

Automated Assessment in Games & Simulations

2015 CRESST Conference

Alan Koenig, Ph.D.

August 19, 2015

National Center for Research
on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies | cresst.org



Automated Assessment

What is Automated Assessment?

A software-based solution to assessing a student's knowledge, skills, and/or abilities (KSA's)

Why use automated assessment?

Games & simulations often require students to engage in cognitively complex tasks, which can involve:

- ❑ multiple, non-trivial steps to complete
- ❑ interdependent tasks
- ❑ multiple pathways to success

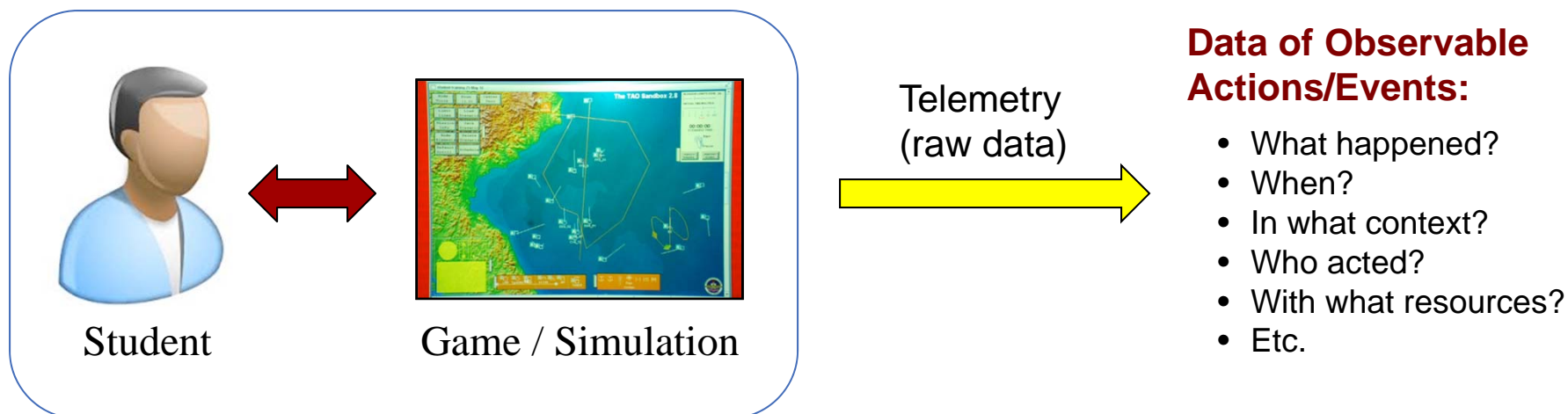
Which means... **assessment can be difficult:**

- ❑ Detecting meaningful actions
- ❑ Making sense of meaningful actions (what, when, how, under what conditions)
- ❑ Inferring latent KSA's from observed actions/events

Automating this process can help clarify and validate human judgment.



The Challenge: Getting what you want out of a game or simulation



But often what we want to know about a student is NOT directly observable. Instead, it has to be inferred.

You did well in the sim... Now what?

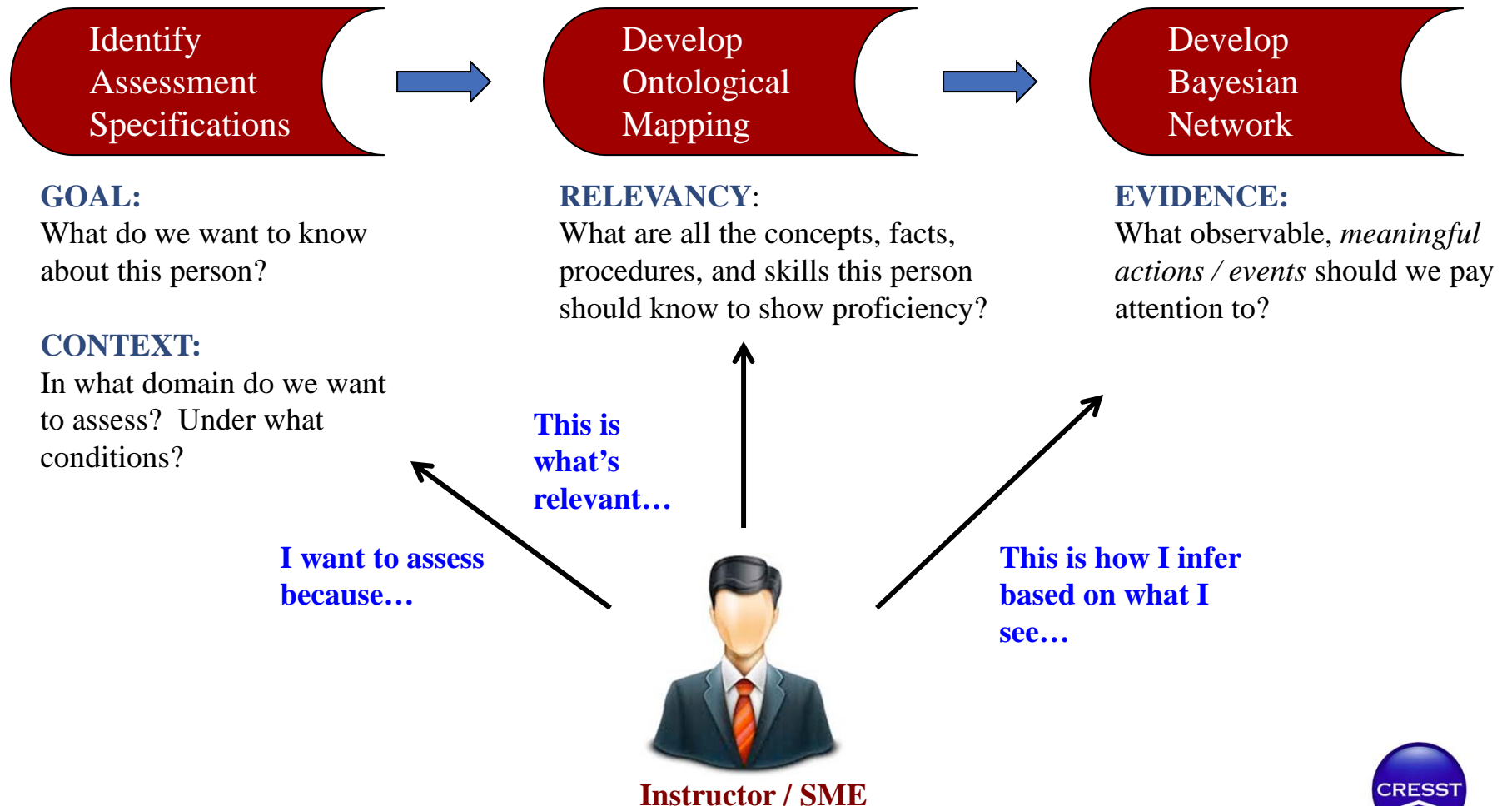
You accrued 136 points,
and got a rating of “Proficient”



But what I *really* want to know is: will you be safe
piloting a DDG in a crowded, chaotic harbor?



Automated Assessment Methodology



Automated Assessment Methodology



Example: Shiphandling Mooring Task:

What's the assessment goal?

Determine a student's shiphandling proficiency with mooring a DDG (Navy Destroyer) to a pier.

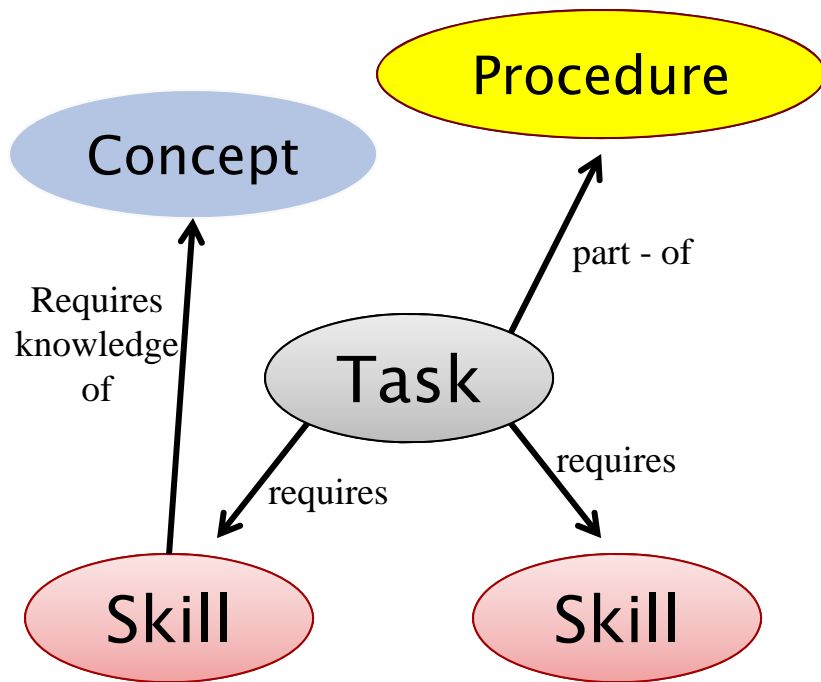
What latent knowledge, skills, and abilities (KSA's) are of interest?

- Proficiency with maneuvering a ship (use of rudder / propulsion / tugs)
- Ability to maintain safe practices

Under what conditions do we want to assess?

- Port of Bahrain, mooring to occur between two docked ships
- Offsetting, light wind & current
- Light harbor traffic
- Daytime

Automated Assessment Methodology



Ontologies are visual representations of domain knowledge

- ✓ Nodes are elements (procedures, tasks, skills, concepts, etc.)
- ✓ Lines are the relationships among elements

Ontologies aid instruction and assessment

- ✓ Support scenario development
- ✓ Support evaluation of performance

Ontology Construction



Concepts / Facts



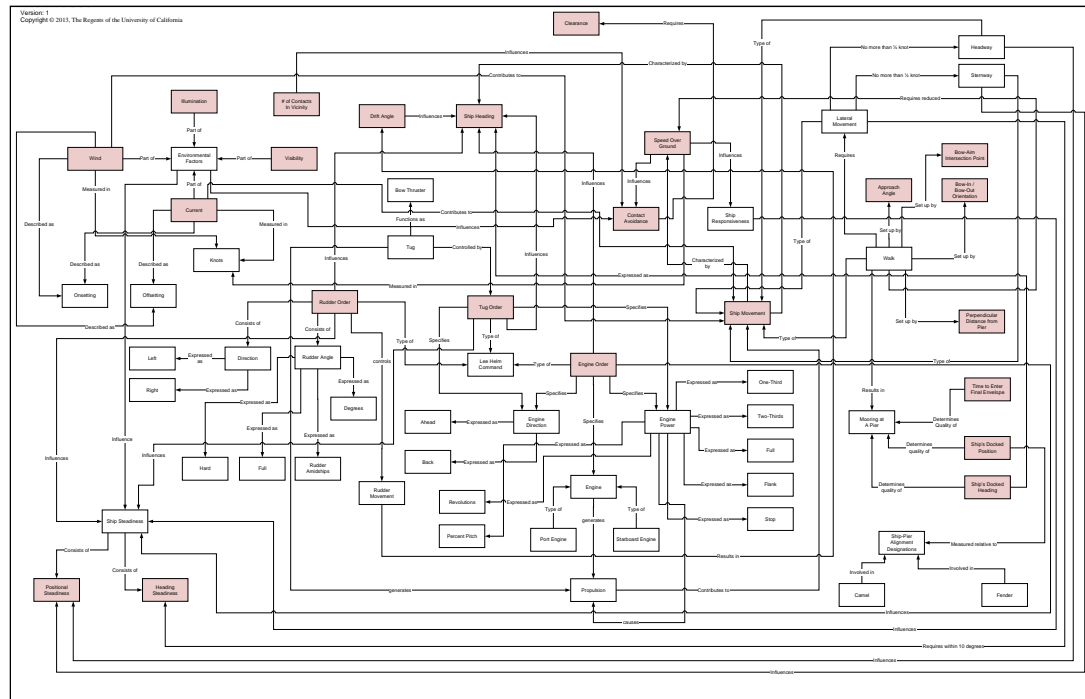
Expert Insights



Navy Doctrine



Procedures / Tasks

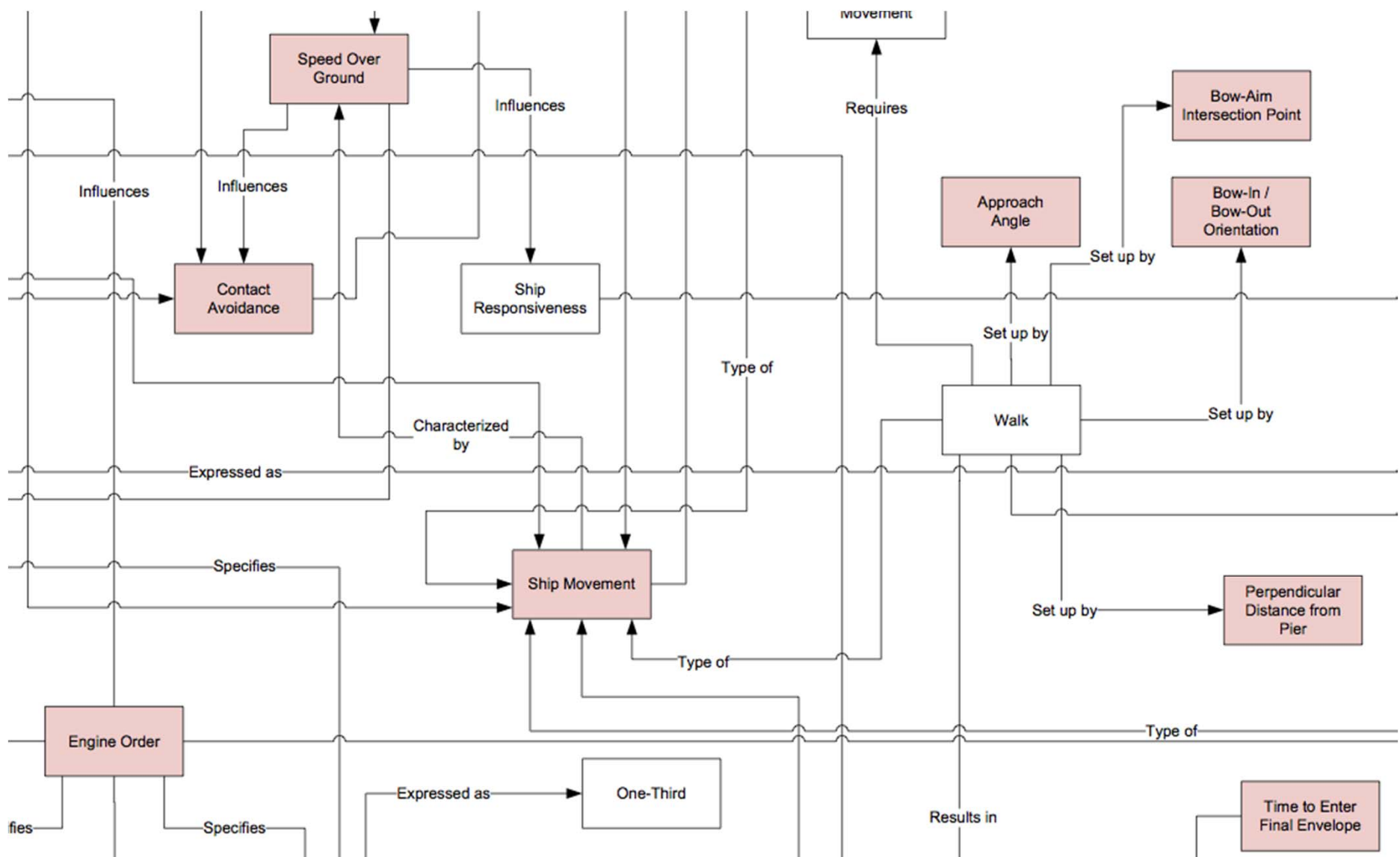


Shiphandling Mooring Ontology

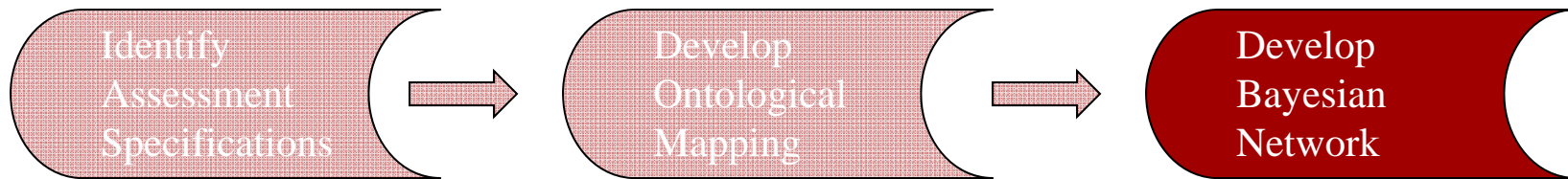
National Center for Research on Evaluation, Standards, & Student Testing



Mooring Ontology (zoomed, partial view)



Automated Assessment Methodology



Steps:

1. Identify meaningful actions / events to detect & observe
2. Construct the Bayesian network
3. Develop scoring rubrics for meaningful actions / events

Identify Meaningful Actions

Based on the *assessment goals* and *ontology*, a set of **observable, meaningful actions** are identified. It is from these actions that inferences to latent skills are made.



What experts pay attention to

Shiphandling Mooring: Meaningful Actions to Consider

- Clearance to Buoys
- Clearance to Other Ships
- Ship's Heading Steadiness
- Ship's Heading from Pier Heading
- Speed Over Ground (fore-aft)
- Speed Over Ground (lateral)
- Approach Track (*within Green Zone**)
- Docked "Bridge-Here" Alignment

Construct the Bayesian Network

- ❑ Constructed using information from the ontology
- ❑ Links represent dependencies between ontology parts. They indicate strength of relationships (stored as conditional probabilities)
- ❑ Probability of mastery of the latent variables is inferred from observable actions

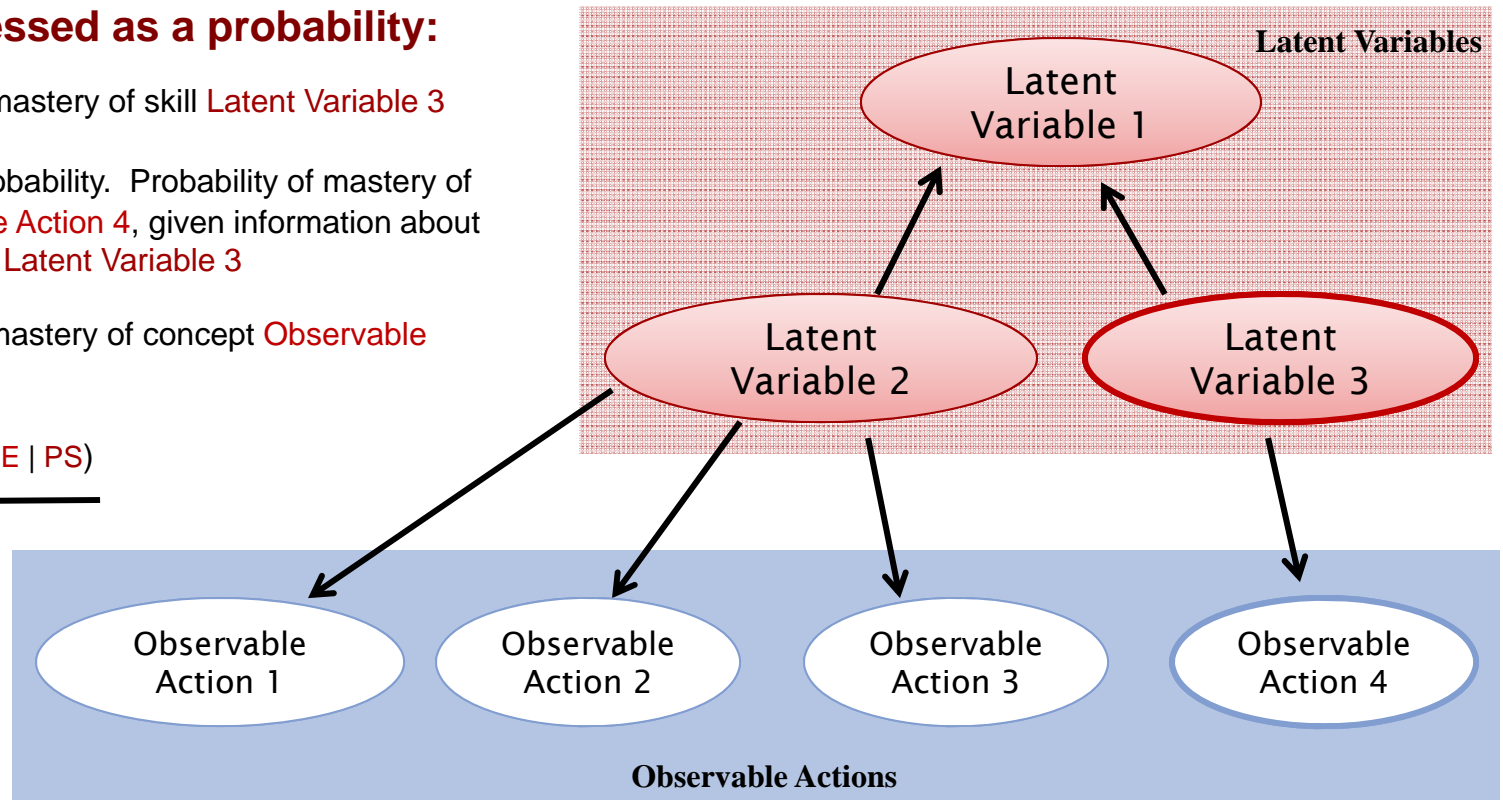
Everything's expressed as a probability:

$P(\text{PS})$: Probability of mastery of skill **Latent Variable 3**

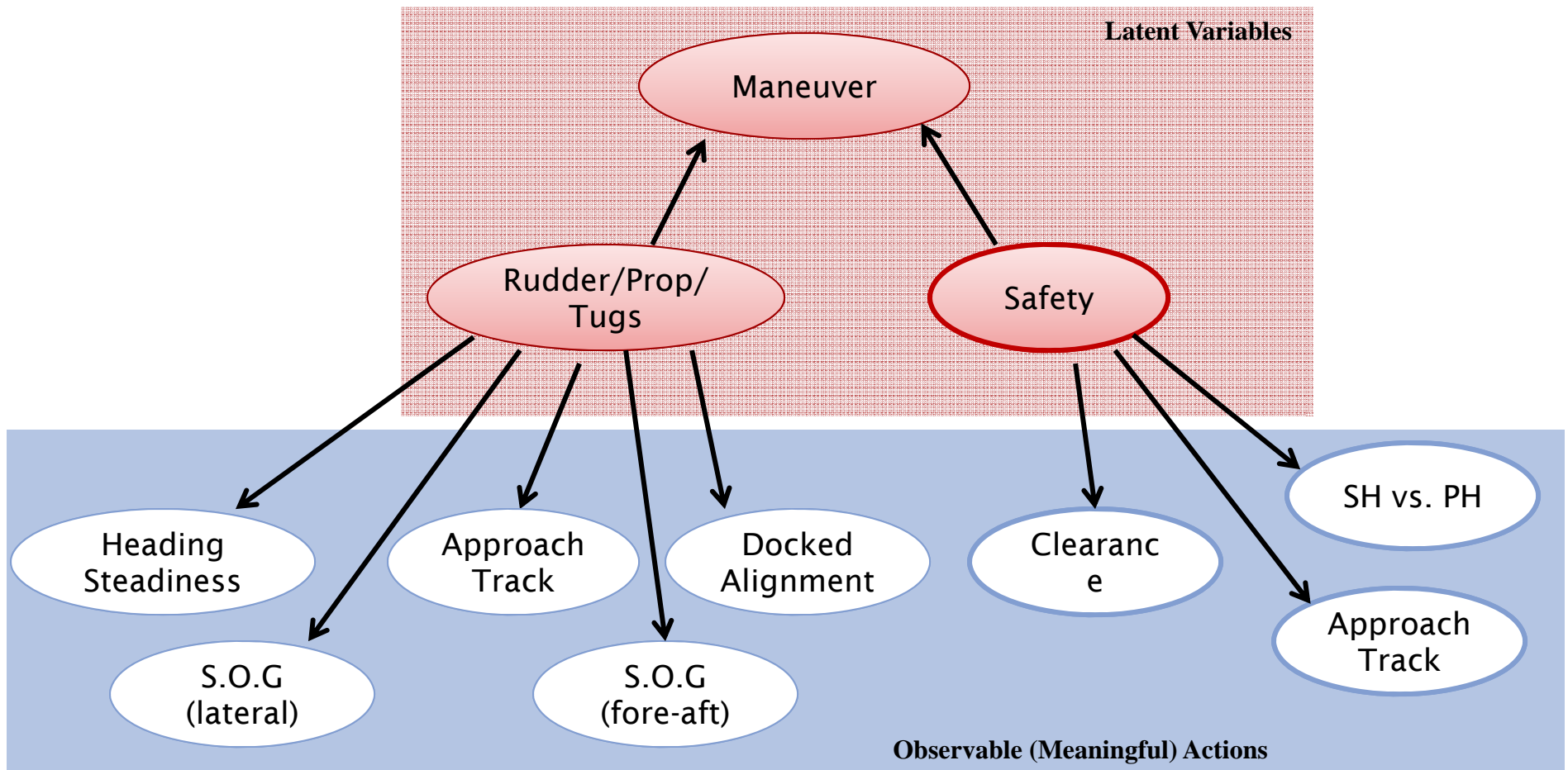
$P(\text{E} | \text{PS})$: Conditional probability. Probability of mastery of skill **Observable Action 4**, given information about mastery of skill **Latent Variable 3**

$P(\text{E})$: Probability of mastery of concept **Observable Action 4**

$$P(\text{PS} | \text{E}) = \frac{P(\text{PS}) \times P(\text{E} | \text{PS})}{P(\text{E})}$$



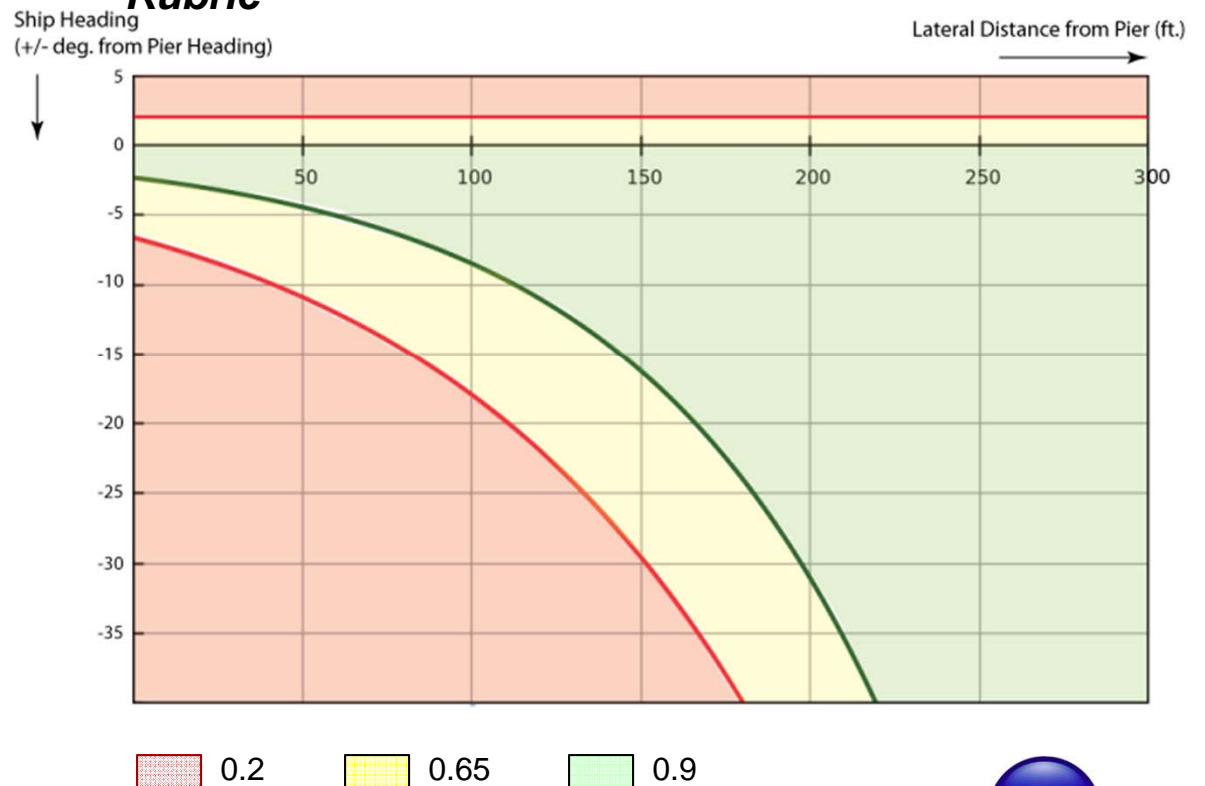
Shiphandling Mooring Bayesian Network



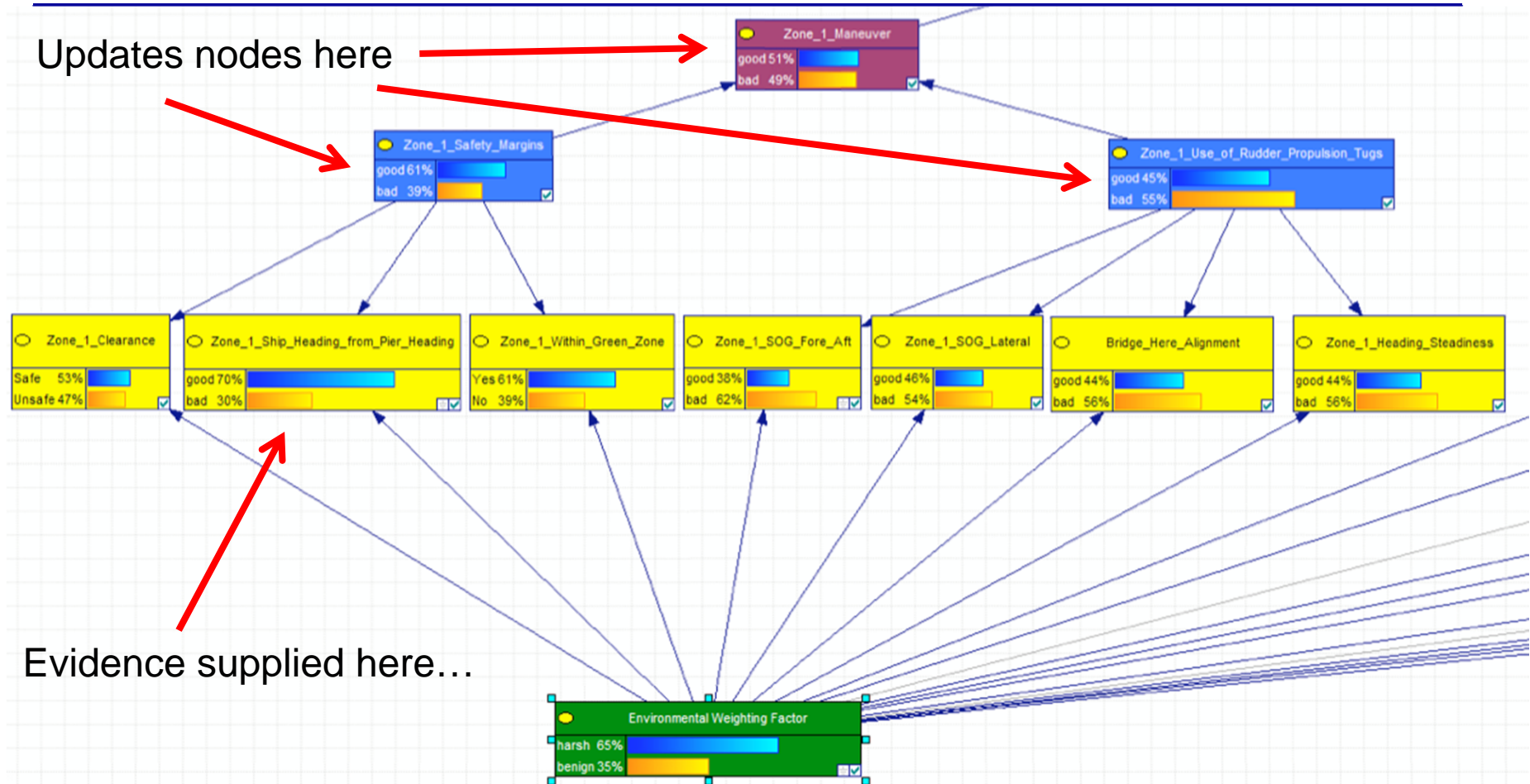
Develop Rubrics for Observable Actions

- ❑ Each observable action (node) in the Bayesian network is “evaluated” using a rubric.
- ❑ Depending on the node, the evaluation can be triggered by:
 - an action / event (i.e. a collision event)
 - a time interval (i.e. evaluate ship’s heading every 20 seconds)
- ❑ The result of each evaluation is a score, which is fed into the Bayesian network.

Example: Ship Heading from Pier Heading Rubric



Bayesian Network (Inferences) Update in Real-Time



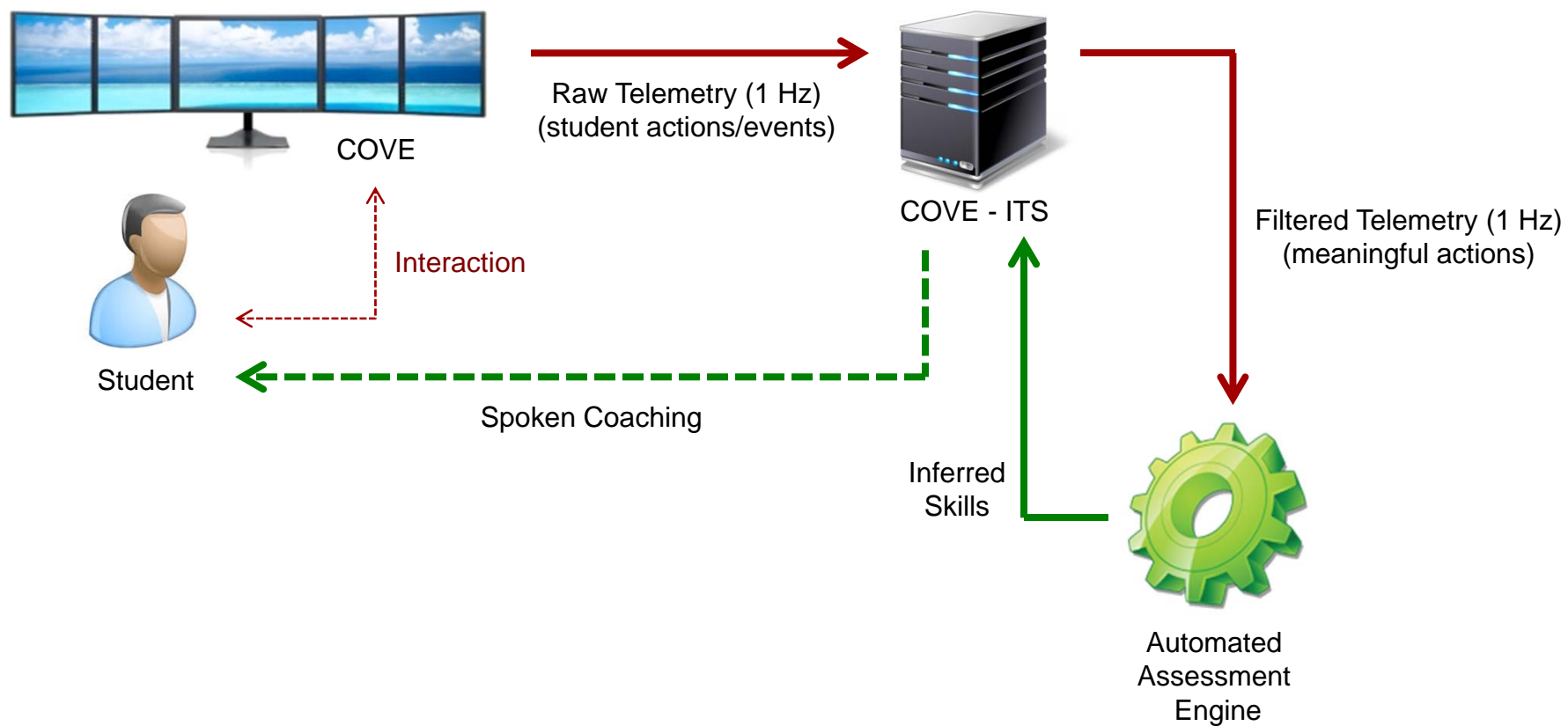
The Automated Assessment Engine in Action



Simulator Components:

1. Conning Officer Virtual Environment (COVE)
A high, fidelity shiphandling simulator used by the Surface Warfare Officer's School (SWOS) in Newport, RI.
2. Intelligent Tutoring System (COVE-ITS)
Provides spoken coaching to student based on observed actions.
3. Automated Assessment Engine
Evaluates observed (meaningful) actions, and infers latent skills of student

COVE Assessment: System Architecture

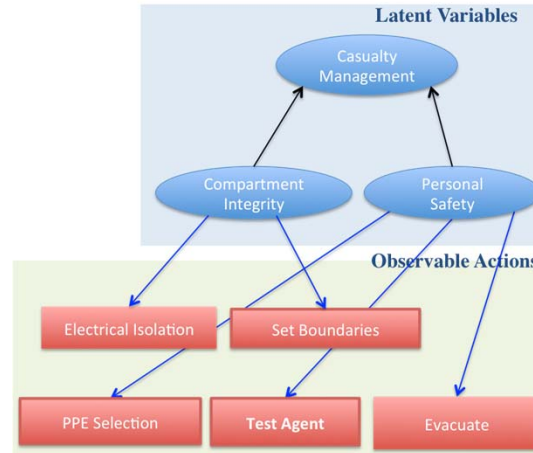


Validation: Does AAE Match Expert Conclusions?

Subjective Scoring

Game	Compartment Integrity (10%)	Personal Safety (20%)	Casualty Management (30%)	SA/DM (30%)	Checksheet Adherence (10%)
G-21_03-06-2012_06-32	9	4	5	5	8
p-02_03-06-2012_06-18	6	5	4	4	6
G-08_03-06-2012_06-24	9	8	8	8	9
G-13_03-06-2012_06-22	9	8	8	7	9
G-03_03-06-2012_06-01	1	1	1	1	9
G-24_03-06-2012_06-22	9	6	7	6	5
G-08_03-06-2012_05-58	6	4	4	5	9
G-04_03-06-2012_06-08	1	5	1	4	4
G-21_03-06-2012_06-16	9	2	5	4	7
G-04_03-06-2012_06-30	9	5	5	5	6
p-18_03-06-2012_06-06	7	7	6	6	4
G-16_03-06-2012_06-26	9	8	7	7	8
G-13_03-06-2012_06-06	6	2	5	4	4

Match??



Bayesian Network Analysis



Automated Assessment Engine



Observed Performance



What *really* matters to experts?




















Expert






Validation: Does AAE Match Expert Conclusions?

Reliabilities of Scoring between Master Mariners and Bayesian Network (BN)

Sample Size: n = 9

		Cronbach alpha		Krippendorff alpha	
		2 master mariners	2 mm +BN	2 master mariners	2 mm +BN
	Maneuver	0.952	0.897	0.905	0.597
	Safety Margins	0.978	0.869	0.957	0.667
	Use of Rudder, Propulsion, and Tugs	0.943	0.883	0.870	0.096
 	Clearance	0.326	0.463	0.157	0.19
 	Ship Heading from Pier Heading	0.906	0.631	0.837	0.385
 	Approach Track	0.662	0.703	0.463	0.161
 	Speed Over Ground (fore-aft)	0.604	0.539	0.464	-0.11
 	Speed Over Ground (lateral)	0.677	0.609	0.534	-0.102
 	Heading Steadiness	0.632	0.569	0.487	-0.145
 	Bridge Here Alignment	0.938	0.931	0.788	0.562

-  Latent (inferred) skill mastery
-  Implicitly scored by instructors using Navy rules/rubrics
-  Explicitly scored by BN based on Navy rules/rubrics



National Center for Research on Evaluation, Standards, & Student Testing

UCLA | Graduate School of Education & Information Studies

VISIT US ON THE WEB
cresst.org

Copyright © 2013 The Regents of the University of California