



# **A Common M&S Credibility Criteria-Set Supports Multiple Problem Domains**

**J.P. Hale**

**Constellation Program Engineering Directorate**

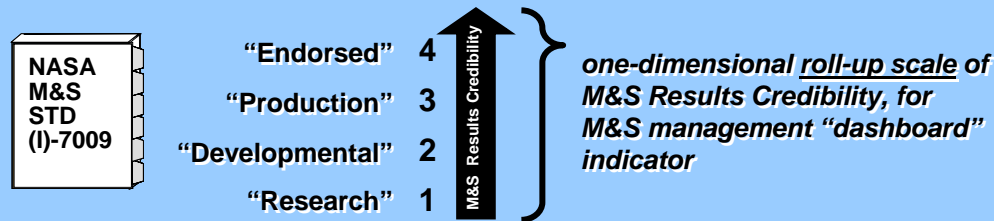
**NASA Marshall Space Flight Center  
Huntsville, AL**

- ❑ NASA Standard for M&S, STD(I)-7009, is an interim standard, undergoing review
- ❑ STD(I)-7009 Appendices A2 and A3 describe initial concepts for M&S Results credibility evaluation
- ❑ A NASA Technical Working Group has been formed with a charter to finalize this work
- ❑ This brief describes some recent thinking in support of that activity

- ❑ Reference: STD(I)-7009 Top-Level Scale for M&S-Results Credibility
- ❑ Define: New, Augmented-AHP Method for Credibility Assessment
- ❑ Reference: Credibility Criteria-Set in STD(I)-7009 Appendix A3
- ❑ Apply: New, Augmented-AHP Method to STD(I)-7009 Appendix A3

# NASA STD-(I)-7009 Definitions for Levels of M&S Results Credibility

## "Levels" of M&S Results Credibility



### **Level 4 – Rigorous - Credibility rigorously established for the current application.**

M&S results were extensively tested and reviewed for the current application. Uncertainty distributions were calculated. Expert users and analysts used formally reviewed and controlled processes. The limitations of M&S were well documented and correct usage was enforced.

### **Level 3 – Production - Credibility tested specifically for the current application.**

M&S results were tested and reviewed for the current application. Quantitative uncertainty bounds were calculated. The users and analysts performed simulations using formal best practices. The limitations of M&S were well documented and any violations were reported.

### **Level 2 – Development - Credibility established for simulation process.**

M&S results were tested and reviewed for problems similar to the current application. Uncertainty was estimated by expert opinion based on the current application. The users and analysts performed simulations using informal repeatable processes. The limitations of the M&S were well documented.

### **Level 1 – Research - Credibility established for model basics.**

M&S results were tested and reviewed for research problems. Uncertainty was estimated by expert opinion based on research problems. The simulations were performed using ad hoc processes. The limitations of the M&S were not well documented



# Common “Measurement Scale” for Augmented-AHP

- Indicates Desired Level or “Threshold Value” -



Accreditation Agent assigns the Acceptability Threshold for this criterion for the Problem at hand

**What is the Desired Level ?**

**Threshold** →

**Criterion Scale**




Common Measurement Scale across all criteria: Indicates Threshold Value, Achieved Score, and Deficiency Flag

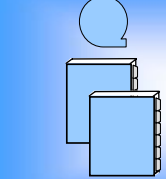
The Threshold Value can provide guidance for the type and amount of evidence to be collected and assessed from M&S artifacts

assessed achievement value

Criterion Data



**Developers**  
Gather Criterion Evidence



**M&S Artifacts**

- M&S Reqts
- M&S Designs
- M&S Code
- M&S Data
- M&S Use

**Collected and Quantified Criterion Evidence from M&S Artifact Data**



Accreditation Agent assigns the Acceptability Threshold for this criterion for the Problem at hand

What is the Desired Level ?

Threshold

Criterion Scale

Common Measurement Scale across all criteria: Indicates Threshold Value, Achieved Score, and Deficiency Flag



V&V Agent assesses M&S criterion achievement value data and determines scores

Criterion Score

What is the Achieved Value ?

assessed achievement value

Criterion Data

Developers Gather Criterion Evidence

M&S Artifacts

- M&S Reqts
- M&S Designs
- M&S Code
- M&S Data
- M&S Use

Collected and Quantified Criterion Evidence from M&S Artifact Data

# Common "Measurement Scale" for Augmented-AHP

- Indicates Deficiency Flags -

Accreditation Agent assigns the Acceptability Threshold for this criterion for the Problem at hand

**What is the Desired Level ?**

**Threshold**

**Criterion Scale**

**Common Measurement Scale across all criteria: Indicates Threshold Value, Achieved Score, and Deficiency Flag**



**Deficiency "Flag"**

**Criterion Score**



V&V Agent assesses M&S criterion achievement value data and determines scores

**What is the Achieved Value ?**

*assessed achievement value*

**Deficiency "Flag" Determination**

**Threshold** - **Score** = Deficiency Flag Type

- OK to use**  
 $0.0 \leq \text{Gap} \leq 0.5$
- Use with Caution**  
 $0.5 < \text{Gap} \leq 1$
- Do Not Use**  
 $\text{Gap} > 1$

**Criterion Data**

**Developers Gather Criterion Evidence**

**M&S Artifacts**

- M&S Reqts
- M&S Designs
- M&S Code
- M&S Data
- M&S Use

**Collected and Quantified Criterion Evidence from M&S Artifact Data**

# Common "Measurement Scale" for Augmented-AHP

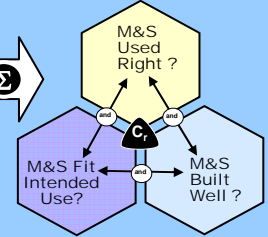
- Uses AHP Weighting Factors to Calculate Aggregate Scores -

Accreditation Agent assigns the **Criterion Weighting** of this criteria compared to other criteria within this criteria domain

## Criterion Scale

**Common Measurement Scale across all criteria: Indicates Threshold Value, Achieved Score, and Deficiency Flag**

$$\text{Criterion Weight} \times \text{Criterion Score} = \text{Weighted Score} \quad (\text{Domain Sub-Scores})$$



**How Important is this Criterion ?**

Threshold

**What is the Desired Level ?**

Accreditation Agent assigns the **Acceptability Threshold** of this criterion for the Problem at hand



Deficiency "Flag"



V&V Agent assesses M&S criterion achievement value data and determines criterion scores

**What is the Achieved Value ?**

## Deficiency "Flag" Determination

$$\text{Threshold} - \text{Score} = \text{Deficiency Flag Type}$$

- █ **OK to use**  
0.0 ≤ Gap ≤ 0.5
- █ **Use with Caution**  
0.5 < Gap ≤ 1
- █ **Do Not Use**  
Gap > 1

assessed achievement value

Criterion Data

**Developers**  
Gather Criterion Evidence

M&S Artifacts

- M&S Reqts
- M&S Designs
- M&S Code
- M&S Data
- M&S Use

Collected and Quantified Criterion Evidence from M&S Artifact Data



Accreditation Agent assigns the Criterion Weighting of this criteria compared to other criteria within this criteria domain

**Criterion Scale**

Common Measurement Scale across all criteria: Indicates Threshold Value, Achieved Score, and Deficiency Flag

$$\text{Criterion Weight} \times \text{Criterion Score} = \text{Weighted Score (Domain Sub-Scores)}$$

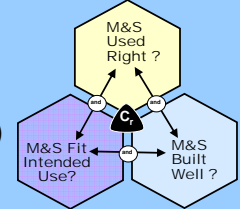
**How Important is it ?**

**Threshold**

**What is the Desired Level ?**



Accreditation Agent assigns the Acceptability Threshold of this criterion for the Problem at hand



**Deficiency "Flag" Determination**

$$\text{Threshold} - \text{Score} = \text{Deficiency Flag Type}$$

- █ **OK to use**  
0.0 ≤ Gap ≤ 0.5
- █ **Use with Caution**  
0.5 < Gap ≤ 1
- █ **Do Not Use**  
Gap > 1



V&V Agent assesses M&S criterion achievement value data and determines criterion scores

**NOTE**

<sup>(1)</sup> If the achieved criterion value exceeds the threshold value, no extra credit is given beyond the threshold value – more value than is required of any criterion does not provide extra credibility for the problem at hand

assessed achievement value

Criterion Data

Developers  
Gather Criterion Evidence

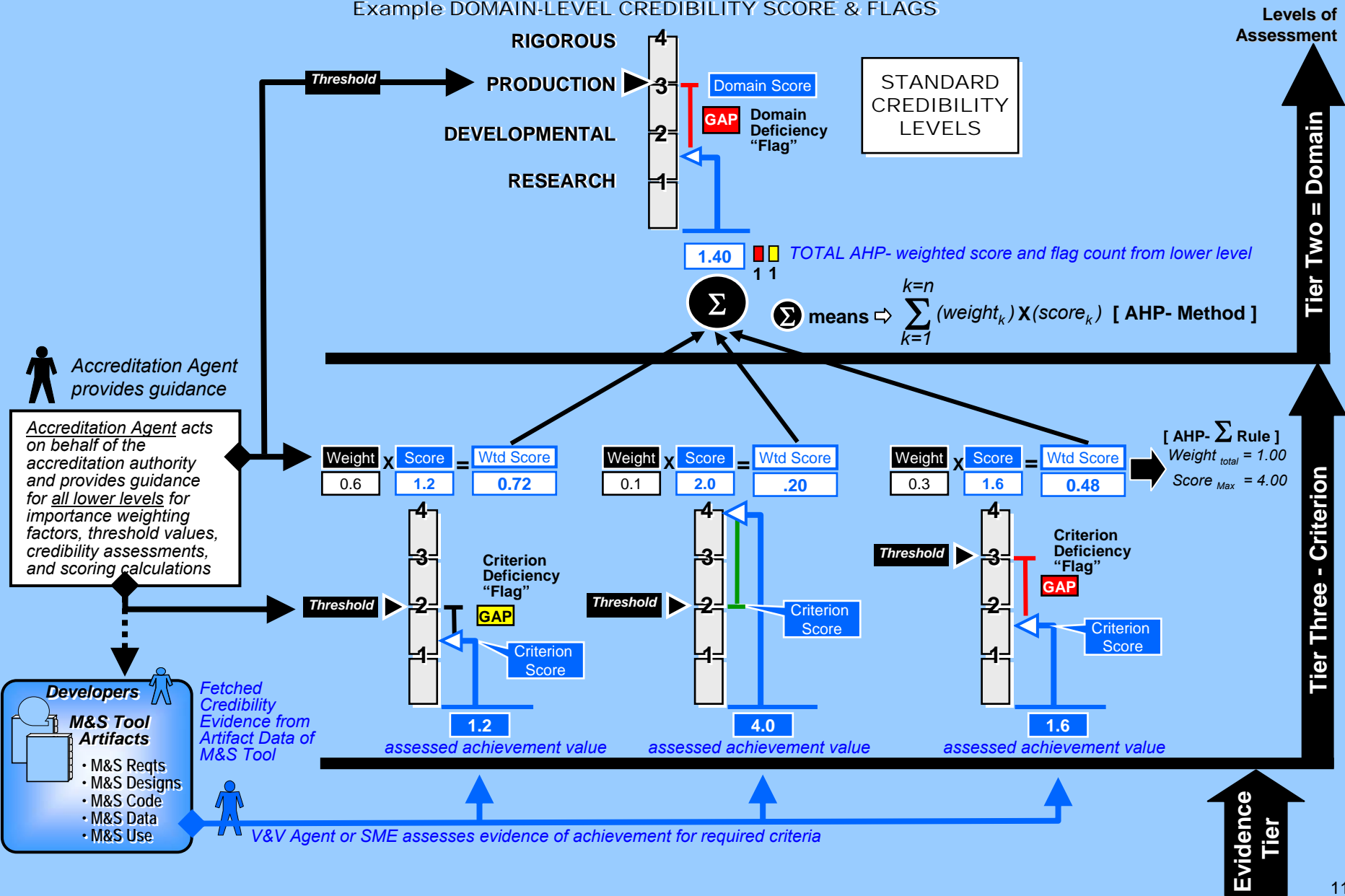
M&S Artifacts

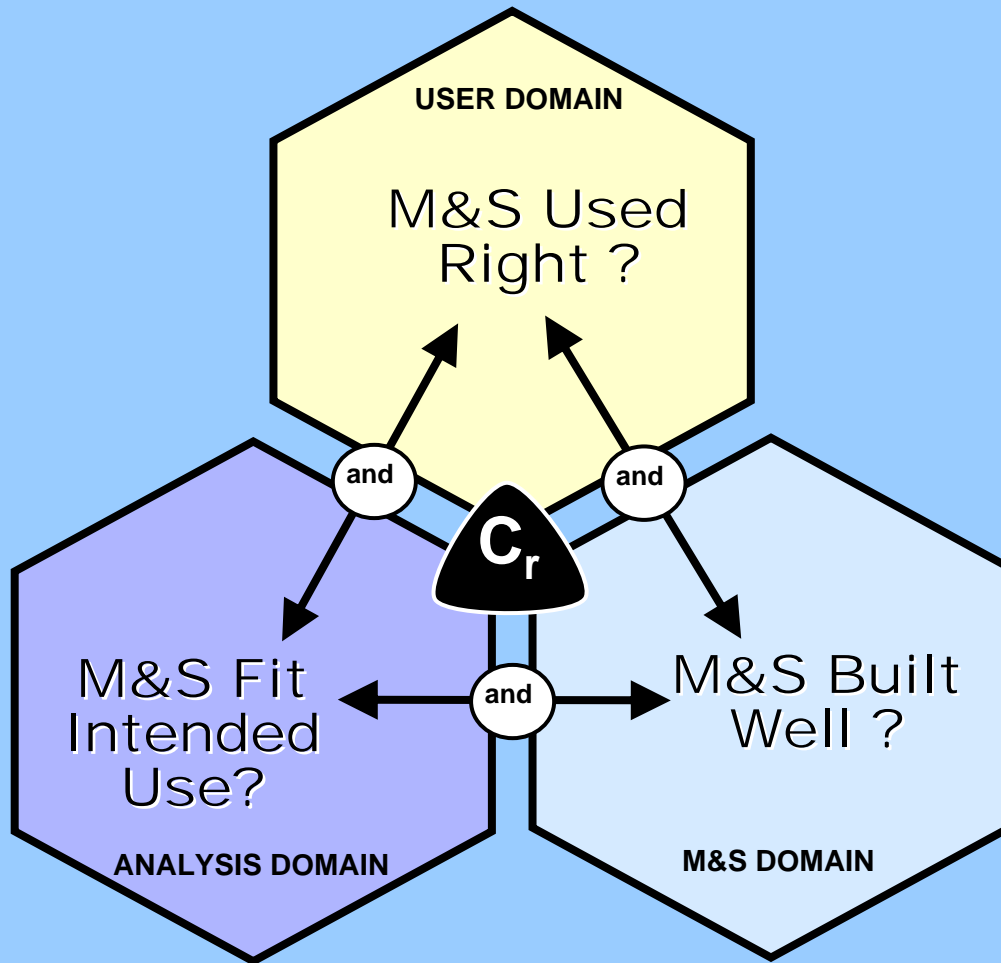
- M&S Reqs
- M&S Designs
- M&S Code
- M&S Data
- M&S Use

Collected and Quantified Criterion Evidence from M&S Artifact Data

# Example Credibility Score-Aggregation to Demonstrate Augmented-AHP Method

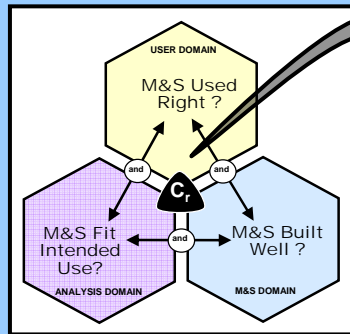
Example DOMAIN-LEVEL CREDIBILITY SCORE & FLAGS





**Credibility of M&S Results =  $C_r$  =**  $\blacktriangleright$  Fit Intended Use (IU), *and*  
 $\blacktriangleright$  Built Well (BW), *and*  
 $\blacktriangleright$  Used Right (UR)

## Key Domains and Criteria for $C_r$



**Credibility ( $C_r$ )  
for  
M&S Results**

**Levels of Credibility  
For M&S Results**

- 4 "Rigorous"
- 3 "Production"
- 2 "Developmental"
- 1 "Research"

**ANALYSIS DOMAIN**

**M&S Fit Intended Use ?**

Right Entities ?

Right Functions & Interactions ?

Right Scope, Scale, & Detail ?

**M&S DOMAIN**

**M&S Well Built?**

Verified Code ?

Numerical Accuracy ?

Validated Outputs ?

Uncertainty Measurement ?

Development Process Maturity ?

Required "ilities" Addressed ?

**USER DOMAIN**

**M&S Used Right?**

Problem Defined ?

Correctly Set-Up ?

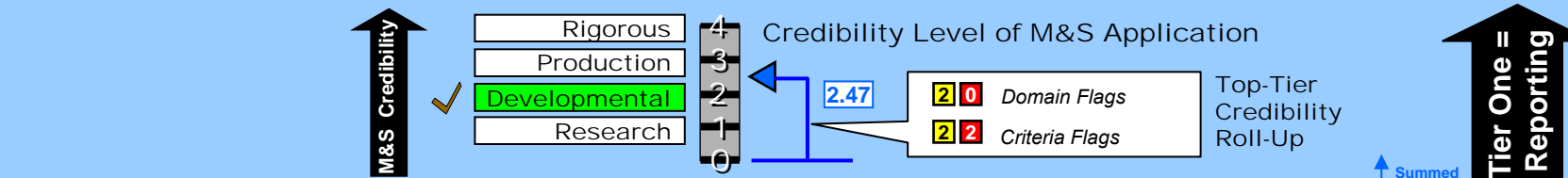
Correctly Executed ?

Correctly Analyzed ?

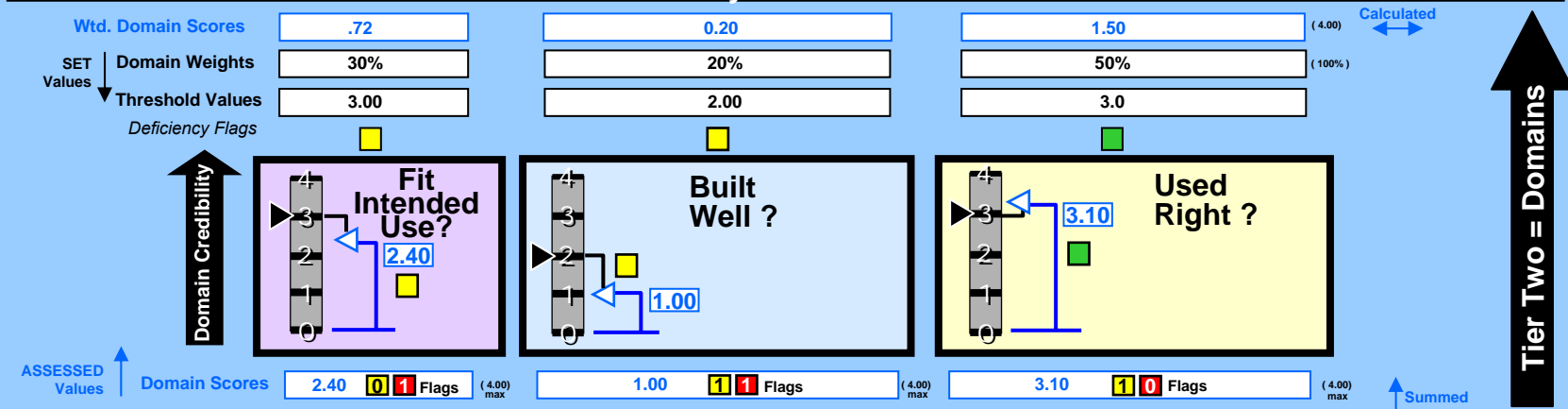
Analysis Traceable to Results ?

Operator / Analyst Qualified ?

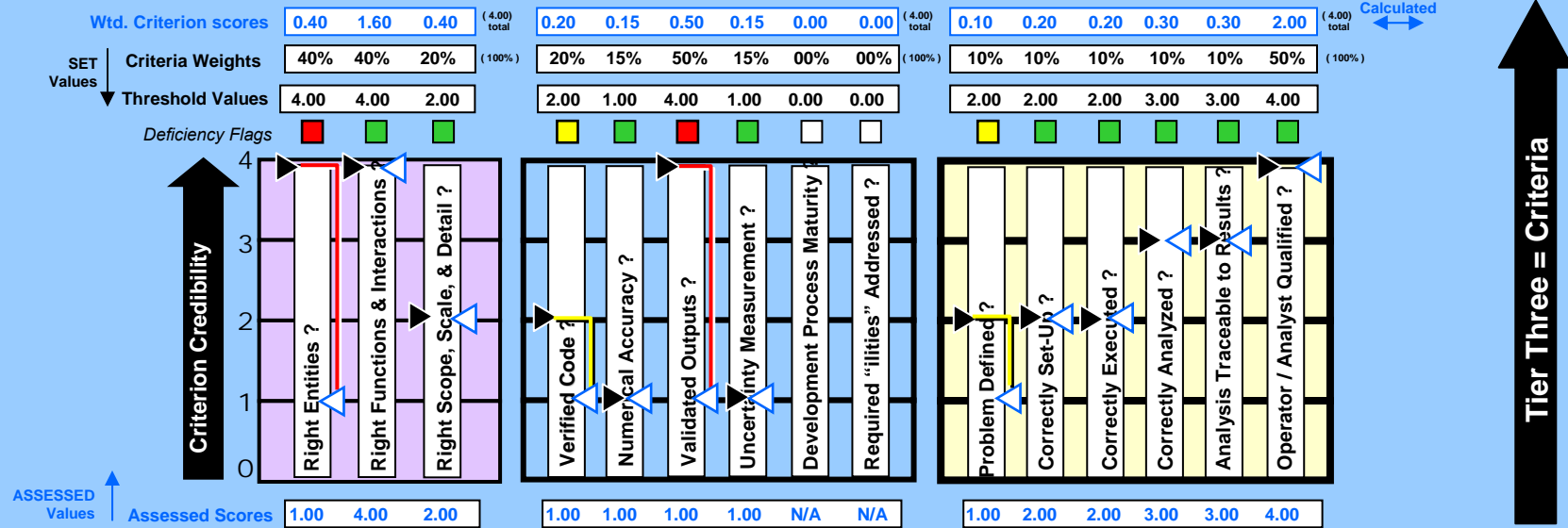
# NASA STD-(I)-7009 Apdx A4 Credibility Assessment Example: Calculated Using Augmented AHP



## Credibility Domain Scores

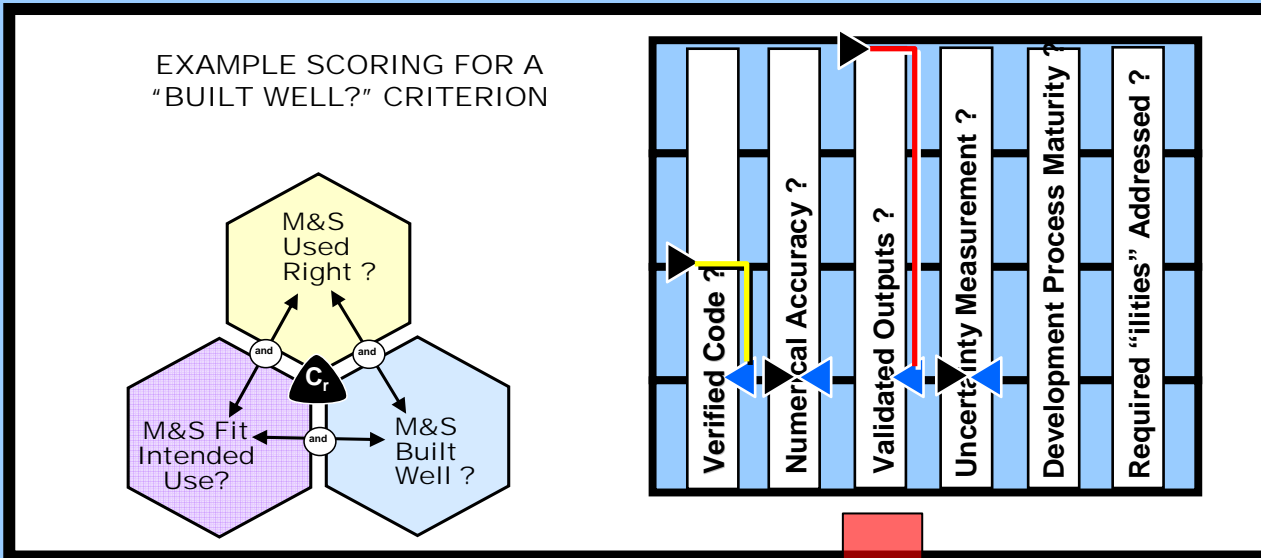


## Credibility Criteria Scores



Levels of Assessment

# Example of Augmented-AHP Scoring for M&S Credibility Criterion from NASA STD (I)-7009

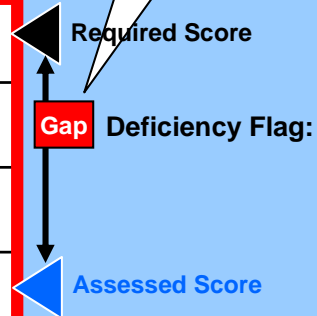


[ Abstract - Credibility Criterion Definitions from NASA M&S Standard STD-(I)-7009 ]

| BW Credibility Criteria | BW Criterion Definitions  | Credibility Levels For BW Criterion   |
|-------------------------|---|---|
| Validated Outputs?      | An assessment of the degree to which the model or simulation outputs are an accurate representation of the real world from the perspective of the intended uses of the model or simulation. | Level 4- Validated by test of the specified range of use<br>Level 3- Peer reviewed for underlying correctness<br>Level 2- Some process formalization<br>Level 1- Working, in progress |

Red Deficiency Flag Means:

**“DO NOT USE:**  
M&S is not validated for the intended use”



## Augmented-AHP Model Roadmap for Assessment of Credibility of M&S Results

**1 NASA M&S STD (I)-7009** - Interim NASA STANDARD FOR M&S - defines Levels of Credibility

**2** - Credibility rigorously established for the current application. M&S results were extensively tested and reviewed for the current application. Uncertainty distributions were calculated. Expert users and analysts used formally reviewed and controlled processes. The limitations of M&S were well documented and correct usage was enforced.

**3** - Credibility tested specifically for the current application. M&S results were tested and reviewed for the current application. Quantitative uncertainty bounds were calculated. The users and analysts performed simulations using formal best practices. The limitations of M&S were well documented and any violations were reported.

**4** - Credibility established for simulation process. M&S results were tested and reviewed for problems similar to the current application. Uncertainty was estimated by expert opinion based on the current application. The users and analysts performed simulations using informal repeatable processes. The limitations of the M&S were well documented.

**5** - Credibility established for model basics. M&S results were tested and reviewed for research problems. Uncertainty was estimated by expert opinion based on research problems. The simulations were performed using ad hoc processes. The limitations of the M&S were not well documented.

