

A Technical Overview of the Evidence Framework Approach: Practical Ways of Thinking About Evidence

Paul V. Pearce Defence Science and Technology Laboratory (Dstl),
pvpearce@dstl.gov.uk

DSTL/CP102977

34th International Symposium on Military Operational Research, July 2017

Abstract

This paper introduces the reader to the basic concepts behind what is called the Evidence Framework Approach (EFA). The EFA consists of three simple and complementary processes to support problem formulation, evidence evaluation and assessment of analytical rigour. The EFA has applicability across a wide range of study types and analysis domains and any study concerned about quality assurance and evidence can benefit its analytical assurance process through the use of the EFA. At its most basic level the EFA is about practical ways to think about evidence and practical ways for improving analytical quality. The EFA can be used to answer one of the questions often posed at the beginning of an analytical study: "How much evidence is enough?"

Background

The drivers for the development of the Evidence Framework Approach (EFA) were initially rooted in the local needs of the Land Environment Operational Analysis (LEOA) project, which formed part of the Defence Science and Technology Laboratory's (Dstl's) Land Environment Decision Support (LEDS) programme.

While the principal aim of the EFA was to support the evidence aspirations of LEDS and LEOA, the EFA was also developed to meet the wider analytical needs of the United Kingdom's (UK) Ministry of Defence (MOD). In particular, the EFA is a response to Levene's Defence Reform recommendations (1) in that it provides effective 'handrails' for assessing evidence quality, thereby exposing evidence based on advocacy; MacPherson's review of analytical models (2) by challenging the appropriate use of models and methods; and more recently Chilcot's Iraq Inquiry (3) which exposed issues in the evaluation, assessment and understanding of evidence in support of decision making.

Following the early developments of the EFA (4,5) the user trials (6,7,8) which established its effectiveness and wider socialisation (9,10,11,12,13,14) of the

EFA, one aspect, the RIGOUR¹ concept (15), was taken forwards into the UK pan-Government-Department Aqua Book² (16). This provided an opportunity to align developing EFA thinking with developing Aqua Book thinking. As a result the EFA augments guidance on analytical quality assurance contained in the Aqua Book.

The current version of the EFA has evolved through application to studies and wider engagement within MOD, industry and academia. It forms a core part of the current LEDS programme, is gaining increasing traction within Dstl's analysis community and more recently within industry, for example aspects of the EFA are included in a revision to the Niteworks® Code of Best Practice for Warfighting Experimentation (17).

It is posited that the EFA has applicability across a wide range of study types and within a variety of other analysis domains and can be used to answer one of the questions often posed at the beginning of an analytical study: "*How much evidence is enough?*" Any study concerned about quality assurance and evidence can benefit its analytical assurance process through the use of the EFA.

The purpose of this paper is to provide a summary of the key aspects of the EFA, to encourage debate and facilitate its application. As such the paper provides an overview of the core aspects of the EFA; introduces each of the core components and how to apply them and provides a simple example of one aspect to aid understanding of its benefit.

Overview of the Evidence Framework Approach (EFA)

At its most basic level the EFA is about practical ways to think about evidence and practical ways for improving analytical quality. Specifically, the EFA aims to:

- Help people become better systems thinkers by embracing complexity within an 'analysis estimate' process.
- Provide a means by which people can consider evidence and its characteristics and engage in discourse about evidence.
- Provide quick methods to support a range of stakeholder interests.

The Core Aspects of the Evidence Framework Approach

To realise these aims the EFA provides three short processes supported by simple tabular tools. The processes and tools focus on three aspects of an analytical study: problem formulation, evidence evaluation and assessment and analytical quality assurance:

¹ Repeatability, Independence, Grounding in reality, Objectivity, Uncertainty managed, Robustness.

² The Aqua Book contains guidance on accomplishing analysis with evidence that is fit-for-purpose, i.e. is of an appropriate quality.

- **Problem formulation:** An enhanced study design/problem exploration approach is proposed to support an Initial Analysis Estimate (IAE). The IAE draws on complexity thinking as currently envisaged within the Cynefin framework (18).
- **Evidence evaluation and assessment:** A means for evaluating and assessing evidence and associated validity aspects is proposed using the Evidence Profile Table (EPT) and the Validation Profile Table (VPT). As a means of bringing both of these aspects together to derive confidence in the findings the Confidence Assessment Table (CAT) is proposed. These three aspects of the EFA draw together thinking from a number of sources (19, 20, 21, 22, 23, 24, 25, 26) to provide a pragmatic way of assessing evidence fitness-for-purpose.
- **Analytical quality assurance:** A means for assessing analytical RIGOUR is proposed using an Evidence Quality Questionnaire (EQQ) appropriate to the current stage of a project. The questionnaires reflect the version of RIGOUR in the Aqua Book (16) which is itself a modification of an early version (15).

Together these processes and tools result in better engagement between stakeholders and the analysis team, by improving the shared understanding of issues and challenges (14); more robustly defensible evidence to support decision making; and a structured means to assess the evidence quality required to support capability development and what a study may achieve. It is important to note that not all three processes need to be applied, analysts can decide on which is more appropriate to their study needs.

Applying the Evidence Framework Approach (EFA)

The Initial Analysis Estimate (IAE)

The IAE is a sensemaking approach which utilises complexity thinking techniques appropriate to the application of Cynefin (18). Trials measuring shared awareness amongst participants (6,14) have shown that the methods improve the quality of engagement for IAE participants compared with unstructured, non-facilitated discussions. The methods lead to a better shared understanding through providing better focus to the conversation between the participants, assisting in the articulation and expression of ideas. The small facilitated groups also encourage a broader range of inputs from different types of participants compared to an unstructured approach. The methods employed in the IAE are shown in Figure 1 with an IAE session typically taking 3-4 hours to complete.

A key benefit of the EFA is in providing guidance in the adoption of complexity thinking. The approach aids stakeholders in categorising the key aspects of a problem to be studied in terms that have a direct bearing on how the aspects should be studied, a key reason for engaging in complexity thinking.

While a particular facilitation method is described it is not necessary to always run

with the same method and users are encouraged to try a range of facilitation methods, varying their selection with the nature of the problem. The key requirement is that they must be compatible with encouraging participants to appropriately explore any issues of complexity.

The IAE activity takes place towards the beginning of the study process as a means to better define the problem and the form a study will take to address the problem. At this stage the Analyst, Analytical Assurer and supporting Subject Matter Experts (SMEs) together with the Commissioner³ discuss the purposes, context and outputs of the study, and consider which methods are appropriate to employ, informing the design of the study. The results of this process are captured in an initial analysis estimate. This is co-ordinated with initial EPT, VPT and CAT assessments and further developed in the study plan. This aspect of the plan is typically captured in a more detailed concept of analysis document or experimental design document.

The Evidence Profile Table (EPT)

The EPT, illustrated in Table 1 is designed for use⁴ in assessing or evaluating the required or achieved quality of a body of evidence to inform decision making. This can be evidence to inform a particular decision, to evaluate a methodology or method(s) to be used for a study, or to evaluate the study as a whole. The EPT is used to assign a level between one and four to each of five factors (see below) that are considered to be generic characteristics of evidence (25). The resulting evidence profile is simply summed⁵ to arrive at a statement concerning evidence warrantability. It may seem counter-intuitive to consider the highest grade as 1 rather than 4. This is a deliberate attempt to force users to think about the evidence and to avoid thinking that not achieving a better score is inherently bad. The nature of the problem, time, cost etc may mean it is not possible or necessary to achieve better scores. The key point is about assessing fitness-for-purpose.

In addition a statement of warrant concerning the evidence is very much a study team view through sharing of the findings, methods used etc amongst peers to enable a judgement to be made. The warrant can then be used to understand or assess the overall evidence position required or achieved for the assertion being made and is used in conjunction with the VPT score to make a judgement using the CAT to determine confidence in the findings informing the decision.

EPT factors – What are they and what do they mean?

The EPT factors are considered generic evidence characteristics and are used to

³ Note that the Analyst, Analytical Assurer and Commissioner are specific roles defined within the Aqua Book.

⁴ The EPT can be used on an individual basis by the Analyst or Analytical Assurer but greatest value will be obtained when the EPT is used as part of the conversation with the Commissioner.

⁵ Currently all the factors are considered to be equally weighted. Further research will be required to determine if this assumption is valid.

structure a conversation on evidence and to assign a profile that can be used to judge the extent of the warrant associated with an assertion or hypothesis. Specifically, the factors are:

- **Comprehensiveness:** Considers the extent of the coverage of the problem space that has or will be explored as an indicator of the coverage of breadth and depth of understanding attainable or that it is assessed could be attainable. It considers the degree to which uncertainties and errors have been or will be assessed and the extent to which coverage of the problem has allowed or will allow the system, its behaviours and its outputs to be understood.
- **Relevance:** Considers evidence drawn from a range of potential sources, e.g. previous studies, literature, data and assumptions and considers their relevance for informing the findings for the current problem. It also considers the extent to which sources have drawn on multiple relevant perspectives and the extent of the inferential gap between assumptions and findings.
- **Objectivity:** Considers the extent to which sources have been challenged and peer-reviewed by the study team prior to wider exposure or socialisation of the findings with customers. These aspects help to determine the extent to which the sources can be relied upon and how much challenge has been given to the findings.
- **Quantity:** Considers the number and variety of sources for generating the evidence, i.e. the methods employed as part of a balanced approach to the generation of evidence or if this is not appropriate the extent of the track record for a particular method where variety is limited or unnecessary. For the former this factor takes into account the number, scale and variety of approaches that have been used to tackle the problem. Where quantity is less of an issue, e.g. it is obvious how to proceed and 'best-practice' is available it considers the extent of the track record for 'best-practice' methods producing evidence. It is not necessary to have a large quantity of sources to score highly if there is a track record of appropriate use.
- **Consistency:** Considers the extent of agreement between multiple sources of evidence and the trends or patterns evident. The terms 'trend' or 'pattern' refer to the broad conclusions drawn. They also refer to the extent of the support for the evidence and what can be said about cause and effect. If multiple sources of evidence on which to base observations on trends and patterns are not available it considers the extent to which support has been tested to judge the level attained, i.e. have alternative accounts for the findings been properly considered?

Each of the statements within a level can be used to determine which cell in the EPT mostly characterises the assertion under consideration. It is not necessary to match every statement within a cell and that assertion which most characterises the situation of interest should be used.

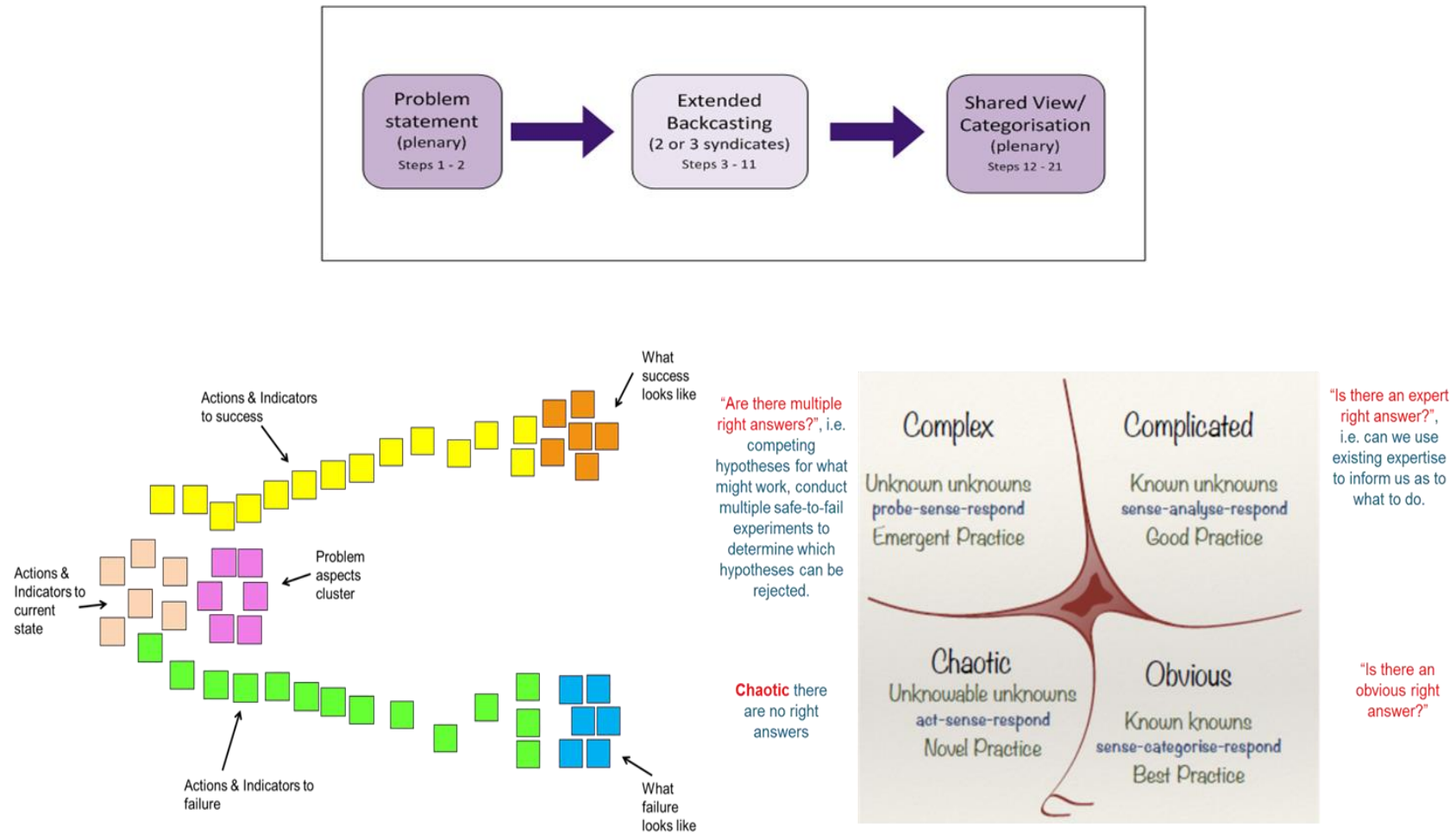


Figure 1 Methods Used Within the Initial Analysis Estimate

UK OFFICIAL

Comprehensiveness Considers the extent of the problem space that has or will be explored for the system under study as an indicator of the breadth and depth of coverage and understanding attainable.	Relevance Considers the relevance of evidence (e.g. source studies, literature, data) and assumptions informing the findings for the problem currently being considered.	Objectivity Considers the extent to which sources of evidence have been peer-reviewed and independent challenge sought.	Quantity Considers the number and variety of sources as part of a balanced approach to the generation of evidence or the extent of the track record where variety is limited or unnecessary.	Consistency Considers the extent of agreement between multiple sources of evidence (trends/patterns) and how it supports the findings and/or the extent to which alternative accounts for the findings are explored.	Profile Level
<p>All key aspects of the problem and related uncertainties have been or will be explored.</p> <p>System outputs and all internal behaviour of the system can be described. All important processes in the system can be explained.</p> <p>Full or partial control can be exercised under normal circumstances and some system behaviour can be predicted or controlled under unusual conditions. This equates to a 'known knowns' perspective on the problem.</p>	<p>Evidence used to inform the findings draws from an extensive number of sources. These provide multiple relevant perspectives for understanding the wider context of the problem.</p> <p>Changes to relevant assumptions which could drive the findings have no impact.</p> <p>There is assessed to be a very small inferential gap between material and findings for the current problem.</p>	<p>Sources of evidence that inform the findings drawn have had extensive challenge.</p> <p>Review and scrutiny has been external to the study programme domain. For example, from across the wider Department or relevant national or international organisations.</p> <p>Relevant caveats and assumptions have been clearly stated. They do not limit the utility of the work for its stated purpose.</p>	<p>The problem is complicated or complex, evidence has or will be drawn from a multi-method approach. This is through the extensive use of combinations of 'hard' and 'soft' methods. These provide multiple lines of enquiry to elicit multiple perspectives.</p> <p>Alternatively, the problem is well understood. Evidence has or will be drawn from a single or limited method approach. This is considered best practice with an extensive track record for addressing problems of this type.</p>	<p>Sources of evidence that inform the findings drawn show strong agreement on trends or patterns across all or the majority of methods employed.</p> <p>There is strong direct support and indirect support for the findings. All relevant alternative accounts and views for the findings have been addressed and eliminated.</p> <p>In terms of cause and effect it is possible to say that A causes B.</p>	1
<p>The majority of the key aspects and related uncertainties have been or will be explored.</p> <p>System outputs and some internal behaviour of the system can be described. Some or all important processes in the system can be explained. Some changes in output or behaviour can be predicted for a limited time.</p> <p>Full or partial control can be exercised under normal circumstances. This equates to a 'known unknowns' perspective on the problem.</p>	<p>Evidence used to inform the findings draws from a good number of sources. These have some relevant perspectives for understanding the wider context of the problem.</p> <p>Changes to relevant assumptions which could drive the findings have some but no significant impact.</p> <p>There is assessed to be a small inferential gap between material and findings for the current problem.</p>	<p>Sources of evidence that inform the findings drawn have had a good level of challenge.</p> <p>Review and scrutiny has been external to the study programme domain. For example, from across other programme domains but not in the wider Department.</p> <p>Relevant caveats and assumptions have been clearly stated. To some extent they limit the utility of the work for its stated purpose.</p>	<p>The problem is complicated or complex, evidence has or will be drawn from a multi-method approach. This is through a good but limited use of combinations of 'hard' and 'soft' methods. These provide alternative lines of enquiry to elicit a variety of perspectives.</p> <p>Alternatively the problem is quite well understood. Evidence has or will be drawn from a single or limited method approach. This is considered good practice with a good track record for addressing problems of this type.</p>	<p>Sources of evidence that inform the findings drawn show moderate agreement on trends or patterns across all or the majority of methods employed.</p> <p>There is moderate direct support and indirect support for the findings. Salient alternative accounts and some non-salient accounts and views for the findings have been addressed and eliminated.</p> <p>In terms of cause and effect it is possible to say that A is very likely to cause B.</p>	2
<p>There are some key aspects and related uncertainties that have not or will not be explored.</p> <p>The nature of the problem space may be considered complex such that aspects are not easily explored. System outputs or some relationships, possibly correlations, between inputs and outputs can be described.</p> <p>Prediction is based on a continuing assumption of outputs being correlated to inputs. Reliable control is not possible. This equates to an 'unknown unknowns' perspective on the problem.</p>	<p>Evidence used to inform the findings draws from a limited number of sources. These provide a limited number of perspectives for exploring the wider context of the problem.</p> <p>Changes to some of the relevant assumptions that could drive the findings have a significant impact.</p> <p>There is assessed to be a large inferential gap between material and findings but it is asserted there is no doubt as to the value of their contribution for the current problem.</p>	<p>Sources of evidence that inform the findings drawn have had some but limited challenge.</p> <p>Review and scrutiny has been external to the study team but within the study programme domain.</p> <p>Relevant caveats and assumptions have been clearly stated. These largely limit the utility of the work for its stated purposes.</p>	<p>The problem is complicated or complex, evidence has or will be drawn from a single method approach. This is through a good but limited use of combinations of techniques within the set of 'hard' or 'soft' methods. This provides few alternative lines of enquiry reducing the variety of perspectives.</p> <p>Alternatively the problem is well or quite well understood. Evidence has or will be drawn from a single or limited method approach with a limited track record addressing problems of this type.</p>	<p>Sources of evidence that inform the findings drawn show some agreement in trends or patterns across all or the majority of methods employed.</p> <p>There is some direct support and indirect support for the findings. Most salient alternative accounts and some non-salient accounts and views for the findings have been addressed and eliminated. Some alternative accounts remain that could support the findings.</p> <p>The problem space is such that cause and effect is difficult to disentangle or multiple competing hypotheses may exist. It is possible to say that A may well cause B.</p>	3
<p>The majority of the key aspects and related uncertainties have not or will not be explored. The nature of the problem space may be chaotic meaning that aspects are difficult to explore.</p> <p>It is very difficult to explain or predict system behaviour and control is not possible.</p> <p>Understanding is absent or very limited and equates to an 'unknowable unknowns' perspective on the problem.</p>	<p>Evidence used to inform the findings draws from a very limited number of sources. These provide a very limited number of perspectives for exploring the wider problem context.</p> <p>Changes to most of the relevant assumptions that could drive the findings have a significant impact.</p> <p>There is assessed to be a very large inferential gap between material and findings and the current problem such that there is significant doubt as to the value of their contribution.</p>	<p>Sources of evidence that inform the findings drawn have had little and very limited challenge.</p> <p>Review and scrutiny has been within the study project team.</p> <p>Relevant caveats and assumptions have not all been clearly stated. This greatly limits the utility of the work for its stated purposes.</p>	<p>The problem is complicated or complex, evidence has or will be drawn from a very limited use of a technique within the set of 'hard' or 'soft' methods. This provides no alternative lines of enquiry and no variety of perspectives.</p> <p>Alternatively the problem is well or quite well understood. Evidence has or will be drawn from a single or limited method approach with no track record addressing problems of this type.</p>	<p>Sources of evidence that inform the findings drawn show little or no agreement in trends or patterns across all or the majority of methods employed.</p> <p>There is little or no direct support or indirect support for the findings. Only some alternative accounts and views have been eliminated. Differently founded accounts are also assessed to have particular merit.</p> <p>The problem space is such that discernible patterns or concepts of system behaviour are difficult to establish or do not exist. In terms of cause and effect it is possible to say that A might cause B.</p>	4

Table 1 Evidence Profile Table V3.0

The Validation Profile Table (VPT)

The VPT, illustrated in Table 2~~Error! Reference source not found.~~, draws heavily on work (21) undertaken in support of the Aqua Book and is designed for use in assessing or evaluating the analytical validity of a body of evidence to inform decision making, i.e. how valid were the methods and the representations of the aspects being studied and measured. The VPT allows a judgement to be made regarding the extent to which the right work is being or has been engaged in, given the purpose and constraints placed upon that work. The key output from the validation process is a judgment concerning the extent to which the work is valid as part of the fitness-for-purpose judgment. It is important to note that a statement of validity should, where possible, be a judgement drawing on the multiple perspectives and views from outside the study team. Where this is not possible the study team is encouraged to form an opinion on validity to inform statements of confidence but the power of this axis is in the wider involvement of the stakeholders. It should not be confused with the objectivity factor within the EPT.

The Validity construct is used to judge the reliability of the warranted evidence using four key validity criteria, Face Validity, Criterion Validity, Construct Validity and Content Validity (21). The VPT is used to assign a level between one and four to each of the four validity factors (see below) that are considered to be generic characteristics of validity. The rationale for the scores is the same as that for the EPT. The resulting validity profile is simply summed⁶ to arrive at a statement concerning evidence validity. The validity together with the warrant can then be used to understand or assess the overall evidence position required or achieved for the assertion being made. Both are used as indicators of fitness-for-purpose and to estimate a position within the CAT to determine a confidence level.

VPT factors – What are they and what do they mean?

The VPT factors are considered generic validity characteristics and are used to structure a conversation on evidence validity and to assign a profile that can be used to judge the extent of the validity. They help understand the strengths and limitations of the analytical approaches (21,26). Specifically, the factors are:

- **Face Validity:** Considers the degree to which the key stakeholders believe there to be an adequate alignment between the characterisation of the issues examined or being examined in the analysis and their understanding of the ‘problem space’. This is essentially about considerations of whether the analysis has engaged with what it purports to have engaged with.
- **Criterion Validity:** Considers the detailed engagement with the issues being examined in the analysis and the extent to which the work actually engages with the issues that it claims to. This is about considering the extent to which the analysis has engaged directly with the relevant variables of interest or if it has used appropriate surrogates.

⁶ As per the EPT currently all the VPT factors are considered to be equally weighted. Further research will be required to determine if this assumption is valid.

- **Construct Validity:** Considers the adequacy (for the purposes of this analysis) of the representation of how the issues being examined are structured. This includes the key factors to which they respond and the mechanisms by which they do this. This is about considering if the analysis has understood and assessed what it purports to have assessed.
- **Content Validity:** Considers the interpretative weight that the work proposed can bear, as a result of its breadth, depth and granularity. This is about considering if the analysis has measured and assessed the relevant aspects at the required level of granularity.

Each of the statements within a level can be used to determine which cell mostly characterises the validity of the assertion under consideration.

The Confidence Assessment Table (CAT)

Whilst the EPT assessment will result in an evidence profile and the VPT assessment a validity profile there is often a need to express this in more simplistic terms and a need to understand the confidence in the findings. This is achieved by using the CAT, Table 3, to assess two criteria, the extent of warrant inferred from the evidence profile and the extent of the validity across the community of interest. Both can be used to make a qualitative judgement about the confidence according to likely confidence bands. The confidence scale⁷ is “Very Low, Low, Medium, High and Very High”.

The summed EPT and VPT assessment scores are used to position the findings along the warrantability axis and the validity axis of the CAT. There is a general rule of thumb beneath each of the warrant criteria and adjacent to each of the validity criteria which provides a more informative statement about any judgement drawn in relation to the findings.

The purpose of the CAT is to take the resulting profile scores and express the target or achieved levels of confidence. It is important to note that for assessing achieved confidence it is possible for the study team to make a judgement as part of a review activity but dialogue with all stakeholders and ‘consumers’ of the evidence is essential to meaningfully judge the validity. Validity and hence confidence is a social construction and is likely to be fluid, i.e. there will be a number of perspectives, views, issues, etc. that need to be considered when seeking to make or understand a judgement about the agreement on the findings, many of which may not be visible to the study team, and its utility. Warrant on the other hand is likely to be fairly stable.

A study can improve the level of warrant and validity by considering the characteristics required within the EPT and VPT and this can be used in dialogue with stakeholders to determine any next steps in the analysis.

⁷ Note, that the confidence shading in the CAT is conceptual in nature to illustrate that boundaries are inherently fuzzy. The greyscale shading is deliberate to avoid any association with the more traditional use of red, amber, green (RAG) type shading with being bad, ok or good as this is about fitness-for-purpose judgements. In addition confidence should not be confused with probability ratings hence there is no quantitative expression of confidence.

Face Validity Considers the extent of the alignment between the issues being examined and how they are characterised in the analysis with the understanding of the 'problem space'. Has the analysis approach engaged or will it engage with the aspects of the problem required by the customer?	Criterion Validity Considers the detailed engagement with the issues within the 'problem space' and the extent to which the analysis actually engages with the issues that it claims to. Is there or will there be good alignment between the things being measured and the things being studied?	Construct Validity Considers how adequate the representation of the issues are within the 'problem space' their structure, the key factors to which they respond and the mechanisms by which they do this. Has the analysis been structured or will it be structured around appropriate concepts relevant to the problem at hand?	Content Validity Considers the interpretative weight that the work undertaken or proposed can bear, as a result of its breadth, depth and granularity. Has the analysis considered or will it consider all relevant aspects for the intended purpose and measured or assessed what it purports to have measured or assessed?	Profile Level
The analysis is thoroughly aligned with the issues and is well structured for the purpose. Relationships to prior knowledge are easily recognised.	Actual variables associated with the issues under study have been directly used in the analysis.	The structure of the functional relationships between issues represented are well described and sufficient for the purpose.	The analysis delivers a thorough understanding of the relevant issues and drivers.	1
The analysis shows good alignment with the issues and is adequately structured for the purpose. Relationships to prior knowledge can be argued for.	Surrogate variables associated with the issues under study have been used and are assessed as being adequate for the purpose.	The structure of the functional relationships between issues represented is adequately described and assessed as being appropriate but not necessarily sufficient for the purpose.	The analysis delivers a good understanding of the relevant issues and drivers.	2
The analysis shows some, but limited alignment with the issues with some concerns about the suitability of the alignment. However, the analysis is largely adequate for the purpose. Relationships to prior knowledge that can be argued for.	Surrogate variables associated with the issues under study have been used extensively but there are concerns about their suitability for the purpose.	The structure of the functional relationships between the issues has been simplistically described but there are concerns about their sufficiency and suitability for the purpose.	The analysis delivers some, but limited, understanding of the relevant issues and drivers.	3
The analysis provides little or no alignment with the issues or prior knowledge and is not adequate for the purpose. Relationships to prior knowledge cannot be argued for.	The linkage between the issues under study and surrogate variables has not been demonstrated hence they are not currently recognised as being suitable for the purpose.	The structure of the functional relationships between the issues represented are inadequately described and are not seen as being suitable for the purpose.	The analysis delivers little or no understanding of the relevant issues and drivers.	4

Table 2 Validation Profile Table V1.0

				Confidence in Findings (Very Low to Very High)																			
				Indicates the confidence band for the findings based on the extent of the Warranty and Agreement.																			
Validity of Findings <i>Considers the fitness-for-purpose of the findings through consideration of Face, Content, Construct and Criteria validity.</i>	The findings are accepted as valid and have significant benefit.	High	04																				
	The findings are accepted as valid and have appreciable benefit.	Strong	05 06 07 08																				
	The findings are accepted as valid and have some or limited benefit.	Moderate	09 10 11 12																				
	The findings are not accepted as valid and have little or no recognised benefit.	Weak	13 14 15 16																				
How to assess Warrant and Validity: All criteria in the EPT and the VPT are equally weighted. Therefore take the sum of the Evidence Profile as an indication of Warrant and the sum of the Validity Profile as an indication of Validity.				20	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05				
				Weak				Moderate				Strong				Proof/BRD							
				Further analysis is likely to change the findings				Further analysis may change the findings				Further analysis is unlikely to change the findings				Further analysis will not change the findings							
				Warrant for Findings (BRD = Beyond Reasonable Doubt) <i>Considers the degree of inference for the findings as a result of the Evidence Profile generated</i>																			

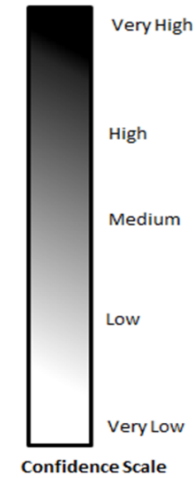


Table 3 Confidence Assessment Table V2.0

The Evidence Quality Questionnaires (EQQs)

There are many different aspects of quality to consider during a study, which can be grouped into three categories; quality of content, quality of process and quality of outcome (21). The EQQs provide a structured framework for the Analyst and Analytical Assurer, to help keep the study process 'on track' in terms of quality, making sure that the analytical team are thinking about the right things in order to generate good quality evidence. The structure and criteria allow users to accurately assess the quality of the evidence generation process throughout the life of a study and identify areas that might affect the quality of the evidence.

There are three separate EQQs, illustrated in Table 4, with a particular questionnaire selected according to the study phase underway, i.e. the Design/Plan phase which has 30 questions; the Execute phase which also has 30 questions; and the Analyse/Exploit phase which has 51 questions. Each questionnaire has six sections structured around the mnemonic RIGOUR (16). The Analyst would consider each statement within a section in turn, deciding for each the extent to which it is true for this particular study, choosing the response that best reflects this judgement.

When using any one EQQ to assess evidence quality for a study, the Analyst reviews the appropriate questionnaire for the current point in the study. This may be done on an individual basis or by the study team as a group to develop shared awareness of the project. Within the LEOA project, for example, the Execute phase EQQ was assessed every six months, took approximately two hours to complete and where appropriate resulted in actions to address any issues emerging. The benefits of an EQQ are gained through the process of completing it; it is not designed to be used to measure projects against one another. In addition assessments support later deliberations regarding EPT and VPT discussions.

			Strongly Agree	Neutral	Strongly Disagree	N/A	Comments
Repeatability	1.	The technique(s) chosen is/are repeatable; that is, for the "same" inputs and constraints, the analysis can reasonably be expected to produce the "same" outputs.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	2.	There are no factors/constraints in the study that prevent or inhibit the repeatability of the technique(s), e.g. resource constraints.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	3.	The study process is clearly defined and methodical.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	4.	Appropriate design documentation has been produced, which may include a concept of analysis, user requirements, design specification, functional specification, data dictionary, and test plan.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
Independence	5.	The design phases involved a range of appropriate military (and wider) stakeholders to ensure the requirement is properly defined and communicated.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	6.	The study team consists of a representative mix of military/scientific specialists, SME members and stakeholders.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	7.	Any potential sources of prejudice have been identified.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	8.	Evidence is being collected from relevant scientific literature review.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	9.	The approach involves a sufficient range of analytical techniques.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	
	10.	A sufficient number of studies will contribute to the conclusions drawn.	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/>	

Table 4 Extract of the Design Phase EQQ

Example Application of the EPT

An example from the LEOA project of applying the EPT is given below in Figure 2. This shows assertions at the study level for which the rationale was communicated through the concept of analysis (COA), for one of the key methods used within the study and for insights produced by this method.

Detail		Comp	Rel	Obj	Qty	Cons	Total	Warrant
Study	COA	4	3	2	3	2	14	Moderate
Method	Wargame 1	2	3	3	2	2	12	Moderate
Insights	Insight 1	2	2	2	2	2	10	Strong
	Insight 2	3	3	3	3	3	15	Moderate
	Insight 3	3	2	4	4	3	16	Weak

Figure 2 Example Evaluation of Evidence Warrant

Essentially the assertion at the study level is that the study will be able to produce evidence of Moderate warrant, i.e. further analysis may change the findings. Time and resources did not permit a comprehensive coverage of the problem space but a focus on drawing together previous research findings into a capping paper was assessed as being likely to support trends and patterns that it was hypothesised might emerge out of the specific focused experiments to be undertaken within this study. Drawing on relevant expertise to challenge and improve the objectivity of the analysis would also support this level of warrant.

One of the methods being used was a wargame that could also be run as a fully constructive simulation post game. It was asserted that the method would also be able to produce evidence of Moderate warrant. Within the questions that provided the focus of the wargame the coverage of relevant factors was assessed as being comprehensive, the method was considered good practice and having a good track record for this type of problem. It was also assessed that the method could provide strong direct support and indirect support for the findings through further analysis from the output of the game. In addition the use of the constructive simulation mode would make it possible to assert that A is very likely to cause B for some of the insights.

Three of the insights that emerged from the gaming are used to illustrate the range of warrants. Following any wargaming event for LEOA it is common practice to produce a short headline paper. A workshop is run as part of the process of reviewing the paper with one of the aims to discuss each insight in more detail, its evidence profile and to assign a warrant to each of the insights. The associated strong warrant for Insight 1 reflects the fact that this insight has arisen in other studies and in other scenarios and that further analysis is unlikely to change the findings. Insight 2 reflects an outcome that has emerged over several LEOA activities and Insight 3 reflects a new or emerging insight that has not been seen before. Discussions with customers on how to improve the level of warrant for Insight 2 and in particular Insight 3 using the EPT were held to shape follow-on analysis activities.

The time taken to undertake these assessments varies with much depending on the availability of evidence from previous studies or the knowledge of those undertaking the assessments. To classify the key insights that were put forwards in the headlines paper, of which there were ten, took a morning of discussion. The insights reported by the LEOA study have stated the warrant explicitly and this has been well received by customers. In particular, the warrant for Insight 1 has resulted in the customer confirming that they now have sufficient evidence to justify further action and that no further analysis will be necessary to explore this issue.

The utility of this aspect of the EFA has been successfully demonstrated and as a result the EFA as a whole will be incorporated into a later version of the Army's Land Handbook for Force Development Analysis and Experimentation.

Conclusions

This paper has given an overview of the EFA and introduced some of the key concepts and ideas. Take up of the EFA as a whole has been modest but the EFA is still in its infancy and is still developing. Particular aspects of the EFA, such as the EPT have become well established within the LEOA project and well received by all project stakeholders. The VPT and CAT are new additions to the EFA portfolio and practical experience in their application is required to determine if they will have similar utility.

It is believed that the application of the EFA will benefit those applying it by helping stakeholders participate in conversations about evidence which enhances understanding. There is little doubt that there is a continuing need for a means of evaluating and assessing evidence and for supporting assessment of the fitness-for-purpose of an analytical process to address a problem. The EFA helps with this, is consistent with the good practice expressed in the Aqua Book and is consistent with the thinking in similar fields of research.

It is posited that routine use of the EFA as part of 'business as usual' analytical quality assurance will enable operational analysis practitioners to improve the robustness of evidence as part of informed decision support activities.

References

1. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/27408/defence_reform_report_struct_mgt_mod_27june2011.pdf - accessed 13th April 2016.
2. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/206946/review_of_qa_of_govt_analytical_models_final_report_040313.pdf - accessed 13 April 2016.
3. Chilcot, J., (2016) "The Report of the Iraq Inquiry Executive Summary" <https://www.gov.uk/government/publications/the-report-of-the-iraq-inquiry> - accessed 21 September 2016.
4. Curram, S., Miles, O., Exelby, D., (2014) "Review of Proposed Approach for a Land Force Development Operational Analysis Evidence Framework - Task 41 – Peer Review of Land Force Development Operational Analysis Evidence Framework", Report No. FATS/4/RED/TA0001-TASK 41, Version 1.0, 24 October 2014.
5. Curram, S., Miles, O., Exelby, D., (2014) "Appendix – Literature Review Summaries for Review of Proposed Approach for a Land Force Development Operational Analysis Evidence Framework - Task 41 – Peer Review of Land Force Development Operational Analysis Evidence Framework", Report No. FATS/4/RED/TA0001-TASK 41, Version 1.0, 24 October 2014.
6. Leggatt, A., Geer, S., (2015) "OAC Task 55 Evidence Framework Approach Trials Report", CR2710 055/TR/2.
7. Burdett, S., Purkess, R., (2015) "OAC Task 55 Evidence Framework Approach Trials Report", CR2710 055/TR/3.
8. Burdett, S., Purkess, R., (2015) "OA Collaboration Task 55 Evidence Framework Approach Guidance Report - Sensemaking & evidence quality assessment in Land Force Development Analysis & Experimentation Projects", CR2710 055/TR/3 v1.0.
9. Pearce, P.V., (2015) "The Evidence Framework", Defence Science Interest Group Meeting, June 2015, DSTL/PUB89554.
10. Pearce, P.V., (2016) "The Evidence Framework Approach: Supporting Evidence Based Decision Making within the UK MOD", Evidence Synthesis by Building a Case (Workshop 2): Amalgamation and the Principle of Total Evidence, 3rd May 2016, Durham University, DSTL/PUB95406.
11. Pearce, P.V., (2016) "The Evidence Framework Approach: Supporting Evidence Based Decision Making within the UK MOD", 84th MORS Symposium, Virginia, USA, 20-22 June 2016, DSTL/CP96100.
12. Pearce, P.V., (2016) "The Evidence Framework Approach: Supporting Evidence Based Decision Making Within MOD", GORS Conference 2016, UK, October 2016.
13. Pearce, P.V., Beaves, N. (2016) "The Evidence Framework Approach: Supporting Evidence Based Decision Making within the UK MOD", DORS Conference, Canberra,

UK OFFICIAL

Australia, 15-18 November 2016,

14. Leggatt, A., Turner, N., Pearce, P.V. (2017) "Planning Studies: Empirical investigation of using team sensemaking for early problem formulation", 13th International Conference on Naturalistic Decision Making 2017, Bath, UK.
15. Mathieson, G., (2000) "Best Practice for using Assessment Hierarchies in Operational Analysis – Principles and Practical Experiences", ICCRTS http://www.dodccrp.org/events/2000_CCRTS/html/pdf_papers/Track_5/025.pdf - accessed 22 September 2016.
16. <https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government> - Accessed 12 April 2016.
17. Jordan, C., (2017) "Code of Best Practice for Warfighting Experimentation", NW/CR/EXP/1924, v4.0, May 2017.
18. Snowden, D., Boone, M., (2007) "A leader's framework for decision making", Harvard Business Review, 85(11), 68, 2007.
19. Solly, R., (2015) "Expressing Deep Uncertainty as Levels of Understanding", DSTL/PUB91926, Deep Uncertainty Conference, 03 November 2015.
20. Solly, R., (2016) "How Well do you Understand your Problem?", GORS Conference, 25 January 2016.
21. Glover, P., (2014) "Verification and Validation for the AQuA Book", DSTL/TR81015, 01 September 2014.
22. IPCC (2010) "Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties", IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties, Jasper Ridge, CA, USA, 6-7 July 2010.
23. Gore, J., Conway, G.E., (2016) "Modeling and Aiding Intuition in Organizational Decision Making: A Call for Bridging Academia and Practice", Journal of Applied Research in Memory and Cognition 5 (2016) 331-334.
24. Mingers, J., (2014) "Systems Thinking, Critical Realism and Philosophy: A confluence of ideas", Routledge, 2014.
25. Reiss, J., (2015) "Causation, Evidence, and Inference", Routledge, 2015.
26. Cartwright, N., (1997) "Where Do Laws of Nature Come From" Dialectica Vol 51, No1 pp 65-78