Guidelines for Conducting PTC User Observations and Evaluations

Recommendation Report

Adam Beitia

December 3, 2018

TABLE OF CONTENTS

TABLE OF TABLES

BACKGROUND

PTC is a safety overlay system that was mandated by the Federal Railroad Administration (FRA) in 2008 with the passage of the Rail Safety Improvement Act and later amended by the Positive Train Control Enforcement and Implementation Act of 2015. In accordance with the technical specifications under 49 CFR part 236, subpart 1 each Class I railroad and each entity providing regularly scheduled commuter rail passenger transportation must implement an FRA-certified PTC system designed to prevent train-to-train collisions, over-speed derailments, incursions into established work zones, and movements of trains through switches left in the wrong position.

Locomotive engineers are trained to operate trains using the standard control apparatus and in accordance with all federal regulations and PTC is a safety overlay system. It is a system designed to correct for human error in high risk scenarios. In addition to its primary functions the system also supplements the train crew's situational awareness.. To ensure adequate training and constructive feedback a set of guidelines for conducting observations of locomotive engineers interacting with the system must be established.

The PTC system operates in conjunction with the locomotive's standard control system and in accordance with all federal regulations as well as the General Code of Operating Rules (GCOR). The system integrates information from a track database with train consist information as well as mandatory directives received from the dispatching center. The onboard PTC screen consists of a main screen which is a real-time display of the train's speed and upcoming targets. In addition to the main screen, PTC has multiple additional screens which can be accessed via the main screen using the eight soft keys (buttons) to view and modify data pertinent to the train's safe movement. To properly interact with the system locomotive engineers must receive focused training and gain experience through correct utilization of the system.

EXECUTIVE SUMMARY

Due to the pace at which PTC has been developed and implemented there exists a need to effectively monitor operators who utilize the system. This document represents an examination of possible, applicable methods for conducting meaningful observations of PTC users and recommends guidelines which could be established to measure proficiency. The analyses contained in this report will examine and recommend protocols for conducting these observations of locomotive engineers interacting with the system in an environment as close to normal working conditions as possible.

There are three physical methods in which the observations can be performed; a direct-participant observer inside the cab, electronically using the event recording devices on each locomotive and finally a mixed or combined observation using both direct-participant observations which would be compared with one or more of the electronic recording methods. The research indicates that direct-participant observations would be the most efficient and effective method to conduct the observations.

The potential observable measures of performance and measures of effectiveness derived from the research can be separated into six main categories. These categories are:

- *System Limitations* – Operator recognizes elements that are beyond the system's capabilities and corrects for them
- *Planned or Anticipated Events* – Expected system behavior
- *Unplanned Events* – Unexpected system behavior
- *Crew Interaction* – Are the conductor and engineer discussing system prompts and events to more efficiently operate their train
- *Error Recognition* – Troubleshooting abilities
- *Target Approach Management* – Ability to navigate the system within tight tolerances

All of these categories present opportunities to measure the operator's ability to recognize the communications being received from the system effectively. These six categories provide general areas of focus which can be applied to specific criteria or situations to effectively measure a PTC user's proficiency interacting with the system in various contexts.

The types of information that the operator will be presented with which can be used as potential individual measures of performance gleaned from the research are the following:

- *State Information-* Situational information; does the operator understand what the screen is displaying
- *Procedural Information-* System displays information which requires the operator to perform certain functions
- *Causal Information –* Does the operator understand the cause for a PTC event?
- *Signal Detection-* Does the operator recognize the various signals and alerts the PTC system is conveying?
- *Chronometric-* Events where time is a factor, does the operator take the necessary action within the time allotted
- *Speed Accuracy-* The time which the operator uses to answer prompts in relation to the accuracy of those inputs

These measurements apply to the information being given to the user by the system and can be quantified in various ways to apply to specific tasks or train operations.

PTC is an evolving technology and the processes for observing and measuring the users must evolve with it. As system limitations are overcome new challenges and opportunities for the operators will present themselves and these observations must be adapted to updates at regular intervals to maintain the highest training and proficiency standards.

Recommendations

It is the final recommendation of this report that structured, direct-participant observations are the most effective method and can be backed up with recorded data for further analysis or explanation if necessary. These observations should focus on those six measurable areas of PTC usage (system limitations, anticipated events, unplanned events, crew interactions/intra-cab communications, error recognition, and target approach management) identified in the study. The complex interactions which the users perform will be measured specifically using the individual measures of performance identified (state information, procedural information, causal information, signal detection, chronometric/timed events, and speed accuracy).

PROBLEM STATEMENT

The development of PTC has required both training and implementation processes to be created and deployed. As a result of these factors there exists a strong need for conducting meaningful observations of locomotive engineers interacting with the PTC system to improve existing processes as well as develop new ones. The results of structured observations will provide a consistent means for evaluating engineer proficiency in the use of PTC and provide a source of data to be used in both the design and improvements of the PTC system itself.

Data collected through observations will be beneficial to supervisors of TE&Y employees and PTC engineering staff. While observations are currently performed for numerous reasons they are not conducted in a structured manner specifically in the use of PTC. This results in an anecdotal collection of observations and the results of those events are often buried in large chunks of text and narratives. The proposed observations would store the data centarally. By centralizing the data, it would normalize groupings of observations, solidly identify trends and separate critical information from less critical.

Observations: The How and Why

The Need for Accurate Observations

Through observations behaviors can be evaluated and predictions can be made. From these predictions service goals can be determined and processes to meet those goals can be developed. The observational process can pave the way toward a clearer understanding of causality and will aid in driving corrective actions for more accurate and efficient utilization of the system.

Observation is more than just the process of watching an individual performing some task. There are two required skills for any successful behavioral observer; the ability to accurately observe events without distractions and the ability to accurately record and document those events/observations. Observational skills are not inherent in every person and should be taught with specific focus on the expected outcomes.

The development of causal explanations is the ultimate intent of the research and must be kept in mind throughout the process of generating the guidelines for the observations. The questions must adequately equip the observers to seek the behaviors that will illustrate outcomes. The behaviors must be traceable to a conclusion in such a way that the observer does not need to interpret the data, only collect it. “Qualitative research is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures, data typically collected in the participant’s setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data.” (Creswell, J. W., & Creswell, J. D., 2017 p.5)

Acute Issues

Engineers experience a range of issues that have the potential to cause significant train delays. Incorrect train consist information (gross weight, operative brakes, length etc.) inputs during initialization, incorrect train-direction input, not adhering to onscreen warnings and prompts, incorrect responses to acknowledgement prompts, delaying track selection or other issues such as not utilizing restricted mode when required must be observed and addressed. These types of issues have the present a multitude of potential situations that could increase safety risks. Trained observers should be taught to focus on the critical concerns most crucial to safe and efficient movement of trains. Every observation will be a series of events which will require a sequence of responses and these responses will be the heart of the observations.

Change in Behaviors

Locomotive engineers are remote operators who receive minimal direct supervision. The process of observation is fundamental to supervision due to the limited opportunities for direct contact between employees and leadership. Observations can be used to determine if change has or has not occurred as well. Training and service plans will be developed from the initial observations. Supervisors and training content developers need accurate information as they consider strategies to balance the needs of the engineers in the cab with those of the company. Follow up observations should be conducted after a yet to be determined timeframe to provide evidence that change has or has not occurred. Furthermore, it should be differentiated whether that change has met a specific service goal or just heightened an engineer’s ability to recognize and correct problematic behavior prior to it causing a train delay.

The Need for Communication

Quality observation improves the quality of communication between system administrators and field managers and by extension also improves communication between field managers and the locomotive engineers whom they supervise. Emphasis on details and accuracy in observations will achieve greater accuracy in the communication of the needs and strategies of the stakeholders. Consistency in training will also be improved through accurate, quality communication.

Improved communication strengthens the process of relationship building and encourages increased involvement. An effective system for communication combined with quality observations will assist in facilitating and building interpersonal interactions between supervisors and system users. Good observation encourages appropriate behavior while interacting with the system and will effectively reinforce those good behaviors. By targeting certain aspects of the engineer's interactions with the system and focusing observations to address those specific elements, more precise supervision of the engineers will be achieved.

Effectiveness

Observation, as an extension of the change process, will be used to evaluate the effectiveness of specific training strategies. The data gathered through accurate initial and subsequent follow-up observations will provide both information and evidence of an engineer's behavioral changes which will in turn drive new intervention and training strategies. A set of standards must be developed from the data collected during the observations to measure the effectiveness of the subsequent training.

The documentation of behaviors during initial observations will be used to establish baseline statistics and trends upon which future observations will be based. The initial observations will focus upon specific behaviors in response to prompts. Subsequent observations will be used to observe the same behaviors yet the results will be analyzed to measure the levels (if any) of improvement and continue driving user training techniques.

METHODS

Introduction

To gain insight and to observe people in their natural work environment is inherently unnatural because the observer's presence has disrupted that natural, normal setting. Because the participants are aware of the observer there will be a level of self-consciousness which is not present when the observer is absent. The presence of the observer will cause inevitable changes in the dynamic of the crew and their behavior which must be minimized to the extent possible. While many of the behavioral changes may be minimal and imperceptible they will still have an impact upon the observation which is why making the observations as non-invasive as possible is of extreme importance.

To properly seek out root causes, perform analyses and make predictions from the observations target behaviors must be identified. “The selection of target behaviors for observation in applied or clinical situations can be viewed as a continuous process (Kanfer, 1985). The process takes a problem or need and then converts it into a research question which utilizes several target behaviors to design the research. These target behaviors will be used to observe engineers at implementation and later after they have had significant experience with the system. This could be used to make determinations of any changes in behaviors. The application of heuristics to the methods of observation to rate their effectiveness and capabilities was explored.

With current technological limitations PTC is not able to operate without human input. “Improper use of the air brake can lead to a train running away, derailment, or unexpected separation. Furthermore, an overly conservative PTC system runs the risk of slowing trains below the level at which they had previously been safely operated by human engineers. Railway speeds are calculated with a safety factor such that slight excesses in speed will not result in an accident. If a PTC system applies its own safety margin then the end result will be an inefficient double safety factor. Moreover a PTC system might be unable to account for variations in weather conditions or train handling and might have to assume a worst case scenario, further decreasing performance.” (Nayak, P. R., Rosenfield, D. B., & Hagopian, J. H. (1983) The requirement for human input means that effective observations of those inputs will also remain a requirement.

Focus Questions

To narrow the potential formats that the observations for primary data collection will employ the following chart (Figure 1) from the University of Portsmouth (2012) was used to answer some basic questions about the needs of the observations.

- Are participant observations possible?
- Are non-participant observations possible ?
- Are covert observations possible?
- In the locomotive cab environment would an unstructured observation be effective?

Types of Observations

Type of Observation	Advantages	Disadvantages
Participant	Close to participants and events being observed More in depth information gathered	Researcher influence too close and wider issues might be missed
Non-participant	Less influence from researcher Overview of a situation/issue more achievable	May miss important insights by not being a participant May influence behaviour if they are aware of being watched by an 'outsider'
Covert	Participants more likely to behave 'normally' Researcher less likely to influence group	Ethical issues with deception, voluntary participation and informed consent Difficult to take notes
Overt	Ethically sound, the researcher can get informed consent Note taking is open and easy to do	Participants may alter behaviour if they know they are being watched
Structured	Easier to note down behaviours if a checklist is available Can be informed by theory Inter-rater reliability can be measured	Researcher may miss other important events or behaviours Limited scope
Unstructured	More scope for identifying unexpected behaviours or events More in depth data	Difficult to take notes More open to individual observers interpretations - less reliable?

(Table 1)

Because this is a recommendation report outlining guidelines for conducting observations they will be structured according to specific needs for consistent means of evaluating engineer's proficiency. An unstructured observation would allow for more in-depth narratives; however, the data would be combined with other aspects of the observation that may be inconsequential or even completely irrelevant altogether. Therefore, an unstructured observation would create extreme difficulties for being able to normalize and organize the collected data effectively due to the amount of time it would require extracting the crucial data. The structured form will give the observer a list of potential work events that the train's crew may encounter throughout the duration of the observation. There will also be a list of potential behaviors associated with each of these work events that could provide a foundation for possible flow charts or the development of other deductive processes.

When considering observational methods, the benefits and negative aspects must be considered when choosing. What will be gained? Will all the required data be available or just a portion of it? Are there additional risks involved? Will efficiency be increased or decreased or affected at all while conducting the observations? To make an effective decision when choosing amongst the possibilities there is a balance between desirable and undesirable consequences which must be factored into the decision-making process. It is this balance that weighs what will be gained from the observations and how much will be gained or lost from each option. The importance of these potential gains and losses can be described in another context which is the preferences and value for the stakeholder associated with different outcomes.

Measurables

The stakeholder's primary intent for the observations are to ensure a consistent means for evaluating engineer proficiency in the use of PTC while also providing a source of data to be used in both the design and improvements of the PTC system itself, including both functionality and the human-machine interface aspect. Therefore, the observations will be structured to ensure that the necessary data is being garnered from the events.

Through these observations, correlations will need to be identified both explicitly and implicitly. "Holism is the defining principle for any observational format. Through a holistic approach "an observer maintains the larger picture, or the total cultural system, as the focal point. All observations and interpretations are attempts to identify relationships between elements and the whole system" (Suen, H. K., & Ary, D., 2014) While some more obvious events will appear one dimensional at first glance the goal is to analyze them to also identify underlying trends and causality. To do this the Measures of Performance (MOPs) and the Measures of Efficiency (MOEs) must be valid measurements of the performance attributes which they are purposed for. Vague references or generalities are less useful therefore the specific observational guidelines must ensure that the MOPs and MOEs are clear, concise and can be reliably assessed. The following heuristics illustrate the principle elements which would be appropriate measures for the observations.

Potential System Measures of Performance

•Learnability

•Ease of Use

•Frequency

•Duration

•Sequence

•Latency

Human-Computer Interaction heuristics:

Potential Individual Measures of Performance

a.	State Information	Situational, conveys status data
b.	Procedural Information	System communicates action to take
c.	Causal Information	Reason for occurrence is communicated
d.	Signal Detection	Does the engineer recognize the alert/signal PTC is displaying
e.	Chronometric	Events where time is a factor
f.	Speed Accuracy	Trade-off; how quickly does the engineer respond to the system and how accurate was the input

(Table 2)

Can Nielsen's heuristics be applied to evaluating an engineer's utilization of PTC?

Because the proposed inspections are not going to involve any redesigns of the PTC system itself a usability inspection needs to be applied to the observations. "Typically, a usability inspection is aimed at finding usability problems in an existing user interface design, and then using these problems to make recommendations for fixing the problems and improving the usability of the design. This means that usability inspections are normally used at the stage in the usability engineering cycle when a user interface design has been generated and its usability (and utility) for users needs to be evaluated" (Mack & Nielsen, 1995). For these observations the primary intent is not to identify problems in the system but rather the aim is to identify individual behaviors while interacting with the system. Essentially it is a reversal of the process because the goal of the observations is to identify issues on the human side of the interface and develop measures to change those behaviors at a cultural level.

When performing heuristic evaluations on software and web sites Nielsen's (1994) "Ten Usability Heuristics" for user-interfaces are commonly applied. Here I assess whether or not these same heuristics can be applied to observing an individual using PTC in a similar fashion? I attempted here to apply these heuristics from the engineer's (and entire train crew's) perspective.

Nielsen's Heuristic 1: Visibility of System Status

This heuristic applies to PTC through the onboard screen display. Location, speed and upcoming target information is not just telling the engineer the status but also assists them in operating the train more efficiently and safely. An observer would have the same information available to him or her.

Nielsen's Heuristic 2: Match between the system and the real world

PTC gives real-time GPS-based information to the train crew. Visually it differs from the real world in that it is a linear display which requires the user to make logical connections from visual cues.

Nielsen's Heuristic 3: User Control and Freedom

Nielsen's third heuristic applies to an "undo" function which is not relevant to PTC, however user control and freedom are still important to system users. If the engineer feels too confined by the system they may become frustrated, detached or indifferent. Even though PTC is a safety overlay it is still important for the user to feel that they are in control of the train and their work environment as well. Part of this is that user's actions require a rapid, concise response. One such PTC feature are secondary acknowledgements with yes and no buttons which leave no room for doubt.

Nielsen's Heuristic 4: Consistency and Standards

The PTC-user interface, as with all interfaces, must be consistent throughout. Industry standards for functionality are adhered to with some room for individual railroads to adjust certain functions at their discretion. The observations should be standard as well.

Nielsen's Heuristic 5: Error Prevention

The prevention of errors requires an individual critically think about their inputs to the system. The observer must be able to note responses in limited time (for example: as the countdown warning banner alerts).

The term error prevention with respect to the PTC onboard screen may include warning messages such as "Braking will occur in --- seconds" or "Have you received permission from the foreman to proceed through the work zone." These messages as well as subsequent acknowledgement prompts assist the engineer in making grievous operational errors.

Nielsen's Heuristic 6: Recognition rather than recall

Instruction and training for use of the system should be received prior to the observation. Manuals should not be relied upon for system use while the train is in motion.

Nielsen's Heuristic 7: Flexibility and efficiency of use

The PTC system should be usable and logical to engineers of different experience levels.

Nielsen's Heuristic 8: Aesthetic and minimalist design

The on-screen interface and inputs should be simple and non-intrusive, thus providing access and logical navigation of the system's menus and functions.

Nielsen's Heuristic 9: Help users recognize, diagnose, and recover from errors

This heuristic is relevant to the system's user interface that can assist the user in the recovery from or prevention of errors. If PTC takes corrective action (train stoppage) it is imperative that the engineer understands the cause by looking at the onboard screen. (related to Nielsen's Heuristic 5: Error Prevention).

Nielsen's Heuristic 10: Help and documentation

Help needed to utilize and engage the system should be primarily provided by formal instruction or remedial tutorials. Minor assistance items can be offered through the pocket reference guide.

Research questions:

What can be evaluated?

What are the implicit heuristics being applied currently to conduct observations at the company?

What are the explicit heuristics being applied currently to conduct observations at the company?

How can things be evaluated?

What usability evaluation measures are being employed while conducting observations?

How are other systems evaluated now?

POTENTIAL METHODS FOR PERFORMING OBSERVATIONS

At the present time all performance evaluations (for train operations) at the company are performed by a direct-observer present inside the cab. There are additional compliance evaluations conducted by reviewing downloadable recordings of the braking systems and control inputs. These remote evaluations are conducted by designated supervisors of locomotive engineers.

Remote Recording to Conduct Observations

The first potential method for conducting the observations would be through the use of remote recording equipment. Human observation is a self-explanatory term. It is the practice of using human observers to collect data through direct observation whereas mechanical observation involves the use of various types of machines to collect the data. Video cameras and other forms of downloadable material that can be reviewed via computer can be used as opposed to a human observer to collect information. That collected data is later interpreted by researchers per the protocols of the observation plan.

All locomotives in which the engineer(s) will be operating from are equipped with multiple forms of recording devices. The F.A.S.T. (Fixing America's Surface Transportation) Act signed into law by President Obama in December 2015 requires all passenger railroads to install both inward-facing and outward-facing cameras, aligning with a previous recommendation from the NTSB. FRA regulations 49 CFR 217 require that a process of random selection of locomotive camera recordings be reviewed as part of the operational testing program. UP's Information Governance Policy and Workplace Recording Rules allow for workplace recordings to only be conducted for legitimate purposes and in compliance with all technical, business and legal requirements. The remote observations could be categorized as part of the Risk Reduction Program which states that “authorized persons will periodically remove or download the LDVR data hard drive/caddy or download the data from LDARS for review of locomotive camera recordings to evaluate potential hazards and operating risks” which would include PTC operations.

UP has testing criteria for conducting structured tests which limit the use of cameras under certain conditions. Per the Union Pacific FTX standards “cameras may be used to conduct real time “” observations”” under the following conditions:

- The testing manager has a radio and must attempt to stop any unsafe behavior observed immediately
- The testing manager must be within proximity to debrief the evaluated employees in a reasonable amount of time
- The testing manager must conduct a face to face debriefing of the event
- Structured tests must not be performed with a camera

Because the observations would not be structured tests as part of the FTX program it would be allowable under the current policy to conduct them remotely. Regarding the potential employee’s knowledge of the observation, informed consent is understood to be a condition of employment while operating a locomotive. Under the above-mentioned regulations and policies the employee(s) are cognizant of the potential to be observed randomly and therefore formal informed consent forms would not be required to conduct the observations

PTC in current form works on a platform which allows for certain data to be collected and reviewed. Additionally, the locomotives on which the system operates are equipped with both inward and outward facing cameras. These will be referred to as automated or mechanical methods for observing engineers and collecting data. The video camera is a commonly used device for mechanically gathering data in an observation. One advantage is that a video camera offers a more exact means of collecting data than what a human observer can record. The second advantage is that the tape can be played back repeatedly and reviewed in detail if necessary.

Reviewing recorded tapes has several practical advantages for an observer. Cost is minimized because the recording equipment is already in place, functional and ready to use. There will be no need to install additional equipment and all of the costs associated with that process. The additional human presence in the cab would also be eliminated. Getting a human observer into the locomotive cab requires scheduling and coordination to achieve. Through the use of recorded data an observer could review an engineer's interaction with the system as needed just by accessing the materials. This would also create a barrier and allow the observer relative anonymity which could reduce or negate altogether certain individual barriers to being observed such as a sense of scrutiny, resentment or embarrassment.

On balance, disadvantages to computer based or video-recorded observations exist which reduce their effectiveness. The alliance that forms between a trusted observer and the individual being observed will be compromised. This lack of a trained human observer will create difficulty in addressing complex issues that may arise during the observation. If technical issues arise the observation may be incomplete or lost altogether. As with any setting in which computers are utilized there exist concerns of privacy and security breaches. There are limitations to even the highest quality automated recordings as well. Because the data is being examined through a remote medium, subtle behavioral aspects may not be readily apparent. The participant's mood, potential state of agitation, and physical state (illness, fatigue level, etc.) cannot always be observed effectively with recorded observations. These observations would be non-participant and could be done either covertly or overtly.

The process of reviewing the data in and of itself can be perceived as a disadvantage. Because the data in any observation must still be reviewed and interpreted by a human a remote observation will increase the time investment of the reviewer. Specifically, the observer would be reviewing downloads and possibly the EMS/PTC interface. A major drawback to this method would be the substantial amount of time required to locate and sort the footage. Additionally the current capabilities of the in-cab recording equipment would not support observations of this type.

Direct-Participant Observations

Systematic observation provides the observer an ethnographic appreciation of the people whom they are studying. A fundamental dimension of any culture is behavior. Direct observation provides total immersion in the situation being watched and behaviors can be witnessed firsthand. The participant observations proposed here will combine observation with informal in-cab interviews. The data from the observations will be recorded into a structured format (form, booklet, etc) to better identify trends and habits/behaviors.

Locomotive cabs are equipped with several observer-friendly features and thus provide ample opportunity for conducting direct and/or indirect observations. This firsthand method of conducting observations would entail the observer acting as a participant by joining the train's crew inside the cab and observing them as they perform work under normal circumstances. "Participant observation is a typical data-collection method used in grounded theory research (Glaser and Strauss 1967, Morse and Field 1996), with data collection occurring in the natural setting in which the participants are located (Adler and Adler 1994). Observations are made of people in the context of their normal environment, setting or field; hence, the term fieldwork or field research (Minichiello et al 1991). Participant observation consists of gathering impressions of the participants' 19ehavior and involves looking, listening and asking" (Lofland 1971). Because locomotive cabs are equipped with additional seating an observer could comfortably conduct their work without any sort of special accommodation.

While having an observer onboard is not an everyday occurrence for train crews it is not so uncommon as to create unusual distractions or disturbances. The term participant is used here only to convey that the observer will physically be inside of the cab, it does not imply that the observer will be performing any duties or functions in the operation of the train or the PTC system. "The researcher as participant observer attempts to assume the role of the individuals under study and to understand his or her thoughts, feelings and actions (Wiersma 2005). According to several authors, participant observation represents an excellent source of qualitative data" (Davis 1986, Morse and Field 1996, Polit and Hungler 1991). This would be an overt form of observation as the observer would inform the participants of the observation prior. Again, as a condition of employment under the FRA's Risk Reduction Program employees are aware that a qualified company officer may board their train at any time to conduct a performance evaluation. This form of direct-observation to assess PTC proficiency would not require any additional arrangements to conduct.

Through the use of direct observation as a participant in the cab there will be ample opportunities for the observer to witness the engineer's processes of logic and inquiry while interacting with the system. The observer's presence inside the cab with the crew will give them the unique advantage of being able to view the interaction through the engineer's perspective and natural work setting.

Another important characteristic of the participant observer is building and maintaining relationships with individuals in the field. It is critical for future observations to be successful that the observers employ a specific strategy for gaining access to the people in the field. A crucial element of the participant methodology is the relationship between those in the field and the observer in the context of human interaction as a whole. The observer's ability to gather truthful, accurate data is heavily dependent upon the quality of this relationship. Even though the observer will not be interacting with the PTC system their interaction with those people being observed will allow them to view the interactions from the role of an insider. This form of observation is designed to spawn practical truths and conclusions about the human-machine interactions which are rooted in the day to day realities of human existence.

While direct observation will be the primary information gathering method, the participant observer should employ additional strategies as well. For example, noting the engineer's experience level with and general mood toward PTC will give the observer valuable context as a baseline. These sorts of subtle points will only be viewable by the observer and while they will not be able to be included in the final data the observer will need training to recognize them. The addition of a questionnaire and possibly a brief, informal interview post-observation could also be included in the process to gain additional insight and conclusions from the subject matter expert's perspective. However, this report's only focus is on different forms of observations used to collect data and therefore interview techniques will not be addressed.

There are disadvantages to being a direct participant and it must be noted that "there are, however, problems associated with being an 'insider'. Lipson (1984) suggests that recognition of patterns of practice might be difficult to identify because the behavior is so familiar and is taken for granted. Routine practice, because it is known, can be missed. Gerrish (1997) warns 'there was the risk that over-familiarization with the setting might lead me to make assumptions about what I was observing without necessarily seeking clarification for the rationale underpinning particular actions'. In my study, interviews were also undertaken with participants as a strategy to avoid missing subtle information." (Bonner, A., & Tolhurst, G. (2002). To avoid such over-familiarizations the observer must be trained to start each observation without preconceptions or assumptions. Some critics of direct observations view them as reductionist and dehumanizing for those being observed. Depending upon the questions being asked there are those that view the observations as incomplete if the scope is too narrow. The scope of this report targeted a broad range of behaviors and conditions to be as inclusive as possible. This report could be the impetus for additional developmental programs and observations which could compare and contrast data sets.

Mixed or Combined Method for Conducting Observations

The third viable option for conducting the observations is through the use of a combination of both direct human observation and remote recording. The findings of a 2003 study conducted by the U.S. Army Research Institute suggest that "even though intelligent agent technologies are evolving rapidly, they are not yet ready to replace expert observation and interpretation of performance. Automated measures are not the only answer. If we discontinue use of expert observers, analysts, and coaches in training, we will lose input that is not available any other way and will also lose the opportunity to allow experts to analyze their own behaviors." (Campbell, C., & U.S. Army Research Institute, 2003) While technologies continue to advance and improve there is still no complete substitute for the intuitive abilities of a live human observer. The relationships that can be established between the engineer and observer cannot be formed through remote recordings. The counter to that is that recording devices can capture mountains of data that a human observer could not capture in the same timeframe.

The Army study implores researchers to incorporate and blend traditional methods with automated measurement tools to achieve more depth in their performance assessments. Trained observers are currently the best way to put automated measures outputs into a meaningful context, but they will require training and guidance to ensure the reliability and validity of their output. Because the observers will be trained and familiar with the system they will also be able to effectively answer the engineer's questions regarding system function and operations.

Throughout the process of observation information can be gained through the asking questions of crew members. To effectively see the full spectrum of the events occurring in the physical environments in which the locomotive engineers are operating trains observers may require more space than what is available to the observer at the user's shoulder. To increase the effectiveness of the observation employing surrogate or additional measures (recorded data illustrating some MOPs) will capture the most information possible. Reliability and credibility will be improved by using multiple types and sources of data. It is so much more than just tracking button pushes the mixed method allows for audio and video recording, eye-tracking, division of attention

EVALUATION CRITERIA

The location or medium through which the observer will conduct the research is one part of the process. The other part is defining what will be or should be observed. While many aspects of the interaction can be observed the key components which will assist the stakeholders are those elements which are measurable and quantifiable. "Interviews with locomotive engineers and conductors indicated that the introduction of PTC systems impacted how they operated the trains. Changes in train handling resulted from a combination of constraints imposed by the PTC braking profile, increases in information and alerts provided by the in-cab displays, and new sources of workload associated with interacting with the PTC system." (Wreathall, 2007) These changes in operating environment are the criteria which must be measured to effectively counter the changes in behavior.

This aspect is essentially an observation in usability of the software with the added element of interfacing it through the train controls and the inputs into the onboard TCU's screen. The observations are not an evaluation of the software or system functionality from a software engineer's perspective but are instead meant to be an evaluation of the user's performance in using the system as it is. The ISO (International Organization for Standardization) recommends certain metrics should be included in usability studies which are efficiency, effectiveness and satisfaction. "According to ISO 9126 usability is divided into five sub-characteristics: understandability, learnability, operability, attractiveness and compliance usability. We evaluate the usability of a software component based on the first three sub-described following characteristics (Santos, Novais, Ferreira, Albuquerque, De Farias, & Furtado, 2016):

> Intelligibility
>
> The component ability to allow the user to understand whether the component is suitable and how it can be used in particular tasks and under specific conditions.
>
> Learnability
>
> The software component's ability to allow the user to learn the application.
>
> Operability
>
> The software component ability to allow the user to operate and control.

These same three components can be applied to an engineer using PTC. However, for performance observations the categories need to be distilled further for context and measurability. The observations will focus on all three aspects with the following subsets of characteristics

What aspects of PTC interaction can be evaluated?

Learnability	• Effectiveness • Efficiency Satisfaction	Reliance/Trust in the System	Efficient/Effective Division of Attention Between Onboard Screen and Locomotive Controls
Operability	• Effectiveness • Efficiency • Satisfaction	Reaction to alerts (planned or intended vs. unplanned)	Acknowledge the signal? Make necessary adjustments to controls in effective timeframe (Timely & Accurate response to prompts & alerts)
Intelligibility		Preparation for planned events	Engineer's actions approaching imminent or anticipated event
Intelligibility Operability		Target Approach Management	*Situational Awareness-Ability to control train with close clearances of system's limitations
		Location Verification-Status Verification	*Situational Awareness
		Failure Modes	Skillfulness in addressing an unexpected system error
		Input of Consist Values/	Accurate inputting and verification of

		Verification Mandatory Directives	train specific information
Operability		Crew Interactions	Sufficient/Accurate Communication among crew members

(Table 3)

Attributes of Locomotive Engineers

User Characteristics	**Work Environment**	**Items to Observe**	**Measures of Performance**
0-3 Months PTC Experience	**Locomotive Cab**	<u>Events</u> • Establishing Track Location-Track Selection • Response to Unanticipated Warnings • Response to System Initiated Prompts Requiring Crew Interaction Exceptions <u>States</u> • System Initialization • Departure Test • Target Awareness in relation to Speed and Location •Entry into Restricted Speed	•Learnability •Ease of Use •Frequency •Duration •Sequence •Latency

		•Operating at Restricted Speed • Mandatory Directives: Receive and Review • Consist: Review / Request New / Modify • Restricted Mode • Energy Management System • Cutting In / Cutting Out • Crew Logoff	
3 or More Months PTC Experience	**Locomotive Cab**	<u>Events</u> • Establishing Track Location- Track Selection • Response to Unanticipated Warnings • Response to System Initiated Prompts Requiring Crew Interaction Exceptions <u>States</u> • System Initialization • Departure Test • Target Awareness in relation to Speed and Location	•Learnability •Ease of Use •Frequency •Duration •Sequence •Latency

		•Entry into Restricted Speed •Operating at Restricted Speed • Mandatory Directives: Receive and Review • Consist: Review / Request New / Modify • Restricted Mode • Energy Management System • Cutting In / Cutting Out • Crew Logoff	

(Table 4)

Because PTC is still being implemented in many areas it is imperative that the depth of the user's experience with the system is evaluated as a factor of the observation. Users with higher levels of experience will be familiar and will have the advantage of their prior cognitions and the decisions they made. A less experienced user will be unfamiliar and unbounded in their reactions and thought processes. No matter the experience level the observer is looking for logical thinking and that the user is making reality-based connections

Procedures

I have been an active observer of system functionality in my role for the past two years which has allowed access to system testing firsthand. While this experience guided my methodologies, it did not guarantee approval and participation. While I have observed PTC training sessions this set of observations would require subsequent examination. I first arranged a preliminary study plan with my supervisor as well as the general director of my department (See Appendix 1). Upon approval of the plan I set out to apply the heuristics from the literature to the potential methods for conducting observations and interviewed several senior company managers of PTC operations as well as managers and directors of operating practices. The questions were asked to focus on specific tasks common to train operation in conjunction with use of the PTC system interfaced with the onboard CDU. The interviews were designed to uncover issues relating to the engineer's ability to effectively interact with the PTC system during dynamic utilization which would be undetectable using other methods.

Interview Questions:

Questions specific to the individual's philosophy:

Who do you think should observe engineers using the system?

What elements of usage do you feel should be evaluated?

When do you think observations should be performed?

Define efficiency as it relates to PTC.

Do you feel that evaluation of an engineer's proficiency is important for effective use of the PTC system? Why or why not?

What do you think the observations should evaluate about an engineer using the system?

Questions specific to the company:

Who performs performance evaluations in this company?

How important do you think efficiency of PTC use is to this company?

When or at what stages of implementation of PTC should observations and subsequent evaluations be performed within this company?

How is efficiency evaluated for engineers using other onboard systems within this company?

Because PTC is on an extended (pending FRA approval) timeline, the production schedule is highly dependent upon the vendor, partner railroads and the FRA in addition to other variables. The time I spent with the senior managers was the end of the third quarter of 2018 with an extension for certain PTC deadlines filed and waiting for approval. The hardware for PTC both onboard and wayside has been installed at the time of this report as well as required employee training.

However, implementation of the full system has yet to be fully enacted on all service routes hence the significance of this study.

Heuristics verified in the study

a.	State Information	Situational, conveys status data
b.	Procedural Information	System communicates action to take
c.	Causal Information	Reason for occurrence is communicated
d.	Signal Detection	Does the engineer recognize the alert/signal PTC is displaying
e.	Chronometric	Events where time is a factor
f.	Speed Accuracy	Trade-off; how quickly does the engineer respond to the system and how accurate was the input

(Table 5)

In the table above (Table 5) letters A, B and C refer to information being given to the engineer through the onboard PTC screen. These are communications from the system which must be translated by the operator. Letters D, E and F are classifications of prompts from the system that will require the engineer's input to avoid a penalty train stoppage. Signal detection can be measured in an operator's response to time to a visual and or audible alert from the system. Chronometric prompts have a time associated with them, the most obvious of these would be the 45-second countdown warning banner which gets displayed when approaching a more restrictive target.

The speed accuracy relationship could apply to an assortment of HCI functions the operator may encounter throughout a typical trip. There are a multitude of factors which could be measured and quantified using speed accuracy variants.

RESULTS

Defining measures of performance will be an evolutionary process that will parallel the development of the PTC system itself. Current system limitations will give way for new opportunities and will also present new challenges for the human actors who will interact with each successive iteration and software version. Currently engineers are required to receive a performance evaluation annually. Based on current standards and training requirements it is the determination of this study that an annual efficiency/performance evaluation of engineers interacting with PTC would be a sufficient timeline to maintain operator standards.

During the interviews two prevailing schools of thought were exposed not of the management but of the operating employees. It was a split between those resisting the system and those adhering to or embracing the system. It was explained that some engineers had difficulty adjusting to the system due to clinging to old habits and behaviors that do not conform with the system's functionality. The other predominant trait witnessed and explained was that of overreliance on the system or more precisely a lack of situational awareness. That lack of situational awareness is due to a deficiency in either the desire or ability to divide their (train crew's) attention between the onboard screen and the rest of the events occurring in their surroundings. After gleaning responses from the participants I compared them to the heuristics from the readings. There were no new heuristics discovered through the course of the study. After compiling the responses from the participants I compared them to the heuristics from the readings. These are the new heuristics identified from the study specific to PTC discovered through the course of the study:

Effective PTC HCI Measurables

System Limitations	Engineer recognizes elements that are beyond system capabilities and corrects for them
Planned or Anticipated Events	Expected system behavior
Unplanned Events	Unexpected system behavior
Crew Interaction	Are the conductor and engineer discussing system prompts and events to more efficiently operate their train?
Error Recognition	Troubleshooting abilities
Target Approach Management	Ability to navigate the system with tight tolerances

(Table 6)

It was also the general consensus of all participants that greater efficiency equated to an engineer's improved ability to problem solve system issues. This goes back to Nielsen's ninth heuristic of assisting users in recognition, diagnosis and recovery from errors. A fundamental measure of performance is the engineer's understanding of an issue and ability to correct it.

FINAL RECOMMENDATION

The respondents were quite willing to participate but the concept of interaction elicited varying degrees of definability. Five of those interviewed put heavy emphasis on the onboard screen prompts stating that users should be focused almost solely on the screen. In contrast four of the respondents stressed the importance of the engineer staying focused on what was happening out in front of the train through the windshield. The remainder of the respondents replied by placing varying degrees of importance upon where the engineer's focus should be while operating with PTC. From an observational standpoint an engineer focused on the CDU screen for the majority of the trip or more focused on what is taking place out in front of the train through the windshield would be equally achievable.

The word interaction when tied to the use of the system was viewed divergently. There were essentially two general schools of thought that were "what are the engineer's missing?" and "what is the system lacking?" both of which encourage further observation.

Ultimately structured, direct-participant observations conducted by fully trained observers would be the most effective method for observing. These observations will always have the capability to be backed up with recorded data if further analysis or explanation is required for the examiner or the employee who has been observed. Those direct observations should focus on the six areas identified in the study; system limitations, anticipated events, unplanned events, crew interactions/intra-cab communications, error recognition and target approach management.

CONCLUSION

Through a quantitative approach to data collection objectivity and the ability to replicate the scenario in virtually any locomotive cab will be achievable. The parameters of the observation should be broad enough so as not to oversimplify the quantifiable variables, thus reducing a complex situation down to few generalized chunks of information. Thorough observations and results could lead to opportunities for future studies as well. "A line (or program) of research is a connected series of studies within a particular problem area that results in progressively more complex research findings regarding the phenomenon under study. These lines of research can cross over into other disciplines and generate new lines of research that diverge from the original line of research." (Teddlie, C., & Tashakkori, A., 2009) The results of quantitative observational research when properly conducted will be independent of the person conducting the observation. This means that any qualified observer who is following the procedures outlined should report similar results. This does not imply that each observation will be identical because they will not be, but the scenario being observed will be similar and the results will be groupable and allow for the identification of trends. This ability to produce similar results gives credibility to the study when seen from an external audience's perspective.

<u>Works Cited</u>

Adler PA, Adler P (1994) Observational techniques. In Denzin NK & Lincoln YS (Eds) *Handbook of Qualitative Research.* London, Sage.

Campbell, C., & U.S. Army Research Institute for the Behavioral Social Sciences. Armored Forces Research Unit. (2003). *Research observations and lessons learned for the future combat systems* (Research product; 2003-04). Alexandria, VA: Armored Forces Research Unit, U.S. Army Research Institute for the Behavioral and Social Sciences.

Campbell, C. H., Deter, D. E., Ford, L. A., Graves, C. R., Campbell, R. C., Pratt, D. M., Jenkins, S. N., & Quinkert, K. A. (1999). Combined Arms Operations at Brigade Level, Realistically Achieved Through Simulation III (COBRAS III): Report on Development and Lessons Learned (ARI RR 1736) [DTIC ADA359920]. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Davis MZ (1986) Observation in natural settings. In Chenitz WC & Swanson JM (Eds*) From Practice to Grounded Theory: Qualitative Research in Nursing*. Sydney, Addison-Wesley.

Jacko, Julie A., and Andrew Sears. *The Human-Computer Interaction Handbook : Fundamentals, Evolving Technologies and Emerging Applications*, Third Editiion. CRC Press, 2003.

Jorgensen, D. (1989). *Participant observation : A methodology for human studies* (Applied social research methods series ; v. 15). Newbury Park, Calif.: Sage Publications.

Lofland J, Lofland LH (1995) *Analyzing Social Settings: A Guide to Qualitative Observation and Analysis*. Third edition. Belmont, California, Wadsworth.

Morse JM, Field PA (1996) *Nursing Research: The Application of Qualitative Approaches*. Second edition. London, Chapman & Hall.

Nayak, P. R., Rosenfield, D. B., & Hagopian, J. H. (1983). *Event probabilities and impact zones for hazardous materials accidents on railroads* (No. DOT-TSC-FRA-83-5). United States. Federal Railroad Administration.

Nielsen, Jakob. (1994). Heuristic evaluation. In Nielsen, J. and Mark, R.L. (Eds.) 'Usability Inspection Methods', New York: John Wiley & Sons.

University of Portsmouth. *Primary Data Collection - Observations*. (2012). Retrieved September 8, 2018, from http://compass.port.ac.uk/UoP/file/664e8001-f121-4e5d-aa06-6c95c797e8af/1/Observations_IMSLRN.zip/page_02.htm

Santos, Novais, Ferreira, Albuquerque, De Farias, & Furtado. (2016). Metrics focused on usability ISO 9126 based. *Information Systems and Technologies (CISTI), 2016 11th Iberian Conference on, 2016*, 1-3.

Suen, H. K., & Ary, D. (2014). *Analyzing quantitative behavioral observation data*. psychology press.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Sage.

Wiersma, W., & Jurs, S. (2005). *Research methods in education : An introduction* (8th ed.). Boston: Pearson/A and B.

Wreathall, J., Roth, E., Bley, D., & Multer, J. (2007). *Human Factors Considerations in the Evaluation of Processor-Based Signal and Train Control Systems: Human Factors in Railroad Operations* (No. DOT/FRA/ORD-07/07). United States. Federal Railroad Administration. Office of Research and Development.

49 U.S.C. § 20157(g)(1), (i)(5); 49 CFR § 236.1005.

APPENDIX 1

Study Plan:

<u>Project Title:</u>

Observations of Engineers engaging the PTC System

<u>Project Duration:</u>

Phase 1: PTC Conference/Interviews: 9/1/2018 – 9/18/2018

Phase 2: Data Analysis and Reporting: 12/1/2001 – 5/08/2002

<u>Purpose of the Study:</u>

The primary purpose of this study is to gain insight into the effective methods for conducting observations of locomotive engineers while interacting with the PTC system. This insight will garner allow for evaluation of those methods as well as research the explicit and implicit heuristics of PTC observation.

<u>Study Design:</u>

The investigator will spend three days face to face with several senior managers of different capacities PTC operations. Additionally, the investigator will spend four days observing train crews and conducting interviews. The company individuals involved will be from the following job capacities: Senior Managers of PTC Operations (SMPO), Director of PTC Ops, Ops Practices, and the General Dir. Of PTC Ops. Participants will be interviewed for 30 minutes each (time permitting) over the course of the three days to assess the benefits and drawbacks to each method. The investigator will be able to determine the viability of each observation method through this process. The interviews will not be tape recorded but detailed notes will be taken. No proprietary information will be recorded or analyzed throughout the process. Prior to any interviews all participants will be fully informed of the scope of the project and no deception techniques will be employed.

<u>Deliverables</u>

All of the data collected will be compiled, analyzed and documented in the form of a written recommendation report. The primary stakeholders will receive a copy of this recommendation report at the end of this project, which is expected to be completed in October, 2018. Additionally, the finished report will be the applied project of the investigator to fulfill the requirements of a Master's degree.

<u>Confidentiality:</u>

No names (aside from the author's) will be attached to the document nor will any names will be used in any of the reports on the data or in relationship to the research. The data will be comingled and therefore reported in aggregates to maintain confidentiality among the participants. No trade secrets or proprietary information will be revealed in the report or the research.

Participation:

All participation in this study is voluntary.