CHAPTER 2

1. The Genesis of Test & Evaluation

1.1 Introduction to Test & Evaluation

We now live in a world where it is well recognised that rapid technological advances are fast outstripping mankind's ability to provide adequate test surveillance using conventional wisdom, tools and techniques. As a result of this shortfall, new measurement tools and techniques are now being urgently developed, and this thrust is being matched by aggressive research, post-graduate education programs, and both national and international test resource development on a very large scale. This has spawned a professional discipline and a multi-billion dollar industry known simply as Test and Evaluation (Crouch, 1992). Test & Evaluation (T&E) is a process for technical and programmatic control of systems acquisition. As the phrase implies T&E is a two part process. The test involves the planning and execution of an experiment in an effort to collect data. Evaluation is the assessment of the collected data, against a known standard, in order to obtain knowledge regarding the quality or goodness of the subject under test (Miller and Sears, 1993), (Dvorak, 1995). At it's most fundamental level T&E is normally conducted to influence some type of a decision. It imparts a known level of confidence (Miller and Sears, 1993) regarding the utility of the subject under test (Dvorak, 1995). T&E consists of structured processes. Mostly these processes involve collection of data describing aspects of the operation of a system which is then compared against criteria, the process of evaluation (Dvorak and Equid, 1994).

1.1.1 What is Test and Evaluation (T&E)

Having briefly introduced test and evaluation and why it is a very necessary part of a two fold process, it is only appropriate to define the two phrases formally. The Concise Oxford Dictionary (Allen, 1992) defines the two phrases as follows:

> *"**Test**. A critical examination or trial of a person's or thing's capabilities; The means of so examining; A standard for comparison or trial."*

> *"**Evaluate**. Assess, appraise; Find or state the number or amount of; Find a numerical expression for."*

Simply stated, T&E is a means of obtaining knowledge about something. To obtain that knowledge we use a two part process integrating (Crouch, 1992):

1. **Testing.** In which we gather data about the thing we are endeavoring to learn more about.
2. **Evaluating**. The analysis and interpretation of the data which enables a conclusion to be made regarding the relative merit of the thing we just tested.

From this two part process we now "know" something about the object, or product under investigation.

A concise definition for T&E is as hard to come by as a concise definition for Research & Development (R&D). The following definitions look at both the Australian and American perspectives.

> ***"Australian*** *T&E is defined as a structured investigation designed to obtain or verify data on which to base an objective assessment (Crouch, 1992)."*
>
> ***"American I*** *T&E is the measurement of the performance of a system, and the assessment of the results for one or more specific purposes (Reynolds, 1994)."*
>
> ***"American II*** *T&E may be defined as all physical testing, modeling, simulation, experimentation and related analyses performed during research, development, introduction and employment of a weapon or subsystem (Defence Systems Management College, 1995)."*

In evaluation of the above three definitions it is evident that (Crouch, 1992) defines T&E as a *structured investigation*, and the *verification of data*, as opposed to (Reynolds, 1994) who sees it as a *measure of a performance (MOP)* of a system, whilst the (Defence Systems Management College, 1995) defines it *as all the aspects of testing and employment of a weapon system*, towards a very military sighted view.

T&E is usually conducted to assist in making engineering, programmatic or process decisions, and to reduce the risks associated with the outcome of those decisions. Following from the above definitions, T&E is seen to be a universal tool that is equally applicable to monitoring R&D efforts as well as monitoring the operational health of systems that have been introduced into the service. Arguably (Crouch, 1992) then, "*the need for rigorous testing should be driven by a passion for success rather than a fear of failure*".

1.1.2 How is T&E used on a Day to Day Basis

Possibly the first thing that comes to mind is a team of skilled engineers crowded around complex high tech. machinery. If you stop to think about it T&E is part of almost everything we do. The following section gives a layman's perspective using two children that have made slingshots, inspired by a similar example in (Dvorak, 1995).

1.1.2.1 Children Making Slingshots

For this example I have chosen two bright young elasto-projectile test engineers Roy Rogers & Steve Stevens. The two boys have challenged each other to an afternoon slingshot challenge down at the riverside valley park, using home made targets. Both boys have done extensive T&E to prepare for their shooting challenge. Both Roy and Steve had identical slingshots made out of wood and rubber bands, with a small leather pouch to hold their ammo. Both Roy and Steve conducted initial testing to perfect their accuracy:

- They picked up stones, tried them out, these seemed to fall short of the target and thus had no choice but to redesign their slingshots.
- They picked smaller rounded stones, which seemed to work better and fly longer, thus were more accurate.
- They used an old tyre tube instead of rubber bands as the sling, tried this out, and managed to hit the target 7 out of 10 times, thus concluded that tyre tubes were - more powerful and happened to work even better.
- After an iterative designing test process the boys came up with the ultimate design, taking into account, projectile type, size, shape, weight, type of sling, type of wood, type of ammo pouch, employment technique, distance, trajectory, etc.

Roy stopped testing when he felt that he had achieved the best design, which was:

- An oak wood slingshot.
- Truck tyre tube as the sling.
- Light small leather ammo pouch and
- Uniform small spherical stones.

Roy was positive that this configuration was sure to reach the target accurately and swiftly every time. That afternoon the boys went down to Riverside Valley Park, set up their targets and began. Roy was shooting down every target with one shot, which was a perfect triumph for his test program. However, Steve was only shooting down every 1 out 3 targets with one shot.

Roy had extended his testing to include the operational environment, that is, he shot at rocks first and noticed the following:

- That they wouldn't fall when hit at the edges
- Changed his targets to potatoes which was a lot better
- Only shot when there was less or no wind
- Prepared a plan

From our example we can see that:

♦ **T&E is a process**, that is:

⇒ Design - Test - Analyse - Fix -Test

⇒ Needed to shoot a number of projectiles to get a significant sample size

⇒ Wood type (material)

⇒ Construction (manufacturing)

⇒ Shooting (employment)

⇒ Logistics

- It involves the collection of data.
- The data relates to aspects of the system operation.
- The data is compared against criteria in a process of *evaluation*.
- Does this type of stone work as well as the last:

⇒ Towards a rock,

⇒ Towards a potato

♦ **An extension of the scientific method, whereby we**

⇒ Identify the problem.

⇒ Hypothesis.

⇒ Experiment.

⇒ Verify Hypothesis.

It is therefore necessary to understand the ***"T&E process***":

- The question often put forth is *"What information is required, not what data can be made available" (Scheikhard, 1991).*
- More targeted testing requirements.
- Better use of test results.
- More effective use of test facilities.

All this leads to research and establishes the profile and the characteristics of the T&E process. The T&E process is a system of documentation coupled to engineering design and project management. Two important facets of the T&E process are:

1. To identify features of the process and how you can improve on the already available documentation, practices, and techniques.
2. How to make all the documentation and further information available in a dynamic manner.

The next paragraph is a view expressed by the Defense System Management College on the T&E process.

> *"**The test and evaluation (T&E) process** is an integral part of the systems engineering process which identifies levels of performance and assists the developer in correcting deficiencies. It is also becoming a significant element in the decision-making process, providing data supportive of trade-off analysis, risk reduction and requirements refinement. Programmatic decisions on system performance maturity and readiness to advance to the next phase of development are becoming more dependent on demonstrated performance. The ultimate customer, the Service-member user, is concerned about neither unit cost nor production schedule. The issue of paramount importance is system performance, i.e., will it fulfill the mission. The test and evaluation process provides data to tell the user how well the system is performing during development and if it is ready for fielding. The program manager must balance the unit of cost, schedule and performance to keep the program on track to production and fielding. The responsibility of decision-making authorities centers on assessing risk trade-offs."*

Hence, test and evaluation is streamlining the process of putting test data into a form that users can analyse quickly and efficiently. Software eliminates the preprocessing phase and allows users to directly access and analyse raw telemetry data streams written in arbitrary and complex formats. Their philosophy *is "record it all and sort it all out later" (Moss, 1993).* Streamlining also results in increased productivity, reduced time for data preparation, access and analysis, and greatly reduced costs when evaluating the total test and evaluation solution.

1.1.3 History of Test & Evaluation

It is said that "One test is worth a thousand expert opinions" (Reynolds, and Damman, 1994). The concept of test and evaluation to determine whether a new device is useful and whether it can accomplish a task that it has been assigned is old as that of invention itself (Stevens, 1986). As inventions and new systems become more complex, this gives rise to a development and testing methodology, hence T&E has evolved and manifested itself into almost everything we do.

1.1.3.1 Prolegomena

The study of test and evaluation has been isolated almost entirely to defence and defence related agencies, namely, the Army, the Navy, and the Air Force. Because of this dilemma very few academic textbooks have been written on this subject, and are not as readily available as textbooks for more traditional research topics. The notable few that the author has discovered so far have been (Stevens, 1986) regarding OT&E and (Rodriguez, 1992) which discusses OT&E suitability related issues, and The Defence Systems Management College textbook, Test and Evaluation Management Guide (1993). These textbooks are oriented toward non-academic applications in defence acquisition (Dvorak, 1995). Thus, due to the scarcity of academic textbooks pertaining to T&E, a majority of the relevant research in this field is in the form of defence related journals or conference papers as well as military standards and instructions as shown in Table 1-1.

DoD ACQUISITION DOCUMENTS	
∗ **DoD DIRECTIVE 5000.1**	Defence Acquisition
∗ **DoD INSTRUCTION 5000.2** ∗ **[Change 1 - February 26, 1993]**	Defence Acquisition Management Policies and Procedures
∗ **DoD 5000.2 MANUAL** ∗ **[Change 1 - February 26, 1993]**	Defence Acquisition Management Documentation and Reports

Table 1-1 (DoD Acquisition Documentation (Damaan, 1993))

The DoD Policy for Acquisition is such that it (Damaan, 1993) *"Establishes a disciplined management approach for acquiring systems and materiel that satisfy the operational user's needs"*. This is applies to all major as well as non-major defence acquisition programs.

1.1.3.2 Test and Evaluation & the Scientific Method

The Test and Evaluation process dates back to scientific principles and foundations. The scientific method (Dvorak and Equid, 1994) is based on a combination of logical reasoning or philosophical assertions with methods for acquiring knowledge. The scientific method can be defined as *"an objective, logical and systematic method (process) of analysis of phenomena for accumulation of reliable knowledge"* (Miller and Sears, 1993).

The Encyclopaedia Britannica (Benton and Benton, 1980a) defines the words science and scientific method in the following manner:

> ***"Science,*** *philosophy of, a discipline in which the elements involved in scientific inquiry - observational procedures, patterns of argument, methods of representation and calculation, and metaphysical presuppositions are analysed and discussed; and the grounds of their validity are* ***evaluated*** *from the points of view of formal logic, practical methodology, and metaphysics"*.

> *"**Scientific Method**, once considered to be a rigorous procedure that included the study of scientific hypotheses, induction, theories, laws, and methods of exploration; now regarded as a family of methods each of which differs according to the subject matter involved. The core of the scientific method, however it is defined, is related to measurement of phenomena and experimentation or repeated observations."*

The measurement of this so called phenomena, repeated observations is conducted in a series of seven steps. These seven steps are defined by Lastrucci (1963) and Fiebleman (1972) which are defined and compared in Table 1-2 below.

THE SCIENTIFIC METHOD SEVEN STEP PROCESS	
LASTRUCCI (1963)	**FIEBLEMAN (1972)**
1. Formulation of the problem (hypothesis)	1. Observation
2. Literature survey	2. Induction
3. Research design	3. Hypothesis
4. Determine "universe" encompassed	4. Experiment
5. Collect data, process for use	5. Calculate (verification)
6. Interpretation of data	6. Prediction (verification)
7. Verification of results	7. Control (verification)

Table 1-2 (The Scientific Method Seven Step Process (based on Miller & Sears, 1993))

A generalised structure of the scientific method compared to the T&E process is presented in the following table (Fiebleman, 1972):

GENERALISED SCIENTIFIC METHOD	TEST & EVALUATION PROCESS
I. DEVELOP HYPOTHESIS	**I. DEVELOP HYPOTHESIS**
1. Identify Question/Problem	1. Develop Test Objectives
2. Formulate Hypothesis	2. Estimate Performance
II. EXPERIMENT	**II. EXPERIMENT**
3. Plan the experiment	3. Develop Method of Test
4. Conduct the Experiment	4. Collect Test Data
5. Analyse the Results	5. Calculate Measures of Performance
III. VERIFY HYPOTHESIS	**III. VERIFY HYPOTHESIS**
6. Check the Hypothesis	6. Compare Results to Thresholds
7. Refine the Hypothesis	7. Retest or Extrapolate

Table 1-3 (Relationship of the Scientific Method vs Test and Evaluation Process (Fiebleman, 1972))

1.1.3.3 Test and Evaluation & Systems Engineering

1.1.3.3.1 Introduction

The discipline of Systems Engineering (Przemieniecki, 1993) first came into being in the late1950s with the advent of the Intercontinental Ballistic Missile program in the United States. The concept of the Intercontinental Ballistic Missile pushed the state of the art in a number of technical areas, resulting in the need to develop engineering specialties to concentrate on these advances. It was important that these engineering specialties worked together in a final product, and the need to balance these specialties created the concept of Systems Engineering. According to MIL-STD-499B (1992), Systems Engineering is defined as follows:

> ***Systems Engineering*** *is an interdisciplinary approach to evolve and verify an integrated and life cycle balanced set of systems product and process solutions that satisfy customer needs. Systems engineering: (a) encompasses the scientific and engineering efforts related to the development, manufacturing, verification, deployment, operations, support, and disposal of system products and processes, (b) develops needed user training equipment, procedures, and data, (c) establishes and maintains configuration management of the system, (d) develops work breakdown structures and statements of work, and (e) provides information for management decision making.*

Systems engineering integrates the total engineering effort to meet cost, schedule, and technical performance objectives (Lacy, 1994), and is both a technical process and a management process.

1.1.3.3.2 Systems Engineering Process

Many definitions of a system have been offered, but, in the broadest sense, "*any two or more objects interacting cooperatively to achieve some goal or purpose constitute a system*" *(Grady, 1993).* According to MIL-STD-499B (1992), the system engineering process is defined as follows:

> ***System Engineering Process*** *is a comprehensive, iterative problem solving process that is used to: (a) transform validated customer needs and requirements into a life cycle balanced solution set of system product and process designs, (b) generate information for decision makers, and (c) provide information for the next acquisition phase. The problem and success criteria are defined through requirements analysis, functional analysis/allocation, and systems analysis and control. Alternative solutions, evaluation of those alternatives, selection of the best cycle balanced solution, and the description of the solution through the design package are accomplished through synthesis and systems and analysis and control.*

A more complete description as depicted by Brook and Arnold (1996) of the system engineering process can be illustrated by the use of system levels, as shown in Figure 1-1. System engineers at the top level define the overall architecture of the complete system in terms of the next components at the next level down. Those at lower levels receive a package of requirements about the architectural element (or group of elements) they are to design, as well as defining any new requirements which appear at that level.

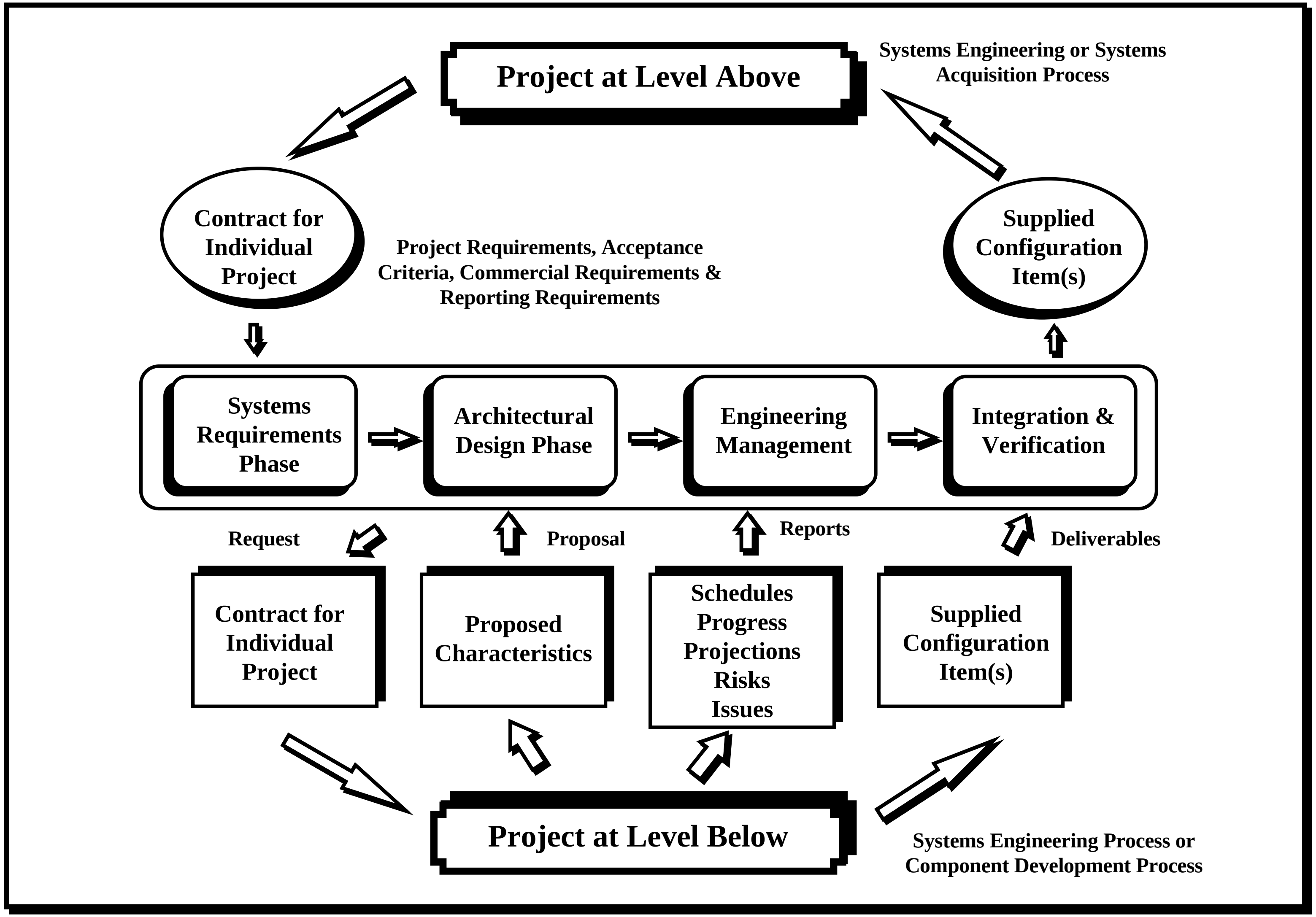

Figure 1-1 (Systems Engineering Process (based on Brook & Arnold, 1996))

It follows then from the decomposition process (Brook and Arnold, 1996) that only the highest systems engineering level responds directly to user requirements and all other levels receive systems requirements from the level above, and then only the subset which is relevant for that level.

Successful OT&E is a systems engineering process in which the system and its testing are approached from an overall systems point of view, and the complete philosophy of the systems approach is brought to bear (Stevens, 1986).

1.1.3.3.3 Test and Evaluation in Systems Engineering

The United States Department of Defence (DoD) Military Standard (MIL-STD) 499B on Systems Engineering depicts T&E as an essential element of the Systems Engineering "engine" (Przemieniecki, 1993). Test and Evaluation must be integrated with the rest of the system engineering effort. The testing program in the Test and Evaluation Master Plan (TEMP) must be consistent with the System Engineering Master Plan (SEMP) (Lacy, 1994). The SEMP is a concise, top-level management plan for integrating all of the system activities. The major objective of the SEMP are to (Lacy, 1994):

- Facilitate communications.
- Integrate all engineering disciplines.
- Ensure the product meets the requirements.
- Establish streamlined checks and controls.
- Define the system engineering process.

The SEMP defines the type of degree of system engineering management, the system engineering process, and the integration of engineering efforts. The plan identifies (Lacy, 1994):

- Organisational responsibilities.
- Authority for system engineering management.
- Levels of control for performance and design requirements.
- Control methods to be used.
- Technical program assurance methods.

- Control procedures to ensure integration of requirements and constraints.
- Schedules for design and technical program reviews.
- A detailed description of the system engineering process to be used.
- Specific tailoring to requirements of the system in-house documentation.
- Trade-off study methodology.
- Types of mathematical and simulation models to be used for system and cost-effectiveness evaluations.

A SEMP is created to structure engineering planning, processes, and outputs. Engineering Management for the DoD is described in MIL-STD-499A. DoD MIL-STD-499A divides systems engineering management into three types of activities (Lacy, 1994):

1. Technical program planning and control
2. System engineering process.
3. Engineering specialty integration.

The structure of MIL-STD-499A activities are shown in Table 1-4 (Lacy, 1994).

SYSTEM ENGINEERING MANAGEMENT PLAN		
TECHNICAL PROGRAM PLANNING AND CONTROL	**SYSTEM ENGINEERING PROCESS**	**ENGINEERING SPECIALTY INTEGRATION**
⇒ Work breakdown structure and specification tree	⇒ Mission requirements analysis	⇒ Reliability
⇒ System test planning	⇒ Functional analysis	⇒ Maintainability
⇒ Decision and control process	⇒ Allocation	⇒ Logistics engineering
⇒ Technical performance parameters (TPP's)	⇒ Synthesis	⇒ Human engineering
⇒ Technical reviews	⇒ Logistic engineering	⇒ Safety
⇒ Vendor reviews	⇒ Life cycle cost analysis	⇒ Value engineering
⇒ Work authorization	⇒ Optimization	⇒ Standardisation
⇒ Documentation controls	⇒ Production-engineering analysis	⇒ Transportability
	⇒ Generation of specifications	

Table 1-4 (The Structure of MIL-STD-499A Activities (Lacy, 1994))

A typical SEMP has a similar format to that of a TEMP (which is addressed in the proceeding chapters) and should contain the information listed in the sample format of Figure 1-2.

Introduction
Part 1 Technical Program Planning and Control
- 1.0 Responsibilities and Authority
- 1.1 Standards, Procedures, and Training
- 1.2 Program Risk Analysis
- 1.3 Work Breakdown Structure
- 1.4 Program Reviews
- 1.5 Technical Reviews

1.6 Technical Performance Measurements
1.7 Change Control Procedures
1.8 Engineering Program Integration
1.9 Interface Control
1.10 Milestones/Schedules
1.11 Other Plans and Controls

Part 2 System Engineering Process
2.0 Mission and Requirements Analysis
2.1 Functional Analysis
2.2 Requirements Allocation
2.3 Trade Studies
2.4 Design Optimization/Effectiveness Compatibility
2.5 Synthesis
2.6 Technical Interface Compatibility
2.7 Logistic Support Analysis
2.8 Producibility Analysis
2.9 Specification Tree/Specificationa
2.10 Documentation
2.11 Systems Engineering Tools

Part 3 Engineering Specialty/Integration Requirements
3.1 Integration Design/Plans
3.1.1 Reliability
3.1.2 Maintainability
3.1.3 Human Engineering
3.1.4 Safety
3.1.5 Standardisation
3.1.6 Survivability/Vulnerability
3.1.7 Electromagnetic Compatibility/Interference
3.1.8 Electromagnetic Pulse Hardening
3.1.9 Integrated Logistics Support
3.1.10 Computer Resources Life Cycle Management Plan
3.1.11 Producibility
3.1.12 Other Engineering Specialty Requirements/Plans
3.2 Integration System Test Plans
3.3 Compatibility with Supporting Activities
3.3.1 System Cost Effectiveness
3.3.2 Value Engineering
3.3.3 TQM/Quality Assurance
3.3.4 Materials and Processes

Figure 1-2 (Typical SEMP Format (based on DSMC, 1990))

1.1.4 Types of Test & Evaluation

In a system T&E requires knowledge of both development activity, which is usually driven by requirements and specifications, as well as the operational environment the system will reside in. This leads to two distinct branches of T&E, namely, Developmental Test & Evaluation (DT&E) and Operational Test and Evaluation (OT&E). DT&E is that T&E that supports the development of a system or process, whilst OT&E is that T&E which assesses (Dvorak, 1995) the effectiveness and suitability for service of a system. The proper scope, structure, and timing of these two types of T&E has yet to be established, a number of T&E practitioners are still in an intense world-wide debate on these issues, namely (Parker, 1993), (Sanders, 1994), (Seglie, 1993a), (Seglie, 1994), (Griffin, 1994), and (Joseph, 1992).

1.1.4.1 Development T&E

The primary focus of DT&E is the identification and verification of system performance specifications. Unfortunately, due to the difficulty in perfectly allocating requirements into systems functionality, a system can often meet all of it's engineering specifications yet still fail to adequately perform its mission (Stevens, 1986). The US Department of Defence Directive 5000.3 defines Developmental Test and Evaluation as follows:

> *"**Development Test and Evaluation** is test and evaluation conducted throughout various phases of the acquisition process to ensure the acquisition and fielding of an effective and supportable system by assisting in the engineering design and development and verifying attainment of technical performance specifications, objectives and supportability."*

Whilst (Joseph, 1992) defines Developmental Test and Evaluation as:

> *"**Development Test and Evaluation** is conducted to assist, the engineering design and development process, and to verify attainment to technical performance specifications and objectives."*

DT&E also includes T&E of components, subsystems, hardware/software, as well as qualification and production acceptance testing. T&E compatibility and interoperability with existing or planned equipment and systems is also emphasised, as well as system effects due to natural and induced environmental conditions. It encompasses the use of models, simulations, and test beds, as well as prototype of full-scale engineering development models of the system (Defence Systems Management College, 1995). According to DoDI 5000.2, the overall DT&E objectives encompass the following:

1. Identify potential operational and technological limitations of the alternative concepts and design options being pursued.
2. Support the identification of cost-performance trace-offs.
3. Support the identification and description of design risks.
4. Substantiate that contract technical performance and manufacturing process requirements have been achieved, and
5. Support the decision to certify the system ready for operational test and evaluation.

1.1.4.2 Production Acceptance Test and Evaluation (PAT&E)

PAT&E is conducted on production items, to ensure systems meet technical specifications and requirements, and is a type of DT&E (Joseph, 1992). It is conducted to assure that production items meet specifications and performance requirements (Defence Systems Management College, 1995).

PAT&E assures that production items demonstrate the fulfillment (Przemieniecki, 1993) of the requirements and specifications of the procuring contract or agreement. The testing also ensures the system being produced demonstrates the same performance as the pre-production models and operates in accordance with the specifications.

1.1.4.3 Contractor Based T&E

In addition to the governmental agencies, the contractor plays a key role in DT&E (Przemieniecki, 1993), especially in the early part of the test program. A contractual system test plan is developed jointly by the Program Officer (PO) and the contractor and it identifies the roles of each participant. The contractor is involved in a range of testing, namely, sub-system testing, operational mock-up testing, and a number of other tests leading up to the first live launch of a rocket or the first test flight of an aircraft.

The Defence Systems Management College (1995) summarises the contractor’s role in testing as follows:

- Deliver Integrated Test Plan (ITP) for approval
- Test Sufficiently before delivery to Government
- Provide Technical support to Government testing
- Correct Problems

- Increase Test and evaluation efficiency

1.1.4.4 Operational T&E

OT&E is conducted to determine a systems operational effectiveness and operational suitability, identify system deficiencies and the need for potential modifications to meet established OT thresholds, and develop tactics (Joseph, 1992).

OT&E has three major distinguishing characteristics:

⇒ It is ***conducted*** in an operationally representative environment.

⇒ It is ***conducted*** on production representative equipment using fleet personnel for operation and maintenance.

⇒ It is ***conducted*** against a threat - representative simulated enemy carrying out threat tactics per the latest threat assessment.

The Defence Systems Management College (1995) defines OT&E as follows:

> *“**Operational Test & Evaluation** is the field test, under realistic conditions, of any item of (or key component of) weapons, equipment, or munitions for the purpose of determining the operational effectiveness and suitability of the weapons, equipment, or munitions for use in combat by typical military users; and the evaluation of the results of such tests”*

The primary focus of OT&E is to ensure that only systems that are operationally effective and suitable will be delivered to the operating forces. The results of OT&E are provided to the appropriate decision makers for decisions on system production and fielding. Therefore, OT&E should be structured to provide inputs at each decision point, including major systems (Defence Systems Management College, 1995).

In general the final evaluation should determine operational effectiveness and suitability. The US Department of Defence (DoD) provides clear definitions of these terms (Seglie, 1993b) (Rodriguez, 1992):

> *“**Operational Effectiveness** is the overall degree of mission accomplishment of a system when used by representative personnel in the environment planned or expected (e.g., natural, electronic, threat) for operational employment of the system considering organisation, doctrine, tactics, survivability, vulnerability, and threat (including countermeasures, initial weapons effects, nuclear, biological and chemical threats).”*

> *“**Operational Suitability** is the degree to which a system can be placed satisfactorily in field use with consideration given to availability, compatibility, transportability, interoperability, reliability, wartime usage rates, maintainability, safety, human factors, manpower supportability, logistics supportability, natural environmental effects and impacts, documentation and training requirements.”*

Operational suitability applies to each level of support. Table 1-5 presents some examples of the levels of support that may be applied, for various weapon systems.

LEVEL	TYPE OF SUPPORT	EXAMPLE		
		A	B	C
1st	Owner or User	Organisational	Crew	Crew
			Unit	Unit
2nd	Supporting Unit(s) with More Capability	Intermediate	Direct Support	Direct Support
				General Support
3rd	Highest Level of Capability	Depot	Depot	Depot

Table 1-5 (Variance in the Definitions of Support Levels (Rodriguez, 1992))

In each of these services, operational testing is conducted under the auspices of an organisation that is dependent of the development agency, in as operationally realistic environments as possible, with hostile forces representative of the anticipated threat and with typical users operating and maintaining the system (Defence Systems Management College, 1995). Often the specific criteria against which to judge the operational effectiveness and suitability are not clearly identified and often gives rise to mandated tests (Seglie, 1993b). In such cases, the objective is implied by the mandate.

1.1.4.5 Differences Between DT&E and OT&E

Development testing is focused on meeting detailed technical specifications, the operational test focuses on the actual functioning of the equipment in a realistic combat environment in which the equipment must interact with humans and peripheral equipment. Where DT&E and OT&E are separate activities and are conducted by different test communities, the communities must interact frequently and are generally complementary. The DT&E provides a view of the potential to reach technical objectives, and OT&E provides an assessment of the system's potential to satisfy user requirements (DSMC, 1995). The key differences are outlined in Table 1-6.

DEVELOPMENT TESTING	OPERATIONAL TESTING
CONDUCT OF TESTS	
Technical, Controlled Environment, Specification Tested, Technical Personnel, "Tweaked System".	Realistic Environment, Fleet Operators and Maintenance, Simulated Enemy Engagements (No Contractors).
SCOPE OF TESTS	
"Black Box", Single Weapon, Generally Only Part of the Complete System.	Total Weapons System Including Operators and Logistics Support.
EVALUATION CRITERIA	
Technical Criteria, Measurable Parameters (Signal Strength, Specifications).	Probability of Mission Accomplishment, P_{det}[1], P_{hit}[2], P_{kill}[3].
MEASUREMENTS & FREQUENCY	
Specific Parameters (Launch Velocity, Load Factor Time To Climb). Test Must be Repeatable.	Generally Specific Measurements Not Tested. Create Combat Conditions and Observe Results. Test Not Repeatable, Interactions Usually Unique.

Table 1-6 (Differences Between DT&E and OT&E (Hoivik, 1995))

In some instances OT&E and DT&E are combined. The following points adhere to carrying out this action (Hoivik, 1995):

[1] Probability of detection.

[2] Probability of a hit.

[3] Probability of a kill.

- Usually conducted to obtain significant cost and time benefits.
- Must provide necessary resources, test conditions, and test data is required by both development agency and operational testing agency.
- Data collected must be sufficient and credible for OT&E agency requirements.
- Separate and independent evaluation of test results are required.

1.1.5 Interest in T&E

It is evident that both the public and the government want credible T&E programs (Reynolds and Damman, 1994), hence their full attention, due to the underlining reasons:

- Consumer concerns about commercial products.
- Government concerns about commercial products.
- Government concerns about governmental products.

There is pressure from consumer groups in areas such as safety in automobile designs and children’s toys, which have alerted considerable interest in the last two decades on the method in which industry tests it’s products. It is not unusual to see impressive, graphic crash tests as part of the marketing material used by new car manufacturers. Fisher-Price has placed lot’s of emphasis on the safety of their toys for kids of different age categories, and has become known as the industry leader in that arena.

Oversight of commercial products by government regulatory agencies has also increased attention to T&E. Take the T&E effects of cigarette smoking on humans for example, this has been a long-term evolving task. Moreover, testing to determine the effects that household aerosol sprays have on the ozone layer, due to the emission of chloro-fluorocarbons, or CFC’s, has been under serious investigation for almost two decades. Another good example is the testing to determine the effects that chopping down trees and gradually wiping-out forests has on the environment. Trees help clean the air by removing poisonous gases and participates such as dust and pollen. Through photosynthesis, trees reduce atmospheric levels of carbon dioxide and release vital oxygen. In addition, well placed trees reduce the need to burn fossil fuels to generate energy for air conditioning. The solution to all these problems and perhaps gradual demise of the planet is rigorous T&E, and verification and validation (V&V) of these test is needed.

1.1.5.1 T&E Education & Training

1.1.5.1.1 In the United States

In the United States alone, both under-graduate and postgraduate T&E studies are available. T&E engineering specialty topics are taught at a variety of civil Universities and tertiary institutes in Georgia, New Mexico and Texas. T&E management is also taught as a disciplined process at Army, Navy, Air Force and Defence colleges; and at the Federal Aviation Academy (Crouch, 1992).

1.1.5.1.2 In Australia - Formation of ACTE as a Support Base for Advancing T&E

In July 1993 interests in large multi-sensor measurement systems engineering and how they are used to obtain effectiveness measures from sensed data were combined with pragrammatic needs at the nearby Aircraft Research Establishment RAAF to form the Australian Centre for Test and Evaluation (ACTE). After two years the T&E research program is becoming more focused from its longer standing Measurement and Instrumentation (M&I) thrust (Sydenham, 1995).

The aim of ACTE is to develop the professionalism and skill level of T&E practitioners by high-level education and training; technology transfer; research and development; consulting and project management. It also provides a resource support for the International Test and Evaluation Association, ITEA (Harris, 1995).

1.1.6 Reasons for Conducting T&E

There are numerous reasons for conducting T&E with the most important objective of T&E being reduction of the risk of doing something. Testing is conducted for many of the following reasons (Dvorak and Equid, 1994):

a) To prove a concept
b) To ensure safety.
c) To ensure adequate human factors.
d) To ensure user requirements are met.
e) To avoid failures in service.
f) To check contract compliance.
g) To support acquisition decisions.
h) To provide feedback to designers.
i) To verify Supportability.

j) To validate models and simulations.

The purpose of test & evaluation as seen by the Defence Systems Management College (1993) is as follows:

> *"**The fundamental purpose of test and evaluation** in a defence system's development and acquisition program is to identify the areas of risk to be reduced or eliminated. During the early phases of development, T&E is conducted to demonstrate the feasibility of conceptual approaches, evaluate design risk, identify design alternatives, compare and analyze trade-offs, and estimate satisfaction of operational requirements. As a system undergoes design and development, the emphasis in testing moves gradually from development test and evaluation (DT&E), which is considered chiefly with attainment of engineering design goals, to operational test and evaluation (OT&E), which focuses on questions of operational effectiveness, suitability and Supportability. Although there are usually separate development and operational test events, DT&E and OT&E are not necessarily serial phases in the evolution of a weapon system. Combined and concurrent development and operational testing is encouraged when appropriate."*

Hence the reduction of risk or complete elimination of these risks in any defence acquisition program can only be accomplished by thorough T&E, and consequently verification and validation of these tests to a well known standard. As implied by the defence systems management college, testing does not stop after the initial developmental phase, simply because the engineering design goals have been achieved and the procurement process has adhered to specifications as set out in the statement of requirement (SOR), but continues right through the operational phase. Even when the desired product is sitting on the shelf ready for sale there is quality control inventory processes that all products having attained a "saleable rank" must adhere to, via random testing and evaluation. This system large volume quality control testing is a mandatory procedure for most (if not all) "commercial of the shelf (COTS)" products.

1.1.7 Importance of Test & Evaluation

Several important features of T&E are (Dvorak and Equid, 1994):

1. T&E is a process,
2. It involves the collection of data,
3. The data relates to aspects of the systems operation, and
4. The data is compared against criteria in a process of evaluation.

In the acquisition of defence systems the importance of T&E as dictated by the United States Government Accounting Office Report, August 7, 1972 was as follows (Reynolds, 1993):

- Establishment of test objectives adequate.
- Most system test plans not adequate.
- T&E in most programs not timely and effective.
- Test results were adequate, but their value was diminished.
- Complete test data not available to decision makers prior to key decision points.

1.1.8 Objectives of T&E

The principal objectives of T&E are to (Crouch, 1992):

- Reduce technical risk
- Find problems while they are cheap to fix
- Give confidence to the decision makers that the:
  - ⇒ Promises of emerging technology are being realised,
  - ⇒ Simulations and models are faithful,
  - ⇒ Systems meet their specifications and that
  - ⇒ Systems behave safely and as predicted over their useful life.

Objectivity enters during test planning, evaluation, and reporting. It has three aspects (Seglie, 1993b):

1. The test should present a balanced spectrum of missions to the system.
2. The evaluation should weigh the pluses and minuses of the system openly and carefully.
3. The report should describe what actually happened in the test and explain the reasons for the judgments that are made.

1.1.9 Need for Conducting T&E

T&E Principles can be applied to a wide range of products. The Defence Department uses disciplined T&E for a range of reasons as follows (Dvorak and Equid, 1994):

1. Systems in procurements/use can be very complex.

2. System usage can be in a dangerous environment or an environment that is not easily simulated for test purposes, and
3. Development and acquisition programs can be phenomenally expensive and T&E is applied to lower technical or program risks.

1.1.10 Foreign Comparative Testing

Foreign Comparative Testing, otherwise known as FCT is a program supported by the United States as a national policy for encouraging international armaments cooperation and helps reduce overall DoD acquisition costs by facilitating the procurement of non-developmental items (NDI). Biskey (1994) defines FCT as follows:

> ***"Foreign Comparative Testing** involves the T&E of selected items of defence equipment developed by US allies, and other nations considered friendly toward the US, to determine whether such equipment can effectively satisfy DoD requirements or correct mission area shortcomings, as cost-effective alternatives to new, and perhaps unnecessary, developmental efforts."*

By identifying foreign alternatives, Biskey states that FCT stimulates competition from US manufacturers; however, safeguards are in place to ensure that US manufacturers are not placed at any disadvantage and that US industrial base issues are considered.

The underlying document for carrying out any FCT program, outlining procedures and formats is the DoD 5000.3-M-2 (1994), instigated by The Under Secretary of Defence (Acquisition & Technology). This standard states that the subsequent acquisition of foreign technology and/or deployment of selected foreign systems evaluated under the auspices of the FCT program results in significant resource savings by avoiding unnecessary duplication of R&D, achieves more timely fielding, and provides viable alternative solutions to component requirements, promoting healthy competition and resultant procurement savings.

Further more, the FCT program directly supports the DoD policy that equipment procured for use by personnel of the Armed Forces of the US stationed in Europe, under the terms of the North Atlantic Treaty, by standardised or interoperable with equipment from other North Atlantic Organisation (NATO) nations.

In the Australian arena, a number of locally grown products are being evaluated under the US FCT program with good prospects for follow-on procurements by the US Services. Walls

(1995) states that these products have successfully competed against international competition before entry into the FCT program. At present, Australian products being tested by the US Navy are as follows:

- Vision System's Laser Airborne Depth Sounder (LADS)
- Ryan Marine's PROPSCAN computerised propeller measurement equipment
- Australian Defence Industries (ADI) Dyad decoy project which relates to the Auxiliary Minesweeping and Surveillance System (AMASS)

An earlier program led to the purchase of Australian transportable recompression chambers from Crown Engineering. Funding for the FCT program has averaged about $US27 million over the last six years. Generally, projects approved for T&E through the FCT program are funded for no more than a two-year effort. However, on an exception basis, funding for T&E of complex systems (such as F/A-18 fighter aircraft) may be provided for a longer period.

1.2 Conclusion

This chapter has reviewed test and evaluation, flight test, past and present. It was found through the comprehensive literature review that, test and evaluation is a process, and merely an extension of the scientific method, that is: design - test - analyses - fix - test. This was substantiated with a layman's example of children making slingshots.

Over the duration of the literature search it was determined that very few academic textbooks had been written on this subject, due to the fact that the subject of this research has been primarily concerned with defence, and in particular the United States Department of Defence.

An outline on the connection between systems engineering and test & evaluation was discussed, and the different types of test & evaluation, namely, developmental test & evaluation and operational test & evaluation. A description on the interests, education and training, reasons for conducting, importance, objectives and the need for conducting test and evaluation was made.

In conclusion, test and evaluation is a very new field of research and open to a vast multitude of exploration in a rapidly maturing environment. The next chapter will look at the genealogy of aircraft flight test and how it is directly interpolated with the role of test and evaluation.