Automated Event-Based Competency Analysis: Detecting Evidence from Training Data

Cristina Boyer, Tess Olson, Nolen Yehlik
The Boeing Company
Berkley, MO
cristina.m.boyer@boeing.com, tess.e.olson@boeing.com, nolen.c.yehlik@boeing.com

ABSTRACT

Evaluation of non-technical competencies is typically done on a more subjective level, relying heavily on instructors to make acute observations during a training scenario. A behavior detection prototype was created by the authors to enable more consistent evaluation of non-technical competencies, addressing the challenge of standardizing training and evaluation in this area. This paper describes how the prototype analyzes training data to detect instances where a student demonstrates behaviors that provide evidence of defined competencies. The prototype was built to focus on discrete events that can occur in a training simulation and assess a student's response to each of the events that transpire. By doing this, the system can detect and assess both planned and unplanned events during a training exercise. A further benefit of the event-based design is that context can be considered when evaluating student responses as evidence of competency, or lack thereof, providing the ability to evaluate soft skills in a more objective way. This allows the system to collect evidence of competencies such as decision making, situation awareness, and workload management, creating objective data that can be used for instructional purposes.

To evaluate the accuracy and functionality of the data generated by the prototype, the authors integrated their system with a desktop trainer simulating a Control Display Unit (CDU) during a non-normal flight scenario. A study was run where instructor participants were shown recordings of student participants to assess their competency performance in the scenario, independently of the prototype. The instructors then analyzed their own data side by side with the prototype output to evaluate and quantify their level of agreement. The results demonstrate the prototype effectively identifies evidence of competency, aligning with traditional instructor assessments, thus improving the data collected during training to enhance the training experience.

ABOUT THE AUTHORS

Cristina Boyer is a Senior Software Engineer on the Advanced Learning team for the Boeing Research & Technology (BR&T) organization in St. Louis, MO. She has more than 13 years' experience at the Boeing Company working on automated performance assessment, adaptive training, and integrated live, virtual, constructive (ILVC). She has filed 3 patents in the area of adaptive training technologies. She received Bachelor of Science degrees in Mathematics and Computer Science from Southern Illinois University Carbondale.

Tess Olson is a software engineer with Boeing Research and Technology on the Advanced Learning team. She has over 5 years of experience in the industry working on various projects ranging from AR and VR training solutions to researching speech recognition solutions that focus on training and communication between human and machine. In addition to these projects, a major focus of hers for the last few years has been a project focusing on a competency-based training system. She has also filed 1 patent in the field of adaptive training technology. She holds a Bachelor of Science in Programming from Columbia College Chicago.

Nolen Yehlik is a Learning Scientist and Human Factors Engineer on the Training and Professional Services team within Boeing Global Services Engineering. He has over 5 years of Human Factors Experience in pharmaceutical and aerospace industries. His field experience includes focuses on research and development, regulatory approval, and training efficacy assessment. Project experience includes automated performance assessment, adaptive training, model based cognitive engineering, human subjects research design, and instructor concordance. He holds a Bachelor of Science in Human Factors Engineering with a minor in Aerospace Life Sciences from Embry-Riddle Aeronautical University – Daytona Beach.

Automated Event-Based Competency Analysis: Detecting Evidence from Training Data

Cristina Boyer, Tess Olson, Nolen Yehlik
The Boeing Company
Berkley, MO
cristina.m.boyer@boeing.com, tess.e.olson@boeing.com, nolen.c.yehlik@boeing.com

INTRODUCTION

When looking at the core competencies required across military and commercial aviation personnel, there are common themes that come up like situation awareness and decision making. These are a sample of the competencies that are essential to effective performance in these fields (IATA, 2023; EASA, 2016; Renier, 2022.; United States Air Force, 2022). Competencies like situation awareness and decision making are also more subjective in nature when being evaluated. For example, while one person may consider an interaction as showing good leadership, another observer may consider the interaction rude. What might be clear communication in the mind of one observer may be lacking timing in the mind of another. It is important that the expectations of these competencies are clearly defined so that they are consistently evaluated and instructed (Goldberg et al., 2021). In addition to this, the observation and evaluation of student performance, especially during complex real-world exercises, is left to a live instructor. With the instructor shortage that the industry is currently experiencing, it is now critical to explore ways of maximizing the number of students a single instructor can teach (Koenig, 2023; Clark, 2020). In this paper we will describe the Behavior Detection System we developed to address these issues by automatically detecting evidence of student task and competency performance during scenario-based training. In addition, the paper will detail how we tested the accuracy of the data.

BEHAVIOR DETECTION SYSTEM

The Behavior Detection System was built to detect discrete events that can occur during a training exercise and analyze student responses to those events. Focusing on individual events instead of the scenario as a whole eliminates the need to define each pathway a student might take through the scenario and allows us to evaluate students on both planned and unplanned events that inevitably occur during complex, real-world exercises. Breaking the evaluation down to the event level also allows us to follow the Event-Based Approach to Training (EBAT) (Fowlkes et al. 1998). Doing this enables the system to look at a student's actions in the context of the event that they are presented with to provide an objective analysis as to whether or not the student performed optimally in response to that event. For example, if a pilot notices a malfunction in the jet they are flying, should this be communicated immediately to the flight lead? Well, that depends. Is the pilot at a critical point in the mission? How dire is the malfunction that has occurred? By defining these parameters, it is possible to clearly state the ideal response to the situation.

The Behavior Detection System incorporates EBAT by defining events with the parameters that describe how and when a learner should respond to that event. Once defined, the learner's expected response can then be associated with one or more observable behaviors that demonstrate desired competencies by their response. The Behavior Detection System then uses these event definitions to track student performance against some of the more subjective competencies like communication. The discrete way these events are defined allows them to be re-used across scenarios.

System Design

This flexibility and scalability are two of the many reasons that the Behavior Detection System was developed using the EBAT approach. The system uses authored events to evaluate actions within the context of scenario data that is gathered or sent out by a training environment (see Figure 1.). The training environment can be a computer-based simulation, physical simulation, or a live environment as long as the data can be accessed. The scenario data includes scenario specific data like weather conditions, time of day, and anything else that could impact how a learner might react to events that occur during a session. The Behavior Detection System is indifferent to where the data comes from as it is detached from and runs independently of any current simulation environment.

This allows the system to evaluate a learner solely based on their actions during each of the events that comprise a scenario. The Behavior Detection System is unaware of which events are going to occur in a scenario prior to their occurrence. It is instead able to detect these events based on established parameters that could happen during a scenario, allowing for the evaluations to be done without any preconceived notions about the content of a session. Once an event is detected, the student's actions are then evaluated against scenario context where appropriate. These evaluations are then formatted into statements following the Advanced Distributed Learning (ADL) Initiative's eXperience Application Interface (xAPI) and stored in a Learning Record Store (LRS) (see Figure 1). xAPI is a JSON based specification that is primarily used for eLearning and an LRS "defines the communication method for sending, receiving and requesting data" (Rustici Software, n.d.; Sottilare et al., 2017). The evaluations that are created and stored include how each of the interactions are related to one or more observable behaviors that demonstrate specific competencies. This allows the student's demonstrated competency performance to be retrieved from the LRS for instructional purposes such as after-action reviews and feedback during a lesson. It also enables the data to be incorporated into implementations of the ADL Initiative's Total Learning Architecture (TLA), providing the ability to analyze the competency data from the Behavior Detection System along with data from other learning activities to evaluate student proficiency (ADL, n.d.).

Figure 1. Behavior Detection System Data Flow

ACCURACY STUDY

The competency performance data generated by the Behavior Detection System is intended to be used for both instructional purposes and for tracking student proficiency in the TLA. Due to this we ran an exploratory study to determine whether the data generated by this system could be trusted to be accurate.

Study Hypothesis

The study focused on evaluating the Behavior Detection System's ability to detect events that occur, and actions performed by a student pilot in a simulated flight scenario. Then log when a student's actions demonstrate observable behaviors of a competency. There are two hypotheses. The first hypothesis states the Behavior Detection System can accurately detect events within the simulation and the actions performed by a student pilot. The second hypothesis posits the Behavior Detection System would detect observable behaviors following a general consensus of instructor pilot subject matter experts (SMEs) when evaluating the same scenario. These hypotheses will be tested through comparative analysis to assess the system's performance and alignment with expert evaluations.

Research Questions

During development of the study, several research questions were proposed to further explore the capability, accuracy, and functionality of the Behavior Detection System. The primary research questions guiding the study were: How accurate is the system's ability to identify the observable behaviors authored for the events that might occur during the training scenario? And how accurate is the system logic used to identify the observable behaviors compared with a live instructor?

In addition to the primary questions above, other questions of interest regarding consistency were identified. How close are the measurements of the same item to each other across simulations when the Behavior Detection system is identifying behaviors demonstrated by student activity? And how do the categorical judgements compare between the instructors when comparing their training observations before and after being shown the results from the behavior detection system.

These research questions were designed to comprehensively explore the accuracy, consistency, and the systems effect on instructor judgement of the Behavior Detection System. By addressing these questions, the study aimed to provide a thorough understanding of the system performance and effectiveness in identifying observable behaviors during student activity within a training simulation.

Research Study Scope

The study had defined the following in-scope elements to focus the approach on the accuracy and functionality of the Behavior Detection System. Specifically, the study aimed to compare the automatic observations of completed or incomplete observable behaviors generated by the Behavior Detection System to the observations of live pilot instructor subject matter experts (SMEs).

However, there are certain elements removed from the scope of the study. These unassessed elements are: The study did not address the training environment itself, including factors such as the design of the simulation or the effectiveness of the training program. The study also did not address student performance of instructor performance in terms of their abilities or skills. Additionally, the study does not delve into grading and evaluation tools, nor does it explore the instructor feedback provided to students. These out-of-scope elements have been intentionally excluded to maintain a focused approach on the comparison between the automatic observations generated by the Behavior Detection System and the instructor SME evaluations.

Research Study Schedule

The data collection took place between August and September of 2023. The representative student participants, who were Boeing employed former airline pilots, completed the study within a single time block of approximately 2 hours. The instructor study participants, who were Boeing employed flight instructors, completed the study within approximately 4 hours. The instructor participants were given the option to complete the study in a single 4-hour time block or divide their participation into two separate time blocks approximately 2 hours each.

The study was conducted virtually to accommodate the recruitment of instructor pilot participants. Multiple Boeing campus locations were selected to run the study including St. Louis, MO., USA; Miami, FL., USA; London, ENG; and Singapore, SG. By conducting the study within the specified time frame and utilizing virtual platforms, the research team was able to collect data from both student and instructor participants across different global locations ensuring a comprehensive evaluation from participants with varied experience, backgrounds, and views.

Study Participants

Two cohorts of participants were recruited for the study, consisting of representative flight student and flight instructor participants. A total of n=2 representative flight students, who were former airline pilots employed in other roles within Boeing, and n=8 Boeing employed flight instructors, were recruited to participate within the study.

Student representative participants were selected based upon their prior training and flight experience. The cohort attributes of the student group were employment status with The Boeing Company, having previously been employed as a pilot, previous experience with the desktop simulation trainer (DST), and their flight experience including aircraft model, recency, role, and flight hours.

Instructor participants were selected based upon their prior instructing experience. The cohort attributes of the instructor group were employment status with The Boeing Company, having previous experience with the desktop simulation trainer (DST), and their instructing experience including, aircraft model and recency.

All study participants were assigned unique participant IDs to ensure anonymity and confidentiality throughout the study, while facilitating data management and analysis. This approach allowed the research team to track and analyze the data collected from each participant, enabling a detailed examination of the system's performance and the participant's responses.

Study Equipment

The study utilized a range of equipment to facilitate data collection and analysis. In terms of hardware, each participant was provided with a computer (desktop or laptop) to run the training simulation locally. A computer monitor, to provide quality visibility of the simulation. A keyboard, mouse or trackpad were necessary to interact with the point and click functionality of the trainer. Lastly a headset or microphone to communicate effectively with the study moderator during the virtual study sessions.

In terms of software, the study employed several tools to support the research objective. A desktop simulation trainer (DST) was used as the test bed for student interaction, instructor observation, and integration with the Behavior Detection System. Screen recording software was utilized to record student use of the simulation and record instructor observation and comment. Video playback software was utilized in displaying student interaction to the instructors for observation. An LRS was utilized to contain the Behavior Detection System output. The Behavior Detection System was used for automatically detecting and identifying observable behaviors during the simulated flight scenario demonstrated by the student participants.

Training Environment

A desktop simulation trainer (DST) was used as the training environment for this study. The simulation provided a graphical user interface (GUI) of an integrated navigation display that included flight management performance functionality. The simulation also provided checklists for the completion of applicable tasks throughout the straining scenario. The simulation was designed to offer experiential training outside of classic training environments. The training scenario presented to the students in our study simulated a fuel leak. The students were expected to interact with the DST GUI through mouse clicks to manipulate the interface and complete the process of recognizing and solving a fuel leak scenario. The high-level steps needed to complete the fuel leak scenario included (1) recognizing the changing fuel flow and level imbalance, (2) close the crossfeed valve between the left and right side of the fuel system, (3) shut down the corresponding engine, (4) complete performance calculations based on the fuel leak and weather conditions, and (5) complete the necessary diversion to the nearest suitable airport. Upon the completion of the scenario, the environment would export the necessary training data.

Methods and Procedures

Competency-based training and assessment (CBTA) was used to define the competency and observable behavior (OB) structure that was used to assess the students' interactions with the DST. CBTA is a type of Evidence Based training (EBT) that has been growing more in popularity across industries including aviation. It is different than current EBT solutions and frameworks because it focuses on more than just the technical skills of an individual (CAE, 2020). Instead, it focuses on both technical skills and non-technical skills like decision-making and situational awareness. The inclusion of the non-technical skills was done to help train and assess the overall performance of an individual to improve safety when performing job operations (IATA, n.d.). When CBTA frameworks are created, a list of core competencies is defined. Each competency in the list is then broken down into several OBs that an individual should show when performing tasks like picking an alternate airport or communicating that a malfunction has occurred (IATA, 2023; EASA, 2016; Renier, 2022.; United States Air Force, 2022). This structure is what we used for assessing students in this study.

The main objective of this study was to evaluate the accuracy of the Behavior Detection System's logic by comparing its recorded observations with those recorded by instructor participants. To achieve this, a two-fold data collection approach was employed. Firstly, video data was collected to capture students' interactions with the DST. This video data served as a crucial resource for instructors to impartially identify evidence of the competencies and observable behaviors (OBs) exhibited by the students. Concurrently, a second data stream was established using Behavior Detection System to automatically detect and document the OBs and competencies demonstrated by the students during their software interactions. Subsequently, the outputs from the Behavior Detection System were comprehensively compared with observations made by instructors during post-observation discussions, thereby addressing the main objective of evaluating the system's accuracy in comparison to live instructors.

Student Procedure

The student procedure began with participants reviewing and, if willing, signing the Informed Consent form. They were then assigned a unique Participant ID for data collection purposes, ensuring their anonymity. Participants completed a demographics survey and received an overview of the study session from the researcher. They were provided with briefing materials and received a verbal briefing from the study moderator to prepare for the simulated flight. Seated at a desk with a computer or laptop, participants had screen recording software activated (without audio/visual recording of themselves). They loaded the desktop simulation trainer (DST) and shared their screen with the moderator and observers. The moderator informed them when to begin the simulation, after which the moderator did not provide assistance. Data collection commenced at this point. Participants completed the simulation using the DST software, with data collected including screen recordings and student click inputs. Once the simulation was completed, the screen recording software was turned off, and participants were instructed that they could exit the DST. They were then asked if they had any further questions regarding the study, and the session was concluded by the moderator.

Instructor Procedure

The instructor procedure began with participants reviewing and signing the Informed Consent form. They were assigned a unique Participant ID for data collection purposes, ensuring their anonymity. Participants completed a demographics survey and received an overview of the study session from the researcher. They were provided with briefing materials and received a verbal briefing from the study moderator to prepare for the simulated flight. Seated at a desk with a computer or laptop, the moderator started the recording of audio and the computer screen (without visual recording of the participant). Participants were shown a list of applicable observable behaviors and asked to list the expected actions they would see during the simulation to demonstrate these behaviors. Instructor participants then loaded the video of the student participant and began assessing the simulation, taking notes. They were also shown a list of applicable observable behaviors and asked to list the actions they observed during the simulation to demonstrate these behaviors. The moderator reviewed the participants' answers and comments, comparing them with the output of the Behavior Detection System. If there were significant discrepancies, follow-up questions were conducted. This process was repeated for one more simulation. Data collected during the session included audio recordings, screen recordings, and questionnaire data. After completion, participants were asked if they had any further questions regarding the study, and the session was concluded by the moderator.

RESULTS AND ANALYSIS

The data collected during the student phase of the study includes (1) the control inputs from the desktop simulation trainer (DST) software collected from the Behavior Detection Software and (2) video data of the students performing the simulation within the DST environment. The student phase of the study provided example training instances for review and evaluation from instructors in the second phase of the study.

The data collected during the instructor phase of the study includes (1) the observational notes from the instructors as they view the video examples, (2) data containing when instructors detected observable behaviors during a particular task (3) data providing how many times an instructor detected an observable behavior (4) differences between the instructors' observations and the detected behaviors from the behavior detection software (5) if the instructors agree or disagree with data provided by the behavior detection software.

Instructor Agreement Level Likert Scale

Table 1. Instructor Agreement Levels Likert Scale

1	2	3	4	5	6
Strongly Disagree	Mostly Disagree	Somewhat Disagree	Somewhat Agree	Mostly Agree	Strongly Agree

After each observation of a student performing the simulation, the instructors were asked to what extent they agreed with the output of the behavior detection system. The instructors would use the Likert scale below to evaluate the observable behaviors the software was pre-programmed to evaluate during the scenario.

Behavior Detection System Instructor Agreement Levels

The following tables list the average level of agreement by the instructor subject matter experts for each observable behavior detected by the behavior detection system. The tables are divided by the observation, the competency, the observable behavior in the competency, and the average level of agreement. If an instructor rated their agreement with the behavior detection system less than three (3), their comment is noted along with the root cause or study artifact.

Application of Procedures (PRO) – Observable Behavior 2 (PRO 02)
Competency - Identifies and applies appropriate procedures in accordance with published operating instructions and applicable regulations.

Observable Behavior **-** Applies relevant operating instructions, procedures, and techniques in a timely manner.

Table 2. Application of Procedures OB 2

Observation #	Level of Agreement (Avg.)
1	4.20
2	4.50
Overall	4.36

There were three disagreements, listed below, to the behavior detection system's observation of this behavior.

Insufficient Time for Assessment (1 comment): An instructor disagreed with the system's observation of PRO 02, contending the desktop trainer did not afford sufficient time to wait and confirm whether the fuel leak occurred in the tank or engine before the student opened the crossfeed valve. Explanation: The comment resulted from a study artifact as the simulation was designed to conclude either upon the selection of a diversion airport or after a predefined time interval had elapsed.

Early Crossfeed Valve Activation (2 comments): Two instructors disagreed with the system's observation of PRO 02, asserting the time when students opened the crossfeed valve did not correspond exactly to the "insufficient fuel" warning based on the statement in a modified quick reference handbook. Explanation: The comments resulted due to the tested configuration lacking a rule within its authoring to account for the precise timing of opening the crossfeed valve leading to an observed discrepancy in the assessment.

Application of Procedures (PRO) – Observable Behavior 5 (PRO 05)
Competency - Identifies and applies appropriate procedures in accordance with published operating instructions and applicable regulations.

Observable Behavior – Monitors aircraft system status

Table 3. Application of Procedure OB 5

Observation #	Level of Agreement (Avg.)
1	5.20
2	5.00
Overall	5.09

There was one disagreement, listed below, to the behavior detection system's observation of this behavior.

Suitable Airport Selection (1 comment): The instructor disagreed with the system's observation of PRO 05, due to a software artifact of the desktop simulation trainer (DST) not offering a more explicitly suitable airport to choose for diversion. Explanation: The preprogrammed weather information had introduced ambiguity in determining the most suitable airport for diversion and did not offer a decisively obvious choice.

Problem Solving and Decision Making (PSD) – Observable Behavior 2 (PSD 02)
Competency - Identifies precursors, mitigates problems, and makes decisions.

Observable Behavior - Seeks accurate and adequate information from appropriate sources.

Table 4. Problem Solving and Decision-Making OB 2

Observation #	Level of Agreement (Avg.)
1	4.80
2	5.20
Overall	5.00

There were two disagreements, listed below, to the behavior detection system's observation of this behavior.

Checking Weather (1 comment): The instructor expressed disagreement with the system's observation of PSD 02, citing their inability to comment because the student did not check the weather during the simulation. Explanation: One student did not check the weather due to a study artifact of the DST not accounting for the route including ETOPS cleared airports.

Suitable Airport Selection (1 comment): The instructor disagreed with the system's observation of PSD 02 due to a software artifact of the tDST simulation's weather report when choosing a diversion airport. Explanation: The weather information preprogrammed into DST may have left some room for interpretation in choosing the best suitable airport as the software did not make have an overwhelmingly obvious diversion choice.

Situation Awareness (SAW) – Observable Behavior 1 (SAW 01)
Competency - Perceives, comprehends, and manages information and anticipates its effect on the operation.

Observable Behavior - Monitors and assesses the state of the airplane and its systems.

Table 5. Situation Awareness OB 1

Observation #	Level of Agreement (Avg.)
1	4.40
2	5.20
Overall	4.81

There were two disagreements, listed below, to the behavior detection system's observation of this behavior.

Early Crossfeed Valve Activation (2 comments): Two instructors disagreed with the system's observation of SAW 01, asserting the time when students opened the crossfeed valve did not correspond exactly to the "insufficient fuel" warning based on the statement in a modified quick reference handbook. Explanation: The comments resulted due to the tested configuration lacking a rule within its authoring to account for the precise timing of opening the crossfeed valve leading to an observed discrepancy in the assessment.

Situation Awareness (SAW) – Observable Behavior 6 (SAW 06)
Competency - Perceives, comprehends, and manages information and anticipates its effect on the operation.

Observable Behavior - Develops effective contingency plans based upon potential risks associated with threats and errors.

Table 6. Situation Awareness OB 6

Observation #	Level of Agreement (Avg.)
1	3.00
2	5.00
Overall	4.09

There were two disagreements, listed below, to the behavior detection system's observation of this behavior.

Suitable Airport Selection (2 comments): The instructors disagreed with the system's observation of SAW 06, due to a software artifact of the desktop simulation trainer (DST) not offering a more explicitly suitable airport to choose for diversion. Explanation: The preprogrammed weather information had introduced ambiguity in determining the most suitable airport for diversion and did not offer a decisively obvious choice.

DICSUSSION

The primary aim of this study was to evaluate the accuracy of the Behavior Detection System's logic by conducting a systematic comparison with assessments made by instructor participants. This comparison was based on a Likert Scale ranging from 1 (strongly disagree) to 6 (strongly agree), where higher ratings indicated greater agreement.

The scatter plot in Figure 2 below depicts instructor ratings on a Likert scale, with the x-axis representing each instructor and the y-axis showing their respective ratings of comparison to the Behavior Detection System. Notably, instructors MI01, MI06, and MI08 each observed only one student participant reducing the number of ratings from ten possible scores down to five. Overall, a majority of the data is concentrated along ratings five and six, but there are some pockets of lower instructor ratings predominantly with instructor MI07.

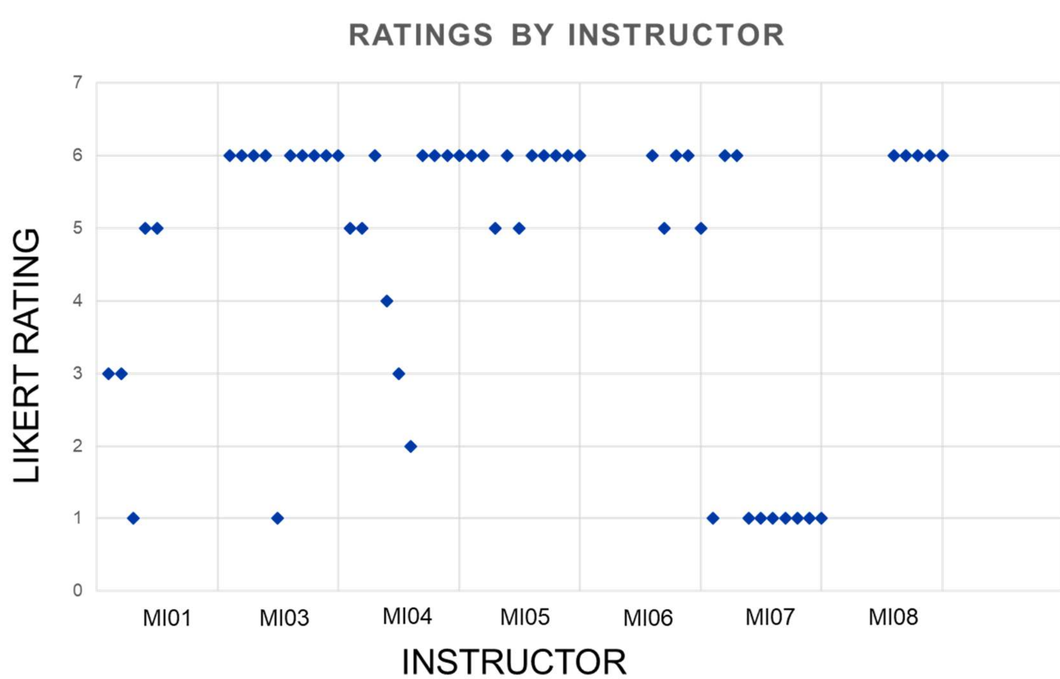

Figure 2. Likert Ratings by Instructor

The bar chart below presents a breakdown of instructor Likert ratings for the observable behaviors programmed within the Behavior Detection System. The x-axis represents the observable behavior (OB), and the y-axis shows the overall count of each level of rating. The key at the bottom depicts each Likert rating on the scale of 1 (strongly disagree) to 6 (strongly agree). Most ratings are grouped around scores of 5 and 6, with a few clusters of lower ratings.

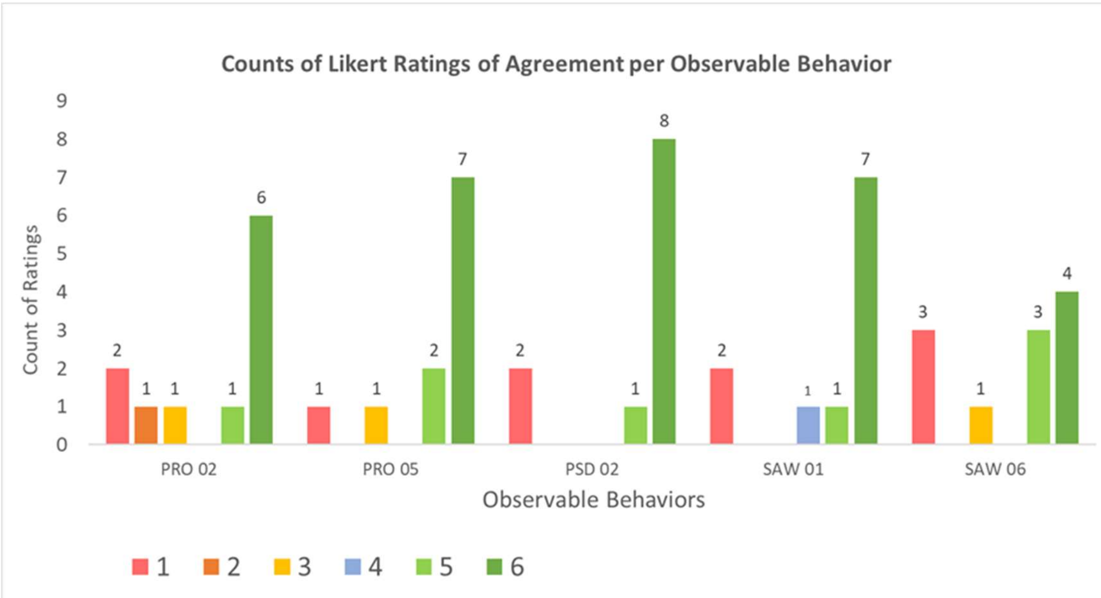

Figure 3. Likert Ratings by Observable Behavior

These findings highlight that the Behavior Detection System's logic aligns with the evaluations performed by the instructor participants, particularly for PRO 05 and PSD 02, which received higher ratings, indicating a strong level of agreement between the system's observations and those of live instructors. This suggests a promising level of accuracy and reliability in the system's ability to detect evidence of observable behaviors.

In this study, the Behavior Detection System demonstrated its capability to accurately identify observable behaviors during student activities, as confirmed by the participating study instructors. The instructors were actively engaged in the process of noting down observable behaviors during the student observation sessions. All predefined OBs incorporated into the Behavior Detection System received consensus approval from most of the instructors during their observations of the students.

Practical Implications

The findings of this study have led to a set of practical implications that can significantly influence the use of the Behavior Detection System in educational and training scenarios. These implications are directly tied to the study's research questions and highlight key considerations for enhancing the accuracy, consistency, and reliability of this system. Implementing these practical insights can lead to more effective and informed use of the Behavior Detection System in training settings.

Simulation Event Definitions: The study highlights the importance of defining the elements of the training events for the best training and learning experience of the student. Target OBs and desired student responses to events need to be identified. Then desired actions can be mapped to target OBs, and the events should be reviewed to ensure incorrect performance is also caught. This enables users to tailor the system to their unique educational objectives and training scenarios within their applicable training environments.

Ongoing Collaboration: The study highlights the significance of ongoing collaboration between technology developers, instructional systems designers, and instructors. This collaboration can foster the evolution of the Behavior Detection System that aligns with the evolving needs of pilot training.

CONCLUSION

As instructor shortages persist it is important to continue to innovate the supporting technologies needed to maximize the number of students they can effectively instruct. The Behavior Detection System offers promising opportunities for enhancing instructor impact by gathering evidence of student competency during scenario-based training. This detailed evidence can also be saved in an LRS and used with the TLA to track student proficiency throughout their career. The following recommendations outline key areas for future research aimed at optimizing the application of the Behavior Detection System in scenario-based training.

Testing for Consistency: The findings highlight the importance of improving consistency in identifying and measuring behaviors of students. Future research could explore consistency (how close are measurements of the same item are to each other) by testing the software with a greater population of representative student participants.

Interrater Reliability: Further studies could determine how interrater reliability compares between instructors before and after being shown the Behavior Detection System results. To accomplish these further studies would determine master ratings to develop a referent rater reliability baseline and then make a comparative analysis of instructor grades after viewing recorded scenarios. These future studies would need to include ratings and grades within the scope.

Efficiency in Assessments: Further research could help determine if implementing the Behavior Detection System could potentially save instructors time and effort in the assessment process.

REFERENCES

ADL. (n.d.). *Total Learning Architecture (TLA)*. Retrieved June 3, 2024, from ADL Initiative: https://adlnet.gov/projects/tla/

CAE. (2020, January 3). *Competency Based Training and Assessment, CBTA, in aviation training*. Retrieved June 10, 2024, from https://www.cae.com/blog/competency-based-training-and-assessment-cbta-in-aviation-training/

Clark, J. (2020). The Pilot Shortage – From Student Pilots to the ATPs. *National Training Aircraft Symposium (NTAS)*. Embry-Riddle Aeronautical University. Retrieved June 10, 2024, from https://commons.erau.edu/cgi/viewcontent.cgi?article=1363&context=ntas

EASA. (2016). *Proposal for a Competency Framework for the Competent Authorities' Inspectors*. Retrieved June 10, 2024, from https://www.easa.europa.eu/en/downloads/22019/en

Goldberg, B., Owens, K., Gupton, K., Hellman, K., Robson, R., Blake-Plock, S., & Hoffman, M. (2021). Forging Competency and Proficiency through the Synthetic Training Environment with an Experiential Learning for Readiness Strategy. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, (pp. 7-10). Retrieved June 10, 2024, from https://gifttutoring.org/attachments/download/4295/21332.pdf

IATA. (2023). *Competency Assessment and Evaluation for Pilots, instructors and Evaluators - Guidance Material, 2nd Edition*. Retrieved June 7, 2024, from https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/competency-assessment-and-evaluation-for-pilots-instructors-and-evaluators-gm.pdf

IATA. (n.d.). *Competency-Based Training and Assessment (CBTA) Expansion within the Aviation System*. Retrieved June 10, 2024, from https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/cbta-expansion-within-the-aviation-system.pdf

Koenig, D. (2023, February 9). Pilot shortage puts pressure on airline operations. *AP News*. Retrieved June 10, 2024, from https://apnews.com/article/business-26cf461c3fd83f19ff5a1e7a9479cff6

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). *Event-Based Approach to Training (EBAT).* The International Journal of Aviation Psychology, 8(3), 209-221.

Renier, Y. (2022). *Competency-Based Training and Assessment (CBTA)*. IATA Operations Safety & Security. Retrieved June 7, 2024, from https://www.icao.int/MID/Documents/2022/CRM/2022.06.20%20YREN%20MEA.pdf

Rustici Software. (n.d.). *xAPI solved and explained*. Retrieved June 3, 2024, from xapi.com: https://xapi.com/

Sottilare, R. A., Long, R. A., & Goldberg, B. S. (2017). Enhancing the Experience Application Program Interface (xAPI) to Improve Domain Competency Modeling for Adaptive Instruction. *L-AT-S*. Cambridge, MA, USA: Association for Computing Machinery (ACM). doi:10.1145/3051457.3054001

United States Air Force. (2022). *AIR FORCE HANDBOOK 36-2647: COMPETENCY MODELING*. United States Air Force. Retrieved June 10, 2024, from https://static.e-publishing.af.mil/production/1/af_a1/publication/afh36-2647/afh36-2647.pdf