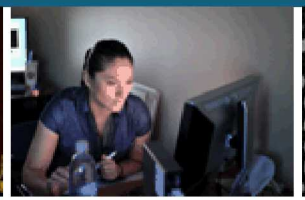
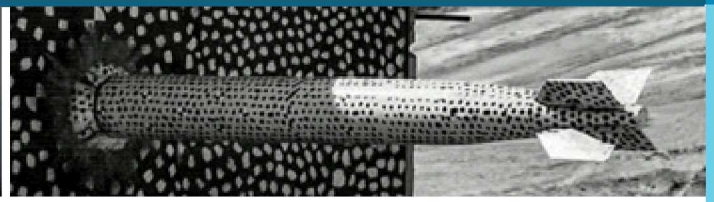
SAND2020-5153C

# Experimental Credibility for Modeling and Simulation





**Blake W. Lance** and Sarah L. Kieweg

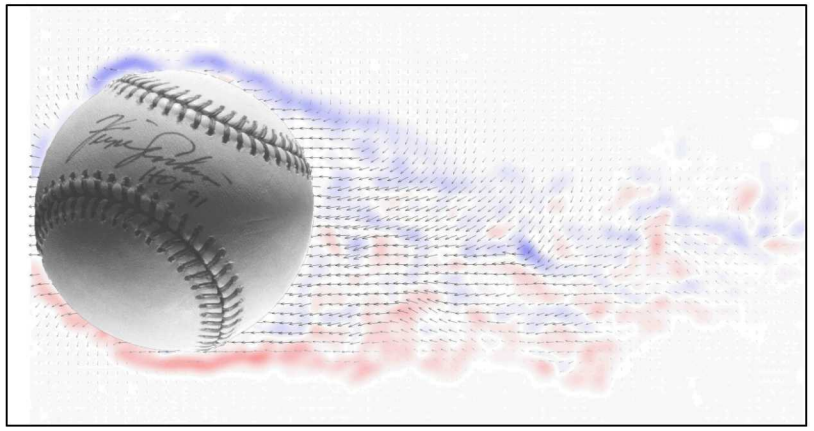
Verification & Validation Symposium, May 20-22, 2020
Online

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

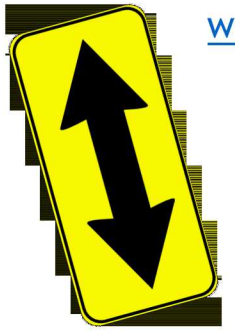
SAND2020-XXXX

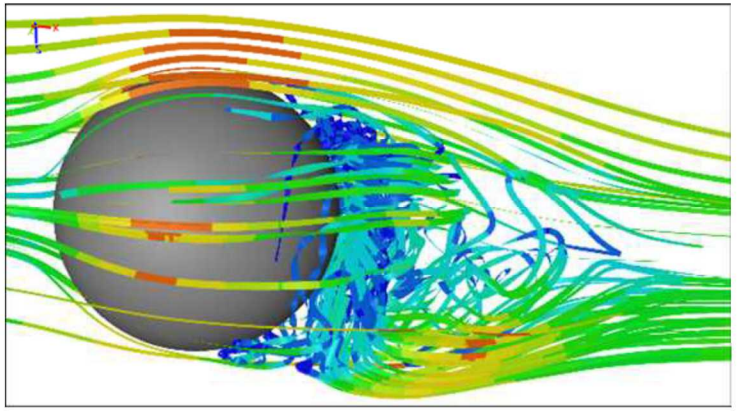
# Experiments and simulations have complimentary strengths

- Experiments provide a real-world view of physics
  - Come with cost and schedule
- Simulations provide rapid insights at lower cost
  - Come with potential model form and other errors
- Experimentalists and modelers should work collaboratively, but good working relationships don't always happen naturally
- The Experimental Credibility process facilitates discussion to align goals and streamline efforts



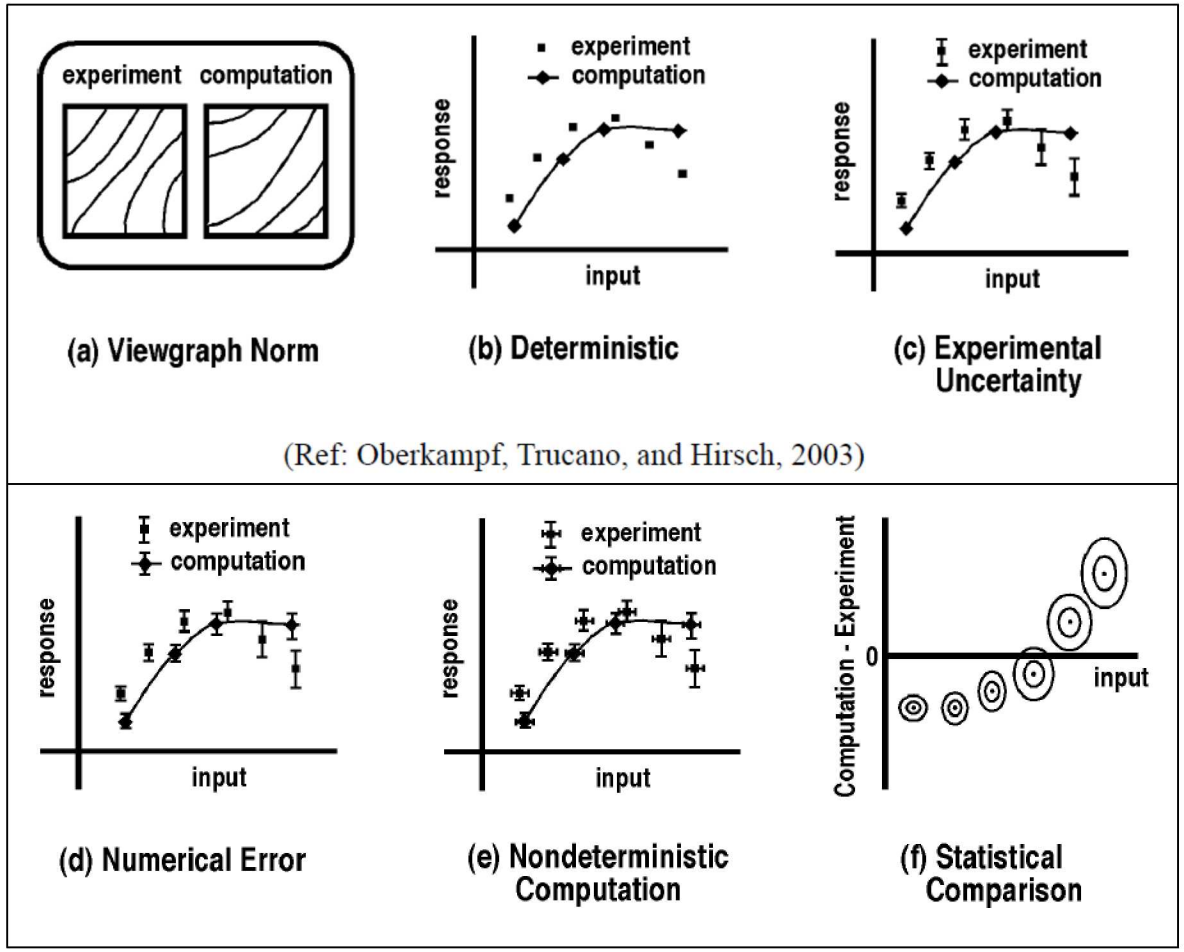
www.baseballaero.com/





# Computer simulations often use experimental data and inherit their quality/credibility

- Data are used to
  - Develop model forms
  - Calibrate model parameters
  - Validate model predictions
- Simulations inherit the quality/credibility of the experiments, including any errors present




Validation comparison levels of rigor

# Methods to evaluate computational simulation (CompSim) credibility are moderately mature

- The Predictive Capability Maturity Model (PCMM) outlines broad assessment methods
  - There is mention of experimental uncertainty in the Model Validation element but little on other aspects of experimental credibility
- The Phenomena Identification and Ranking Table (PIRT) has a column for validation

| ID | Phenomena | Importance | Adequacy for Intended Use: Math Model | Code | Validation | Model Parameter |
|---|---|---|---|---|---|---|
| A | **Radiative Heat Flux** | | | | | |
| A1 | Gray emission/absorption, non-scattering | H | H | H | L | M |
| A2 | Non-gray, non-homogeneous radiative transport | M | L | L | L | L |
| B | **Radiative Source Terms** | | | | | |
| B1 | Emission: small scale, ~1cm, turbulent mixing (affects flame area) | H | M | L | L | M |
| B2 | Emissive flux: combustion chemistry (affects soot temperature) | H | M | L | L | M |
| B3 | Emissive flux: fine scale, ~1mm, turbulent strain (affects soot temperature) | U | L | L | L | L |
| B4 | Emissive flux: soot diffusive transport (affects soot mass fraction) | H | L | L | L | L |
| B5 | Emissive flux: soot formation chemistry (affects soot mass fraction) | H | M | L | L | L |
| B6 | Emissive flux: soot oxidation chemistry (affects soot mass fraction) | H | M | L | L | L |
| B7 | Emissive flux: gas emission | M | M | L | L | L |
| B8 | Absorption | M | H | L | L | M |
| B9 | Scattering | L | H | L | L | M |
| C | **Convective Heat Flux** | | | | | |
| C1 | Large-scale turbulent mixing affecting flame geometry and radiative view factors | H | M | M | L | L |

"How to PIRT", SAND2016-6465 TR

| MATURITY / ELEMENT | Maturity Level 0<br>Low Consequence, Minimal M&S Impact, e.g. Scoping Studies | Maturity Level 1<br>Moderate Consequence, Some M&S Impact, e.g. Design Support | Maturity Level 2<br>High-Consequence, High M&S Impact, e.g. Qualification Support | Maturity Level 3<br>High-Consequence, Decision-Making Based on M&S, e.g. Qualification or Certification |
|---|---|---|---|---|
| **Representation and Geometric Fidelity**<br>What features are neglected because of simplifications or stylizations? | • Judgment only<br>• Little or no representational or geometric fidelity for the system and BCs | • Significant simplification or stylization of the system and BCs<br>• Geometry or representation of major components is defined | • Limited simplification or stylization of major components and BCs<br>• Geometry or representation is well defined for major components and some minor components<br>• Some peer review conducted | • Essentially no simplification or stylization of components in the system and BCs<br>• Geometry or representation of all components is at the detail of "as built", e.g., gaps, material interfaces, fasteners<br>• Independent peer review conducted |
| **Physics and Material Model Fidelity**<br>How fundamental are the physics and material models and what is the level of model calibration? | • Judgment only<br>• Model forms are either unknown or fully empirical<br>• Few, if any, physics-informed models<br>• No coupling of models | • Some models are physics based and are calibrated using data from related systems<br>• Minimal or ad hoc coupling of models | • Physics-based models for all important processes<br>• Significant calibration needed using separate effects tests (SETs) and integral effects tests (IETs)<br>• One-way coupling of models<br>• Some peer review conducted | • All models are physics based<br>• Minimal need for calibration using SETs and IETs<br>• Sound physical basis for extrapolation and coupling of models<br>• Full, two-way coupling of models<br>• Independent peer review conducted |
| **Code Verification**<br>Are algorithm deficiencies, software errors, and poor SQE practices corrupting the simulation results? | • Judgment only<br>• Minimal testing of any software elements<br>• Little or no SQE procedures specified or followed | • Code is managed by SQE procedures<br>• Unit and regression testing conducted<br>• Some comparisons made with benchmarks | • Some algorithms are tested to determine the observed order of numerical convergence<br>• Some features & capabilities (F&C) are tested with benchmark solutions<br>• Some peer review conducted | • All important algorithms are tested to determine the observed order of numerical convergence<br>• All important F&Cs are tested with rigorous benchmark solutions<br>• Independent peer review conducted |
| **Solution Verification**<br>Are numerical solution errors and human procedural errors corrupting the simulation results? | • Judgment only<br>• Numerical errors have an unknown or large effect on simulation | • Numerical effects on relevant SRQs are qualitatively estimated<br>• Input/output (I/O) verified | • Numerical effects are quantitatively estimated to be small on some SRQs<br>• I/O independently verified | • Numerical effects are determined to be small on all important SRQs<br>• Important simulations are independently reproduced |
| **Model Validation**<br>How carefully is the accuracy of the simulation and experimental results assessed at various tiers in a validation hierarchy? | • Judgment only<br>• Few, if any, comparisons with measurements from similar systems or applications | • Quantitative assessment of accuracy of SRQs not directly relevant to the application of interest<br>• Large or unknown exper-imental uncertainties | • Quantitative assessment of predictive accuracy for some key SRQs from IETs and SETs<br>• Experimental uncertainties are well characterized for most SETs, but poorly known for IETs<br>• Some peer review conducted | • Quantitative assessment of predictive accuracy for all important SRQs from IETs and SETs at conditions/geometries directly relevant to the application<br>• Experimental uncertainties are well characterized for all IETs and SETs<br>• Independent peer review conducted |
| **Uncertainty Quantification and Sensitivity Analysis**<br>How thoroughly are uncertainties and sensitivities characterized and propagated? | • Judgment only<br>• Only deterministic analyses are conducted<br>• Uncertainties and sensitivities are not addressed | • Aleatory and epistemic (A&E) uncertainties propagated, but without distinction<br>• Informal sensitivity studies conducted<br>• Many strong UQ/SA assumptions made | • A&E uncertainties segregated, propagated and identified in SRQs<br>• Quantitative sensitivity analyses conducted for most parameters<br>• Numerical propagation errors are estimated and their effect known<br>• Some strong assumptions made<br>• Some peer review conducted | • A&E uncertainties comprehensively treated and properly interpreted<br>• Comprehensive sensitivity analyses conducted for parameters and models<br>• Numerical propagation errors are demonstrated to be small<br>• No significant UQ/SA assumptions made<br>• Independent peer review conducted |

"Predictive Capability Maturity Model for Computational Modeling and Simulation", SAND2007-5948

# Previous frameworks have been published to evaluate experiments and their results to inform simulations

- Oberkampf and Smith merged concepts from PCMM and strong-sense model validation experiments and applied it to computation fluid dynamics validation experiments

**Table 1 Example of a MVEC table after assessment of an experiment**

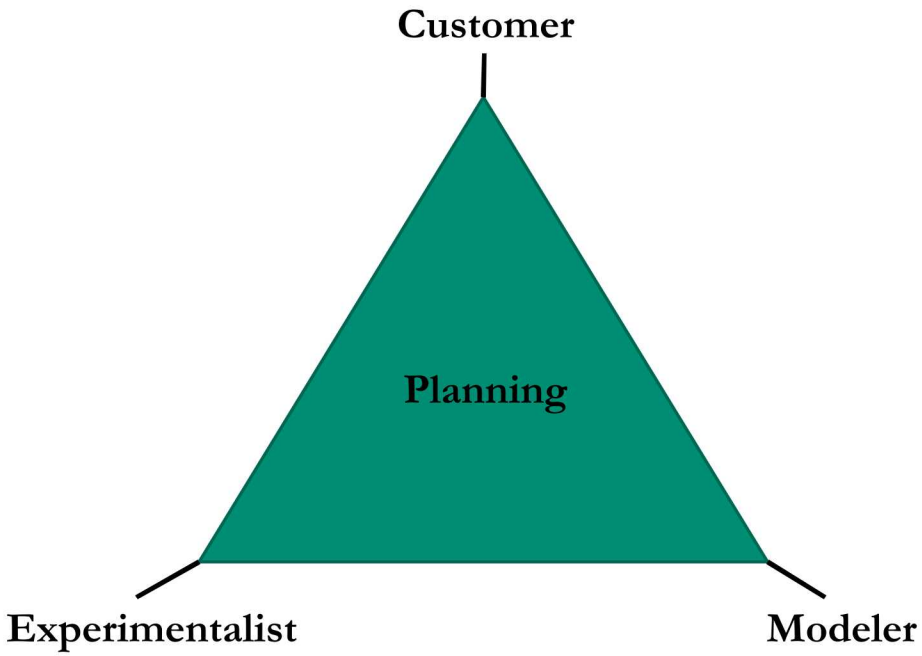
| | Completeness | | | | |
|---|---|---|---|---|---|
| Attribute | Completeness level 0 | Completeness level 1 | Completeness level 2 | Completeness level 3 | Attribute score |
| Experimental facility | | Assessed | | | 1 |
| Analog instrumentation and signal processing | | | Assessed | | 2 |
| Boundary and initial conditions | | Assessed | | | 1 |
| Fluid and material properties | Assessed | | | | 0 |
| Test conditions | | | Assessed | | 2 |
| System response quantities | | Assessed | | | 1 |

Oberkampf and Smith, 2017, "Assessment Criteria for Computational Fluid Dynamics Model Validation Experiments", ASME JVVUQ.

- Elele et al. presented several thoughts about experimental facility review for validation and proposed a risk-based approach.
  - "Applying Modeling and Simulation Verification, Validation, and Accreditation (VV&A) Techniques to Test and Laboratory Facilities", V&V Symposium, 2012.
- V&V-40 on medical devices discusses the rigor of validation comparisons in terms of output quantity, relevancy, and uncertainty.

# Experimental Credibility is a process

- The Experimental Credibility process provides a structured method to assess experiments used for simulations
  - Correctness
  - Completeness
  - Applicability to intended use
- It encourages
  - Early planning of experiments
  - Communication between stakeholders: experimentalists, modelers, system integrator
  - Documentation of experimental credibility that aids simulation credibility



# Plan and Assess Experiment Overview

- Spreadsheet with seven elements
- Usable with any application
- Scoring is removed as it may detract from the purpose of documenting and communicating credibility evidence
- The prompts are open-ended questions that contain best practices
- Team of experts and users
  - Modelers
  - Experimentalists
  - Customer
  - V&V partner
- Team assessment steps
  - Discuss prompts, strengths/weaknesses
  - Write assessment commentary
  - Identify action items

| **Plan and Assess Experiment** | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Describe how CompSim will be involved in the planning of this experiment? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use? How?<br>• What documentation and general/specific understanding do you have of the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could be changed to improve the applicability? | |
| **Experimental Verification** | • What methods will be used to verify testing apparatus control/code is performing as desired?<br>• How is the post-process of the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Describe how the test conditions will be characterized for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Describe how the output measurements will be characterized for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, will the validation metrics and criteria be specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Plan and Assess Experiment Elements

- Planning
- Intended Use
- Sample, Geometric, Material Fidelity
- Experimental & Environmental Fidelity
- Experimental Verification
- Uncertainty Quantification
- Peer Review and Documentation

| Plan and Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Describe how CompSim will be involved in the planning of this experiment? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use? How?<br>• What documentation and general/specific understanding do you have of the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could be changed to improve the applicability? | |
| **Experimental Verification** | • What methods will be used to verify testing apparatus control/code is performing as desired?<br>• How is the post-process of the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Describe how the test conditions will be characterized for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Describe how the output measurements will be characterized for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, will the validation metrics and criteria be specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Element 1: Planning

- Test Purpose
  - What is the overall goal for the experiment and simulation for all the stakeholders?
  - Who is the end user?
  - What is the intended use?
- How will modeler be involved in test planning?
- How much communication between
  - Experimentalist
  - Modeler
  - Customer

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| Element | Prompts to Consider | Assessment Commentary |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Element 5: Intended Use

- Possible uses
  - Materials characterization
  - Calibration
  - Validation
- For the intended use
  - Describe the degree test conditions will be known
  - Describe how the measurements provide the required information
  - For validation, were metrics and acceptance criteria specified?
  - Could a range of conditions be helpful?

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Element 2: Sample, Geometric, and/or Material Fidelity
# Element 3: Experimental & Environmental Fidelity

- Sample, Geometric, Material Fidelity
  - How representative is the test article to the application?
  - Can you describe the test article pedigree and any pre-processing?
- Experimental and Environmental Fidelity
  - How relevant are the test conditions to the application?
  - Are improvements possible/needed to improve applicability?
  - Are all simulations inputs measured?

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Element 4: Experimental Verification

- How do you know you measured what you think you measured?
- Describe any testing of experimental control software
- How are test equipment calibration and quality implemented?
- Describe any testing of data post-processing codes
- Could instrumentation affect test conditions?
- How could repeatability be confirmed?

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Element 6: Experimental Uncertainty Quantification (UQ)

- To what degree will the uncertainty be quantified for:
  - Test conditions
  - Measurements of outputs
- What types of uncertainty measurements would be helpful for the UQ calculations/simulations?
- Can level of repeatability be quantified?

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework. | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# "Assess Experimental Uncertainty" Tool

| **Assess Experimental Uncertainty** | Read these prompts, discuss with team, and write a response assessment for each element. Use this when assessing the Uncertainty Quantification element of the Assess Validation Experiment tool. | | Assess the pros and cons of experiment in terms of quantified uncertainty |
|---|---|---|---|
| **Element** | **Prompts to Consider** | **Best Practices** | **Commentary** |
| **Pre-test planning:** | • Was there pre-test planning between experimentalist and end-user? Was there discussion on use of data and documentation needs? | • Discussion initiated pre-test.<br>• Decide who will do which parts of data analysis and UQ.<br>• Agree upon level of documentation on data pedigree and UQ<br>• Clearly define end use of experiment. | |
| **Pre-test: Define measurand(s) needed to obtain QOI(s)** | • Are the Quantities of Interest (QOIs) defined and specified how will be measured and/or quantified?<br>• How do measurands relate to QOIs? Require post-processing? | • Discussion/activity initiated pre-test<br>• Plan to measure range of local and globally integrated quantities<br>• Specify and document functional relationship between measurand(s) and final QOI(s), and how data processed and/or reduced.<br>• Document other unmeasured quantities used to calculate QOI. | |
| **Pre-test: Measurement process and management of** | • Is the measurement and calibration process well described?<br>• Were expected uncertainties considered in experimental | • Define test objectives • Map measurement parameters and nominal level to what calibrations and instruments will determine each.<br>• Identify correlated errors (e.g. measurements that come from same calibration/instrument) • Specify required uncertainty for each measurand so that final result has required uncertainty | |
| **Pre-test/Post-test: Expected and Estimated Uncertainties** | • Is there an uncertainty inventory for all conditions and measurements?<br>• What is missing or a limitation for use of test (e.g. UQ and validation)? | • Should be done both pre-test (expected) and post-test.<br>• For each measurand in test, complete spreadsheet of (expected) uncertainties.<br>• Consider all possible sources of uncertainty.<br>• Consider documentation, calibration histories, previous tests with similar instruments, previous uncertainty analyses, expert judgement. | |
| **Pre-test/Post-test: Uncertainty Propagation and Sensitivity Analysis** | • What uncertainty sources are small compared to others?<br>• Which uncertainties are not well characterized and can something be done to improve that?<br>• What could be done now or in future to reduce predicted or measured uncertainties? | • Propagate estimated (or actual) measurement uncertainties into the expected (or actual) range of results for the QOI(s).<br>• Identify which measurand(s) have greatest impact on uncertainty of result.<br>• Identify if there is a better measurement technique to use.<br>• Communicate between experimentalist and analyst on whether expected result uncertainty will be adequate for intended use.<br>• If multiple tests, repeat calculation of results and find uncertainty of the result directly, and compare to propagated uncertainties from each measurement; extract info about zeroth and first order replication level analysis (e.g. infer sample-to-sample variability with multiple tests). | |

# Element 6: Experimental Uncertainty Quantification

Pre-Test

**Define Measurand(s) and Connections to QOI(s):**

- How are they related?
- Do they require post-processing?

**Define Measurement Process and Manage Uncertainties:**

- How are the measurement and calibration methods described?
- Were expected uncertainties considered in test design?

Pre-Test *and* Post-Test

**Expected and Estimated Uncertainties**

- Is there an uncertainty inventory for sensors and expected values?
- Are the estimated uncertainties small enough to meet test requirements?

**Uncertainty Propagation and Sensitivity Analysis**

- What uncertainty sources are large compared to others?
- If total uncertainties are too large, can the largest sources be reduced?
- Could some uncertainties be better characterized?

# Element 7: Peer Review and Documentation

- Assessed for each of the Elements of this tool.
- What additional peer review or documentation will be needed?
  - What level of rigor is expected in the documentation?

**Peer Review**

- Which of Elements 1-6 will be reviewed by subject matter experts?

**Documents**

- Will all of the Elements be documented?
- Are tabulated experimental results tied to their description and archived?

| Assess Experiment | Read these prompts, discuss with team, and write a response for each element. Use this when assessing and communicating credibility evidence for computational simulation (i.e. **CompSim**) that uses this **experiment**. Complete during pre-test planning and again during post-test analysis. | How did these elements impact the strength and weakness of this test for the purpose of CompSim intended use? |
|---|---|---|
| **Element** | **Prompts to Consider** | **Assessment Commentary** |
| **Planning** | • Is purpose of the test known to the experimentalist and end-user (e.g. CompSim analyst)? What is the intended use of test?<br>• How much communication will there be between the experimentalist, customer, and end-use analyst during both the planning and post-test stages? Does this create any strengths or weaknesses to the outcomes?<br>• Will CompSim be involved in the planning of this experiment, and in what way? | |
| **Sample / Geometric / Material Fidelity** | • Is the sample, geometry, and/or material relevant to the specified requirement and/or intended application? Is the proximity sufficient for this type of test and intended use?<br>• Do you know the pedigree?<br>• Is there any pre-processing of the sample/material that could impact applicability? | |
| **Experimental / Environmental Fidelity** | • How relevant is the environment and test conditions to the requirement and/or application? Is the proximity sufficient for this type of test and intended use?<br>• What could have been changed to improve the applicability? | |
| **Experimental Verification** | • Was the code that controls the testing apparatus verified?<br>• Was the code that post-processes the raw data verified?<br>• Are the test facility and equipment documented well and calibrated?<br>• How do you know you measured what you think you measured? Do you have any confirmation of the measurements? What is the evidence that the test peformed correctly? | |
| **Intended Use (e.g. validation, calibration, materials characterization** | • Will the test conditions be characterized well enough for the intended use? Will any conditions be missing, not well-characterized, or in doubt?<br>• Will the output measurements be characterized well enough for the intended use? Will enough quantities of interests be measured, and will the right ones be measured?<br>• For validation, were validation metrics and criteria specified before the testing, or after? | |
| **Uncertainty Quantification** | • This includes uncertainty on both test conditions and outputs - did the test provide the uncertainty on both needed for the intended use?<br>• To assess the uncertainty quanitification, use the elements of the "Assess Experimental Uncertainty" framework | |
| **Peer Review and Documentation** | • Which of the above elements of the test will be reviewed by subject matter experts? Which elements will not, and of those, which may need further review and why?<br>• Which of the above elements will be documented? Will the documentation serve the needs of the intended use, and help write the credibility evidence for the CompSim? Or is there anything missing that would have improved the validation process? | |

# Conclusions

- Experimental Credibility is a process that
  - Encourages early planning
  - Facilitates communication between experimentalists, modelers, and customers
  - Enables documentation
  - Suggests best practices for validation and calibration experiments
  - Aids CompSim credibility
- It includes a UQ spreadsheet that presents best practices and aids documenting credibility evidence
- The impact can be accelerated development and more unified project teams

- Future Work
  - Continue revisions with additional feedback
  - Integrate into VVUQ workflow such as Sandia Analysis Workbench (SAW)
  - Consider updating name – Experimental Assessment and Simulation Integration (EASI)?