Empirical Validation as a Tool

for Evaluating
Forensic Science Evidence

From foundational validity to reliable application

Pate Skene
pate@skenelaw.edu
skene@duke.edu

Empirical validation and Standards

2009 National Academy of Science report

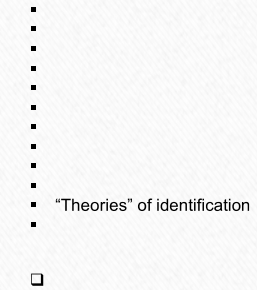
2016 PCAST report

2014 - 17 National Commission on Forensic Sciences

2014 - NIST/OSAC

Organization of Scientific Area Committees on Forensic Sciences (OSAC), National Institute of Standards and Technology (NIST)
https://www.nist.gov/osac

Empirical validation studies



Why do you need empirical validation?

What kind of validation do you need?

What does the validation actually tell you?

How do you challenge?

* Created by Nina Chernoff, CUNY law school; adapted by Anthony Monaghan

Rule 702

- **Rule 702(a)**

If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion, or otherwise, if all of the following apply:

(1) The testimony is based upon sufficient facts or data.

(2) The testimony is the product of reliable principles and methods.

(3) The witness has applied the principles and methods reliably to the facts of the case.

Rule 702

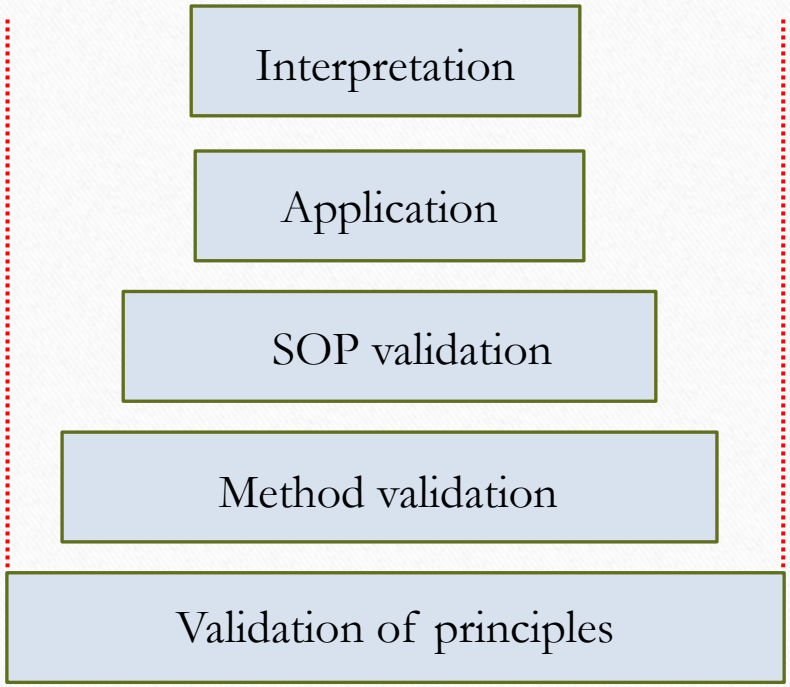
- **Rule 702(a)**

If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion, or otherwise, if all of the following apply:

Reliability

(1) The testimony is based upon sufficient facts or data.
(2) The testimony is the product of reliable principles and methods.
(3) The witness has applied the principles and methods reliably to the facts of the case.

Foundational validity and Reliability



Is the witness's opinion within the limits/scope of the validated principles and methods?

Does the relevant SOP allow analysts to beyond those limits? (i.e., vague procedure, judgment w/out specific criteria, or vacuous standards)

How well do the validated principles and methods define the scope, scale, and power/resolution of the method?

Rule 702

- **FRE 702 (with proposed amendments)**

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if the proponent has demonstrated by a preponderance of the evidence that:

(a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
(b) the testimony is based on sufficient facts or data;
(c) the testimony is the product of reliable principles and methods; and
(d) the ~~expert has reliably applied~~ expert's opinion reflects a reliable application of the principles and methods to the facts of the case.

Foundational validity and Reliability

Interpretation

Application in a specific case

SOPs

Method validation

Validation of principles

Foundational validity and Reliability

Interpretation

Application in this case

SOP

Method validation

Validation of principles

Rule 702

- **Rule 702(a)**

Relevance

If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion, or otherwise, if all of the following apply:

Reliability

(1) The testimony is based upon sufficient facts or data.
(2) The testimony is the product of reliable principles and methods.
(3) The witness has applied the principles and methods reliably to the facts of the case.

Foundational validity and Relevance

Relevance

Reliability

Foundational validity and Relevance

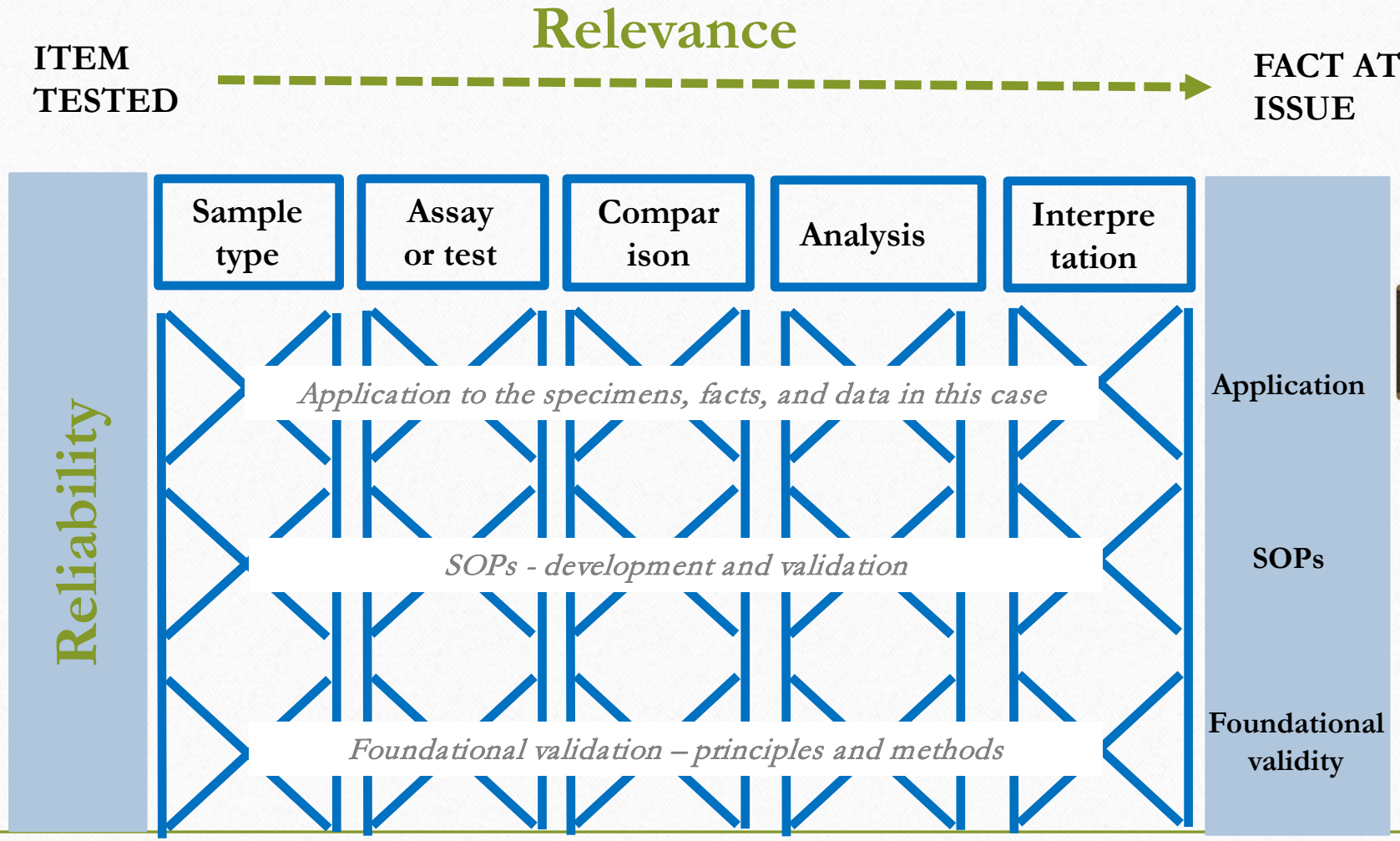
Relevance

Reliability

"when a trial court 'conclude[s] that there is simply too great an analytical gap between the data and the opinion proffered,' the court is not required "to admit opinion evidence that is connected to existing data only by the ipse dixit *of the expert."*

McGrady at 889, quoting *Joiner,* 522 U.S. at 146.

Foundational validity and Relevance



Why empirical validation?

- "Theories" of identification

Why empirical validation?

McGrady

Whether expert witness testimony is admissible under Rule 702(a) is a preliminary question that a trial judge decides pursuant to Rule 104(a).

…To the extent that factual findings are necessary to answer this question, the trial judge acts as the trier of fact. The court must find these facts by the greater weight of the evidence.

McGrady at 890.

State v. McGrady, 787 S.E. 2d 1, 368 N.C. 880 (2016)

Why empirical validation?

Human factors research in performance-based professions:

- medicine
- aviation
- engineering
- management
- martial arts
- music

CONFIDENCE ≠ COMPETENCE

Experience → confidence (subjective feeling)

\+

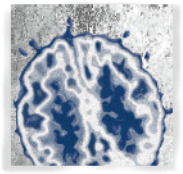
Empirical feedback → competence (accurate performance)

Prediction error and learning

Basic research

Dopamine reward prediction error coding

Wolfram Schultz, MD, FRS



Reward prediction errors consist of the differences between received and predicted rewards. They are crucial for basic forms of learning about rewards and make us strive for more rewards—an evolutionary beneficial trait. Most dopamine neurons in the midbrain of humans, monkeys, and rodents signal a reward prediction error; they are activated by more reward than predicted (positive prediction error), remain at baseline activity for fully predicted rewards, and show depressed activity with less reward than predicted (negative prediction error). The dopamine signal increases nonlinearly with reward value and codes formal economic utility. Drugs of addiction generate, hijack, and amplify the dopamine reward signal and induce exaggerated, uncontrolled dopamine effects on neuronal plasticity. The striatum, amygdala, and frontal cortex also show reward prediction error coding, but only in subpopulations of neurons. Thus, the important concept of reward prediction errors is implemented in neuronal hardware.

 Dialogues Clin Neurosci. 2016;18:23-32.

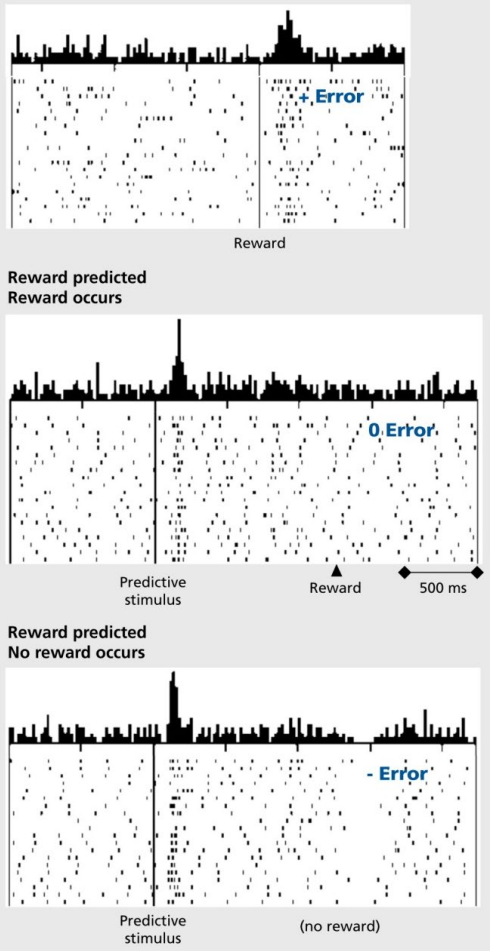
Keywords: *neuron; substantia nigra; ventral tegmental area; striatum; neurophysiology; dopamine; reward; prediction*

Introduction

I am standing in front of a drink-dispensing machine in Japan that seems to allow me to buy six different types of drinks, but I cannot read the words. I have a low expectation that pressing a particular button will deliver my preferred blackcurrant juice (a chance of one in six). So I just press the second button from the right, and then a blue can appears with a familiar logo that happens to be exactly the drink I want. That is a pleasant surprise, better than expected. What would I do the next time I want the same blackcurrant juice from the machine? Of course, press the second button from the right. Thus, my surprise directs my behavior to a specific button. I have learned something, and I will keep pressing the same button as long as the same can comes out. However, a couple of weeks later, I press that same button again, but another, less preferred can appears. Unpleasant surprise, somebody must have filled the dispenser differently. Where is my preferred can? I press another couple of buttons until my blue can comes out. And of course I will press that button again the next time I want that blackcurrant juice, and hopefully all will go well.

Author affiliations: Department of Physiology, Development and Neuroscience, University of Cambridge, United Kingdom

Address for correspondence: Wolfram Schultz, Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, United Kingdom
(email: ws234@cam.ac.uk)



Prediction error and learning

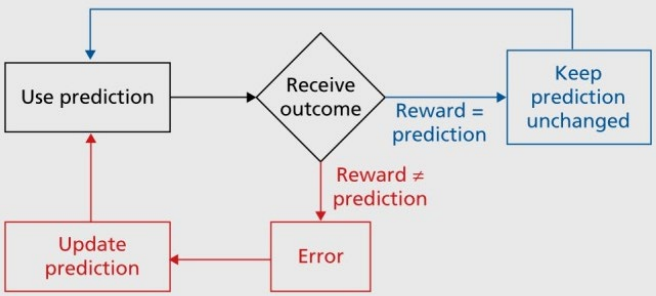
Basic research

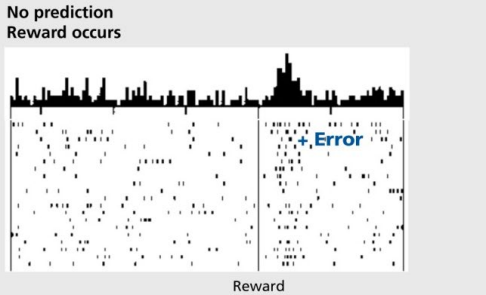
Dopamine reward prediction error coding

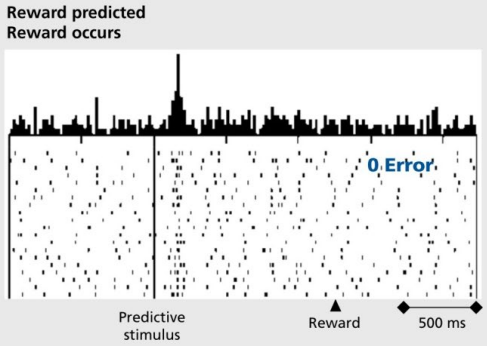
Wolfram Schultz, MD, FRS

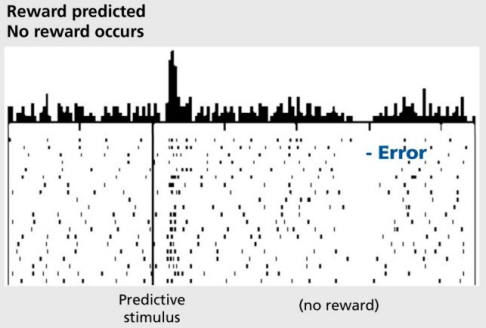
Reward prediction errors consist of the differences between received and predicted rewards.

They are crucial for basic forms of learning about rewards and make us strive for more rewards—an evolutionary beneficial trait…

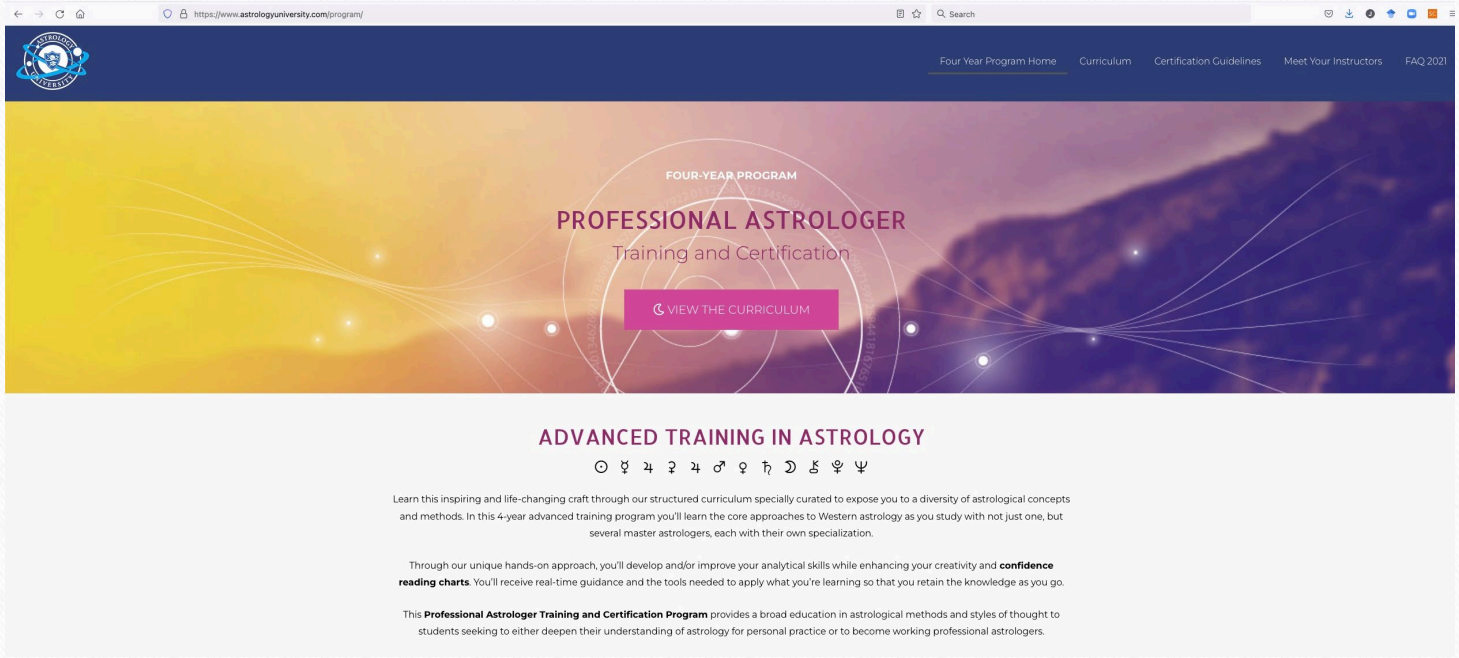








Why empirical validation?



"Nor, on the other hand, does the presence of Daubert's general acceptance factor help show that an expert's testimony is reliable where the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology or necromancy."

Kumho Tire 526 U.S.137 at 151.

Why empirical validation?

ABOUT THE PROGRAM

Over the course of your studies you will **learn the core building blocks of Western astrological practice** and will be introduced to several approaches to Western astrology including: Psychological, Archetypal, Mythic, Ancient/Traditional, Humanistic, Evolutionary and Transformational.

You will be exposed to a diverse set of teachings that will ultimately map out an approach to **astrology that is empowering and life-affirming while being respectful of the power of our free will** to intentionally engage with life in a co-creative process.

TWO TRACK OPTIONS

As a student of this program you will have the option to study on one of two tracks: **Certification-Track** or **Audit** (**Non-Certification Personal Development**.) Students on the certification track will complete written assignments and testing and receive feedback and mentoring support. If you are not interested in certification, you have the option to drop in to just the courses you're interested in.

The complete program is offered over 4 years and a new cohort starts every August. You can choose to study on a schedule with a group of students participating in weekly live Q&A sessions, or you may complete the requirements on your own time at your own pace. Either way you'll still get access to real-time support with your questions as you move through the material.

Certification-track students must complete all course requirements to sit for our certification exam. At the end of successful completion of the four year program, students will receive a diploma. Students who complete the requirements will also be able to sit for our certification exam.

Why empirical validation?

Why empirical validation?

Astrology?

- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓
- ✓

- "Theories" of identification

- ❑

Created by Nina Chernoff, CUNY law school; adapted by Anthony Monaghan

What counts as validation?

- KEY element is empirical feedback
- NOT unique to science

  MOST kinds of human expertise involve empirical testing
- **Forensic sciences has been the aberration**

What counts as validation?

- KEY element is empirical feedback

  What specifically in your proficiency testing or experience would show you if you were wrong?

- Forensic sciences have been the aberration

What counts as validation?

- Black box studies
- Scientific research on relevant principles and methods

- Proficiency testing
- Experience

(but, wait, didn't I just tell you that these are NOT adequate?)

- *Fact of testing or experience tells you (almost) nothing*
- *Dig deeper – WHAT about the proficiency testing or experience can show that it is or is not a reliable basis for an opinion?*

What counts as validation?

What specifically in your proficiency testing or experience shows how reliable it is as a basis for your analysis in this case?

PROFICIENCY TESTING

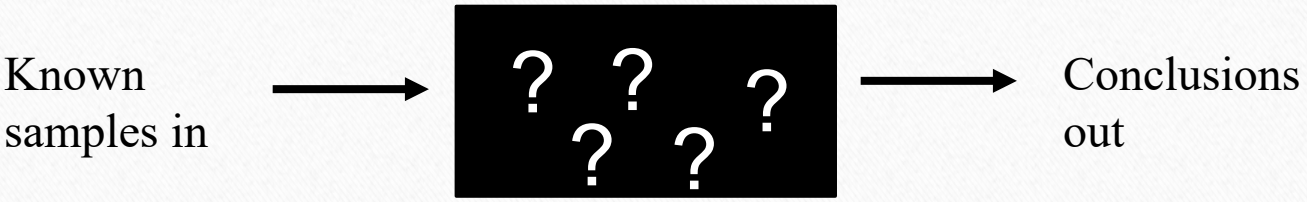
- In forensic sciences, typically tests for threshold competence
  easy examples = exaggerated confidence
- **Safety-critical professions** (aviation, engineering, medicine)
  test to limits of the method

TRAINING and EXPERIENCE

- How many times have you done an analysis like the one in this case under conditions that could show if you were wrong?
- What specific criteria do you use to verify that you are doing you analysis consistently and accurately?

What counts as validation?

Black box studies

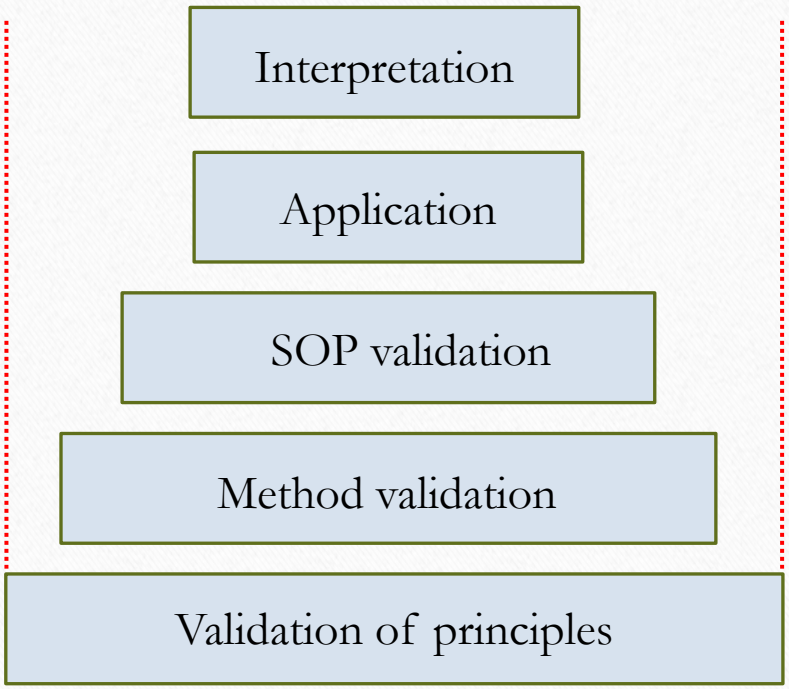


Testing well-specified methods

- Effects of various test parameters (temp, reagents, flow rates, cycle number, etc…)
- Effects of sample conditions, amount, contaminants
- Sensitivity, specificity – limits of detection, limit of quantitation
- Criteria and decision thresholds (instrument or human judgment)
- Foreseeable failure modes
- Theoretical limits – limits defined by principles on which the method is based

What do you need to look for?

- KEY purpose is defining limits



How well is the method specified?

What specifically shows whether the witness reliably applied the method as it was tested?

Types, amount of specimens, facts, or data needed for reliable results

Defining the limits of error

"Error rates"

Confidence/credibility intervals

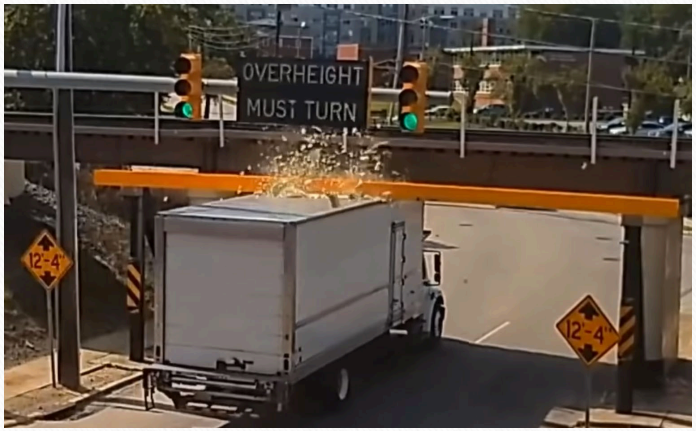
Resolution,
ability to discriminate differences

Uncertainty in measurements, discrimination

What do you need to look for?

Limits of testing

Limits defined by testing



Black box ⟷ **Defined methods**

- Was "the method" adequately specified?
- Range of difficulty / resolution?
- Representative participants?
- Test conditions -- Hawthorne effect

- Effects of variation/deviation from method
- Type, amount, condition of samples
- Controls, calibration, known failure modes
- Specified criteria for steps involving human judgment

Subjective methods

Black box

- Was "the method" adequately specified?
- Range of difficulty / resolution?
- Representative participants?
- Test conditions -- Hawthorne effect

How well does the validated method specify what constitutes reliable application?

***State v. McPhaul*,** 256 N.C. App. 303 (2017)

Wood testified that she uses the same examination technique as is commonly used in the field of latent print identification, and she employed this procedure while conducting her examination in this case. However, when Wood testified to her ultimate conclusions, she was unable to establish that she reliably applied the procedure to the facts of this case.

Subjective methods

How well does the validated method specify what constitutes reliable application?

***State v. McPhaul*,** 256 N.C. App. 303 (2017)

Black box

- Was "the method" adequately specified?
- Range of difficulty / resolution?
- Representative participants?
- Test conditions -- Hawthorne effect

Can the expert show that her subjective process was sufficiently similar to the validation studies to be reliable?

e.g., same criteria for marking features, similar comparison thresholds

What do you need to look for?

Forensic Science International 318 (2021) 110457

Testing the accuracy and reliability of palmar friction ridge comparisons – A black box study

Heidi Eldridge[a,b,*], Marco De Donno[b], Christophe Champod[b]

[a] RTI International, 3040 E. Cornwallis Rd., Research Triangle Park, NC 27709, USA
[b] University of Lausanne, Batochime Quartier Sorge, Lausanne-Dorigny, VD, CH-1009, Switzerland

226 latent print examiners
12,279 decisions
12 false identifications

0.7% false positive rate

328 participants enrolled
226 completed at least one trial
133 completed all trials

Palm marks from 50 individuals

What do you need to look for?

Forensic Science International 318 (2021) 110457

Testing the accuracy and reliability of palmar friction ridge comparisons – A black box study

Heidi Eldridge[a,b,*], Marco De Donno[b], Christophe Champod[b]

[a] RTI International, 3040 E. Cornwallis Rd., Research Triangle Park, NC 27709, USA
[b] University of Lausanne, Batochime Quartier Sorge, Lausanne-Dorigny, VD, CH-1009, Switzerland

328 participants enrolled
226 completed at least one trial
133 completed all trials

Palm marks from 50 individuals

Representative of all active examiners?

Representative of casework?

How does ability to discriminate within small group related to ability to discriminate across entire population?

Subjective methods

Black box

- Was "the method" adequately specified?
- Range of difficulty / resolution?
- Representative participants?
- Test conditions -- Hawthorne effect

Well-specified methods

Foundational studies <u>define the limits</u> of validity

Well-specified methods

- Sample requirements (exclusion criteria)
- Foreseeable failure modes
- Theoretical limits
- Interpretation criteria / decision thresholds
- Statistical analysis, likelihood ratios

Foundational studies <u>define the limits</u> of validity

Interpretation

Application

SOP validation

Method validation

Validation of principles

Well-specified methods

Did the application stay within bounds of the validated method?

Is the opinion based on validated principles and methods?

Table 1. ICP-MS test results compared to laboratory thresholds for significant levels

Element	Threshold for significant level	Test results (Range for Items 1-4)
Antimony (Sb)	≥ 0.03 ppm	0.010 – 0.022 ppm
Barium (Ba)	≥ 0.2 ppm	0.017 – 0.043 ppm
Lead (Pb)	≥ 0.3 ppm	0.063 – 0.161 ppm

SOP: When the levels of Sb, Ba, and Pb do not meet the thresholds described, STOP.

SEM/EDX results

SOP: 3-4 particles minimum for determination
Result: 1 and 3 particles, respectively, on items

Opinion: Particles indicative of GSR detected

Well-specified methods

Foundational studies define the limits of validity

Application in this case

SOP

Method validation

Validation of principles

How do you use this?

Black box

Most common with pattern comparison

- Latent prints
- Firearms and toolmarks
- Fire investigation
- Digital devices

Defined limits

Most often chemical/biological methods

- DNA
- Gas chromatography –FID or -MS
- Gunshot residue

How do you use this?

Black box

Most common with pattern comparison

- SOPs leave wide discretion
- Failure to mitigate cognitive biases
- Lack of real-time documentation
- Close-nonmatches from database hits
- Failure to consider alternatives

Defined limits

Most often chemical/biological methods

- Sample amount, quality
- Peak vs background calls
- Likelihood, random match, statistics
- Opinions outside the scope of expertise

How do you use this?

What kinds of experts can help?

scientific study design ("meta-science", "science of science"
- defining scope and limits of the method

statistics
- likelihood, probabilities
- statistical confidence, credible intervals

human factors
- cognitive factors in expert analysis (perception, biases, memory)
- juror understanding and (mis)interpretation of expert testimony

cross-examination
- insist on specific, step-by-step description of their analysis
- probe for specific basis of comparison for subjective judgments

How do you use this?

What do you ask?

- How does the SOP set limits on the method?
- Did you stay within the SOP: real-time documentation
- Step-by-step logic: is every step supported

- How does your method compare alternative explanations?
- What in your method shows if something is wrong?

Foundational validity and Relevance

Relevance

Reliability

"when a trial court 'conclude[s] that there is simply too great an analytical gap between the data and the opinion proffered,' the court is not required "to admit opinion evidence that is connected to existing data only by the ipse dixit *of the expert."*

McGrady at 889, quoting *Joiner,* 522 U.S. at 146.