THE USE OF HEART TO ASSESS THE RISK OF REMOTE CONTROL LOCOMOTIVE OPERATIONS: A TALE OF TWO CITIES

Stephen J. Reinach
Foster-Miller, Inc.
Waltham, MA

Steven Fadden
Landmark College
Putney, VT

Frederick C. Gamst
University of Wyoming
Laramie, WY

Sarah A. Acton
SA Technologies, Inc.
Marietta, GA

In an effort to reduce operating costs and increase safety and efficiency, U.S. Class I freight railroads have begun to use remotely controlled locomotives in and around railroad switching yards. To better understand the safety implications of implementing this technology, a human reliability assessment was conducted to compare remotely controlled locomotive operations with conventional (engineer onboard) yard switching operations. This paper discusses application of the Human Error Assessment and Reduction Technique (HEART) with 2 yard switching employee subject matter experts. Each was asked to assess 11 conventional scenarios and 11 nearly-identical remote control scenarios. Human error probabilities were calculated for each scenario. The HEART assessment revealed no overall difference in human error probabilities between the 2 methods of operation. Additional analyses suggest significant variability between the two assessors. This paper explores differences in how assessors used HEART, including differences in selection of generic task types and error-producing conditions.

INTRODUCTION

In an effort to reduce operating costs and increase safety and efficiency, U.S. Class I freight railroads have begun to implement remote control locomotive (RCL) operations in and around railroad switching yards. U.S. railroads are permitted to use RCL operations as long as they follow all relevant Federal Railroad Administration (FRA) safety regulations. RCL operations consist of three components: the locomotive, an onboard control computer that interfaces with the locomotive's controls (and is usually mounted somewhere inside or on the locomotive), and a portable operator control unit used to control the locomotive. A remote control operator typically wears the portable unit on a safety vest so that the unit rests in front of the operator's chest or stomach (see Figure 1). The operator can control the RCL from the ground, from as far away as 1 mile.

Figure 1. RCO

To better understand the safety implications of RCL operations, FRA's Office of Research and Development Human Factors Program and Office of Safety initiated a multi-study RCL operations research program in early 2002, just as RCL operations began on a large scale in the United States. This paper discusses results from a comparative risk assessment of RCL and conventional yard switching operations.

Given the myriad tasks and dynamic environment of railroad yard switching, multiple data collection methods were employed to support the risk assessment. First, a hierarchical task analysis was conducted to delineate the specific tasks involved in RCL and conventional yard switching operations. Next, a preliminary hazard analysis was conducted to assess the overall risk of each method of yard switching operation to help prioritize the most risky operating scenarios. Finally, a human reliability assessment (HRA) was conducted based on the riskiest operating scenarios under each method of operation. Operator reliability is also referred to as human error probability (HEP), since human error is quantified as a probability (likelihood of an error). Operator reliability, thus, refers to the likelihood of an error occurring.

Two complementary HRA techniques were used: HEART (Williams, 1988) and absolute probability judgment. This paper focuses on the application of HEART; see Reinach, Fadden, Gamst, Acton, and Bartlett (2006) for further discussion of other components of the overall research effort.

Originally developed as a less complex and research-intensive technique for the process control industry, HEART is an economical and flexible technique that can be adopted for use in domains that involve human operators performing tasks under different kinds of conditions. HEART was also developed for use by practitioners who do not have extensive training and experience in the technique. Significant benefits of HEART are that (1) HEP values are already provided in the methodology so that subject matter expert (SME) assessors (e.g., a remote control operator or other yard operating employee) do not need to generate reliability estimates on their own, (2) HEPs are based on established human performance data, and (3) it does not require a significant amount of time, resources, or training to use.

METHODS

The HEART process consists of an iterative procedure where an SME assessor examines a scenario to identify the general type of task being performed, along with the conditions that may influence performance in the specific scenario. The HEART process consists of the following steps:

1. Review a description of the scenario to be evaluated.
2. Identify one predefined Generic Task Type (GTT) that best approximates the task described in the scenario.
3. Identify all predefined Error Producing Conditions (EPCs) that influence task performance in the scenario. EPCs are also referred to as performance shaping factors.
4. Assign an Assessed Proportion of Affect (APOA) for each EPC to indicate the extent to which the EPC influences task performance. Assessors could assign a small or large influence for each EPC (i.e., forced choice).
5. Calculate the final HEP. After the SME assessor has selected the GTT, EPCs, and APOA, a research team member calculates the final HEP.

Eleven sets of yard switching scenarios were developed to capture moves that could be performed under either conventional or RCL operations. Scenarios include a variety of EPCs that could contribute to, or degrade, the reliability associated with the system being assessed. For each set, one scenario was developed to be evaluated in the context of a conventional yard switching operation, and a second scenario was developed to be evaluated in the context of an RCL operation. Differences between the two scenarios in each set were due only to operational differences (e.g., the presence or absence of a locomotive engineer).

A pilot study was conducted first to evaluate the feasibility of HEART and to provide an opportunity to modify and streamline the HEART method. Next, the HEART assessment was conducted. Typically, HEART is used with only one SME assessor. This is how HEART was designed to be used. To sample from a broad range of operating experiences, environmental conditions, and operating characteristics, however, separate HEART assessments were conducted with two different SME assessors. One SME came from, and worked in a railroad yard in, a large Southern city in the United States (location A), while a second SME came from, and worked in a railroad yard in, a large Midwestern city (location B). Each SME had both RCL and conventional yard operations experience, and each evaluated both the RCL and conventional scenarios.

RCL operations were expected to result in greater HEPs since they are a new method of operation and have the concomitant challenges associated with new operations, such as the use of inexperienced operators and unforeseen consequences of introducing new technologies.

RESULTS

To examine HEP data, a criterion of acceptable precision was adopted from Kirwan (1996). This criterion is based on the general practice used by analysts in the HRA field, where “most estimates are within a factor of 10 of the true value” (Kirwan, 1996, p. 362). Specifically, HEPs that were within 1 order of magnitude (OOM) were considered equal, while HEPs that were 1 OOM or more apart from each other were considered different.

Table 1 presents the resulting HEP data from the HEART sessions. The table identifies HEP estimates generated by each participant, listed by scenario and type of operation (conventional (CONV) or RCL). Within each scenario and participant, the less reliable HEP is presented in a grey shaded cell if the HEP is at least 1 OOM larger than the comparative HEP. Considerable variability exists in the HEP data. Using the 1 OOM criterion, both participants identified 4 conventional scenarios as more unreliable than their RCL counterparts, 3 RCL scenarios as more unreliable than their conventional counterparts, and the remaining scenarios with no difference.

It is possible that the HEART data demonstrate that each method of operation is associated with some degree of risk, and that this risk depends on the scenario. To provide insight into this possibility, HEPs were compared between participants for each common scenario to determine the degree of agreement between the two HEART assessors. Comparisons are shown on the right-hand side of Table 1.

For both types of operation, participants were within an acceptable degree of precision for about half the scenarios (5 of 10 HEPs for conventional, and 5 of 11 HEPs for RCL). Given that participants generated HEPs that were within 1 OOM for only about half of the scenarios, this raises a question regarding the reliability of the HEART participants and possibly that of the HEART method.

Because the HEART HEPs are dependent on the selection of tasks and conditions from specific lists of information, the selection of specific GTTs and EPCs was further examined for conventional and RCL scenarios to shed light on possible sources of variability observed.

Table 2 shows which GTTs each HEART assessor (from location A or B) selected for each scenario. It illustrates that assessors did not agree about the GTT assignment for a substantial number of scenarios. In the table, A is used to indicate the GTT selected for each scenario by the first assessor, while B is used to indicate the GTT selected for each scenario by the second assessor (the fourth scenario for conventional operations was not evaluated by the first assessor). Because GTT selection is a major driver of the final HEP values, an imbalance in GTTs represents a large contributor to variability in the HEP data. The second assessor (B) selected a larger range of GTTs than the first assessor (A), who restricted his selections to only the first 2 GTTs presented on the list. Of the 22 total scenarios that were evaluated, assessors selected the same GTTs for only 8 scenarios, with 3 conventional scenarios being assigned the same GTT and 5 RCL scenarios being given the same GTT.

Assessors' selection of specific EPCs was also examined. While a total of 39 EPCs were available for consideration in each scenario, assessors never chose from among 14 of the EPCs. The remaining 25 EPCs and their selection by scenario and HEART assessor are illustrated in Table 3. In contrast to the EPCs that were never selected by participants, HEART assessors repeatedly selected 5 EPCs (top of Table 3), each of which was selected at least 10 times by each assessor. Selection of these 5 EPCs by both assessors a number of times supports the notion that both assessors saw the applicability of each EPC. However, review of the selection patterns of these 5 EPCs reveals that they were not selected with equal frequency—each of the 5 EPCs was selected more often by one participant than the other. This imbalance in selection frequency is even more pronounced in the remaining 20 EPCs: 15 of the 20 remaining EPCs were selected by one HEART assessor or the other, but not both.

DISCUSSION

The HEART assessment revealed no overall difference in HEPs between the two methods of operation, i.e., no clear pattern in favor of one type of operation over another. It is possible that RCL and conventional operations are both associated with some risks resulting in greater HEP for one type of operation over the other, depending on the scenario.

Analyses suggest significant variability, however, between the two SMEs regarding the selection of GTTs and EPCs for the different scenarios. The substantial number of differences between SME assessors' data suggests that their understanding of the scenarios, and how the GTTs and EPCs applied to scenarios, was inconsistent, leading to inconsistent HEPs. Assessors appeared to be poorly calibrated to one another, possibly as a result of (1) insufficient training on HEART, (2) poor understanding of the HEART method and/or the scenarios, (3) lack of fit between the HEART method and railroad yard operations, or (4) a problem with the reliability of the HEART method. It is also possible, of course, that two assessors are simply too few to yield consistent results, and with a greater number of HEART participants, greater consistency may emerge among scenario assessments. Typically, however, HEART is used with only one assessor.

The inconsistency in the use of HEART highlights a source of concern associated with the HEART approach and suggests that more research is needed to determine the reliability of the HEART risk assessment method and its suitability for use in certain complex operations such as railroad yard operations that involve myriad tasks and dynamic variables. This need for more research into the application and reliability of HEART is especially important, given that HEART is ordinarily used with only one assessor, and therefore, the reliability and suitability of HEART and its results may not lend themselves to such evaluation. In the meantime, researchers who wish to use HEART should consider employing a controlled experimental design with multiple participants and applying parametric and/or nonparametric statistical tests to analyze results. This approach will help to minimize any variability introduced by the HEART method.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Thomas Raslear, FRA Office of Research and Development Human Factors Program, and Mr. John Conklin, FRA Office of Safety, for supporting this research. The authors would also like to thank participating railroad employees for sharing their railroad operations expertise. This research was performed under FRA Office of Research and Development Contract DTFR53-01-D-00029. The views and opinions expressed here are solely the authors and do not necessarily represent those of the sponsoring agency or participants.

REFERENCES

Kirwan, B. (1996). The validation of three human reliability quantification techniques–THERP, HEART and JHEDI: Part 1–technique descriptions and validation issues. *Applied Ergonomics, 27 (6)*, 359-373.

Reinach, S., Fadden, S., Gamst, F., Acton, S., & Bartlett, S. (2006). *A comparative risk assessment of remote control locomotive operations versus conventional yard switching operations.* (DOT/FRA/ORD-06/09). Washington, DC: Federal Railroad Administration.

Williams, J. (1988). A data-based method for assessing and reducing human error to improve operational performance. Proceedings of the *IEEE Fourth Conference on Human Factors and Power Plants*. Monterey, CA, 5-9 June, pp. 436-450. IEEE: New York.

Table 1. HEART HEP data using GTT nominal human unreliability values

Scenario	**Location A**		**Location B**		**Agreement between participants? (common HEPs within 1 OOM)**	
	CONV	**RCL**	**CONV**	**RCL**	**CONV-CONV**	**RCL-RCL**
S1	1.000	0.240	0.032	0.452	No	Yes
S2	0.066	0.001	1.000	1.000	No	No
S3	0.002	0.005	1.000	1.000	No	No
S4	(missing)	0.0007	1.000	0.0006	N/A	Yes
S5	0.0009	0.013	0.002	0.0008	Yes	No
S6	0.004	0.002	0.098	0.00006	Yes	No
S7	0.006	1.000	0.0004	0.036	Yes	No
S8	1.000	0.022	0.009	0.004	No	Yes
S9	0.340	0.014	0.015	0.013	Yes	No
S10	0.011	0.118	0.0002	0.716	No	Yes
S11	0.002	0.001	0.011	0.0004	Yes	Yes
Number of scenarios with greater HEPs (shaded cells; at least 1 OOM)	4 CONV worse than RCL 3 RCL worse than CONV 3 scenarios with no difference		4 CONV worse than RCL 3 RCL worse than CONV 4 scenarios with no difference		5 of 10 estimates for CONV agree	5 of 11 estimates for RCL agree

Table 2. HEART GTT assignments by scenario

GTT description	**Conventional scenarios**											**RCL scenarios**										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Completely familiar, well-designed, highly practiced…		A	A		A B	A	A B	B	B	A	A B		A	A	A B	A B	A	B	A	A B	A	A B
Complex task requiring high level of comprehension…	A		B	B				A	A			A B	B	B				A				
A task that requires a response to a system…										B							B		B			
Totally unfamiliar, performed at speed…		B																				
Fairly simple task performed rapidly…						B																
Routine, highly-practiced, rapid task…																					B	
A task to restore or shift system…	B																					

Table 3. HEART EPC selection frequency

EPC description	Total	CONV total	RCL total	A participant total	B participant total
Operator inexperience	25	16	9	16	9
Task pacing caused by the intervention of others	25	14	11	15	10
Unfamiliarity with a situation which is potentially important but which only occurs infrequently or which is novel	19	10	9	12	7
The need to transfer specific knowledge from task to task without losing (forgetting) that knowledge	13	1	12	10	3
Task performance standards are unclear or confusing	10	6	4	4	6
Poor or confusing information conveyed by procedures and person-to-person interaction	8	5	3	0	8
Disruption of normal work-sleep cycles	6	4	2	5	1
A need for making decisions which are beyond the capabilities or experience of an operator	6	2	4	4	2
A poor or hostile environment (below 75% of health or life-threatening severity)	6	6	0	1	5
A danger that the physical abilities to perform task will be exceeded	5	1	4	0	5
A shortage of time available for error detection and correction	5	2	3	0	5
Inconsistency of meaning between displays and procedures	4	2	2	0	4
No clear or timely confirmation to the operator of a performed action	4	1	3	0	4
High-level emotional stress	4	1	3	4	0
A mismatch between perceived and real risk	4	1	3	1	3
No means of conveying technical information (position, location, system state) to operators in a form which they can readily understand	2	0	2	0	2
Alarms or signals are not easily heard over other noises present in the environment	2	1	1	1	1
Low workforce morale	2	1	1	2	0
Age of personnel performing perceptual tasks	2	2	0	0	2
A mismatch between the educational achievement level of an individual and the requirements of the task	2	0	2	0	2
Unclear allocation of function and responsibility	1	1	0	0	1
No obvious means of reversing an unintended action	1	1	0	1	0
Noticeably unreliable equipment (i.e., faulty gauges, inaccurate feedback, no feedback)	1	1	0	0	1
Poor, ambiguous, or ill-matched system feedback	1	1	0	0	1
Little or no intrinsic meaning in a task	1	0	1	0	1