Available online at www.sciencedirect.com

SCIENCE DIRECT®

Safety Science 44 (2006) 313–334

SAFETY SCIENCE

www.elsevier.com/locate/ssci

HEPI: A new tool for human error probability calculation for offshore operation

Faisal I. Khan [a,*], Paul R. Amyotte [b], Dean G. DiMattia [b]

[a] *Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, Nfld, Canada A1B 3X5*

[b] *Department of Chemical Engineering, Dalhousie University, Halifax, NS, Canada B3J 2X4*

Received 2 March 2005; received in revised form 20 July 2005; accepted 7 October 2005

Abstract

Human error plays a sometimes overriding role in accident causation through either direct action or poor design. Offshore operations are particularly susceptible to these human errors due to the complex working environment. Offshore platforms have significant potential for severe ramifications and thus present a challenging scenario for human error prediction and mitigation. The burgeoning offshore industry in the North Atlantic region presents new territory and opportunity to further the case for human error identification and prevention. Due to the relatively slow progress in the field of quantification of human reliability, there is a need to advance this area of research and provide techniques that are useful in human error quantification and that can be embedded in the main framework of quantitative risk assessments.

Recently, a new human error probability index (HEPI) has been developed based on the SLIM (success likelihood index methodology) approach. The application of HEPI could be used to limit the opportunities for human error occurrence and mitigate the results of such errors through changes in training, design, safety systems and procedures, resulting in a more error tolerant design and operation. This paper aims to present a brief description of HEPI for the offshore muster process. Application of the developed index is illustrated through a case study of a previous incident.

Keywords: Human error; Human error probability; Offshore emergency operation; Mustering operation

---

[*] Corresponding author. Tel.: +1 709 737 8939; fax: +1 709 737 4042.
E-mail address: fkhan@engr.mun.ca (F.I. Khan).

0925-7535/$ - see front matter
doi:10.1016/j.ssci.2005.10.008

1. Introduction

Human factors play a major role in offshore emergency operations (muster, evacuation, etc.) and their successful outcome. The importance of human factors in offshore operations has been recognized through several reports published by the UK Health and Safety Executive dealing with the inclusion of human factors in the offshore industry (Widdowson and Carr, 2002), and the human factors assessment of safety critical tasks in the offshore industry (Johnson and Hughes, 2002). These reports provide guidance for the integration of human factors principles into offshore system design, development and operation. However, initiatives have not been developed to quantify the human error probabilities (HEPs) associated with the major actions that take place during a platform emergency situation (such as a muster). On a regulatory basis, there is generally no clear definition or specific requirement for the inclusion of human error considerations in management systems or risk assessments. This is likely due to the lack of readily available human reliability assessment (HRA) tools. The current work was therefore undertaken with the objective of providing such a tool for HEP calculation and consideration of the risks arising from such errors.

The probability of an error (incorrect action) is dependent on the relevant performance shaping factors (PSFs). HRA tools should allow the analyst to take into account characteristics of the task, physical environment, organizational environment and operator characteristics. The application of PSFs in human reliability analysis of emergency tasks is seen as an important factor in HRAs (Kim and Jung, 2003). This necessitates the meaningful collection of human performance data by specifying the contextual information that needs to be gathered to allow aggregation of error into categories with underlying commonalties. Databases tend to focus on the types of injuries and operations that were carried out at the time of the accident (Gordon et al., 2001).

Although there has been some degree of research applied to the quantification of human error probabilities, only a few of these techniques have actually been applied in practical risk assessments (Embrey et al., 1984). Techniques that have been proposed as a means of estimating HEPs, and that have enjoyed the widest application, include HEART (human error assessment and reduction technique), THERP (technique for human error rate prediction), and SLIM (success likelihood index methodology). These are the methodologies that human error and human factors texts most often reference (Bedford and Cooke, 2001; Embrey et al., 1994).

Extensive reviews of many of these techniques have been conducted (Kirwan, 1994, 1998; Kirwan et al., 1988). Several studies (Williams, 1983; Apostolakis et al., 1988; Kirwan and James, 1989; Zamanali et al., 1998; Spurgin and Lydell, 2002) have been presented that focus on the SLIM procedure along with other expert judgment methods (e.g. THERP and HEART). The systematic human actions reliability procedure (SHARP) provides a methodology by which human reliability assessment techniques (e.g. THERP) can be applied to risk assessments (Miller and Swain, 1987). The SHARP framework has been considered as a basis from which an industry standard may be set. As described in subsequent sections, the current work extends the SHARP framework by including recommendations to mitigate the probability of human error and reassess the initial risk level based on the implementation of these recommendations.

The SLIM technique has evolved and taken on several forms since its initial development and follow-on modification (Embrey et al., 1984). An example is the failure likelihood index method (FLIM), which utilizes a failure likelihood index (FLI) as opposed to a success likelihood index (SLI; Chien et al., 1988). Because of the central role of SLIM in the current work, a brief review of notable applications of SLIM and its derivatives is presented below.

Dougherty and Fragola (1988) developed time reliability correlations (TRCs) to predict the probability of failure of an action. A TRC is a probability distribution based on the time to complete an action and the action's likelihood of success (Bedford and Cooke, 2001; DiMattia et al., 2005). Dougherty and Fragola's approach was based on the premise that if a correct diagnosis is not made within a critical period of time, then a failure occurs.

Kirwan (1998) conducted an exhaustive review and evaluation of a wide range of human error identification (HEI) techniques. SLIM was treated as a means of quantifying HEPs but was not evaluated alongside other techniques such as THERP. Kirwan noted that communication in emergency and routine situations is often a contributor to, or a cause of real events. Kirwan's work also reinforces the importance and role of human error in risk assessment and stresses the need to adopt a scientific approach toward predicting and managing human error. Zamanali et al. (1998) applied SLIM through a team that included operators, to predict human error rates (HERs) for a processing plant. PSFs were treated as directly acting and indirectly acting. Weighting factors were calculated by expert judgment via a pairwise comparison of the importance of each PSF relative to the other. The PSFs used in their study were rushing, training, experience, plant conditions, personnel availability, communication, action consequences, confusion and equipment location.

Spurgin and Lydell (2002) reviewed both SLIM and FLIM along with HEART and THERP. A noteworthy comment by Spurgin and Lydell is that there still exists a significant gap between academic research and practical HRAs. The human error probability index (HEPI) developed in the current work is an attempt to help bridge this gap and provide meaningful human error reduction suggestions.

The ability to use SLIM in a what-if approach is a useful and powerful aspect of this technique that permits a variety of emergency scenarios to be analyzed efficiently. Since the comparative work by Kirwan et al. (1988), SLIM has evolved into a widely recognized expert judgment technique that employs judges to provide numerical feedback that is used as input to formulate the probabilities associated with human error.

Expert judgments are primarily of interest when data are lacking. In the case of offshore platform emergency situations, there are few data directly pertaining to human errors that may occur during a muster, especially for musters that occur under severe circumstances (e.g. fire and explosion). It is through the application of SLIM that HEPs were estimated under such circumstances in the current work. Before describing the development and use of HEPI for muster operations, a brief description of the SLIM approach employed in the current work is presented in the following section.

2. SLIM approach and its application in offshore emergency operations

The success likelihood index methodology is a rating oriented model originally developed with the support of the US Nuclear Regulatory Commission, and has been successfully applied to other industries including chemical manufacture and transport (Embrey et al., 1994; Bea, 2001; DiMattia, 2004). The SLIM technique is intended to be applied to tasks at any level of detail, making it applicable to a range of operations including musters and evacuation. The development of SLIM was prompted by a lack of HEP data (Embrey et al., 1984)—an issue that still exists today in the public domain. With SLIM, errors can be quantified at various stages, including whole tasks, sub-tasks, and task steps. The premise behind SLIM is that the probability of an error associated with a task, sub-task, or task step is a function of the performance shaping factors associated with the task. Performance shaping factors (PSFs) are those parameters influencing the ability of a human being to complete a given task. In reality, a large number of PSFs can affect the probability of failure (POF). Through hierarchical task analysis (discussed in DiMattia, 2004 and DiMattia et al., 2004, 2005), a list of 11 PSFs were initially considered; subsequently using pairwise comparison, the six most relevant PSFs were identified. The PSFs considered in the current SLIM analysis of offshore platform emergency musters are:

- Stress,
- Task complexity (Complexity),
- Level of training (Training),
- Level of experience (Experience),
- Factors associated with the muster initiator (Event Factors), and
- Factors associated with the weather/environment (Atmospheric Factors).

As described in detail by DiMattia (2004) and DiMattia et al. (2004, 2005), three reference muster scenarios of varying severity (man overboard, gas release, and fire and explosion) were studied. The critical first step was the formation of the team of judges who were to generate the relevant data (selection, weighting and rating of PSFs) for the muster scenarios. A grouping of five judges, known as the core review team, was selected for the initial tasks of deciding on the muster scenarios themselves, the specific muster actions, and the set of performance shaping factors to be used (see for details DiMattia, 2004; DiMattia et al., 2004, 2005). The following selection criteria were used for the core review team:

- Actively involved in offshore activities as a member of a producing company or regulator.
- Actively participated in platform musters or involved in the design or evaluation of platform safety systems.
- Participated or led risk assessments in offshore related activities.
- Minimum of 10 years of industrial experience in hydrocarbon processing.
- Capable of dedicating the required time to perform evaluations and committed to participate as required.
- Does not work directly for any other member of the CRT or with any member of the CRT on a daily basis.
- Available to meet in person during work hours.

The elicitation of PSF weights and ratings which are used in human error probability calculation was conducted by the elicitation review team consisting of the five members of the core review team and an additional 19 judges. (PSF weight is the relative importance of a given PSF in comparison to the PSF judged to be the most important; PSF rating is a measure of the quality of a given PSF.) The elicitation review team was thus composed of

24 judges whose primary job functions were: engineering (14 members), operations (6), health and safety (3), and administrative (1). Further details on judges' qualifications and backgrounds are given by DiMattia (2004) and DiMattia et al. (2004, 2005).

Once elicited, the PSF weight data were subjected to statistical analysis (ANOVA and Kruskal–Wallis) to test various null hypotheses aimed at determining whether, for example, the muster scenarios affected the judges' PSF weights for each muster action. Such quantitative and other qualitative tests are documented in detail in DiMattia (2004). The conclusion reached is that the elicited PSF weight data are rationally explainable and show no significant biases arising from the team of judges that provided the data (e.g. due to sample size, background qualifications, etc.). The PSF weight and rating data were subsequently processed by means of SLIM to calculate the success likelihood index (SLI) and probability of success (POS) for eighteen muster actions ranging from point of muster initiator to the final actions in the temporary safe refuge (TSR). Subtraction of the POS from unity resulted in a calculated HEP value for each muster action.

The full set of calculated human error probabilities for the three reference muster scenarios (man overboard, gas release, and fire and explosion) represents the foundation of HEPI. Confidence in these predicted HEP values arises due to the rigorous and scientifically validated process of data elicitation and analysis afforded by SLIM. (Again, all supporting details may be found in DiMattia (2004) and DiMattia et al. (2004, 2005).) In particular, the relatively large number of judges (24) used in the current work for data elicitation helps to overcome the concern expressed by Apostolakis et al. (1998) regarding the use of SLIM. These authors argued that the use of mean values of PSF weights and ratings (based on expert judgment) as point estimates, does not account for data uncertainties. It has been shown, however, by Johnson et al. (2001) that the analytic value of averaged probability judgments increased in accuracy as the number of judges increased; hence our use of a judging group with many members. These judges exhibited a wide range of years of experience and training, and the elicitation was conducted on an individual basis over an extended period of time, thus lowering the possibility of joint work (i.e. eliminating conditional dependence).

The remainder of this paper deals with the extension of the SLIM-generated HEP data to the more widely applicable framework of HEPI. This was accomplished by development of a generalized muster questionnaire and reference graphs for PSF weights, PSF ratings, and POS and SLI values, coupled with incorporation of a consequence table, risk matrix, risk reduction measures, and a risk re-ranking procedure. Details on all of these features follow in subsequent sections.

3. HEPI methodology

The human error probability index provides a proactive, quantitative approach for the inclusion of human factors in risk assessments. It was developed to provide a methodology for the identification, assessment and mitigation of risk due to human error during offshore platform emergency musters. Risk is estimated by calculating human error probabilities and their consequences for the various muster actions. Fig. 1 outlines the main steps of the HEPI process. Steps 1–5 can be completed by an individual performing a risk analysis and are quantitative in nature. An experienced team performing a risk analysis should complete steps 6–8. These steps require the team to draw upon experience to answer pertinent questions. The results of this index provide a clear indication of situations that

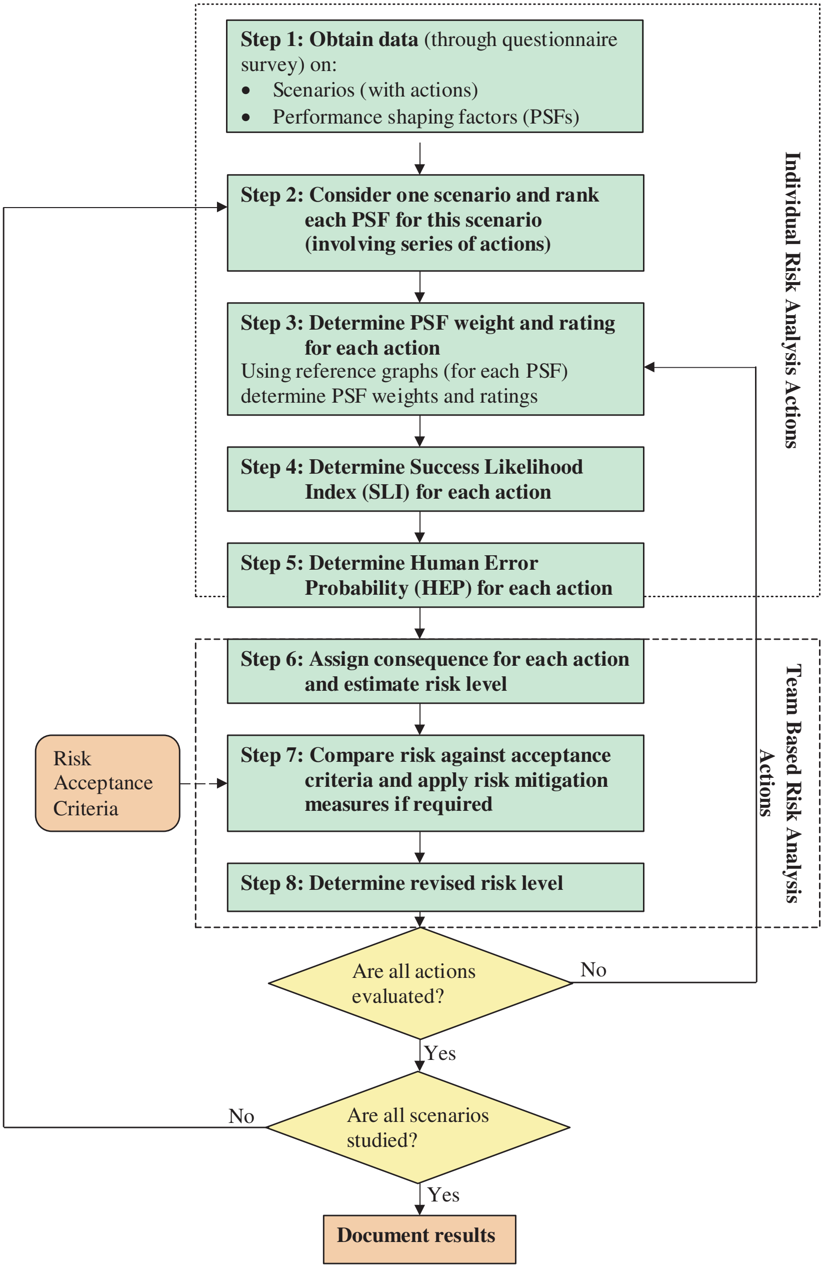

Fig. 1. HEPI methodology.

provide the highest risk and the muster actions that are most likely to end in failure. Through HEPI, error reduction recommendations for training, procedures, management systems and equipment are brought forward, allowing operators and designers to effectively apply human factors in a proactive approach to high risk scenarios. The framework of HEPI has universal application in the offshore industry providing a human factors tool to engineers, operators and health and safety personnel.

4. HEPI for offshore muster operations

The muster sequence considered in HEPI involves 18 distinct actions as shown in Fig. 2. (It should be noted that action 13—collect personal survival suit if in accommodations at

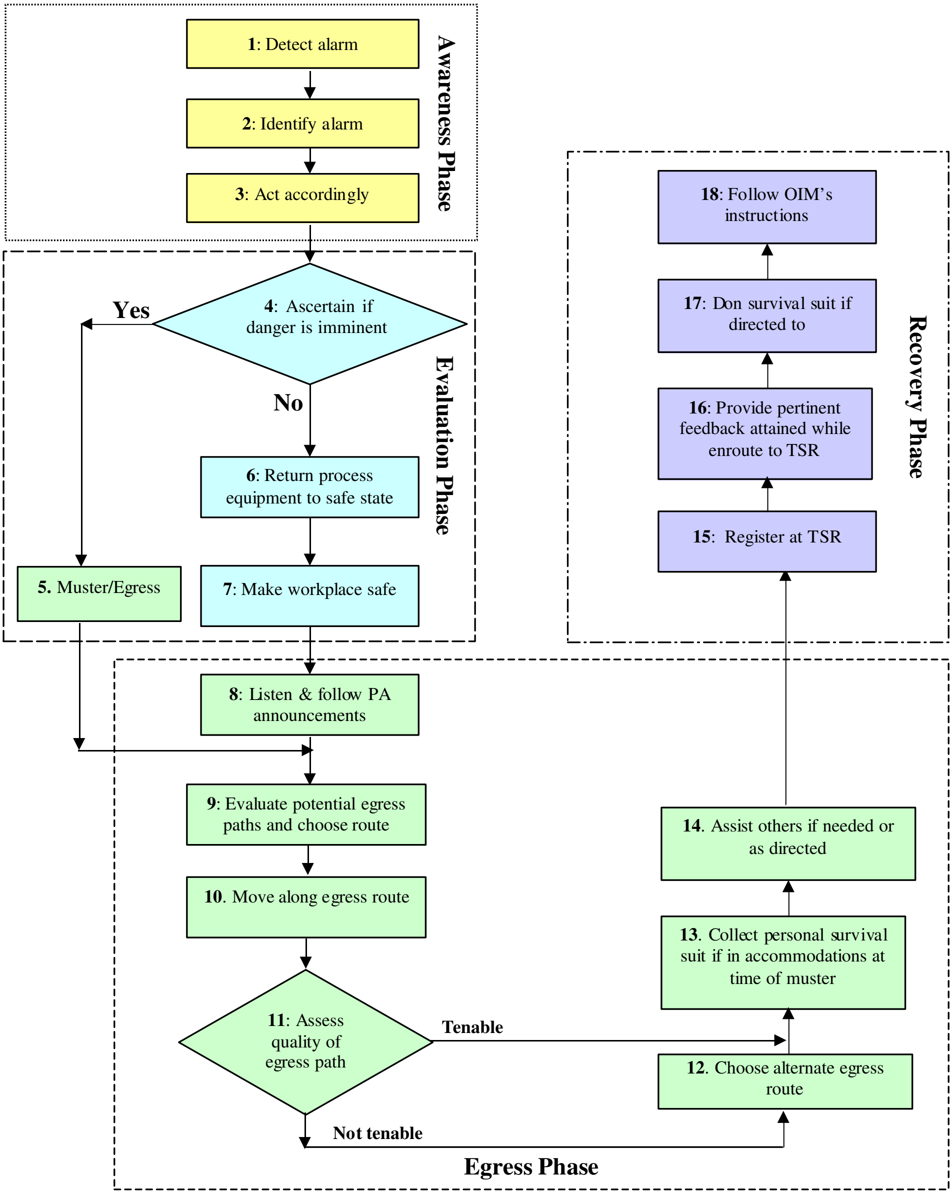

Fig. 2. Muster actions broken down by muster phase (number denote sequence order for actions).

time of muster—was not part of the reference muster scenarios used to elicit the HEP data, and is therefore outside the scope of the current discussion.) Whether these muster actions are successfully completed is a function of a number of factors—for example, the severity of the muster initiator and its effects on the various performance shaping factors. As previously discussed, and as described in detail by DiMattia et al. (2004, 2005), the reference muster scenarios were purposely set up to cover a range of severity in terms of their impact on the six performance shaping factors. This enabled the development of a set of reference graphs for PSF weights and ratings such that these parameters could be determined for any credible muster scenario—i.e. not just the three reference muster scenarios. These reference graphs thus represent the generalization of the SLIM-generated HEP data and are the foundation of HEPI.

Accordingly, each muster action has six reference curves (one for each PSF) to determine the weights and ratings. These curves have been placed on a single graph resulting in 17 PSF weight reference graphs (i.e. one for each muster action) and 17 PSF rating reference graphs. (Recall that although there are 18 muster actions, action 13—collect personal survival suit if in accommodations at time of alarm—is not part of the three reference musters.) Examples of the PSF weight and rating reference graphs are given by Figs. 3 and 4, respectively, for muster action 1—detect alarm. (Figs. 3 and 4, and subsequent figures and tables, also show data for the specific case study of the Ocean Odyssey to be covered later in the paper). DiMattia (2004) contains similar plots for the other muster actions shown in Fig. 2 (again, with the exception of action 13 which was not included in the reference muster scenarios). The ordinate values in Figs. 3 and 4 are, respectively,

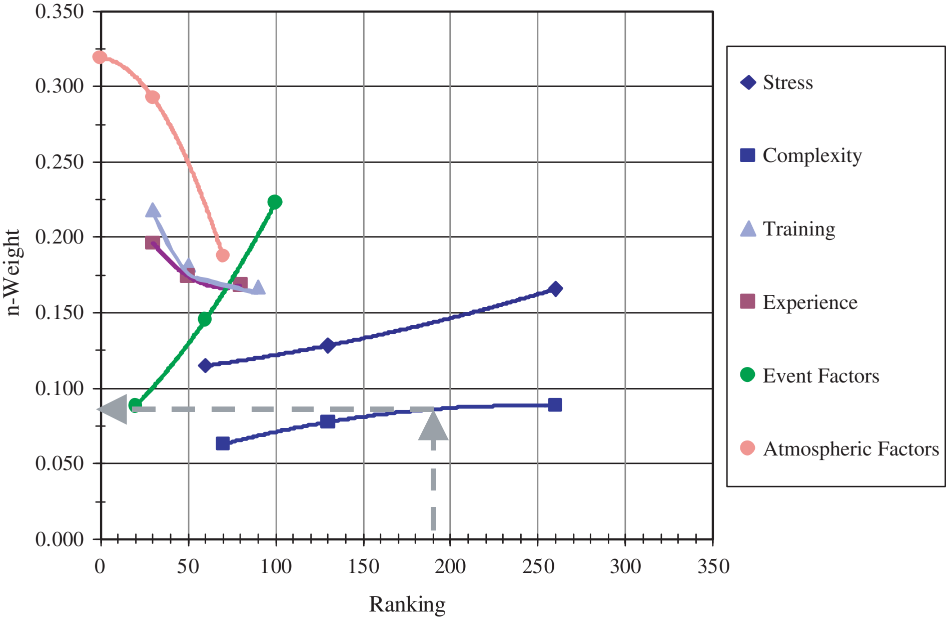

Fig. 3. Reference graph of PSF *n*-weights for muster action 1. (Dashed lines depict data for Ocean Odyssey case study.)

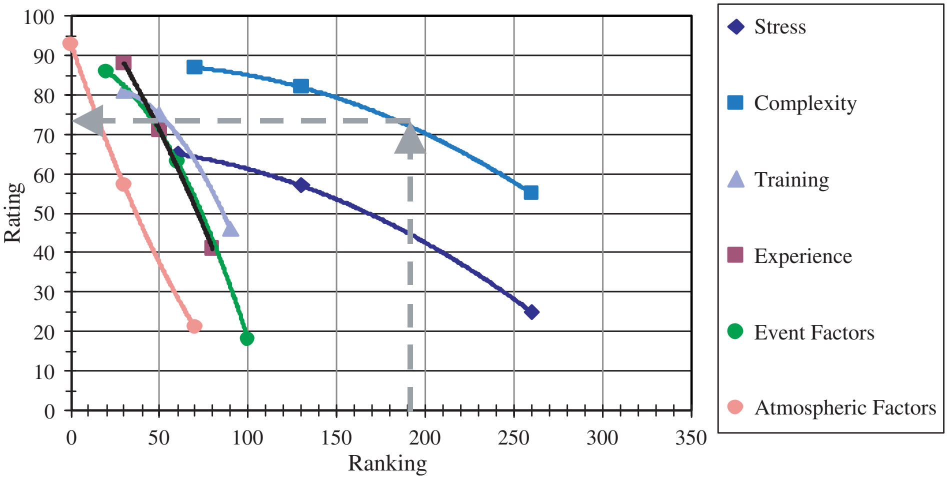

Fig. 4. Reference graph of PSF ratings for muster action 1. (Dashed lines depict data for Ocean Odyssey case study.)

PSF normalized weights (or *n*-weights, which are defined as a given PSF weight divided by the sum of the all the PSF weights for a given action) and PSF ratings. The abscissa values in Figs. 3 and 4 are PSF "rankings", which are determined by the process outlined in the section immediately following. The three data points for each PSF in each of Figs. 3 and 4 correspond to the three reference muster scenarios (man overboard, gas release, and fire and explosion). The best-fit curves joining the data points permit interpolation of PSF *n*-weights and ratings in a consistent manner. These curves are a key component of the mechanism by which the SLIM-elicited data points for the three reference musters are generalized to other muster scenarios.

Step-wise details of HEPI implementation are now presented according to the framework given in Fig. 1.

4.1. Step 1: obtain data (through muster questionnaire)

The first step in the HEPI process is to develop a muster scenario by means of the HEPI muster ranking questionnaire (Table 1). By asking relevant questions, this questionnaire enables eventual determination of the probability of human error during key muster actions. As shown in Table 1, each question has one or more performance shaping factors associated with it. In this manner, the answers to each question will impact on the PSF weights and ratings in subsequent steps of the HEPI process. For example, the relevant PSFs for question 4—What is the time of day when the muster is initiated?—are stress and complexity. If the muster occurs at night, stress levels will be higher than a muster during daylight hours. Similarly, the complexity of the muster actions will be affected by the time of day. For example, egress phase actions are more complex at night due to a lower visibility or due to an individual's level of alertness if woken from sleep. The 12 questions in Table 1 frame the muster scenario by identifying key aspects such as the muster initiator, the mustering individual and the conditions at the time of the muster.

Table 1
HEPI muster ranking questionnaire

Muster Ranking					
(1) What is the muster initiator? (event factors, stress, complexity)					
(i) Drill	0	(iv) Fire	30	(vii) Spill	20
(ii) Man overboard	10	(v) Fire & explosion	**30**	(viii) Helicopter crash	20
(iii) Gas release	20	(vi) Ship collision	20	(ix) Man down	20
Total 30					
(2) What is the immediate risk from the muster initiator? (event factors, stress, complexity)					
(i) No effect on platform integrity; does not impede muster progress; no threat to personnel					0
(ii) Can impede muster progress with potential injury to personnel					10
(iii) Threatens integrity of platform; impedes muster progress; potential for loss of life					**30**
Total 30					
(3) What is the weather at the time of muster? (atmospheric factors, stress, complexity)					
i(a) Sun/cloud	0	i(b) No wind	0	i(c) <−30 °C	30
ii(a) Rain	**10**	ii(b) Windy	**10**	ii(c) −21 to −30 °C	20
iii(a) Snow/sleet	20	iii(b) Significant wind	20	iii(c) −20 to 0 °C	**10**
iv(a) Snow storm	30	iv(b) Hurricane	30	iv(c) 1–30 °C	0
v(a) Heavy fog	20	v(b) Tornado	30	v(c) >30 °C	10
Total 30					
(4) What is the time of day when the muster is initiated? (stress, complexity)					
(i) Daytime	**0**	(iii) Crew change			20
(ii) Night time	20	(iv) Night time after 12 am and before 6 am			30
Total 0					
(5) What is the individual's description (job type)? (training)					
(i) Operator	**10**	(iii) Maintenance	20	(v) Kitchen staff	30
(ii) Engineer	20	(iv) Administration	30	(vi) Cleaning staff	30
Total 10					
(6) What is the level of the individual's offshore experience? (experience, training)					
i(a) <6 months	40	i(b) Regular hitches	**10**	i(c) No musters	20
ii(a) 6 months to 3 years	30	ii(b) Irregular hitches	20	ii(c) 1–5 musters	10
iii(a) 4 to 10 years	**20**	iii(b) Rare hitch	40	iii(c) >5 musters	**0**
iv(a) >10 years	10				
Total 30					

(7) How familiar is the individual with respect to their job at the time of muster? (stress, experience, complexity)					
(i) Routine task that is familiar	**10**	(iii) New task; never done before	40		
(ii) Task that is infrequently performed	20				
Total 10					
(8) How complex is the individual's job at the time of muster? (complexity)					
(i) Not complex	10	(iii) Very complex and highly procedural	30		
(ii) Somewhat complex and procedural	**20**	(iv) Very complex; highly procedural; team required	40		
Total 20					
(9) What is the level of criticality of the job at the time of muster? (stress)					
(i) Job will not result in escalation of muster initiator	**0**	(iii) Job will escalate muster initiator	20		
(ii) Job has potential to escalate muster initiator	10				
Total 0					
(10) What is the location of the individual in relation to the muster initiator? (stress, complexity, event factors)					
i(a) Event initiated on different deck or platform	10	i(b) Event does not affect egress route	0		
ii(a) Event initiated on same deck	**20**	ii(b) Event may affect egress route	10		
iii(a) Event initiated in close proximity to individual	30	iii(b) Event affects egress route	**20**		
Total 40					
(11) How many personnel are onboard (POB) at the time of muster? (stress, complexity)					
(i) <25% POB	10	(ii) 25 to 75% POB	20	(iii) 76 to 100% POB	**30**
Total 30					
(12) Does the individual have any specialized level of training? (training, complexity, stress)					
(i) Not trained—First aid	10	(iii) Not trained—rescue	10		
(ii) Not trained—gas detection	10	(iv) Not trained—fire fighting	10		
Total 0					

Values in bold with a grey background are responses of Ocean Odyssey case study.

4.2. Step 2: rank performance shaping factors

Performance shaping factors are ranked by summing the values obtained from the questions in the HEPI muster ranking questionnaire (Table 1) that are relevant to a particular PSF. For example, the ranking for the PSF training would be the sum of the *Total* values from questions 5, 6, and 12. This rank for each PSF is then used to determine the performance shaping factor weights (*n*-weights) and ratings for each muster action, as described in step 3.

4.3. Step 3: determine performance shaping factor weights and ratings

The next step in the HEPI process is to determine the PSF *n*-weights and ratings based on the rankings determined in step 2. To determine the PSF *n*-weight or rating for a given action, the value is interpolated according to the ranking for that PSF. This involves use of the previously described reference graphs, of which Figs. 3 and 4 are examples for muster action 1 *n*-weights and ratings, respectively. Best-fit curves were utilized in the reference graphs so that interpolation of PSF *n*-weights and ratings could be done in a consistent manner. Any PSF ranking which falls beyond the range of the reference muster rankings (i.e. the end-points of the best-fit curves), retains the boundary value to prevent extrapolation of the curves. The values for the PSF *n*-weights (σ) and ratings (δ) beyond these boundaries are considered unknown and cannot be predicted by the current methodology.

The *n*-weights and ratings are thus recorded for each muster action. These data are used in step 4 to obtain the success likelihood index data for the various muster actions. Again, it should be noted that there are no *n*-weights or ratings for action 13 because the reference muster scenarios used to elicit the HEP data did not include this action. Additionally, as a procedural point, it is noted that the PSF rankings for these reference scenarios—man overboard, gas release, and fire and explosion—were determined by using the same questionnaire (Table 1) as employed in HEPI.

4.4. Step 4: determine success likelihood index values

Step 4 in the HEPI process requires the calculation of the success likelihood index (SLI) for each muster action. First, the SLI values (Ψ) are calculated as the product of the *n*-weight and rating for each PSF in a given muster action, as shown in Eq. (1). For example, to calculate the SLI for stress in action 1, the *n*-weight determined in step 3 for stress in action 1, would be multiplied by the rating (as found for stress in action 1).

$$\mathrm{SLI}(\Psi) = n\text{-weight}(\sigma) \times \mathrm{rating}(\delta) \tag{1}$$

The total SLI (Ω) for a muster action is then the sum of the SLIs for the six PSFs, as shown in the following equation.

$$\Omega = \sum \Psi \tag{2}$$

It is this value of SLI, Ω, that is used in step 5 to estimate the likelihood of successfully completing a given muster action (or alternatively, of not successfully completing the muster action).

4.5. Step 5: determine human error probability values

In step 5, the log probability of success (POS) value is determined for each muster action from one of three SLI reference graphs. These SLI reference graphs were developed from the three reference musters as part of the preliminary SLIM data elicitation and analysis process (DiMattia, 2004). The SLI reference graphs cover a range of SLI values: 76–88 (developed from the man overboard scenario), 45–72 (developed from the gas release scenario) and 20–47 (developed from the fire and explosion scenario). There are therefore some regions of gap and overlap in the SLI reference graphs. If a determined SLI value falls between the reference bounds, it is recommended that the log POS be estimated based on the more conservative (i.e. lower) SLI range. Fig. 5 shows the reference graph for the mid-range set of SLI values ($\Omega = 45–72$). Again, this graph (Fig. 5)—and the two others for the upper and lower ranges of Ω—depicting the linear relationship between log POS and SLI, were developed using the data collected through the elicitation process, details are available in DiMattia (2004).

The inverse (antilog) of the log POS is performed to determine the probability of success (POS). Subsequently, human error probability, HEP, is calculated for each muster action using the following equation:

$$\mathrm{HEP} = 1 - \mathrm{POS} \tag{3}$$

The HEPI user now has an estimated probability of failure for each of the 18 muster actions (except action 13) shown in Fig. 2. These HEP values are derived from the SLIM-elicited and analyzed data which have been generalized to the muster scenario defined by the user in Table 1. The analysis to this point can be conducted by a single person because of the analytical rigour of the SLIM process that is the foundation of the HEPI procedure up to and including step 5.

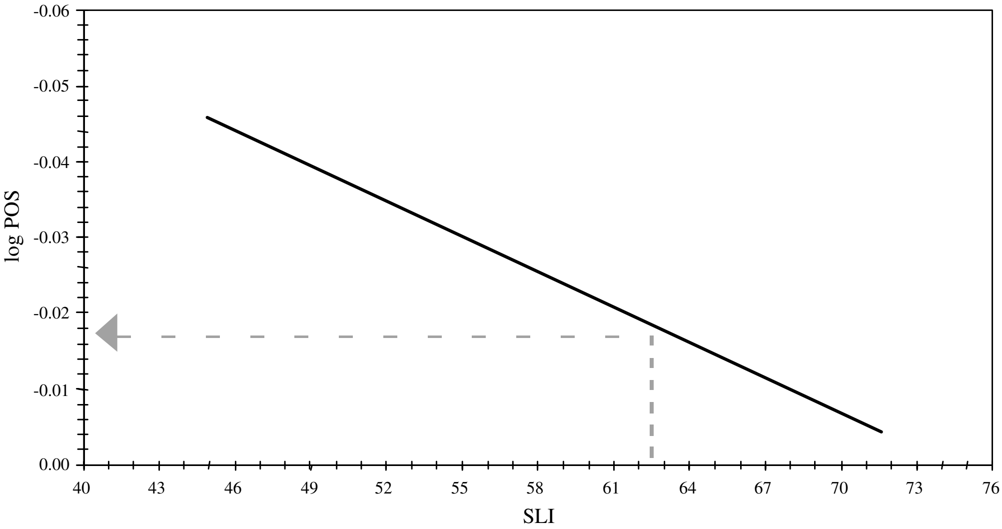

Fig. 5. Reference graph to determine probability of success for mid-range set of SLI values. (Dashed lines depict data for Ocean Odyssey case study).

One component of risk—probability—has been estimated. To enable a comprehensive assessment of risk, it is now necessary to estimate the second component—consequence severity. The HEPI procedure from this point forward is best conducted by a team of knowledgeable individuals because of the somewhat subjective nature of the judgements required.

4.6. Step 6: assign consequences and estimate risk level

The next step is to relate the failure to complete an action with its consequence. As described above, the HEP, in conjunction with an assigned consequence, provides a means to characterize the risk of not completing a muster action. A common industry practice is to relate the severity of an incident to relevant categories through a consequence table (Table 2). The advantage of a consequence table is its flexibility to suit any relevant consequence category with a wide range of severity. The consequence categories for HEPI, as shown in Table 2, are based on four key muster categories; the potential consequences range from simple time delays to loss of life.

Empirical data are required to more accurately assign consequences to human error; observations taken from muster drills can help facilitate this need. In the absence or limited availability of such data, an experienced team should be assembled to assign conse-

Table 2
HEPI consequence table

Severity	Egressability	Other POB	Muster initiator	Health
Critical (1)	Can no longer reach TSR or any other safe refuge. Can no longer have a dry evacuation	Prevents one or more persons from reaching TSR or any safe refuge. Prevents others from having a dry evacuation	Raises muster initiator severity to a level where muster is no longer possible	Results in loss of life
High (2)	Can no longer reach TSR or complete actions in TSR	Prevents one or more persons from reaching TSR or prevents others from completing actions in TSR	Raises muster initiator severity to a level where muster is in jeopardy	Results in significant physical injury
Medium (3)	Moderate to significant delay in arriving at TSR. Moderate to significant delay in completing TSR actions	Moderately to significantly delays others from reaching TSR or their actions in TSR	Raises muster initiator severity to a level that produces moderate to long delays in reaching TSR	Potential for minor to moderate injuries
Low (4)	Minor delay in reaching TSR or in performing actions in TSR	Minor delay for others reaching TSR, or in others completing actions in TSR	Is not likely to raise muster initiator severity and does not affect time to muster to any significant level	No injuries likely

Grey background depicts results from Ocean Odyssey case study.

quence rankings based on first-hand muster experience. As mentioned, the consequence table (Table 2) used by HEPI has four consequence categories and a severity ranking should be assigned to one of these four categories based on the main consequence arising from the action being analyzed.

Once consequence severities have been assigned, the next step is to translate the results into a risk ranking through the use of a risk matrix. Table 3 is the risk matrix that is used in HEPI to relate HEPs and assigned consequences. The dark grey blocks in Table 3 indicate the highest risk, followed by lower risk signified by light gray shading with the lowest risk being associated with the non-shaded (white) blocks. Risks that fall in the dark and light grey shaded areas should be mitigated to the non-shaded areas if possible.

4.7. Step 7: apply risk mitigation measures to reduce risk

In step 7, mitigating measures are suggested for each action based on the format provided by Kennedy (1993). This allows the user to reduce the risk level if deemed necessary from step 6. Tables of risk mitigation measures (RMMs) are available within HEPI to provide suggestions to lower risk through improvements in training, procedures, management systems and equipment. Table 4 provides a list of suggested RMMs for action 1; RMMs for the other muster actions are given by DiMattia (2004). HEPI does not restrict RMMs solely to these categories, as the user may apply additional categories and risk mitigation measures. Work is currently underway within our research group to identify further RMMs that are explicitly based on the principles of inherent safety (see, for example, Khan and Amyotte, 2003; DiMattia et al., 2005).

Based on an action's RMMs, new PSF ratings are established. The revised HEP and consequence severity lead to the determination of a new risk level, as explained in step 8.

4.8. Step 8: determine revised risk level

The *n*-weights signify the importance of each PSF based on each muster action; thus PSF *n*-weights are not recalculated if RMMs are applied. The RMMs do, however, enhance the quality of the PSFs and therefore their rating. Fig. 6 provides an illustration of a suggested procedure to obtain a new PSF rating based on a percent improvement of the original rating for a given action. The percent improvement is based on the difference between the PSF rating determined through the HEPI ranking process, and the optimal PSF rating. This optimal PSF rating is simply the highest rating value given in the reference graphs used in step 3 of the HEPI procedure.

Table 3
HEPI risk table

Human error probability	Consequence severity			
	Critical (1)	High (2)	Medium (3)	Low (4)
A: 0.10–1.0	1A	2A	3A	4A
B: 0.01–0.10	1B	**2B**	**3B**	4B
C: 0.001–0.01	1C	2C	3C	4C

Values in bold are for Ocean Odyssey case study.

Table 4
Possible risk mitigation measures for action 1

Action	Training	Procedures and management systems	Equipment
Detect alarm	1. Familiarization of personnel with alarms 2. Muster training at infrequent intervals 3. Enlisting feedback after training exercises on alarm effectiveness 4. Behavioural studies to determine panic potential 5. Training of control room operators to limit and remove inhibits as soon as possible 6. Training of experienced personnel to assist others as identified	1. Regular preventative maintenance of alarm system 2. Regular testing of alarm system 3. Survey of alarm effectiveness in severe weather conditions 4. Limiting number of alarm types that can be enunciated to lessen potential confusion 5. Identification of new personnel with different coloured clothing 6. Buddy system for new personnel 7. Location board in control room identifying work locations and personnel 8. Equipping all personnel in process units with two-way radios 9. Push buttons in strategic process locations	1. Strategic placement of alarm systems to ensure coverage in all areas 2. Alarm redundancy through both audio and visual enunciation 3. Review of alarm system and comparison with advances in technology 4. Review of applicable regulations and standards

The example in Fig. 6 is a re-rating of the PSF stress for muster action 1. As shown in Fig. 4, the optimal (i.e. highest) rating for this PSF in action 1 is 65. If the HEPI-determined rating was, for example, 50, then a 20% improvement in stress rating would result in a re-rated PSF of 53 according to the following calculation: $50 + [(65 - 50) \times 0.2] = 53$. Steps 4–6 would then be repeated with the new PSF ratings and the consequences reassessed and risk re-ranked.

5. Application of HEPI to offshore muster operations (case study)

To illustrate the use of HEPI, an example is performed using the above discussed procedure. The case is that of the semi-submersible drilling rig, Ocean Odyssey (Odyssey), which suffered an explosion and caught fire following a well blowout in September, 1988. Most of the survivors evacuated via lifeboat, with several individuals jumping directly into the sea (Robertson and Wright, 1997; DiMattia, 2004).

The Odyssey event occurred in the North Sea at mid-day. The muster initiator was a serious fire and explosion on the drilling rig. The weather was moderate at the time with

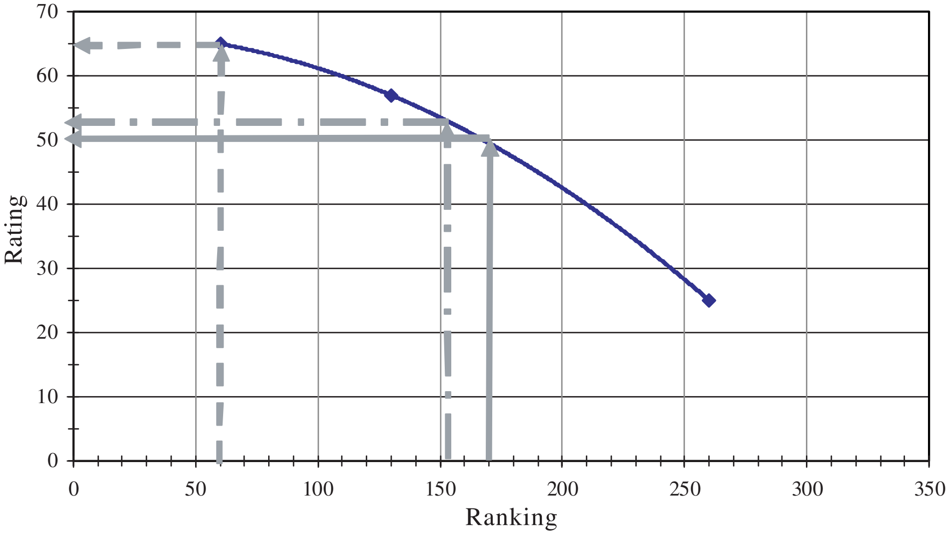

Fig. 6. Re-rating of PSF stress for action 1 based on a 20% improvement through application of RMMs (solid line is the original PSF rating (50), dashed/dotted line is the re-rated PSF (53), and dashed line is the optimal PSF rating (65) for action 1).

wind speeds of 12–18 knots and a visibility of over one mile. There were 67 personnel onboard (POB), of which 58 evacuated by the totally enclosed motor propelled survival craft (TEMPSC) and eight jumped directly into the sea. One person onboard, the radio operator, remained on the rig and perished (DiMattia, 2004).

The event was not a complete surprise, as well-instability was recognized and had occurred over a prolonged period of time. This permitted individuals to mentally prepare for musters and potential evacuations. When the event occurred, the noise from the escaping gas made PA announcements very difficult to hear, with only four of the 22 individuals who were interviewed stating that they heard any PA announcements. Fifteen of the interviewees also stated that they heard no muster alarm at the time of the event, and that the signal to muster was through word of mouth from other POB (Robertson and Wright, 1997; DiMattia, 2004).

Fifteen of the POB went directly to the muster station by a route practiced during their muster exercises, while five survivors were prevented from taking their first choice of egress route by gas, smoke or fire. (Three of these five eventually jumped directly to the sea.) These personnel did not muster immediately with “non-essential” personnel and as such received greater exposure to event factors that delayed their arrival at the TSR. This delay caused several individuals to actually miss the launch of the lifeboats.

The registration at the TSR was chaotic and may have actually occurred on the TEMPSC and not at the muster point. The muster lists were out of date and the confusion of the event resulted in some POB being accounted for twice. The Odyssey was equipped with single-sized survival suits and many of the survivors indicated that the suits were difficult to put on. Three of the POB had no experience in donning these suits and this was further complicated by the loss of dexterity because of the integrated mitts.

The Odyssey's muster was inefficient and personnel relied heavily on one another in all phases of the muster. If the muster initiator had escalated more quickly, the loss of life may have been much greater. The incident report (Robertson and Wright, 1997) provides little information on the single fatality, although it appears that he was on-duty at the time of the muster. Further, there is no information concerning the Odyssey's POB capacity and little information is given in the report of any work activities at the time of muster that may have affected the initiating event or impeded egress routes.

For the purposes of this example (which is restricted to action 1—detect alarm), a likely scenario for the Odyssey is an operator who has a reasonable amount of experience (3–10 years) and is conducting a routine task that has no effect on the initiating event. The first step (HEPI step 1) was to answer the questions in the muster setup table (Table 1, with the Odyssey values in boldface). This enabled the performance shaping factors to be ranked (HEPI step 2), as shown in Table 5. The fire and explosion scenario on the Odyssey was severe; thus, for example, the event factors PSF ranking reflects this fact with a value of 100. Following the HEPI process (step 3), the PSF rankings in Table 5 were used in conjunction with the reference graphs given by Figs. 3 and 4 (with the Odyssey example shown for the PSF complexity). This enabled the PSF *n*-weights and ratings for action 1 to be determined as shown in Table 6.

As per HEPI step 4, the individual PSF SLI values were calculated by Eq. (1) and then summed according to Eq. (2) to yield the total SLI for action 1—a value of 62.3 (see also Table 6). Then, in accordance with HEPI step 5, Fig. 5 (SLI reference graph) was used to determine the log POS value for the SLI value of 62.3. Eq. (3) was then used to arrive at the HEP estimate of 0.04.

In HEPI step 6, the consequence associated with a failure to perform action 1 was assigned via the HEPI consequence table (Table 2). The consequence was deemed to be primarily related to the egressability category and to have a High (2) severity as indicated by the light-shaded box in Table 2. (Clearly, failure to detect the muster alarm may result

Table 5
Ranking of performance shaping factors for Ocean Odyssey case study

Question (Table 1)	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors
PSF rankings						
1	30	30			30	
2	30	30			30	
3	30	30				30
4	0	0				
5			10			
6			30	30		
7	10	10		10		
8		20				
9	0					
10	40	40			40	
11	30	30				
12	0	0	0			
Total	**170**	**190**	**40**	**40**	**100**	**30**

Table 6
PSF *n*-weights, ratings and SLI values for Ocean Odyssey case study

Action	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors
PSF n-weights						
1 (detect alarm)	0.14	0.08	0.19	0.18	0.22	0.29
PSF ratings						
1 (detect alarm)	50	74	79	79	19	55

Action	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors	SLI total
SLI values							
1 (detect alarm)	7.0	5.9	15.0	14.2	4.2	16.0	62.3

in an inability to reach the TSR.) The risk associated with this HEP (0.04—i.e. B) and consequence (High—i.e. 2) is level 2B, as shown in bold-face and light shading in Table 3.

In an effort to reduce the risk associated with this action (re. HEPI step 7), risk mitigation measures were considered (e.g. Table 4). Table 7 provides new (improved) PSF ratings based on a percent improvement of the original ratings for this action. The percent improvement, as previously described, is based on the difference between the PSF rating determined through the HEPI process (Table 6) and the optimal PSF rating (Fig. 4).

Based on the adoption of appropriate RMMs in Table 4, the percent improvements assigned to the PSF ratings for action 1 were: stress—30%, complexity—30%, training—60%, experience—20%, event factors—20%, and atmospheric factors—30%. Stress can be improved (i.e. lowered) through better training methods and the assistance of more experienced personnel. Complexity can be improved (i.e. reduced) through improved

Table 7
Reference table for re-rating of PSFs—action 1

Detect alarm	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors
Optimal rating	65	87	81	88	86	93
Original rating	50	74	79	79	19	55
Improved ratings[a]						
10%	52	75	79	80	26	59
20%	53	77	79	81	32	63
30%	55	78	80	82	39	66
40%	56	79	80	83	46	70
50%	58	81	80	84	53	74
60%	59	82	80	84	59	78
70%	61	83	80	85	66	82
80%	62	84	81	86	73	85
90%	64	86	81	87	80	89
100%	65	87	81	88	86	93

Ocean Odyssey case study.

[a] Percent improvement in PSF rating is the difference between the original PSF rating found in Table 6 and the action's optimal PSF rating found in Fig. 4. Example: 30% improvement in stress $= 50 + ((65 - 50) \times 0.3) = 55$.

Table 8
New PSF ratings based on RMMs and new SLI values based on new PSF ratings

Action	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors
New ratings						
1 (detect alarm)	55	78	80	81	32	66
Percent improvement over original PSF ratings in Table 6	30	30	60	20	20	30

Action	Stress	Complexity	Training	Experience	Event factors	Atmospheric factors	SLI Total
New SLI values							
1 (detect alarm)	7.7	6.2	15.2	14.6	7.0	19.1	69.8

Ocean Odyssey case study.

training and alarm redundancy in the form of lights and audio signals. Training can be improved by scheduling drills at infrequent intervals and enlisting feedback after training exercises. Experience can be somewhat improved through more realistic drills and a higher frequency of testing. Event factors can be mitigated through superior equipment and maintenance, thus improving the reliability and availability of safety systems. Atmospheric factors can be lessened through a reduction in the effect of severe weather by ensuring the ability of the safety systems to effectively annunciate the muster alarm to all personnel onboard in any location.

The re-rated PSFs are recorded in Table 8 along with the new PSF SLI values. Based on the re-rated PSFs, the total SLI for action 1 was improved from 62.3 to 69.8. A new log POS value was determined, and the probability of success was increased from 0.96 to 0.98. Consequently, the HEP was reduced by 50% to 0.02 from 0.04. The new HEP remained within the same probability range shown in Table 3 (i.e. B: 0.10–0.01), but the consequence was deemed to have been lowered from High (2) to Medium (2). (See dark-shaded box in Table 2.) This results in a 3B risk level, as shown in bold-face and dark shading in Table 3. The assigning of a reduced consequence because of the implementation of risk mitigation measures follows industry practice and requires experienced judgment. This process can be repeated for the remaining muster actions that apply to this muster scenario.

6. Concluding remarks

A human error probability index (HEPI) has been proposed that applies the estimated probability of failure for each muster action through a series of reference graphs. The HEPI methodology helps promote a consistent approach to the assessment of human factors in offshore platform musters. The index employs a consequence table and risk matrix to assess the ramifications of human failure for a range of muster severities. Risk mitigation measures are provided for each muster action, permitting a re-evaluation of risk.

The application of these HEP predictions in the form of an index (HEPI) brings forward a human reliability assessment (HRA) tool that is accessible and practical. It is hoped that HEPI will help to bring human error out of the post-accident blame context into a proactive risk reduction process. The ability to adequately prepare for severe muster scenarios is achievable through advancements in training, procedures, management systems and equipment. As no system can ever be free of human error, preparing to deal with this eventuality by making systems more error tolerant will help to lower risk.

Tools such as HEPI can help bring forward a common understanding of human error, the mechanisms that cause error, and the modes under which human failure occurs. The collection of human error data through drills and actual events can lead to the formation of a human error database. HEPI can be enhanced as more data are gathered, improving the quality of HEP predictions and risk mitigation measures.

6.1. Future scope of improvements

- It is recognized that the approach to risk re-ranking is subjective in nature and can possibly lead to differences in interpretation. Further empirical data are required for a more rigorous treatment of RMMs. Until these data are available, RMMs could be assessed through further elicitation of PSF ratings, qualified by the adoption of these measures.
- Other expert judgment techniques could be used to compare with the HEP estimates obtained from the current approach (SLIM).
- A probabilistic approach to determine success likelihood index values based on the product of the PSF normalized weight (*n*-weight) and rating distributions may add value. Monte-Carlo simulations could be used to determine the SLI values, such that the success likelihood index becomes the most probable value as opposed to a mean value.
- Use of fuzzy logic would help in better describing the PSF weight and rating boundaries and the SLI regions for the reference muster scenarios.
- Use of extreme (best and or worst) value instead of mean value approach.

Acknowledgements

The authors gratefully acknowledge the financial support of Petroleum Research Atlantic Canada (PRAC) and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Apostolakis, G.E., Bier, V.M., Mosleh, A., 1988. A Critique of Recent Models for Human Error Rate Assessments. Reliability Engineering and System Safety 22, 12–17.

Bea, R.G., 2001. Risk Assessment and Management of Offshore Structures. Progress in Structural Engineering and Materials 3, 180–187.

Bedford, T., Cooke, R., 2001. Probabilistic Risk Analysis: Foundations and Methods. Cambridge University Press, Cambridge, UK.

Chien, S.H., Dykes, A.A., Stetkar, J.W., Bley, D.C., 1988. Quantification of Human Error Rates Using a SLIM-Based Approach. In: Proceedings of the Institute of Electrical and Electronics Engineers 4th Conference on Human Factors and Power Plants, pp. 297–302.

DiMattia, D.G., 2004. Human Error Probability Index for Offshore Platform Musters. Ph.D. Thesis, Department of Chemical Engineering, Dalhousie University, Halifax, NS, Canada.

DiMattia, D.G., Khan, F.I., Amyotte, P.R., 2004. Determination of Human Error Probabilities for Offshore Platform Musters, Paper D-111, Presented at Bhopal and its Effects on Process Safety, International Conference on the 20th Anniversary of the Bhopal Gas Tragedy, December 1–3, 2004, Kanpur, India.

DiMattia, D.G., Khan, F.I., Amyotte, P.R., 2005. Determination of Human Error Probabilities for Offshore Platform Musters. Journal of Loss Prevention in the Process Industries 18, 488–501.

Dougherty, E.M., Fragola, J.R., 1988. Foundations for a Time Reliability Correlation System to Quantify Human Reliability. In: Proceedings of the Institute of Electrical and Electronics Engineers 4th Conference on Human Factors in Power Plants, pp. 268–278.

Embrey, D.E., Humphreys, P.C., Rosa, E.A., Kirwan, B., Rea, K., 1984. SLIM–MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment. Report No. NUREG/CR-3518 (BNL-NUREG-51716), Department of Nuclear Energy, Brookhaven National Laboratory, Upton, NY.

Embrey, D.E., Kontogiannis, T., Green, M., 1994. Guidelines for Preventing Human Error in Process Safety. Center for Chemical Process Safety, American Institute of Chemical Engineers, New York.

Gordon, R., Rhona, F., Mearns, K., 2001. Collecting Human-Factors Data from Accidents and Incidents. Society of Petroleum Engineers Production and Facilities, 73–83.

Johnson, R., Hughes, G., 2002. Evaluation Report on OTO 1999/092, Human Factors Assessment of Safety Critical Tasks, Report No. 33, Health and Safety Executive, UK.

Johnson, T.R., Budescu, D.V., Wallsten, T.S., 2001. Averaging Probability Judgments: Monte Carlo Analyses of Asymptotic Diagnostic Value. Journal of Behavioral Decision Making 14, 123–140.

Kennedy, B., 1993. A Human Factors Analysis of Evacuation, Escape and Rescue From Offshore Operations, Report No. OTO 93 004, Health and Safety Executive, UK.

Khan, F.I., Amyotte, P.R., 2003. How to Make Inherent Safety Practice a Reality. Canadian Journal of Chemical Engineering 81, 2–16.

Kim, J.A., Jung, W., 2003. A Taxonomy of Performance Influencing Factors for Human Reliability Analysis of Emergency Tasks. Journal of Loss Prevention in the Process Industries 16, 479–495.

Kirwan, B., 1994. A Guide to Practical Human Reliability Assessment. Taylor and Francis, London.

Kirwan, B., 1998. Human Error Identification Techniques for Risk Assessment of High Risk Systems—Part 1: Review and Evaluation of Techniques. Applied Ergonomics 29, 157–177.

Kirwan, B., James, N., 1989. The Development of a Human Reliability Assessment System for the Management of Human Error in Complex Systems. Reliability 5A (2), 1–11.

Kirwan, B., Embrey, D.E., Rea, K., 1988. Human Reliability Assessors Guide, Report No. RTS 88/95Q, Safety and Reliability Directorate, Warrington, UK.

Miller, D.P., Swain, A.D., 1987. Human Error and Human Reliability. In: Salvendy, G. (Ed.), Handbook of Human Factors. Wiley, New York, pp. 219–250.

Robertson, D.H., Wright, M.J., 1997. Ocean Odyssey Emergency Evacuation, Analysis of Survivor Experiences, Offshore Technology Report—OTO 96009, Health and Safety Executive, Suffolk, UK.

Spurgin, A.J., Lydell, B.O.Y., 2002. Critique of Current Human Reliability Analysis Methods. In: Proceedings of the Institute of Electrical and Electronics Engineers 7th Conference on Human Factors and Power Plants, pp. 3-12–3-18.

Widdowson, A., Carr, D., 2002. Human Factors Integration: Implementation in the Onshore and Offshore Industries. HSE Books, Sudbury, UK.

Williams, J.C., 1983. Validation of Human Reliability Assessment Techniques. Fourth National Reliability Conference 2 (2), 1–9.

Zamanali, J.H., Hubbard, R.R., Mosleh, A., Waller, M.A., 1998. Evolutionary Enhancement of the SLIM–MAUD Method of Estimating Human Error Rates. Transactions of the Institute of Electrical and Electronics Engineers, 508–514.