和一线作战能力对接的测试和评估手段, 也被视为美国国防部采办过程中的“期末考试”。因此, 作战测试与评估在美国国内获得了高度重视, 其组织运行的具体要求得到了美国国会的直接授权并纳入了《美国法典》。《美国法典》对作战测试和评估做出如下定义:

在现实的作战条件下, 对武器、设备或弹药或关键组成部分进行现场试验, 以确定其典型军事用途, 并测试其作战效能和作战适用性; 并对测试的结果进行评估。

此外, 《美国法典》还指出: 该术语不包括仅基于计算机建模、仿真的评估活动, 也不包括仅基于设计规范中所包含信息进行的分析评估。这就意味着, 作战测试和评估并不是闭门造车, 需要依托于实际作战模拟环境进行。

相较于开发测试和评估, 美国国防部格外强调作战测试和评估的独立性。要求测试和评估机构与武器装备系统研发单位、采购机构以及武器装备具体用户相独立。之所以如此强调, 是为了回避在测试和评估活动中, 测试机构与其他机构之间可能存在的利益往来关系。正是因为是独立的, 无论最终得出了什么样的作战测试和评估结果, 测试机构的利益不会因此受到影响, 也就确保了作战测试和评估的结果是客观可信的。

作战测试和评估主要关注两个方面的问题: 一个是作战效能, 另一个是作战适用性。其中作战效能是指系统在预设的作战情况下, 在仿真的作战环境中, 面对假想的能够执行适当的应对策略的威胁力量时有效执行预期功能的能力。作战适用性是指在特定的机组人员以预设的数量和预期的经验水平的条件下, 对系统进行运行和维护时, 系统能够保持可靠性、可维护、可备战以及在部署完成、可兼容、可互操作和保证安全的情况下具备后勤保障的能力。

以“确定声纳在自然声学环境中检测、分类

和跟踪的能力”作为评估对象为例，作战效能和作战适用性评估需考虑以下要素：

表 1 作战效能和作战适用性评估要素

作战效能评估要素	作战适用性评估要素
检测概率	可靠性
检测范围	可维护性
检测时正确分类的概率	可用性
将威胁分类为非威胁的概率	后勤保障难度
检测和分类之间的时间	兼容性
分类范围	互操作性
建立轨道的可能性	培训
检测和轨迹建立之间的时间	安全
轨道建立的范围	文档
	运输能力
	战时使用率
	可操作性
	软件支持

2.3 美国国防采办测试与评估的组织结构

在美国国防采办体系框架内，测试和评估独立形成了一套功能完善、职责分明的组织结构。如图 3 所示，美国国防采办测试和评估体系分为上下两个层级。其中上层是依托于国防部内部职能框架的国防部测试和评估层级，它依据国会核准的预算和总统下达的指示，负责全军测试和评估工作，并负责制定相关政策措施；下层则是各

军种间相互独立的军兵种测试和评估体系，职能范畴局限于军兵种采办项目之内。

在国防部测试和评估层级内，国防部长是美国国防采办体系的最高决策者，接收所属国防部官员、现役军官的咨询建议。其中，隶属于负责采办、技术与后勤国防部副部长(Deputy Secretary of Defense, Under Secretary of Defense for Acquisition, Technology and Logistics, USD(AT&L))的负责开发测试和评估的助理国防部副部长(the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation, DASD(DT&E))是具体负责开发测试和评估的国防部高级官员，其职权包括：审查批准《测试和评估总体规划》；监管国防采办中的开发测试和评估事务；为副部长提供相关决策支持等^[26]。作战测试与评估主任是国防部内主管作战测试评估的高级文职人员，也是美国国防部长关于作战测试与评估的首席顾问，行政级别与助理国防部长相当。他负责美国国防部内测试和评估的政策制定、评估和监督，其独立于其他国防部官员，直接向国防部长和国会报告，而掌管国防信息系统局的国防部首席信息官则负责汇总相关测试和评估信息，向总统、国防部长汇报^[27]。

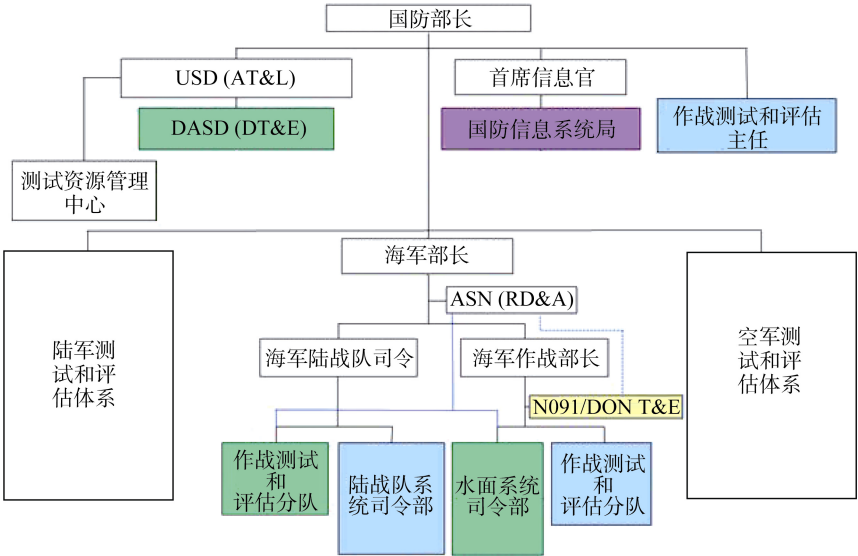


图 3 测试和评估体系组织架构

军种层级之间, 也有相对独立的测试和评估管理机构 and 专业的测试和评估力量。以海军为例, 海军部长是海军内部测试和评估事务最高决策者, 分管研发和采办的助理海军部长 (Assistant Secretary of the Navy for Research, Development, and Acquisition, ASN (RD&A)) 负责管理海军内部测试评估事务, 测试、评估和技术要求主任 (N091) 则具体负责对测试和评估的协调指挥。同时, 海军两个兵种设有独立的测试和评估部队, 负责对海军所开展的采办项目组织测试和评估。

2.4 美国国防采办测试与评估的实施步骤

在整个国防采办体系当中, 测试和评估对于

整个项目而言, 既是识别技术风险的一种工程手段, 也是用于验证项目性能的一种具体方法。一般情况下, 一次独立的测试和评估分为五个具体步骤来实施, 这样可以为决策制定者提供涵盖项目全生命周期、涉及诸多关键设计和决策问题的参考依据。值得注意的是, 测试和评估不应当独立于采办项目之外, 在项目初具雏形时就需要开始着手测试和评估的计划制订, 并且与决策者的信息需求、程序规定、采办策略等其他需要分析的因素相结合, 以确保测试和评估结果的时效性和可靠性。测试和评估的具体步骤如图 4 所示。

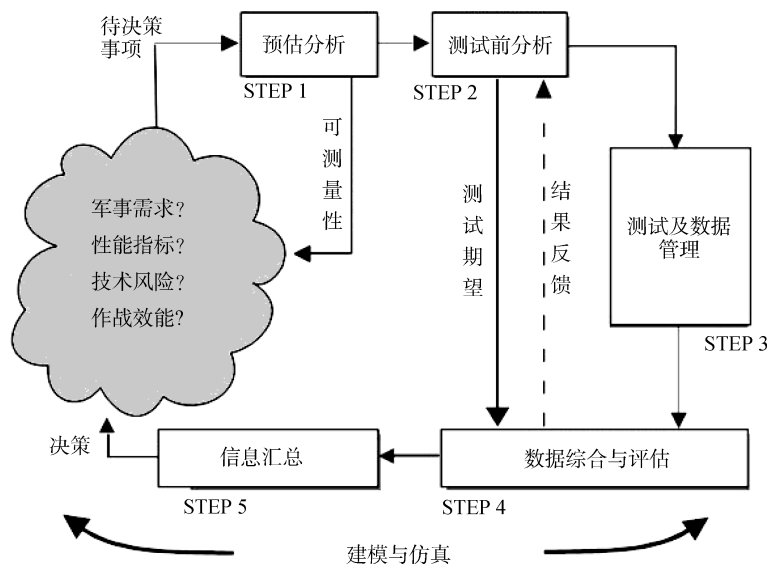


图 4 测试和评估步骤

步骤 1 是对测试和评估的基本数据和信息进行预估分析。在国防采办过程中, 决策者会对具体性能、指标的要素提出测试和评估需求, 需要测试和评估人员对这些要素的可测量性进行充分评估, 对测试和评估过程中可能出现的关键评估问题以及测试数据所需的测试条件 and 环境作出反馈。国防采办一般从项目初期就开始对测试和评估进行粗略规划, 这有助于进一步完善项目基础性文件, 从而为决策者提供足够的信息支撑, 确保项目采办过程的稳步推进。第一步中所提出的需求通常集中在相对基础的领域, 具体可能包括

一些基础的技术概念、项目原型种类或者系统的生产形式; 所明确的信息则包括对项目潜在作战效能和作战适用性的初步评估, 以及项目在未来满足决策者不同需求的时间节点。

步骤 2 是基于步骤 1 中对评估目标的预测试分析, 被称为测试前分析, 以确定测试和评估所需的数据具体类型和数量、开展测试和评估所需的具体分析工具并对测试结果做出预测。在测试前分析中通常会使用经过验证的模型和仿真系统, 有助于为具体项目确定设计测试场景、设置测试环境, 并进一步讨论如何开展测试和评估,

如何控制调配测试和评估资源, 如何使排序测试试验顺序最优化以及如何评估测试结果。

步骤 3 是展开测试活动和组织数据管理, 是对采办项目涉及系统开展实际测试和评估的步骤。在这一步骤中, 测试和评估人员会根据步骤 2 中所提出的数据需求, 依照预定方法对项目系统组织测试和评估, 并对所取得的数据进行初步整理。测试和评估经理则会判断已有数据中哪些是有效数据, 哪些还需要进一步组织测试以获得新的数据。这一步骤需要计划并执行大量必要的测试环节, 以获取并收集足够多的数据来支持决策者做出分析。在进入步骤 4 之前, 测试评估人员需要对数据进行筛选以确保其完整性、准确性和有效性。

步骤 4 是数据综合与评估, 在这一步骤中, 测试评估人员通过一定的技术分析手段将步骤 3 中获得的实际测试数据与步骤 2 中的预期数据进行对比分析, 并将分析结果整合为有价值信息供决策者参考。当测试数据与预期结果不同时, 测

试评估人员需要重新检查测试条件和测试的组织程序, 以排除数据误差中可能存在的人为因素, 如计算机模拟失真、测试条件未满足、仪器误差或操作过程失误。因此就要求在步骤 2 组织测试之前, 必须仔细选择测试方法、充分描述和记录测试、操作环境、系统性能参数和相关预想的假定。在数据分析期间通常可以使用仿真和建模 (Modeling and Simulation, M&S) 手段来进一步扩展作战效能和作战适用性的评估。

步骤 5 则是决策者将测试和评估信息与其他信息相汇总时, 最终决定适当的行动方式。此过程可能提出新的测试和评估需求, 并再次重复测试和评估的过程。

以上分析的是一个独立的测试和评估活动开展步骤, 而事实上, 测试和评估是内嵌于整个采办体系的, 因此在整个采办体系之中, 不同的测试和评估类型也有相对固定的组织顺序, 具体如图 5 所示。

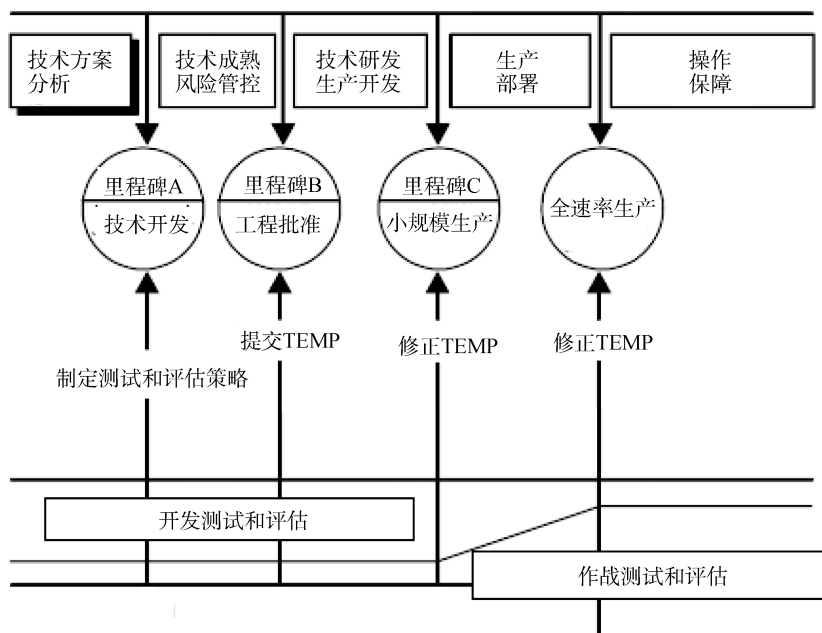


图 5 采办进程中测试和评估的组织实施

如图 5 所示, 在前三个阶段主要开展开发测试和评估活动, 用来协助工程设计和项目的开发过程顺利实施, 并验证技术性能指标和目标的达

成情况。而在生产和部署以及操作和保障阶段则主要进行作战测试和评估活动, 以确定系统的作战效能和作战适用性, 识别系统缺陷, 找出为了

满足既定的作战指标而需要进行的改进之处,并为被测系统制定适宜的战术。

可以看出,测试和评估活动贯穿于整个国防采办进程。在军事技术由知识形态逐渐物化成为武器装备的过程中,不同类型的测试和评估手段按照预定计划由简到繁、按部就班地组织开展,既着眼于装备体系的总体规划,又扎根于采办过程各个不同阶段,在全盘统筹的同时,严谨细致地考察技术指标、装备效能等要素,实现了开发测试和评估同作战测试和评估之间的有序衔接。

3 美国国防采办测试和评估体系的特点

3.1 集中管理分散组织

美国国防部统管国防采办测试和评估工作工作。测试和评估的总体规划、政策制定、预算编订均是由国防部牵头组织,负责对测试和评估的建设和资源的利用进行统一组织管理,并形成了统一规范的测试和评估体制机制^[28]。国防部具体负责审查和管理军事技术采办过程中的测试和评估进程;指导并管理各个军种的测试和评估活动。在国防部体系框架下,美国各军种都构建了完备的测试和评估管理体系,设立了负责测试和评估事项的专职人员,并在每个项目中组建测试和评估协调小组,负责具体项目测试和评估活动的协调沟通。

3.2 法规制度体系完善

法规制度是装备测试和评估活动的参考依据,因此在测试和评估体系的建设过程中,美国十分重视相关法规制度体系的建设,形成了层次分明、门类齐全的测试和评估法规体系。除国会批准的法案和国防部颁布的适用于全军的命令外,各军兵种还根据各自的特点,制定和颁布适用于自身的测试和评估的法令条例和规章制度,如美国空军发布的《AFPD99-1》《AFI99-103》等一系列测试和评估政策和相关指令,建立了主要由空军部主导、各兵种司令部参与、空军测试和

评估部队为主力的测试和评估系统。

3.3 人才队伍健全

测试和评估是一个知识密集型的研究方向,需要依托大量高素质科技人才支撑推动其发展。美国军方高度重视测试和评估人才队伍的建设工作,并成立了国防采办学院以培养测试和评估专业人员。此外,美国国防部定期组织测试和评估人员培训,如为期四天的作战测试主任(Operational Test Director, OTD)课程,课程内容包括一些作战测试与评估的最新情报,例如分析、测试方法和最新技术进展。为了帮助测试和评估人员开展好作战测试与评估工作,美国还定期出版《测试和评估指导手册》,作为开展测试和评估活动的指导性手册。

4 对策和建议

科学、全面地开展军事技术评估,关系到国家战略能否顺利施行,是与国家战略目标的实现和国家战略利益拓展直接相关的。经过几十年的艰苦奋斗,在技术领域上,我国即将实现从跟跑并跑到并跑领跑的转变。但是,随着我国军事技术发展体量的增大,其发展风险亦随之增加,这也对我国的军事技术评估体系建设提出了更高的要求,因此极有必要进一步完善军事技术评估体系。

结合我国军事技术发展实际和技术评估特点,提出以下五条对策和建议。

4.1 制定完善技术评估指导规范

虽然目前我国行政机构十分重视技术评估在军事技术转化应用中的作用,但是目前并没有形成针对军事技术评估的指导性规范,对于军事技术的评估活动大多参照一般技术评估规范执行;一般的技术评估规范中,存在“技术评估”和“技术评价”两种概念,概念之间界限模糊,容易产生误导;国家标准《科学技术研究项目评价通则》中主要考察技术水平,评估标准单一,对军事技术研发所面临的风险因素以及所带来的政治、经

济、文化和军事影响未做考量;现存法规也存在一定程度的兼容性问题。科学而合理的指导规范是建立军事技术评估体系的基础。需进一步根据军事技术的特点完善其评估的指导规范,为我国的军事技术创新活动提供及时的方向引导和航向修正。

4.2 成立独立技术评估机构

虽然目前正在建设以市场为导向的技术评估机制,但是在评估机构设置上还是缺乏独立性。技术评估机构应当是独立于研发单位、采购机构和具体用户的,其职能是紧贴国家战略需求,为技术用户提供专业技术评估服务。国家军事技术评估机构可以接受财政预算资金,但需要与研发单位、采购机构以及具体用户保持独立平行。同时,该机构作为独立法人,应实行市场化运营和管理,严格执行“去行政化”原则。在充分享受自治权的同时,该机构的运营目标必须始终与国家战略相一致;国家科技管理部门在不干扰正常研究秩序的前提下,应对军事技术评估机构所开展的评估活动进行监督和指导,为机构的长期发展规划提出意见和建议。

4.3 建立全生命周期的军事技术评估体系

军事技术的研制过程周期长、风险高、投入大,通过事前、事中、事后的节点式评估手段难以满足军事技术评估的现实要求。需要建立全生命周期的评估体系,组织专业评估人员全程参与到军事技术的基础研究、应用研究和开发研究环节,及时发现潜在风险,通过技术手段加以修正;需要针对不同的研究阶段细化军事技术评估的类别,如前文所述,美国军方就根据技术发展的不同阶段,将军用技术的评估活动分为:开发测试评估、生产验收测试评估以及作战测试评估,以此提高军事技术评估活动的针对性,有益于提高研究和成果转化效率。

4.4 建立军事技术预警机构

目前,军事技术的发展呈现出全方位、深层

次发展的态势和多点突破、深度融合、广泛渗透的特征,对国家军事实力、综合国力以及国家安全和战略主动权产生重要影响。可以在国家科技管理部门成立军事技术预警机构,具体负责承担军事技术发展方向研判工作,来确保军事技术的发展方向同国家战略诉求保持一致。机构需要设立信息收集、筛选、评估、决策等业务部门,组织各领域专家组建专家咨询委员会,使机构具有独立完成技术预警、技术研判的能力;系统还需要建立常态化的值班机制,及时跟踪国内基础科学最新研究成果和国际军事技术发展动向,确保预警反馈的时效性,助力形成技术突袭能力;一些发达国家大肆宣传一些似是而非的技术概念,有意将对手引入歧途,分散对手人力物力,达成技术欺骗的目的,需要注重提高机构的分析判断能力,避免受到战略欺骗。

4.5 加强相关学科和人才队伍建设

军事技术评估是一个综合了多学科、多专业的综合性业务,对人才需求十分紧迫。研究机构要在已有人才队伍建设计划的基础上,有步骤、有层次、有重点地尽快推进和落实高层次人才队伍建设工作,在原有理论研究人才队伍的基础上,吸纳更多具有技术应用研究经验的研究人员加入军事技术评估队伍,改善军事技术评估人员的知识结构;科技主管部门要发挥好指导作用,通过组织高水平、高层次的交流、合作与培训活动,拓宽人才自身的视野,强化军事技术评估的综合水平。

参考文献

- [1] U.S. DEPARTMENT of DEFENSE. Department of Defense Instruction 5000.2 [Z]. 2017-2-2.
- [2] 徐 峰. 美国科技管理体制的形成与发展研究[J]. 科技管理研究, 2006, 26(6): 13-16.
- [3] 黄先智. 美国科技政策的演变及特点[J]. 云南科技管理, 2003, 16(2): 50-52.
- [4] PURKS S K, HAMILTON J C. Test and Evaluation Management Guide [M]. Fort Belvoir: Defense Systems Management College Press, 1988: 3.

- [5] JOSEPH K W. A Comparative Analysis of Ship Operational Test and Evaluation in the United States Navy and the Royal Australian Navy [D]. Monterey: Navy Postgraduate School. 1992: 18.
- [6] CONGRESSIONAL RESEARCH SERVICE. Congressional Research Service and the American Legislative Process [R]. 2011:2.
- [7] 郭 华, 孙 虹, 阚 为 等. 美国科技评估体系的研究和借鉴[J]. 中国现代医学杂志, 2014, 24(27): 109-112.
- [8] 崔 侃, 王保顺. 美军装备试验与评估发展[J]. 国防科技, 2012, 33(2): 17-22.
- [9] WENDEL L. Has the Navy's Operational Test and Evaluation Force become antiquated? [J]. Theses and Dissertations, 1986: 35.
- [10] PONTURO J. Analytical Support for the Joint Chiefs of Staff: The WSEG Experience, 1948-1976[R]. Alexandria: INSTITUTE FOR DEFENSE ANALYSES, 1979: 3.
- [11] Senate U S C. A History of the Department of Defense Federally Funded Research and Development Centers[R]. Washington: Office of Technology Assessment, 1995:4.
- [12] Thomas H. Research, development, test and evaluation in the Navy.[D]. California: Naval Postgraduate Schpppl, 1961: 1.
- [13] U.S. GENERAL ACCOUNTING OFFICE. Information Report on RDT&E Organizations in the Department of Defense[R]. Washington: U.S. GAO, 1967:7.
- [14] U.S. DEPARTMENT of DEFENSE. Department of Defense Directive 5129.1 [Z]. 1959-2-10.
- [15] ANONYMOUS. Role of Director of Defense Research and Engineering under the Department of Defense Reorganization Act[J]. Science, 1958, 128(3320):348.
- [16] Pugh G E. Operations Research for the Secretary of Defense and the Joint Chiefs of Staff[J]. Operations Research, 1960, 8(6): 839-846.
- [17] WENDEL L. Has the Navy's Operational Test and Evaluation Force become antiquated? [J]. Theses and Dissertations, 1986: 26.
- [18] DEFENSE SCIENCE BOARD. Summer Study Report on Acquisition Cycle[R]. Washington: U.S. DBA, 1977: 7.
- [19] BLUE RIBBON DEFENSE PANEL. Staff Report on Operational Test and Evaluation: Report to the President and Secretary of Defense on the Department of Defense[R]. Washington: U.S. BRDP, 1970 : 59.
- [20] U.S. DEPARTMENT of DEFENSE. Department of Defense Directive 5000.3[Z]. 1973-02-19.
- [21] U.S. DEPARTMENT of DEFENSE. Department of Defense Directive 5000.1[Z]. 1971-05-13.
- [22] GENEVIEVE J K. Technology Assessment in Congress: History and Legislative Options[R].2005:1.
- [23] U.S. DEPARTMENT of DEFENSE.Department of Defense Directive 5000.1.[Z]. 2011-08-20.
- [24] 陈 鹏. 美武器装备采办管理改革透析[J]. 外国军事学术, 2004(9): 46-48.
- [25] 杨武成. 美国武器装备采办系统探析[J]. 科技信息, 2011(29): 64-65.
- [26] U.S. DEPARTMENT of DEFENSE.Department of Defense Instruction 5134.17[Z].2011-12-25.
- [27] U.S. DEPARTMENT of DEFENSE.Department of Defense Instruction DoDD 5141.02[Z]. 2011-02-02.
- [28] 董 菲. 美国国防采办体制研究[D]. 南京: 南京航空航天大学, 2008: 23.

The Historical Context, Basic Framework and Characteristics of the Test and Evaluation System For American Defense Acquisition

Yang Peiyu, Liu Jifeng

(College of Humanities and Social Sciences, National University of Defense Technology, Changsha 410073, China)

Abstract: The United States has built a solid bridge between military technology and army's combat effectiveness by establishing an institutionalized defense acquisition test and evaluation system, which ensures the leading position in technology of its army for a long time. By analyzing plenty of related journals, this article reviews the historical context of test and evaluation system for American defense acquisition, and summarizes its basic types, organization structure, implementation steps and main characteristics. This article also provides some useful reference for the construction of China's military technology test and evaluation system.

Keywords: defense acquisition; testing and evaluation; technology assessment