Validation and Performance Metrics

Stephane Cotin
cotin.stephane@mgh.harvard.edu

CIMIT Simulation Group / Harvard Medical School
http://simcen.usuhs.mil/miccai2003

Validate, but what?

- Validate training systems as a whole
  - Skills trainers
    - » Define objective measures for technical skills
    - » Classify types of skills to teach
  - Procedural training systems
    - » Full scale validation studies
    - » demonstrate validity & reliability
- Validate components (e.g. tissue models)
  - Use quantitative measurements when possible
  - Define a reference model and a set of metrics
- Define standardized methods of validation
  - Truth Cube (www.truthcube.org) for soft tissue models

Types of Validity*

- Face validity
  - The chosen tasks resemble those that are performed during a surgical procedure
- Content validity
  - The test for measuring a specific skill actually measures this skill and not another one
- Construct validity
  - Degree to which the test actually captures the skill level it was designed to measure (discrimination between experts and novices)
- Concurrent validity
  - Extent to which a test yields the same results as other measures of the same phenomenon
- Predictive validity
  - Relates to transfer in the operating room

* Reber, AS "Dictionary of Psychology"

Defining Performance Metrics for Skills Training

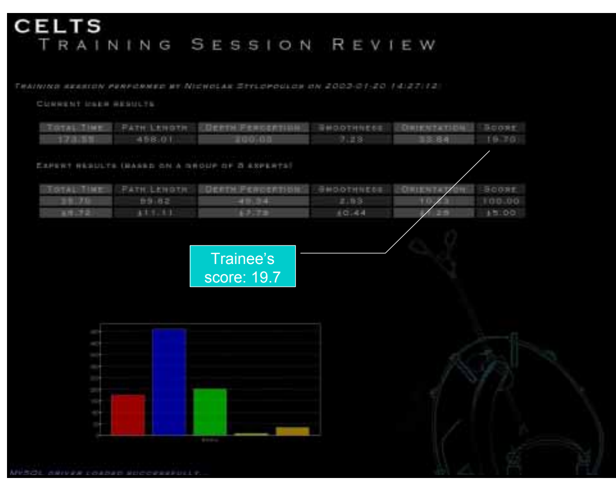
- "CELTS" Computer Enhanced Laparoscopic Skills Training System
- Use kinematics analysis to define a set of parameters that describe specific skills
- Define metrics by measuring the "distance" (z-score) between values obtained by the trainee and an expert group



Performance Assessment using Kinematic Analysis

- Depth Perception
- Smoothness of motion
- Instrument orientation
- Path length
- Grasping
- Time
- Task outcome

$$\textit{Smoothness of motion} = \frac{\sqrt{\frac{1}{2}\int_0^T \left(\frac{d^3x}{dt^3}\right)^2 dt}}{T}$$



Validity of CELTS

- Face validity
  - Tasks were chosen by surgeons and provide realistic visual and haptic feedback
- Content validity
  - Tasks are designed to incrementally teach laparoscopic skills, from depth perception to suturing
- Construct validity
  - Validation study demonstrated very significant differences in scores between experts and novices
- Concurrent validity
  - Unknown
- Predictive validity
  - Unknown

Validation

- Validation of the whole system
  - Validate the training system in terms of educational content and transfer to the real world
  - Need to assess the reliability of the training system
- Validation of the components of a system
  - Occurs during development phase
  - Define specific metrics for assessment of major components of the system
  - Example: "Truth Cube" for assessing soft tissue models

Validation & Reliability

- Types of reliability
  - Inter-rater reliability
    - » Extent to which two different evaluators score the same test
    - » Used to validate a training system by comparing the score provided by the simulator with the (usually qualitative) assessment provided by an instructor
  - Test-retest reliability
    - » If tested twice on the same the trainee should get a similar score
    - » However when learning occurs scores should improve

Validation

- Requires rigorously defined validation studies
- Must be tailored to each system and its purpose
- CELTS as an example: validation studies need to answer those 3 questions
  - Q1: does the scoring system differentiate expert performance from non-expert performance?
  - Q2: does the scoring system correlate with qualitative performance assessment from a panel of experts?
  - Q3: How can we track performance over time (learning curve)?

Other validation system: Blue DRAGON

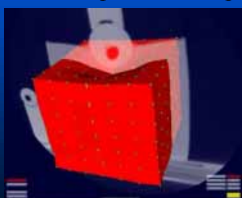
- Track the motion of and forces applied to surgical tools during live procedures
- Methodology based on a finite state Markov model
- Each state corresponds to a fundamental tool/tissue interaction based on tool kinematics and Force/Torque signatures
- Analysis of the data shows that skills differences are related to:
  - type of tool/tissue interactions being used
  - transitions between tool/tissue interactions being applied by each hand
  - time spent while performing each tool/tissue interaction
  - overall completion time

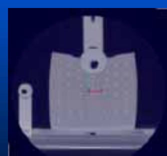
Component Validation

- Validation of tissue models
  - Many tissue models available (FEM, mass-spring, …)
  - Which one works best overall and for a specific application?
  - What to compare a model to?
- What else needs to be validated?
  - Haptic feedback
  - Curriculum
  - …

“Truth Cube”

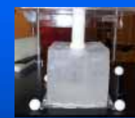
- Concept of a “gold standard” for assessing soft tissue models
- Comes from the “real world” to avoid introducing assumptions or simplifications





More information at:
www.truthcube.org

“Truth Cube” setup



- Material: two-part Silicone Rubber mixed in 30:70 proportion
- Beads: 1.58 mm diameter Teflon balls
- Stiffness of rubber with beads = Stiffness of rubber without beads
- Poisson's Ratio assumed to be 0.5
- Young's modulus = 14.9 kPa
- Tests performed under nearly frictionless condition
- Imaging performed on General Electric *LightSpeed* CT scanner

References

- Richards, C., Rosen, J., Hannaford, B., MacFarlane, M., Pellegrini, C., Sinanan, M., *Skills Evaluation in Minimally Invasive Surgery Using Force/Torque Signatures*, Surgical. Endoscopy., 14(9):791-8.
- Phil Wallace, *The Role of Simulation in Skill Learning*, SimTect Medical Symposium, 2001
- Patricia Youngblood and Parvati De, *An Evaluation Framework for the Validation of Surgical Simulators,* SimTect Medical Symposium*,* 2001
- N. Stylopoulos, S. Cotin, S. Dawson, M. Ottensmeyer, P. Neumann, P. Jackson, D. Rattner, *CELTS: A clinically-based Computer Enhanced Laparoscopic Training System*, Proc. MMVR, 2003
- A. Kerdok, S. Cotin, M. Ottensmeyer, A. Galea, R. Howe, S. Dawson, *Truth Cube: Establishing Physical Standards for Real-Time Soft Tissue Simulation*, Medical Image Analysis 2003; 7:283-291.

Conclusion

- The medical community is expecting a lot from medical simulation
- Negative training must be avoided by developing realistic training systems based on real-world data
- Validation is the key to the acceptance of computer-based simulation by the medical community
  - Validate components of simulators like tissue models
  - Validate the final training system