

Development of an Online Checklist for the Assessment of Alarm Systems and Alarm Management in Process Control

Martina Bockelmann^(✉), Peter Nickel, and Friedhelm Nachreiner

Gesellschaft für Arbeits-, Wirtschafts- und Organisationspsychologische
Forschung e.V. (GAWO), Oldenburg, Germany
martina.bockelmann@gawo-ev.de

Abstract. In the past, poorly designed alarm systems and inadequate alarm management contributed to the emergence of critical events with partially serious consequences. Based on the experience gained from these incidents, guidelines were created with recommendations for the design of alarm systems and alarm management. A comprehensive checklist has been developed to analyze the current design quality of alarm systems and alarm management and has been used in various control rooms across different branches of industry in Germany. Using the checklist, design deficiencies can be identified and action needs can be derived. So far, the results also show systematic differences in the application of the checklist between individual assessors and between groups of assessors.

Keywords: Alarm system · Alarm management · Ergonomic design · Checklist · Process control · Industrial safety

1 Introduction

Investigation reports from the 1970s until today inform about critical events, with some disclosing serious consequences for employees, companies, the environment and the public. Details convey that poorly designed alarm systems and alarm management jointly contributed to incidents such as at Three Mile Island nuclear power plant (1979), in the oil refineries of Texaco (1994) and BP (2005) and on the oil rig Deepwater Horizon (2010). As causal factors, reports present issues like non-response of alarms, high alarm rates, poor prioritization of alarms, non-ergonomic and inappropriate design of displays and controls and the lack of systematic training of the control room operators in dealing with critical situations, to name but a few [e.g. 1–3].

Parallel to and as a consequence of the events, several guidelines and standards have been evolved presenting requirements for an ergonomic design of alarm systems and alarm management [e.g. 4–8]. In addition, the events triggered research into safety and human factors and also lessons learnt from experience in the field (best practice) resulting in publications with available findings and recommendations [9–11].

Well-designed alarm systems and adequate alarm management, which fulfil ergonomic requirements, should therefore have an impact on system availability, system

reliability and operational safety – especially in the case of critical process trends and process conditions.

However, it is unclear to what extent these design requirements and recommendations have been applied in operational practice. Therefore, a research project has been conducted to address the following questions:

1. How can the ergonomic design quality of alarm systems and alarm management be easily, consistently and reliably assessed?
2. What is the current quality of the design in control rooms in process industries? [not covered by this paper]
3. What are important ergonomic recommendations to further improve existing and future alarm systems and alarm management? [not covered by this paper]

For this purpose, a checklist has been developed in order

1. to analyze and evaluate the design quality of alarm systems and alarm management in different control rooms and within various sectors of industry and
2. to derive hints for potential improvements or needs for action to implement design requirements and recommendations where appropriate.

2 Methods

2.1 Checklist Development

The design of the computerized checklist was based on the results of a feasibility study carried out in 2008/2009 in six control rooms from three chemical companies [12, 13].

For the present study, an extended knowledge base has been further established based on a systematic review of individual design requirements from relevant ergonomics literature as well as relevant guidelines, normative provisions and specifications [e.g. 2, 5–8, 14, 15]. Potentially relevant design requirements were collated, summarized and structured into thematic areas, such as design of alarm systems (e.g. prioritization), design of operator requirements (e.g. operator performance limits) and design of alarm management (e.g. performance monitoring and improvements).

A complete evaluation of all potentially relevant characteristics for an alarm system was not possible for technical, temporal and financial reasons; an operationalization of all existing requirements is further impossible due to the singularities of each control system and the process under control. Therefore, a sample of prominent and substantial characteristics from the knowledge base was selected by expert reviewers in a multi-staged process and transferred into easily usable items. All items were transferred into questions and were supplemented by examples and notes. The answer categories in the checklist were either pass/fail decisions (yes/no) or decisions for traffic light categories ('Green' = good design, requirement fulfilled, 'Yellow' = design basically OK, but better solutions would be conceivable, 'Red' = unsatisfactory design solution calling for improvement).

The computerized checklist was implemented as (1) an offline version using a portable computer and (2) an online version provided by a browser via internet.

2.2 Checklist Suitability

The usability of the checklist was tested in a multi-step procedure. A draft version of the checklist was presented to experts in ergonomics and human factors. The review resulted in a final draft version with several amendments (e.g. new and deleted items, rephrasing of questions and examples).

This final draft of the checklist was tested for usability in operational use by a senior member of staff from a chemical company and by two human factors and ergonomics experts (HF/E experts). The senior staff member was asked to go through the checklist, read each checklist characteristic carefully and write a comment in the event of ambiguities, misleading wording, doubling etc. The HF/E experts were required to apply the checklist in a typical control room within the chemical industry under realistic investigation conditions and comment about structure, content and checklist design.

Comments and suggestions for improvement by both parties were subsequently discussed by the above mentioned expert group. Final adjustments (e.g. reduced ambiguity of questions through rephrasing, refining and supplementing examples) resulted in the final version for the presented study.

The final version of the checklist contains 148 items, assigned to the following design areas:

1. Alarm generation/alerting
2. Alarm presentation
3. Alarm prioritization
4. Alarm system functionalities and technical measures
5. Operator performance limits
6. Action guidelines and system interactions
7. Control and feedback
8. Alarm culture and alarm philosophy
9. Continuous improvement
10. Documentation
11. Training

2.3 Checklist Application

Application of the checklist in control rooms required several methods of data collection, such as observation, visual inspection, interviews with control room operators and supervisors, physical measurements and document analyses.

In total, alarm systems have been investigated at 15 workplaces in different control rooms in 12 companies from three industrial sectors, i.e. electrical power generation and distribution, food industry and chemical industry.

Each alarm system was evaluated by two HF/E experts and – where possible – by two experienced practitioners (e.g. technicians, system engineers, safety experts etc.) from participating companies independently on different days, thus providing for different kinds of expertise. This allows tests of different aspects of the usability of the checklist, e.g. by users with different educational background and know-how, as well

as a verification of rater effects, e.g. status effects (HF/E experts vs. experienced practitioners) and effects of individual raters within these groups.

An alarm system investigation carried out by HF/E experts lasted between 7 and 10 hours and varied according to the complexity of the process under control, the process control system under investigation, the type and extent of the alarm management activities and the events specific to the days of assessment. A shift change was intentionally included, if possible, in order to be able to observe different operators in interaction with the alarm system and by doing so, to reduce operator-specific variance. No information is available about the time required for investigations carried out by the engineering staff of participating companies.

2.4 Statistical Analysis

After a first descriptive analysis of the data, observer agreement was determined in order to find out whether the raters had assessed the items of the checklist identically or at least similarly in the assessment of the systems. Cohen's kappa (κ) and, in addition, weighted kappa (κ_w , only for polytomous system of answer categories, traffic light categories, $n = 123$) were used as (rough) indices of observer agreement.

The advantage of weighted kappa is that the extent of disagreement in non-identical judgements is taken into account in the calculation of the index [16, 17]. In the present investigation, this means that differing judgements in the form of "good design" (= 'Green') and "unsatisfactory design solution" (= 'Red') would achieve a higher weight than a deviation between "good design" (= 'Green') and "design basically OK" (= 'Yellow').

If two practitioners – in addition to two HF/E experts – are also available for the assessment of each alarm system, six kappa coefficients can be calculated for each system. For 15 investigated systems, it results in a maximum of 90 indexes.

The classification of Landis and Koch [18] was used to classify the kappa coefficients (see Tables 1 and 2).

Kappa or weighted kappa can, however, only be considered as a first, global measure of observer (dis)agreement, since no conclusions can be made as to the possible causes of the variability or disagreement in the ratings [19]. On the other hand, it provides some first impressions on the level of agreement in using the checklist and thus, its usability for the purpose intended. Since not all data have been collected in the present study, some considerations will be given about potential reasons for relatively higher or lower levels of agreement.

3 Results

At present, 30 assessments done by the two HF/E experts and 21 assessments done by experienced practitioners are available, i.e. nine ratings by practitioners are still missing.

Based on the data available at the time of submitting this report, the results obviously identify differences in design quality of alarm systems and alarm management as well

as requirements for design improvement (e.g. in the area of design of human-machine-interface, prioritization of alarms, alarm management and concerning operator training in alarm systems and the handling of alarms).

Table 1. Distribution of kappa coefficients (classification according to Landis & Koch, 1977 [18])

	“slight” (0,00–0,20)	“fair” (0,21–0,40)	“moderate” (0,41–0,60)	“substantial” (0,61–0,80)	“(almost) perfect” (0,81–1,00)
HF/E expert – HF/E expert			5	10	
HF/E expert – Practitioner	3	23	16		
Practitioner – Practitioner		4	4		1

(modes in bold)

In general, the HF/E experts tended to rate the systems more strictly than the practitioners from the companies, as shown in the example in Fig. 1. In this example, the assessments of the HF/E experts agree quite well. In comparison to the HF/E experts, the experienced practitioners judged much milder – especially assessor EP1; i.e. they classified design aspects more often as “good design”. Moreover, the assessments of the practitioners also differ from each other.

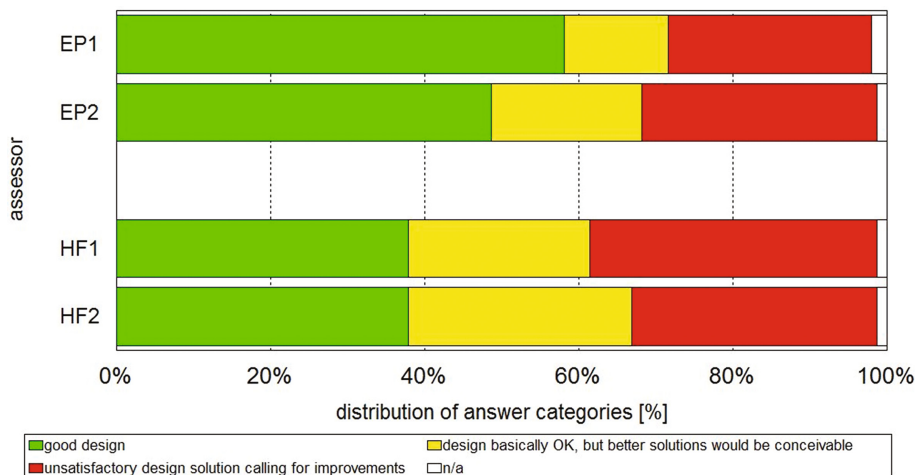


Fig. 1. Relative frequencies of answer categories per assessor at workplace 1 (Color figure online)

Regarding the inter-rater agreement, 66 out of a maximum of 90 possible kappa coefficients have been calculated so far, since the system assessments by nine practitioners are not yet available.

The kappa coefficients within the HF/E experts group range from moderate to good agreement, which is higher than in the other two groups (see Table 1). In particular, the

kappa coefficients between HF/E experts and experienced practitioners show a wide range (slight to moderate) and are generally rather low. The values within the practioners group are between fair to moderate, with one exception (almost perfect). However, here some doubts can be raised, since both raters shared a room, filled in and finished their checklists at nearly the same time and wrote similar comments, so that they might have discussed their assessments.

The weighted kappa coefficients show a somewhat higher level (see Table 2). With one exception, the inter-rater agreements between the HF/E experts are substantial to almost perfect. The inter-rater agreements of the comparisons between HF/E experts and practitioners are significantly lower. The spectrum ranges from fair to substantial agreement. The weighted kappa coefficients between experienced practitioners are widely distributed, ranging from fair to almost perfect.

Table 2. Distribution of weighted kappa coefficients (classification according to Landis & Koch, 1977 [18])

	“slight” (0,00–0,20)	“fair” (0,21–0,40)	“moderate” (0,41–0,60)	“substantial” (0,61–0,80)	“(almost) perfect” (0,81–1,00)
HF/E expert – HF/E expert			1	11	3
HF/E expert – Practitioner		10	24	8	
Practitioner – Practitioner		1	4	3	1

(modes in bold)

Higher weighted kappa values, as compared to kappa, suggest that deviations in the form of one scale unit (e.g. “good design” and “design basically OK”) are more frequent than discrepancies in the form of two scale units (“good design” and “unsatisfactory design solution”).

4 Discussions

The calculation of kappa coefficients (κ/κ_w) already indicate some patterns of agreement and disagreement. The results of these analyses obviously indicate some differences, and most probably systematic differences between individual assessors and types of assessors (HF/E experts vs. experienced practitioners) in using the checklist and its assessment criteria. There can be several different reasons for such a result.

For example, the concepts used in the checklist were not clear to the practitioners (the HF/E experts were engaged in the description of the concepts, so they should be aware of their content, which is supported by their higher agreement), which would be supported by the disagreement between the groups of raters. This would have to be addressed by a reformulation of the concepts, basic training in using the checklist and the concepts behind the items or additional information material (e.g. examples) to explain the item content.

A further reason for deviation in evaluation could be the fact that the raters observed different situations leading to differences between the investigation objects. It was not

unusual that the practitioners carried out or finished their assessments weeks or months later than the HF/E experts for company reasons. Meanwhile, the design quality of the alarm system and alarm management could have changed.

On the other hand, there could be difficulties with individual items and their scaling. This cannot be analyzed satisfactorily by using kappa coefficients. However, Generalizability Theory [G-theory; 20, 21] can provide the relevant information about the (absolute and relative) contribution of several systematic error terms (rater, group of raters, item, control system under inspection and their interactions) to the error of measurements. This can be done by performing an analysis of variance and estimating the size and proportion of all variance components indicating error of measurement (i.e. the systematic error components plus random error) simultaneously. Such a statistical analysis is in progress, but can only be finished after the data collection and processing have been completed. The assessment of the psychometric properties [according to ISO 10075-3; 22] of the checklist using a G-theoretical approach is still in progress.

With a view to the severe consequences which may be associated with poorly designed alarm systems and inadequate alarm management, the importance of well-designed and well-managed systems becomes particularly apparent. Therefore, a systematic and continuous alarm system and alarm management analysis is an important element in the safety concept of a company to ensure the functionality of an alarm system and the requirements for continuously monitoring, maintaining or improving its performance and, finally, to keep a plant in a safe state [e.g. 11, 14].

For this purpose, an objective, reliable, valid, sensitive, diagnostic and easy-to-use instrument would be desirable, with which the design state of alarm systems, including the alarm management, can be evaluated in order to identify potential design deficiencies and to implement appropriate work-design measures if necessary.

Acknowledgements. The research for the study was partly funded by the Forschungsgesellschaft für angewandte Systemsicherheit und Arbeitsmedizin e.V. (FSA) [<http://www.fsa.de/en/home/>].

References

1. Kemeny, J.G., Babbitt, B., Haggerty, P.E., et al.: Report of The President's Commission on the Accident at Three Mile Island (1979). <http://www.threemileisland.org/downloads/188.pdf>. 20 December 2016
2. Health & Safety Executive (HSE): HSE Information Sheet, Chemicals Sheet No. 6. Better alarm handling (2000). <http://www.hse.gov.uk/pubns/chis6.pdf>. 20 December 2016
3. U.S. Chemical Safety and Hazard Investigation Board (CSB): Investigation Report: Refinery Explosion and Fire (15 Killed, 180 Injured). BP, Texas City, Texas; March 23, 2005. Report No. 2005-04-I-TX, March 2007 (2007). <http://www.csb.gov/bp-america-refinery-explosion>. 20 December 2016
4. ANSI/ISA 18.2: Management of Alarm Systems for the Process Industries. ISA, Research Triangle Park (2009)
5. DIN EN 62682: Alarmmanagement in der Prozessindustrie. Beuth, Berlin (2013, 2016)
6. EEMUA 191: Alarm systems. A guide to design, management and procurement. EEMUA, London (2013)
7. NA 102: Alarmmanagement. NAMUR, Leverkusen (2008)

8. YA-711: Principles for alarm system design. Norwegian Petroleum Directorate, Stavanger (2001)
9. Bransby, M.L., Jenkinson, J.: The Management of Alarm Systems. HSE Books, Sudbury (1998)
10. Ivergård, T., Hunt, B. (eds.): Handbook of Control Room Design and Ergonomics: A Perspective for the Future. CRC Press, Boca Raton (2009)
11. Hollifield, B.R., Habibi, E.: Das Alarm-Management Handbuch. Ein umfassender Ratgeber. Nino Druck GmbH, Neustadt (2012)
12. Bockelmann, M.: Entwicklung und Überprüfung eines Prototyps eines Instrumentes zur Beurteilung und Optimierung des Gestaltungszustandes von Alarmsystemen – eine Machbarkeitsstudie (unpublished diploma thesis). Carl von Ossietzky Universität, Oldenburg (2009)
13. Bockelmann, M., Schütte, M., Nachreiner, F.: Entwicklung und Überprüfung eines Prototyps eines Instrumentes zur Beurteilung und Optimierung des Gestaltungszustandes von Alarmsystemen – Ergebnisse einer Machbarkeitsstudie. In: Gesellschaft für Arbeitswissenschaft (ed.) Neue Arbeits- und Lebenswelten gestalten, pp. 905–908. GfA-Press, Dortmund (2010)
14. Health & Safety Executive (HSE). Human Factors Briefing Note No. 9. Alarm Handling (n.d.). <http://www.hse.gov.uk/humanfactors/topics/09alarms.pdf>. 20 December 2016
15. VDI/VDE 3699-5: Prozessführung mit Bildschirmen – Alarmer/Meldungen. Beuth, Berlin (2014)
16. Bortz, J., Lienert, G.A., Boehnke, K.: Verteilungsfreie Methoden in der Biostatistik. Springer, Berlin (2000)
17. Bortz, J.: Statistik für Human- und Sozialwissenschaftler. Springer Medizin, Heidelberg (2005)
18. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977)
19. Schütte, M.: Reliabilitätsbestimmung arbeitsanalytischer Verfahren. In: Nickel, P., Hänecke, K., Schütte, M., Grzech-Šukalo, H. (eds.) Aspekte der Arbeitspsychologie in Wissenschaft und Praxis, pp. 79–102. Pabst, Lengerich (2004)
20. Cronbach, L.J., Gleser, G.C., Nanda, H., Rajaratnam, N.: The Dependability of Behavioral Measurements. Theory of Generalizability for Scores and Profiles. Wiley, New York (1972)
21. Brennan, R.L.: Generalizability Theory. Springer, New York (2001)
22. ISO 10075-3: Ergonomic principles related to mental workload – Part 3: Principles and requirements concerning methods for measuring and assessing mental workload. ISO, Geneva (2004)