UCLA | CRESST NATIONAL CENTER FOR RESEARCH ON EVALUATION, STANDARDS, AND STUDENT TESTING

CRESST SHIPHANDLING AUTOMATED ASSESSMENT ENGINE: MOORING AT A PIER

Alan D. Koenig
John J. Lee
Markus R. Iseli

MAY 2016

CRESST REPORT **852**

CRESST Shiphandling Automated Assessment Engine:
Mooring at a Pier

CRESST Report 852

Alan D. Koenig, John J. Lee, and Markus R. Iseli
CRESST/University of California, Los Angeles

May 2016

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported by grant number N00014-13-C-0331 from the Office of Naval Research with funding to the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

To cite from this report, please use the following as your APA reference: Koenig, A. D., Lee, J. J., & Iseli, M. R. (2016). *CRESST shiphandling automated assessment engine: Mooring at a pier* (CRESST Report 852). Los Angeles, CA: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

CRESST SHIPHANDLING AUTOMATED ASSESSMENT ENGINE: MOORING AT A PIER[1]

Alan D. Koenig, John J. Lee, and Markus R. Iseli
CRESST/University of California, Los Angeles

Abstract

To meet the challenges of training shiphandling skills more effectively, the U.S. Navy seeks to automate the assessment of shiphandling skills to allow for less supervised practice and, therefore, reduced instructor load. As part of a broader initiative at the Surface Warfare Officers School (SWOS), CRESST has been working to develop this capability using an automated assessment engine (AAE), which infers student shiphandling proficiency based on meaningful, observable actions. This report describes the rubrics used in the AAE, as well as the inference model used therein. Plans for a future validation study are also outlined.

Introduction

Despite its critical importance to the U.S. Navy, shiphandling training is becoming increasingly more challenging as Surface Warfare Officers are spending less time at sea, and consequently experiencing fewer opportunities to control (conn) a ship at sea while under the apprenticeship of a more experienced master mariner. This can lead to less confidence and diminished shiphandling competence among Surface Warfare Officers, which in turn increases the Navy's risk for accidents (collisions with other vessels, allisions with fixed objects like piers or buoys, groundings, etc.) when carrying out mission-critical tasks.

To address these concerns, the Navy has invested in the Conning Officers Virtual Environment (COVE) and the associated intelligent tutoring system—COVE-ITS (Wong, Kirschenbaum & Peters, 2010)—that can be paired with it. Together, these synthetic training systems provide opportunities for students to practice shiphandling tasks with spoken coaching and feedback, based on their actions. However, in its current form, the COVE-ITS is unable to provide deeper, more aggregated assessments of student performance, such as those that could (1) identify skill areas that require remediation, and (2) identify root causes for misconceptions or gaps in knowledge as they contribute to diminished skills.

[1]We thank our Office of Naval Research Program Manager, Ray Perez; our Assistant Director at CRESST, William Bewley; the following people from the Surface Warfare Officers School (SWOS) for their tremendous support—Bud Weeks, Phil LaPointe, current SWOS Executive Richard Callas, former SWOS Executive George Ponsolle, Jim Marion, and other SWOS N72 staff; our consultant and former SWOS Executive Director David Monroe; Jason Wong, Susan Kirschenbaum, and Lauren Ogren (Naval Undersea Warfare Center); and COVE-ITS developers Stanley Peters, Elizabeth Bratt, and Jia Huang (Stanford).

Based on earlier work with shipboard damage control (Iseli, Koenig, Lee, & Painess, 2010; Koenig, Lee, Iseli, & Wainess, 2009), the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles (UCLA) has developed an assessment engine to complement and enhance the spoken coaching and feedback capabilities of the COVE-ITS. The integration of COVE-ITS and the CRESST automated assessment engine enables instructors (and students) to more accurately and efficiently identify strengths and weaknesses within the shiphandling domain, and guide remediation (when necessary) to focus on root cause misconceptions and/or gaps in knowledge.

The automated assessment engine (AAE) is a software module that receives telemetry from the COVE-ITS and, from this information, assesses (and infers) student shiphandling proficiency. The AAE makes use of a probabilistic graphical model (i.e., a Bayesian network) that represents "assessment relevant" constructs and variables, along with their conditional dependencies to one another. This assessment framework makes probabilistic inferences of student mastery of various shiphandling skills. It accomplishes this by evaluating observable student actions and decisions in their situational contexts (via a data feed from the COVE-ITS), and by propagating this information through the network—linking these observable actions to associated, but more latent variables (i.e., shiphandling skills).

Once every second, the COVE-ITS sends information about all relevant, observable simulation states (such as ship heading, distance from pier, and clearance) to the AAE, which uses a probabilistic graphical model to infer student proficiency probabilities from these observations. Thus, the AAE is capable of answering questions like the following:

> If we observe the student do X, what does that tell us about the student's understanding/mastery of Y (or Z, or …)?

The AAE expresses its findings as probabilities of proficiency. For example, a Maneuver score of 0.83 means:

> Based on what the system has observed, the system estimates an 83% probability that the student is proficient in the Maneuver skill area.

Specific Skills Assessed

As a proof of concept, CRESST was initially tasked with assessing the mooring evolution[2] (i.e., landing at a pier) for the DDG (guided missile destroyer class ship). The DDG is the most common ship in the Navy, with 62 currently active in the fleet (Petty, 2016). The specific skills

[2]SWOS uses the term *evolution* to describe the collection of shiphandling skills that pertain to a broader task.

assessed by the AAE are in Table 1. Both Safety Margins and Rudder/Propulsion/Tugs are higher level constructs, with Maneuver being the highest level.

Table 1

Shiphandling Skills Assessed in the Automated Assessment Engine

Skill	Definition
Clearance	Ability to maintain safe distances from ships, buoys, piers, and anything else that poses a collision hazard to the ship
Ship Heading From Pier Heading	Ability to keep the ship's heading appropriately oriented relative to the pier heading
SOG Steadiness	Ability to consistently maintain appropriate speed over ground during the evolution
Heading Steadiness	Ability to set and maintain a controlled, stable heading during the evolution
Ship's Track	Degree to which the ship's position stayed within acceptable bounds during the evolution
Alignment to Bridge Here Sign (Mooring only)	Extent to which the ship is appropriately aligned with the Bridge Here sign as it completes its mooring to the pier
Safety Margins	Overall ability to keep the ship safe throughout the entire evolution
Rudder/Propulsion/Tugs	Overall proficiency with the use of rudder, propulsion, and tugs (if applicable) throughout the entire evolution
Maneuver	Overall proficiency with shiphandling maneuver skills throughout the entire evolution

AAE Probabilities Vs. COSA Scores

The Conning Officers Shiphandling Assessment (COSA) tool, developed earlier by CRESST, is currently used at the Surface Warfare Officers School in Newport, Rhode Island. COSA is a paper-based form used by SWOS instructors to provide point values describing the student's observed level of achievement. Conversely, the AAE is a computer-based system that provides probabilities (based on observations) that the student is proficient in one or more skill areas.

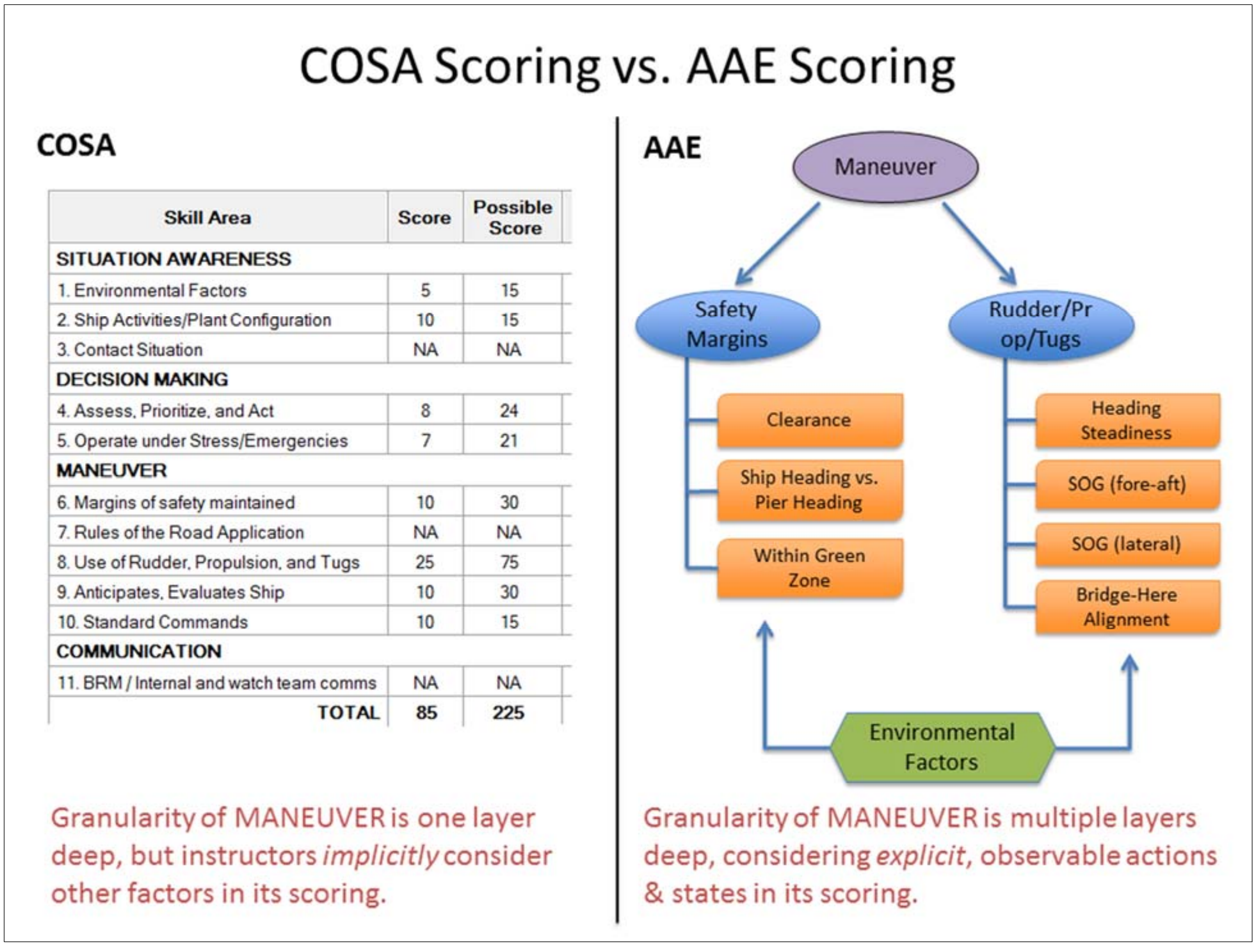


Figure 1. COSA and AAE comparison.

COSA provides scoring only for higher order shiphandling skills, such as Safety Margins, or Rudder, Propulsion, and Tugs (see Figure 1). These higher order skills usually cannot be measured directly, but instead must be inferred from observing and aggregating the student's actions (i.e., lower order skills) over time and in context. In COSA, this type of evaluation is performed *implicitly* by the expert instructor where all of the observed student actions are evaluated subjectively, the conclusions of which drive the COSA scoring.

In contrast, with the AAE, this process is done *explicitly* via computer using streaming telemetry of moment-to-moment student actions provided by the COVE-ITS, and thus scoring does not require a human evaluator. So, while COSA requires some subjectivity on the part of instructors to assess lower order skills, the AAE uses objective rubrics for scoring lower level skills (observables), and Bayesian statistics for deriving higher order scores (i.e., probabilities). Consequently, the granularity of a COSA score (like the Maneuver skill) is only one layer deep, requiring instructors to implicitly consider other factors when scoring. In contrast, the granularity of the Maneuver skill in the AAE is multiple layers deep, with explicit, observable actions and states taken into consideration when scoring.

Finally, to convert AAE scores to COSA scores, the grain size of the score must be reduced through a standard-setting process. For example, an AAE-produced probability in the range of 0.75 to 1.0 might correspond to a score of "proficient" in COSA.

New Constructs

In order to fully quantify and categorize the observable student action/event data arising out of the mooring evolution, it was necessary to create two new constructs (not currently represented in the COSA scoring). These two constructs are (1) the notion of a green zone, and (2) the concept of approach zones.

Green Zone

For this work we introduced a new construct, called the green zone, which is depicted in Figure 2. This construct is considered a part of situational context when scoring observable actions/states.

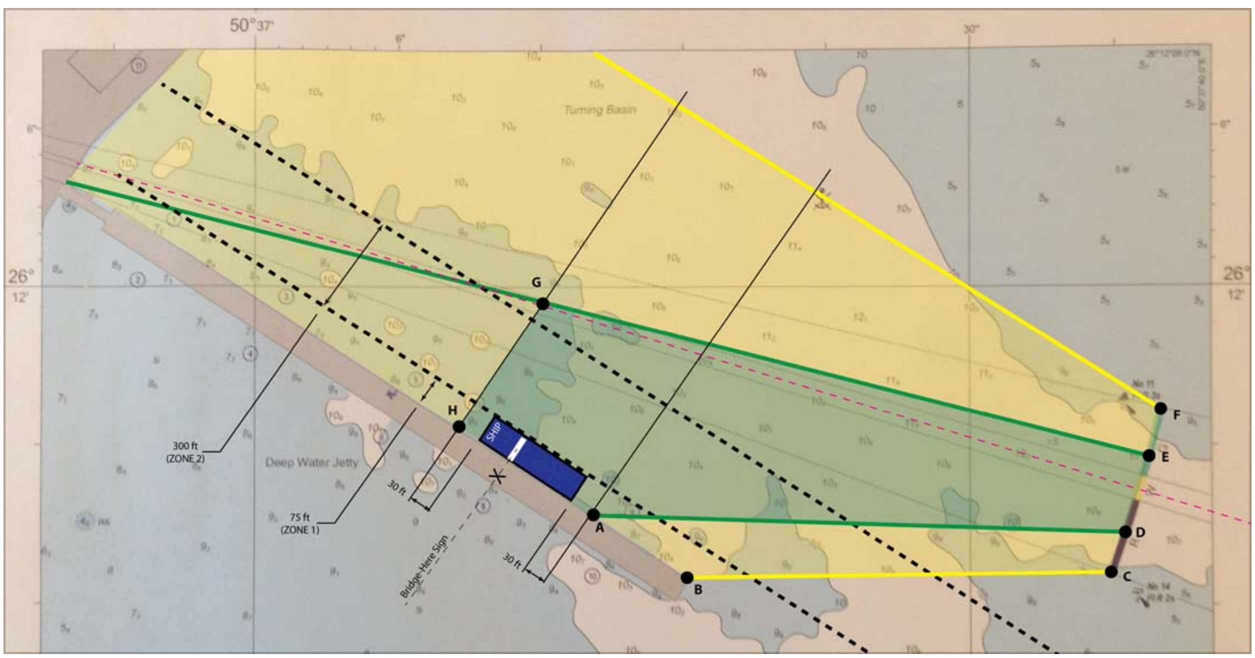


Figure 2. Green zone (approach track) for mooring at a pier.

The green zone is constructed from pier geometry and defines a region of "acceptable" positioning of the ship as it approaches the pier. In the figure, Points D and E represent the buoys marking the entrance to the harbor. A line is drawn from the top buoy at Point E to the point where the pier meets the quay wall. Another line is drawn from the lower buoy at Point D to Point A. Point A represents the maximum acceptable distance astern of the ship based on the ship's ideal docked position. Ideally, when docked, the ship's bridge should align with the Bridge Here sign marked by an X. A line is also drawn 30 feet in front of the ship (maximum acceptable distance) once it reaches its landing position out to the line drawn earlier from the first buoy to where the pier meets the quay wall (G to H). The resulting polygon, ADEGH, is defined as the green zone. Ideally the ship should remain in the green zone and avoid shoal water (shallow water where the ship could run aground) during its entire course toward the pier.

Approach Zones

We also defined a new construct to represent the zones that the ship passes through on its way toward the pier. This construct is called approach zones. This concept was derived based on discussions with subject matter experts (SMEs), who are both master mariners and resident shiphandling instructors at SWOS.

There are three approach zones (see Figure 3), each based on perpendicular distance from the pier. Zone 3 is the outermost zone, spanning the point where the tug initially meets up with the ship (usually 1000 feet from the pier) to the outermost edge of Zone 2, which is set at 300 feet from the pier. Zone 2's other bound is at 75 feet from the pier. Finally, Zone 1 is defined as 75 feet or less from the pier.

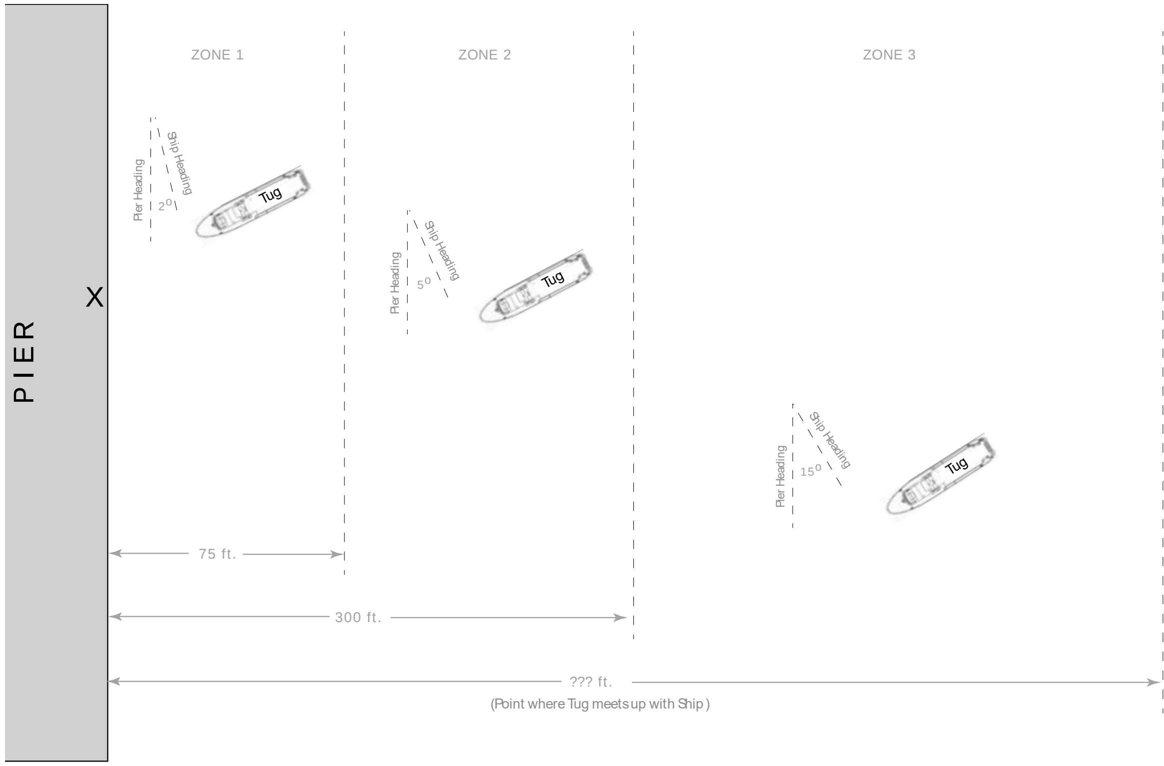


Figure 3. Approach zones for mooring.

As the ship moves through each zone and gets progressively closer to the pier and other docked ships, its situational context changes. So when assessing performance, it's important for the AAE to consider the observed actions and events in the appropriate situational context in which they occur and to determine if the student has good ship control or is weaving between zones. For this reason, each construct (or shiphandling proficiency) measured and inferred by the AAE is analyzed using rules and rubrics that vary slightly, depending on which zone the ship was in when the assessment occurred. For example, the construct of Heading Steadiness will use slightly different assessment criteria for Zones 1, 2, and 3.

AAE Visualization Tool

The AAE visualization tool was developed to show a real-time update of the resulting BN values (i.e., inferred shiphandling proficiencies) for each skill area over time. This interactive tool is used for detailed analysis. It can display moment-by-moment scoring of both observed and inferred skills, and enables investigation of single points in time across multiple screens.

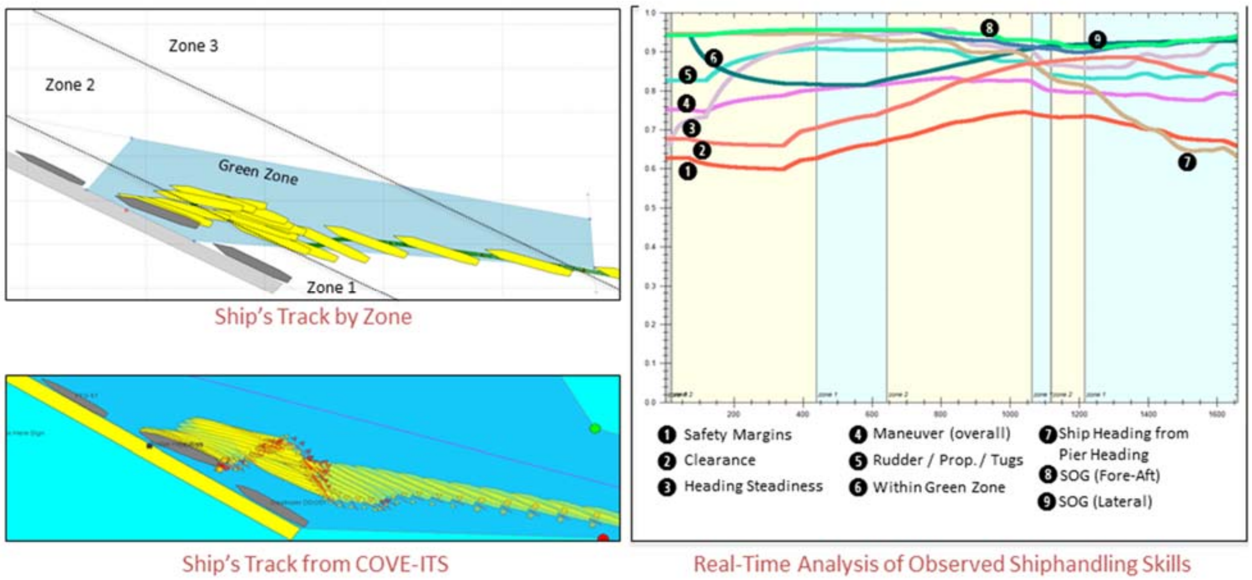


Figure 4. AAE visualization tool.

Figure 4 shows excerpts from the AAE visualization tool. The graphics on the left show the ship's track by zone (top left) and the ship's track from COVE-ITS (bottom left). In the scenario above, the student-controlled ship must land between two docked ships—analogous to parallel parking, but with a much larger vessel. The chart to the right shows the real-time analysis of observed shiphandling skills. The actual output is delineated by color, and hovering the mouse over a line shows the associated skill. The shaded regions in the chart delineate movements in and out of different approach zones. The y-axis shows Bayesian network probabilities computed by the AAE and the x-axis represents evolution time (in minutes).

The Data Model: Under the Hood of the AAE

The underlying data model for the AAE is a Bayesian network (BN), which is a directed, acyclic graph, composed of nodes and links. Nodes represent assessment-relevant student skills and links define their interrelationships. At any time during the student's engagement with the simulator, each node contains a value (between 0 and 1) that represents a probability of the student's proficiency in the node's associated skill.

The design of the AAE inference engine presented two main challenges: (1) Since observed student action measures (e.g., the ship's heading relative to the pier) could not be scored as right or wrong but rather had to be evaluated on a continuum, we established mathematical functions (rubrics) to map action measures to continuous scores, which were then fed to the BN. (2) Since we did not have enough training data, the BN's conditional probability tables had to be defined iteratively with the input of SWOS SMEs. Note: The structure of the BN was derived manually as is described below.

We considered two types of nodes (skills) for the BN described in this report: those skills that are *directly observable* (as explicit telemetry from the COVE-ITS), and those skills that are *latent*, meaning they are not directly observable but instead must be inferred from other nodes.

The structure of the BN for the Maneuver skill is shown in Figure 5. The node Maneuver is dependent on the Level 2 nodes Safety Margins and Rudder, Propulsion, and Tugs. As such, the nodes Safety Margins and Rudder, Propulsion, and Tugs are factors of, and therefore influence, the top-tier latent node Maneuver. Each of these Level 2 higher order skills, in turn, depends on lower order skills. Safety Margins, for instance, depends on Clearance, Ship Heading From Pier Heading, and Approach Track and the latent node of Rudder, Propulsion, and Tugs depends on observing the ship's Speed Over Ground (SOG) (both fore-aft and lateral), Heading Steadiness, and Alignment to Bridge Here Sign. This conceptual hierarchy is shown in Figure 5 below. These skills depend on context (i.e., the ship's current approach zone) so a separate, but identically structured, Bayesian network is created to determine conceptual proficiency for each of the three zones.

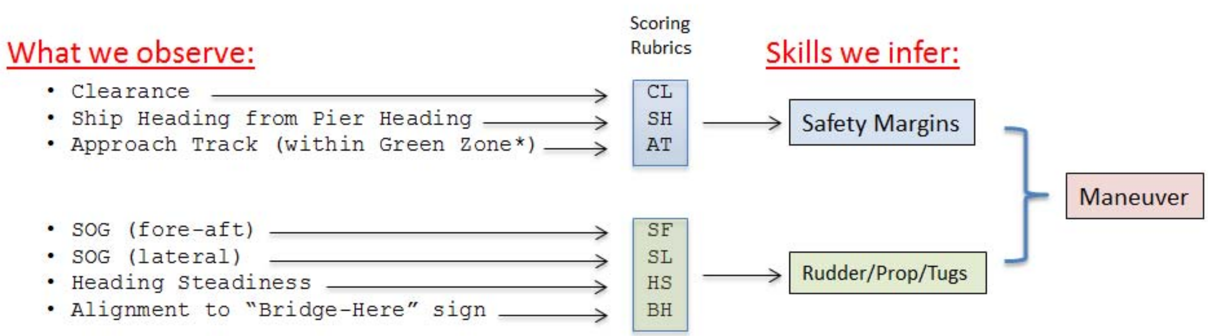


Figure 5. Data model from observables to inferred skills.

Scoring Rubrics Associated With Observable Variables

Scoring of meaningful observed actions/events (i.e., the directly observable nodes described above) requires the use of rubrics. Each action has its own rubric, and each rubric takes into account the situational context—or state of the world—during which the action occurred. As previously described, the COVE-ITS scores observable actions (using the rubric) every second. These scores, expressed as continuous values between 0 (bad performance) and 1 (perfect performance), are passed into the BN where they are associated with their corresponding observable nodes. As this occurs, the probabilities of all other connected nodes in the network (both directly observable types and latent types) are immediately updated to reflect this new evidence about the student. In this way, as the student progresses through the simulation, a stream of constantly updating inferences is made regarding the student's shiphandling skill/proficiency.

Each scoring rubric was developed by CRESST based on repeated consultation with subject matter experts at SWOS. The following sections describe the specifics of each rubric.

Clearance Rubric

The Clearance rubric evaluates how close the ship is relative to other objects that the ship could potentially collide with, such as buoys, other ships, and the pier or land. This rubric uses a construct called the danger zone radius, which is defined as the distance away from the ship's current position that the ship will be in 20 seconds, assuming it maintains its current speed and heading.

Figure 6 shows the Clearance calculation along with a table of scoring rules used to update the Clearance node BN probability. In essence, if another ship (or buoy/pier) enters the danger zone radius, meaning the student-controlled ship is getting too close, the table in the figure is used to determine the BN probability. If the ship collides with another ship (an allision), then the BN probability is automatically set to 0.1, no matter what else occurs, representing a single-point failure. In the COVE, when a collision or allision occurs, the scenario is over, and the student is not yet ready, since they have not demonstrated proficiency with the complex skills involved.

Clearance Rubric

Criteria to Assess Clearance:

1. Calculate "danger zone radius" (i.e. distance away from current position that ship will be in 20 seconds) = COG * 20.
2. Check if any ships are within "danger zone radius" around the current position.
3. If one or more ships inside "danger zone radius", use the table below to determine BN score based on the ship with the shortest distance from ownShip. Otherwise, DO NOT set evidence.

COG Speed of OwnShip (v)	Optimal (0.9)	Adequate (0.65)	Poor (0.1)
< 0.5 kts (slow)	> 50 ft	10 ft < dist. <= 50 ft	< = 10 ft
0.51 kts < v <= 3 kts	> 100 ft	20 ft < dist. <= 100 ft	< = 20 ft
> 3 kts (fast)	> 150 ft	30 ft < dist. <= 150 ft	< = 30 ft

NOTES:

1. For buoy's, if clearance > 25 ft., score is 0.9, otherwise, score is 0.1
2. For pier, table values don't apply, but if ship hits pier, give score of 0.0
3. If a score of 0.1 is received, overall score (i.e. computed running average score) is 0.25 no matter what else occurs.
4. If score of 0.0 is received, overall score (i.e. computed running average score) is 0.1 no matter what else occurs.

Figure 6. Clearance rubric.

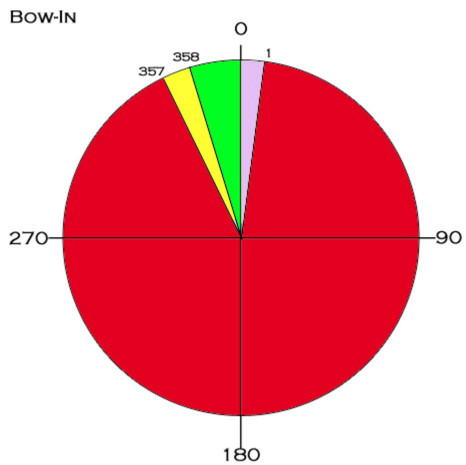
Ship Heading From Pier Heading Rubric

The Ship Heading From Pier Heading rubric takes into account maintaining a ship heading that's close (but slightly offset) to the pier heading, with progressively stricter precision required the closer the ship gets to the pier. To reflect this, a graph rubric is used to create regions where the BN values assigned are based on the lateral distance from the pier in relation to the ship heading relative to the pier heading. When carrying out the mooring maneuver, two factors must be considered to determine the correctness of the ship's heading relative to the pier heading. These factors are landing and bow orientation. Landing refers to the side of the ship that is closest to contacting the pier—that is, port or starboard. Bow orientation refers to whether the ship approaches the final landing position with the bow aimed out to sea (i.e., away from land), which would be bow-out, or if the ship's bow is aimed toward the shore, which would be bow-in.

In either case, the bow of the ship should be angled in toward the pier as the ship approaches. As the ship gets closer to the pier, the shiphandler must maneuver the ship such that the heading eventually perfectly matches that of the pier. Figure 7 below shows the scoring used for assessing the relative ship's heading to the pier heading as the ship approaches the pier.

Notice that the absolute angles considered differ based on bow-in vs. bow-out, and on port-side landing vs. starboard-side landing. These scores are supplied to the BN as input (observable) values, and correspond to the regions listed in the graphs.

Bow-In, Port-Side Landing

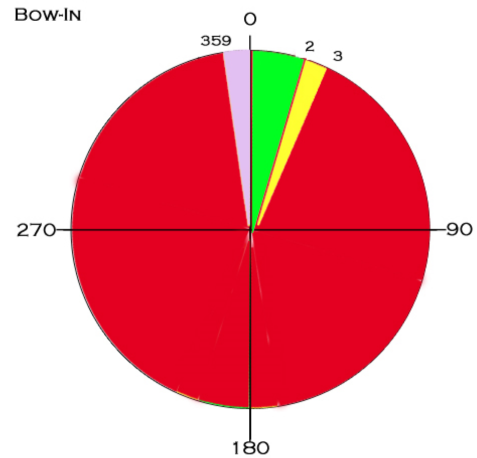


Degree values represent transformed ship heading (TSH), assuming pier heading of zero degrees.

Zone 1 Scoring:

- BN score = 0.9 for 358 <= TSH <= 0
- BN score = 0.65 for 357 <= TSH < 358
- BN score = 0.4 for 0 < TSH <= 1
- BN score = 0.1 for 1 < TSH < 357

Bow-In, Starboard-Side Landing

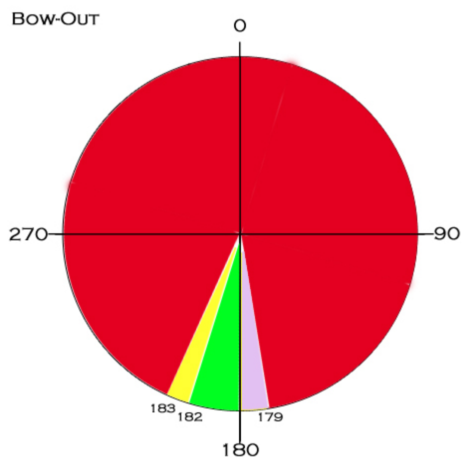


Degree values represent transformed ship heading (TSH), assuming pier heading of zero degrees.

Zone 1 Scoring:

- BN score = 0.9 for 0 <= TSH <= 2
- BN score = 0.65 for 2 < TSH <= 3
- BN score = 0.4 for 0 < TSH <= 359
- BN score = 0.1 for 3 < TSH < 359

Bow-Out, Port-Side Landing

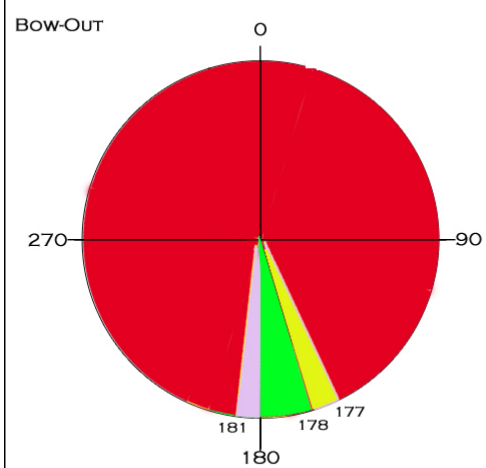


Degree values represent transformed ship heading (TSH), assuming pier heading of zero degrees.

Zone 1 Scoring:

- BN score = 0.9 for 180 <= TSH <= 182
- BN score = 0.65 for 182 < TSH < 183
- BN score = 0.4 for 179 <= TSH < 180
- BN score = 0.1 for 0 < TSH < 179 OR 183 < TSH < 360

Bow-Out, Starboard-Side Landing



Degree values represent transformed ship heading (TSH), assuming pier heading of zero degrees.

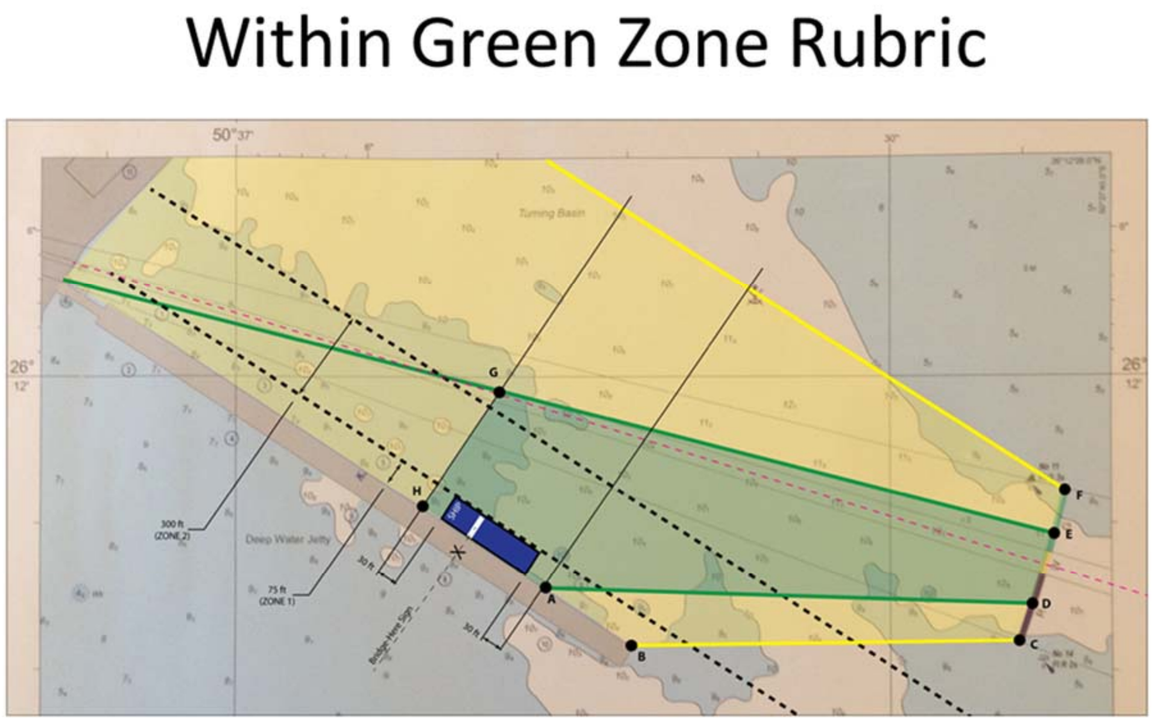
Zone 1 Scoring:

- BN score = 0.9 for 178 <= TSH <= 180
- BN score = 0.65 for 177 <= TSH < 178
- BN score = 0.4 for 180 < TSH <= 181
- BN score = 0.1 for 0 <= TSH < 177 OR 181 < TSH <=360

Figure 7. Ship Heading From Pier Heading rubric.

Within Green Zone Rubric

For the Within Green Zone rubric, the probabilities are based on the coordinates of the ship, and whether the ship was (1) not in the green zone at all, (2) partially in the green zone, or (3) fully inside the green zone (see Figure 8). These correspond to probabilities of 0.2, 0.65, and 0.9, respectively, as shown in the figure.



ZONE	Optimal (0.9)	Adequate (0.65)	Poor (0.2)
Zone 1	Within Green Zone	Partially in Green Zone	Outside Green Zone
Zone 2	Within Green Zone	Partially in Green Zone	Outside Green Zone
Zone 3	Within Green Zone	Partially in Green Zone	Outside Green Zone

Figure 8. Within Green Zone rubric.

Speed-Over-Ground (Fore-Aft and Lateral) Rubrics

For the Speed Over Ground (SOG) fore-aft and lateral rubrics (Figure 9 and Figure 10), the tables shown in Figure 9 were developed from subject matter expert (SME) input. In them, optimal, adequate, and poor distinctions are made based on the speed of the ship, and its proximity to the pier (i.e., the zone). However, in practice, scoring based on these parameters is not optimal when using an automated assessment system because the boundaries between what is considered optimal and adequate, or adequate and poor result in abrupt “jumps” in scoring as the ship transitions between zones. To address this, the rubric was transformed to make use of the graph look-ups of curved lines (shown in Figure 10), in which the y-axis indicates ship's speed, and the x-axis indicates ship's distance from the pier. This produced smoothed, continuous scoring, more closely matching the way human experts consider performance.

As the graphs indicate, slowing down as you get closer to the pier is desired, so that a collision with the pier will not occur. Conditions like wind and current can make this more difficult.

Ship SOG (fore/aft)

	Optimal	Adequate	Poor
ZONE 1	0.0 - 1.2	1.3 - 2.5	> 2.5
ZONE 2	1.0 - 3.0	0.0 to 1.0 or 3.1 to 5.0	> 5.0
ZONE 3	2.0 - 4.0	0.0 to 1.9 or 4.1 to 5.0	> 5.0

Ship SOG (lateral)

	Optimal	Adequate	Poor
ZONE 1	0.0 - 0.5	0.51 - 1.0	> 1.0
ZONE 2	0.2 - 0.8	0.0 to 0.19 or 0.81 to 1.2	> 1.21
ZONE 3	0.2 - 1.2	0.0 to 0.19 or 1.21 to 2.0	> 2.0

Figure 9. Speed Over Ground (fore-aft and lateral) rubrics, as provided by SMEs. Probability values of 0.9, 0.65, and 0.2 are assigned to optimal, adequate, and poor performance, respectively.

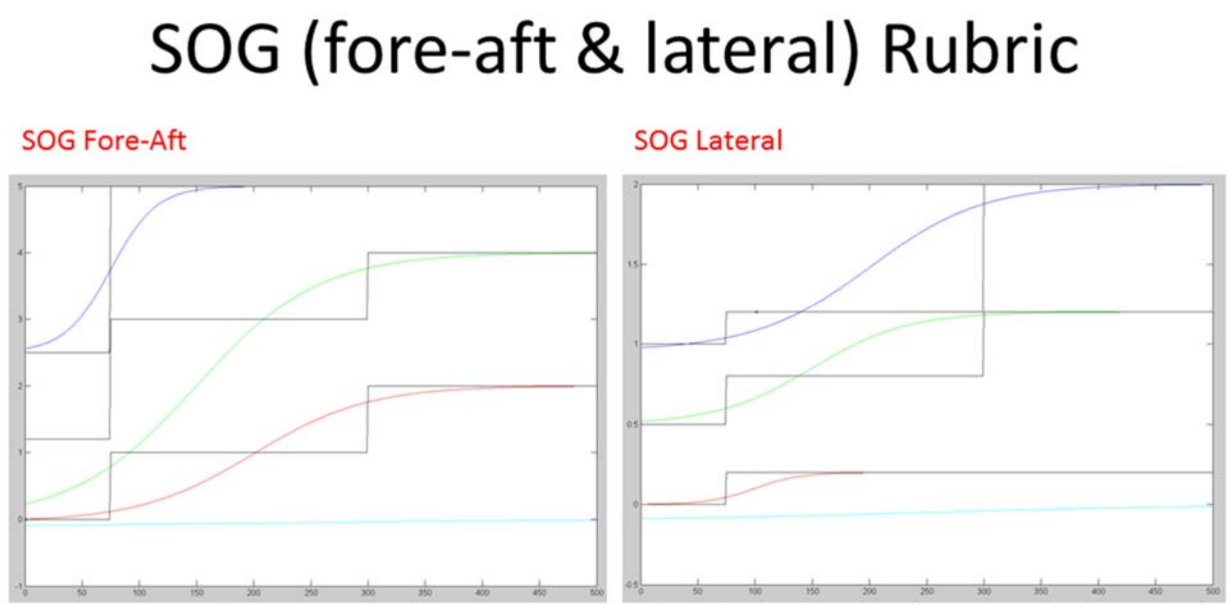


Figure 10. Speed Over Ground (fore-aft and lateral) curves—speed over ground (y-axis) vs. distance to pier (x-axis). Colored curves are smoothed versions of black step functions, which represent the values from Figure 9. SOG values above the blue curve are considered poor performance, between the green and blue curves are adequate, between the green and red curves are optimal, between the red and cyan curves are adequate, and below the cyan curve are considered poor performance.

Heading Steadiness Rubric

For Heading Steadiness, we examine how much the ship's heading changes over a 10-minute interval (see Figure 11). The less the heading changes, the more steadily the ship is moving. More efficient and skilled shiphandlers use fewer commands to maintain the ship's heading because they are able to anticipate the momentum and inertia of the ship's movement through the water, and plan their helm orders accordingly. This minimizes the need for overcorrecting. The graph, shown in Figure 11, shows BN probabilities on the y-axis, and the standard deviation of the heading (values here consist of 5 minutes of ship's heading telemetry, with ship's heading readings calculated once per second within that time span) on the x-axis. For example, Grouping 1 consists of values with time t = 0 to time t = 300 sec. Grouping 2 consists of values with time t = 1 to time t = 301 sec. Grouping 3 consists of values with time t = 2 to time t = 302 sec., etc. These plots are then used to assign probabilities based on the standard deviation of the heading (*SD*). The BN probability is calculated using a logistic function that maps the domain of *SD*s onto the range of probabilities between zero and one with the inflection point at *SD* = 5.5 degrees being scored with a probability of 0.5 (neither good nor bad performance): $1/[1+e^{(SD - 5.5)}]$.

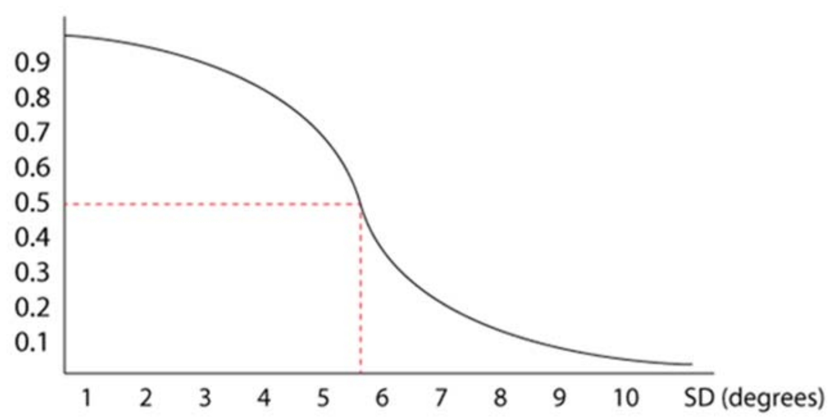


Figure 11. Heading Steadiness rubric—BN score (y-axis) vs. Heading standard deviation (deg.).

Alignment to Bridge Here Sign Rubric

The Alignment to Bridge Here Sign rubric (Figure 12) ascertains the shiphandler's ability to align the bridge of the ship to the Bridge Here sign that's posted on the pier. The tolerance on either side is plus or minus 30 feet. Note from the table in the figure that this is only evaluated in Zone 1 and the BN probabilities are based on the closeness of the bridge's alignment with the sign.

Bridge-Here Sign Alignment Rubric

Use the table below to determine the BN score based on stated +/- tolerance of ship's longitudinal (fore/aft) alignment of the bridge with the Bridge-Here sign.

ZONE	**Optimal (0.9)**	**Adequate (0.65)**	**Poor (0.2)**
Zone 1	Within 15 ft.	Within 30 ft.	> 30 ft.
Zone 2	N/A	N/A	N/A
Zone 3	N/A	N/A	N/A

Figure 12. Alignment to Bridge Here Sign rubric.

Bayesian Network for Mooring

A Bayesian network model was created for the mooring exercise using GeNie software (Druzdzel, 1999), and is shown in Figure 13. At the highest level (top middle of the figure) is the Maneuver node. Each zone consists of a Maneuver node (for that zone) as well as Level 2 nodes (Safety Margins and Rudder, Propulsion, and Tugs), which link to observable nodes (Level 3 nodes in yellow). For example, the Safety Margins node in Zone 1 is linked to the Zone 1 Clearance, Ship Heading from Pier Heading, and Within Green Zone nodes. All the observable nodes are also linked to the environmental factors (i.e., amount of wind and current present), reflecting that environmental factors can increase the difficulty of mastering all the observable nodes. In addition, weighting is applied in the BN to put more emphasis on the Zone 1 performance, as this zone is the most critical portion of the exercise since that is when the ship makes its landing at the pier.

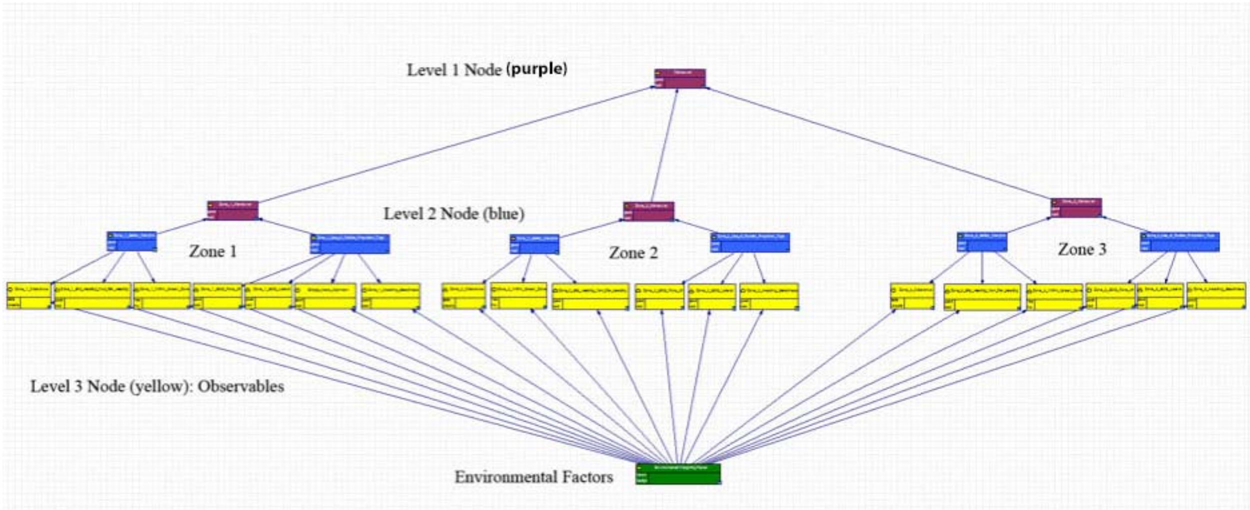


Figure 13. Mooring Bayesian network.

Planned Validation of Bayesian Network

CRESST is planning a validation study to compare the results of the Bayesian network with that of subject matter experts (SMEs). The plan is to collect 30 or more student runs (with an equal representation of bow-in, bow-out, port-side landing, and starboard-side landing). The AAE will score each run across each construct (i.e., Maneuver, Clearance, etc.). Two SWOS SMEs will jointly be presented with the BN scores along with a playback of the corresponding runs, as well as a bird's eye graphical view of the ship's track and orientation as it proceeded through the mooring process. Together, the two SMEs will use a Likert scale to express their degree of agreement (one agreed upon rating between the two of them) with the BN score for each construct. The results will be analyzed for statistically significant correlations, and where disagreement exists, the BN will be revised accordingly.

Discussion

Our methodology shows good promise. Automating the scoring of constructs from observables and aggregating them up to latent constructs is viable from a technical perspective. The validation study is needed to confirm the utility of this methodology. This has implications for real-time, automated assessment using BNs in the future since the use of Bayesian networks for assessment has proven difficult in the past and these networks have not yet been applied to an open, unconstrained, and dynamic simulation environment (e.g., see Conati, Gertner, & VanLehn, 2002; Martin & VanLehn, 1995; Remolina, Ramachandran, Stottler, & Howse, 2004). CRESST is currently working on revising the architecture of the automated assessment engine to make it accessible as a web service that the COVE-ITS (and any other future client applications) can communicate with in real time. This will provide a flexible approach for systems to receive assessed performance telemetry, which is potentially applicable to other fields as well, such as medicine (Koenig, Iseli, Wainess, & Lee, 2013) and others.

The planned next steps include a review of AAE-scored mooring runs across a range of performance, and to get the level of agreement (using a seven-point Likert scale) of the master mariners with the AAE calculated values. This will be done for the aggregated skills of Maneuver, Safety Margins, and Rudder, Propulsion, and Tugs. Work is also underway to integrate the results of the automated assessment engine with the COVE-ITS to provide more targeted spoken coaching to meet the needs of each individual shiphandler.

References

Conati, C., Gertner, A., & VanLehn, K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction, 12*(4), 371–417.

Druzdzel, M. J. (1999). SMILE: Structural Modeling, Inference, and Learning Engine and GeNIe: A development environment for graphical decision-theoretic models (Intelligent Systems Demonstration). In *Proceedings of the Sixteenth National Conference on Artificial Intelligence (AAAI-99)* (pp. 902-903). Menlo Park, CA: AAAI Press/The MIT Press.

Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010, August). Automated assessment of complex task performance in games and simulations. In *Proceedings of the Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL.

Koenig, A., Iseli, M., Wainess, R., & Lee, J. J. (2013). Assessment methodology for computer-based instructional simulations. *Military Medicine, 178*(10S), 47–54.

Koenig, A. D., Lee, J. J., Iseli, M., & Wainess, R. (2010). *A conceptual framework for assessing performance in games and simulations* (CRESST Report 771). Los Angeles: University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Martin, J., & VanLehn, K. (1995). Student assessment using Bayesian nets. *International Journal of Human-Computer Studies, 42*(6), 575–591.

Petty, D. (2016). *The United States Navy Fact File: Destroyers—DDG*. Retrieved from http://www.navy.mil/navydata/fact_display.asp?cid=4200&tid=900&ct=4

Remolina, E., Ramachandran, S., Stottler, R. H., & Howse, W. R. (2004). *Intelligent simulation based tutor for flight training.* Paper presented at the Industry/Interservice, Training, Simulation & Education Conference, Orlando, FL.

Wong, J. H., Kirschenbaum, S. S., & Peters, S. (2010). *Developing a cognitive model of expert performance for ship navigation maneuvers in an intelligent tutoring system.* Newport, RI: Naval Undersea Warfare Center Division.