

The Evidence Framework Approach:

**Analyst's Note for Applying the EPT, VPT and
CAT**

Introduction

Reference AN1 provides a full technical description of the Evidence Framework Approach (EFA) which is consistent with the UK pan-Government-Department Aqua Book [Reference AN2] which was published in early 2015 as guidance on accomplishing analysis with evidence that is 'fit-for-purpose', i.e. is of appropriate quality. It uses specific descriptors for the key stakeholders referring to the Commissioner, the Analyst and the Analytical Assurer. The first two roles map on to the more generic roles of Customer and Analyst. The third role is that of quality assurance for the study.

This document provides guidance on applying the Evidence Profile Table (EPT), the Validation Profile Table (VPT) and the Confidence Assessment Table (CAT) as part of an evidence evaluation and assessment activity. Materials to support activities within this document are available via the Dstl [SharePoint site](#) or on CatalyST.

The EPT, VPT and the CAT can be used during problem formulation to facilitate the discussion between Commissioner, Analyst and Analytical Assurer around the broad level of evidence expected from or achieved by a study or artefact to be assessed and the level of validity, all are used as part of determining appropriate 'fitness for purpose'. By providing a structured basis for this conversation, it helps to align expectations so that the study generates evidence that meets the needs of the Commissioner and is achievable within the means available to the Analyst. The EPT, VPT and CAT can be used at regular intervals throughout a project.

This Analyst's Note will help you apply the EPT, VPT and CAT. While this may appear to be quite a time consuming task the more practice you have in applying the EPT, VPT and CAT the better. Most assessments are able to be undertaken within a few minutes and add real value to the analytical product.

EPT Introduction

The EPT is designed for use 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. The resulting evidence profile is summed to arrive at a statement concerning evidence warrantability. It is important to note that 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 validity profile 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 (see AN1) and are used as handrails around which to structure a conversation on evidence and to assign a profile that can be used to judge the extent of the warrant we can associate with it. Specifically, the factors are:

- **Comprehensiveness:** This factor or evidence characteristic considers the extent of the problem space that has or will be explored for the system under study as an indicator of the depth of understanding attainable or that could be attainable. It considers the extent of the problem space explored or to be explored and indicates the ‘completeness’ or coverage of the problem space and the degree to which uncertainties and errors have been or will be assessed. It considers the extent to which coverage of the problem has allowed or will allow the system, its behaviours and its outputs to be understood.
- **Relevance:** This factor considers evidence drawn from a range of potential sources, e.g. previous studies, literature, data and assumptions and considers their relevance for informing the findings¹ for the current problem. It also considers the extent to which sources have drawn on multiple relevant perspectives and the extent of any inferential gap between assumptions and findings.
- **Objectivity:** This factor considers the extent to which sources have been challenged and peer-reviewed prior to ‘publication’ by the study team. 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. Note that this is peer-review and challenge prior to wider socialisation of the findings with customers.
- **Quantity:** This factor considers the number and variety of sources² 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 consider 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 use.
- **Consistency:** This factor refers to the extent of agreement between multiple sources of evidence. The ‘trend’ or ‘pattern’ refers to the broad conclusions drawn. It also refers 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 consider the extent to which support has been tested to judge the level attained, i.e. have alternative accounts for the findings been discounted?

¹ The terms findings is used to cover many relevant aspects, e.g. results, conclusions, recommendations etc.

² Reference is made to ‘Hard’ and ‘Soft’ methods in the EPT. The Hard methods typically are associated with quantitative and objective data in the main. The Soft methods typically produce qualitative and subjective data in the main.

EPT factors - How do I select a level?

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

VPT Introduction

The VPT is designed for use in assessing or evaluating the validity of a body of evidence to inform decision making and is based on the work within AN3, i.e. *“validation is literally to make valid, through the agreement of those judged competent to take such views”*. 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 concerning the evidence is a judgement about the multiple perspectives and views from outside the study team or that which might be expected concerning the findings, methods etc. It should not be confused with the objectivity criteria within the EPT.

Validity is used to judge the reliability of the warranted evidence using four key validity criteria, Face Validity, Construct Validity, Content Validity and Criteria Validity. 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 resulting validity profile is summed 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 (see AN3) and are used as handrails around which to structure a conversation on evidence validity and to assign a profile that can be used to judge the extent of the validity we can associate with it.

- **Face Validity:** This factor 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.
- **Construct Validity:** This factor considers the adequacy (for the purposes of this analysis) of the representation of how the issues (phenomena) 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.

- **Criterion Validity:** This factor considers the detailed engagement with the issues (phenomena) being examined in the analysis and the extent to which the work actually engages with the issues (phenomena) 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.
- **Content Validity:** This factor 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.

VPT factors - How do I select a level?

Each of the statements within a level can be used to determine which cell mostly characterises the validity of the assertion under consideration. It is not necessary to match every statement within a cell and you should pick that which most characterises your assertion.

CAT Introduction

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 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 scale is “Very Low, Low, Medium, High and Very High”. Note, that the confidence shading is conceptual in nature to illustrate that boundaries are inherently fuzzy. In addition confidence should not be confused with probability ratings hence there is no quantitative expression of confidence.

To determine the level of warrant simply sum the evidence profile score resulting from the EPT assessment. To determine the level of validity simply sum the validity profile score resulting from the VPT assessment. This can be used to position the findings along the warrantability axis and the validity axis of the CAT. There is a general rule of thumb associated with each warrant or validity criteria which provides a more informative statement about judgement in relation to the findings.

The purpose of the CAT is 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 within dialogue with stakeholders to determine any next steps in the analysis.

Using the EPT, VPT and CAT to establish a requirement

At the initiation of a study, the Commissioner, Analyst and Analytical Assurer should meet to discuss the need for analysis to support the Commissioner's question or problem. In this meeting, it is helpful to set an evidence profile and validity profile requirement that describes what the profile for each could be at the end of the study.

It is important in this discussion for the Analyst to gain a good understanding of the context of the question and the purpose of the study, i.e. the decision(s) the study is to inform. You should talk through the five factors of the EPT and the four factors of the VPT, including the description against each level and what it might look like in the context of this particular study. Aim to reach an agreement about the evidence and validity profiles required or achievable to support a decision, bearing in mind that these should be based on 'fitness-for-purpose' considerations informed by the ability to understand the nature of the problem itself, the time and resources available, and the 'business criticality' or impact of the findings on the decision.

Often, the Commissioner might come to this meeting with an idea in mind for the target profile they would like for the study. This forms a useful starting point for agreed profile requirements. Where it is difficult to decide between levels it will be helpful to consider what exactly is meant by the generic descriptions in the EPT and VPT in the context of the particular study and to differentiate the levels by this means.

It is quite likely that the requirement will not be the same for all factors. Although a score of one for any factor represents the highest level that can be achieved, this level may not be suitable for any of the EPT or VPT factors for an individual study. Higher levels are likely to be obtained at a higher cost in time and resources, so there will often be a trade-off to be made. It is entirely possible that the nature of the problem itself is such that it precludes obtaining scores at the higher levels.

At this stage, you could also use the EPT and VPT to 'score' any existing body of evidence, with respect to how ready it would be to support the decision or problem without any further work being done. This requires information about relevant existing sources that can be exploited, such as the results of previous analysis, experiments, or other relevant reports.

The difference between the scored profiles and the required profiles will indicate the gap in evidence available to support the decision, and inform a discussion on how to bridge that gap.

The profiles should be scored and the results used to find the degree of warrant and validity within the CAT. This then establishes the scored and required levels of warrant and validity and can be used to state how the findings may be affected by the levels. Once the warrant and validity is established the confidence on the findings can be established as a target for the study, identifying what the study realistically or practically is able to achieve or aspire to and what this means in terms of confidence.

If undertaken at the start of an Initial Analysis Estimate it is worthwhile revisiting the profile requirements and the warrant, validity and confidence assessments at the end of the Estimate event to refine it. This provides an agreed requirement to take forward into the production of an analysis plan. It is helpful to present the plan together with an EPT, VPT and CAT forecast, which communicates the evidence profile and associated warrantability, the validity profile and associated validity and confidence that you would expect to be obtained if the Commissioner goes ahead with the proposal set out in the plan.

Using the EPT, VPT and CAT to assess a study

You may decide to use the EPT, VPT and CAT throughout the lifecycle of a project. Regular comparison between the current position and the evidence requirement will help identify where there is a risk of falling short of the requirement, and where steps should be taken to avoid these shortfalls. These should be discussed with the Commissioner at review meetings. For a more detailed quality assurance assessment, the Evidence Quality Questionnaire is available for use by the Analyst and Analytical Assurer.

One particular application is to use the EPT, VPT and CAT as a means of framing study conclusions and recommendations, using the statements within the factors to convey uncertainty concerning the study findings. In addition a statement of warrant and validity can be assigned to each individual conclusion or finding together with a confidence rating if generated as a further means of expressing the ‘fitness-for-purpose’ of the evidence. This differentiates those findings that may require further work from those that do not.

The EPT, VPT and CAT are also useful at study (project) closure, where the Commissioner and Analyst meet to reassess the body of evidence, and identify the degree to which the study met the requirements initially anticipated for it. The descriptions of each level in the tables, the warrant, validity and confidence provide a basis for discussion around these points, so that both parties have a clear means to articulate aspects of the study’s conduct and findings, and why the study progressed in the way it did. It is especially important that the Commissioner fully understands the limitations and caveats on the results of the analysis so that they are interpreted correctly. The EPT, VPT and CAT form an important part of the quality assurance of the study, and their use during the course of the study will provide a strong mitigation against the study’s findings being rejected by the Commissioner at any stage, particularly at its conclusion.

References

- AN1. The Evidence Framework Approach: Practical Ways of Thinking About Evidence, Pearce, PV, DSTL/TR098527 1.0, 31 October 2016.
- AN2. The Aqua Book: Guidance on producing quality analysis for government. HM Treasury, 2015.
- AN3. Verification and Validation for the AQuA Book, Glover, P, DSTL/TR81015, 01 September 2014.