

**The Efficacy of Peer Checking as an Error Capture and Recovery Measure**

Regulatory Research Summary Paper

ONR-RRR-096

#### June 2023

## The Efficacy of Peer Checking as an Error Capture and Recovery Measure

#### Regulatory Research Summary Paper

ONR-RRR-096

#### CM9 2023/12844

Contents

[The Efficacy of Peer Checking as an Error Capture and Recovery Measure 2](#_bookmark0)

[Introduction 4](#_bookmark1)

[Purpose of the Research 4](#_bookmark2)

[Peer Checking in the Nuclear Context 4](#_bookmark3)

[Review 5](#_bookmark4)

[Research Questions 5](#_bookmark5)

[Method 5](#_bookmark6)

[Results and Discussion 6](#_bookmark7)

1. [To what extent is peer checking an effective capture/prevention tool? 6](#_bookmark8)
2. [Which psychological phenomena/theories help to explain the effectiveness (or](#_bookmark9) [otherwise) of peer checking? 8](#_bookmark9)
3. [What performance influencing factors positively or negatively impact the](#_bookmark10) [effectiveness of peer checking? 9](#_bookmark10)
4. [What are the characteristics of a high reliability checking system? 10](#_bookmark11)

[Limitations & Future Research 11](#_bookmark12)

[Conclusion 12](#_bookmark13)

[Further Information 13](#_bookmark14)

# Introduction

### Purpose of the Research

The Office for Nuclear Regulation (ONR) commissioned research to consider the efficacy of peer checking as an error capture and recovery measure in high-hazard industries. This work was prompted by the publication of papers in the medical sector that have questioned the value of peer checking as a means to reduce errors and improve reliability. Given the widespread use of peer checking within the nuclear industry, it is important that ONR has a scientifically informed and evidence-based position on its effectiveness.

A literature review was conducted to examine the available evidence across industries, to determine the relevance to the nuclear sector and to take learning from this. Supplementary insights were gathered through industry engagement sessions with operations professionals from a range of high hazard industries, as well as a group of Human Factors (HF) specialists. The review has been used to establish whether there are grounds for concern in relation to existing nuclear practice, and to identify any future research needs.

The use of peer checking is not limited to the nuclear sector. Checking tasks are common practice across many high hazard industries. For example, in healthcare, double checks are routinely applied in the administration of medicines, and in the aviation industry, peer checks are used by flight and maintenance crews to conduct pre-take off checks.

The prevalence of peer checking across high hazard industries is based on the commonly held assumption that peer checks serve to identify and reduce the potential for error.

However, some researchers have questioned the tangible evidence base for this practice:

*“There is insufficient evidence to either support or refute the practice of double checking the administration of medicines.” – Alsulami, Conroy & Choonara (2012); British Medical Journal, Quality and Safety.*

This paper summarises the main findings of this initial review, and draws tentative conclusions where appropriate, taking cognisance of the extent and quality of the available evidence. The paper also sets out ONR’s regulatory position in relation to the use of peer checking at nuclear licensed sites, and outlines avenues for further research for duty holders to consider.

### Peer Checking in the Nuclear Context

Peer checks are routinely deployed in nuclear facilities to identify and recover from errors that may arise during safety important tasks. These are often formally claimed and designated as operational safety measures in safety cases, including as fault prevention, fault recovery, or consequence mitigation measures.

Where safety case arguments claim peer checks as a means to improve reliability, relevant good practice requires proportionate substantiation of the claim. This is important to provide the necessary qualitative evidence to support claims on human performance, and should be

based on assessment of the factors that influence human reliability, in the context of the task.

Other applications in nuclear include maintenance tasks, whereby checks are used to ensure engineered safety measures are safe to work on, and are correctly returned to service. Quality management systems also utilise checks (for example, product quality checks), which, if inadequately performed, can have significant safety impacts and/or reputational consequences.

Checking tasks can take many forms, including self-checks, peer-to-peer checks, and independent checks. Checks vary in the level of independence and diversity they provide, based on factors such as the person, time, location and parameter that feature in a task’s design. It is commonly assumed that the higher the independence of a check, the more likely it is that errors will be correctly identified and recovered, which in turn, increases the reliability of the tasks or action being checked.

# Review

### Research Questions

The review and analyses were structured around the following research questions:

1. To what extent is peer checking an effective error capture/prevention tool?
2. Which psychological phenomena/theories help to explain the effectiveness (or otherwise) of peer checking?
3. What are the performance influencing factors that positively or negatively impact the effectiveness of peer checking?
4. What are the characteristics of a high-reliability peer checking system?

### Method

##### Literature Review

The research was predominantly based on the literature review. A literature search for papers relating to the efficacy of peer checking was conducted across multiple high hazard industries, using specified search terms1. A total of 40 papers were selected for detailed review, based on a set of selection criteria2.

The papers were analysed to determine the extent and quality of the available evidence, and to identify and capture pertinent information relating to the research questions. The methodology and key research findings for each paper were summarised in a literature

1 Search terms included: peer check, cross check, double check, independent verification.

2 Selection criteria included: number of citations, date of publication, credibility of the research papers, relevance to research questions.

review table. This was used as a basis for identifying the key themes and learning points arising from the literature findings, in addition to limitations of the extant literature, and avenues for future research.

Peer checking proved a challenging concept to research, due to many different conceptualisations and interpretations of the term being used in the literature. No dominant academic taxonomy for types of checking was identified through the review, with peer checking used as an umbrella term for a wide range of checking tasks with different characteristics and task design. Self-checks were found to be commonly examined in the literature, and therefore considered worthy of inclusion in the review. Consequently, the following taxonomy was used to structure the analysis:

* **Self-check:** Check performed by the same person who undertook the task
* **Peer-to-peer check:** Check performed by a peer (e.g. another person with similar training and background, often in the same team or organisational structure)
* **Independent check:** Check performed by an individual independent of the task (e.g. a supervisor).

##### Industry Engagement Sessions

To supplement the evidence from the literature review, additional insights were collected from a series of industry engagement sessions.

Seven industry operations professionals, who regularly use peer checking as part of their daily tasks, were interviewed. These individuals came from the healthcare, oil and gas, defence, and nuclear sectors. A focus group of HF specialists working across multiple industries was also held, to gather HF Subject Matter Expert (SME) insights on the topic. Taken together, these sessions provided supplementary insights derived from practical experience of designing and applying peer checking systems.

# Results and Discussion

### To what extent is peer checking an effective capture/prevention tool?

Unfortunately, whilst there were numerous, valuable insights to be gained from the literature, there were few structured and systematic studies that compared different types of check and their efficacy for error identification and recovery. As such, it is difficult to draw conclusions with regard to their relative effectiveness.

The majority of papers reviewed originated from the healthcare sector (82.5%). The nuclear sector accounted for 10% of the papers, with the remaining 7.5% originating from aviation, rail, oil and gas, or manufacturing.

Based on the evidence across all sectors, there was limited conclusive support for the

efficacy (or otherwise) of peer checking. Of the literature reviewed, 12.5% provided supportive evidence, 65% provided mixed, and 22.5% provided counter evidence. This is a notable finding considering the high prevalence and wide acceptance of peer checking as a valid error identification and recovery measure.

In considering these findings, it is important to note that the data is based predominantly on the healthcare context, with limited breadth in cross-sector evidence. This limits the cross- sector conclusions that can be drawn, and the direct applicability of the findings to the nuclear industry. Whilst similarities are undoubtedly present, ONR notes that the context in which checks are deployed, and the task design of checks, can differ significantly across industries. Authors rarely specified the precise characteristics of the checking task being examined in their study/review (beyond the broad check type), which limited the opportunity for similarities between medical and nuclear checking tasks to be identified.

The available literature predominantly related to peer-to-peer checking in healthcare. Confusingly, these were often referred to as independent checks, however they were typically performed by peers, similar in training, background and organisational status, with close proximity to the task. Some papers referenced self-checking, and a very limited number referred to independent checking (as conceptualised in nuclear applications).

Whilst peer-to-peer checking and self-checking are used in the nuclear industry, nuclear checking tasks are commonly characterised by a higher proportion of ‘Independent Checks’, for example, those carried out by a Duly Authorised Person (DAP) or task supervisor. This is particularly true for checks in relation to safety important tasks and administrative controls, where higher levels of reliability are sought.

Limitations were also identified in the methodological rigour of the papers reviewed. For example, many papers utilised self-report or solely qualitative data to inform the conclusions drawn. Whilst these methods have their value, few studies applied an experimental design that objectively tested the efficacy of checking tasks under controlled conditions. Consequently, there is currently a small, empirically derived quantitative evidence base to supplement the qualitative and self-report findings. This makes generalising the findings difficult, and hence further, structured studies that control and test specific variables (e.g. independence factors such as person, time, location, parameter) are recommended.

The industry engagement sessions identified that most participants perceived that peer checking can be an effective measure for error identification and recovery. However, this was caveated by the view that the check must be proportionately applied, well designed, taking account of the context-dependent performance influencing factors that may affect the reliability of the check, and should ideally not be used as a single line of defence.

Participants recognised how much checking is relied upon and ingrained in safety systems, and the reassurance that this measure can provide (albeit sometimes potentially false assurance). It is possible that variation in check design quality may be one of the factors contributing to the mixed literature findings on the efficacy of checking.

### Which psychological phenomena/theories help to explain the effectiveness (or otherwise) of peer checking?

The literature review and industry engagement sessions provided insights into psychological phenomena and theory that are potentially relevant to peer checking efficacy. It is noted that most papers and SME discussions did not make explicit reference to psychological theory. For the purposes of this review, the reviewers inferred the most relevant theory using expert judgement, based on the factors discussed in the papers. The methodological limitation of this approach is noted, recognising the potential for confirmation bias, as psychological theory more familiar to the reviewer’s may have been more readily identified.

Similar psychological theories were distilled through both the literature review and SME engagement sessions. The dominant theories and phenomena identified through the review were as follows (in order of prevalence):

* 1. **Diffusion of responsibility** - Where no single individual takes full responsibility for a task, as the responsibility is shared between two or more individuals, resulting in a decreased feeling of responsibility.

In the context of checking tasks, this may result in a task being performed less rigorously, in the knowledge that performance will be checked by another individual. Checks may also be performed less rigorously, due to an over-confidence that the task has been suitably performed.

* 1. **Habituation –** The diminishing of a response to a repeated stimuli.

In the context of checking tasks, there is a risk that the individual becomes overly accustomed to the checking process itself, but less attuned to identifying discrepancies, meaning recovery action is less likely to be instigated.

* 1. **Deference to authority –** The act of submitting to the judgement of those in more senior positions of influence to one’s own.

In the context of checking tasks, checks performed by subordinates may be less effective than checks performed by more senior or experienced members of staff, due to the assumed reliability of the task performer’s actions.

* 1. **Confirmation bias –** The tendency to utilise information to support or reinforce one’s pre-existing ideas or beliefs.

In the context of checking tasks, the anticipated outcome of a check may override the salience of the actual outcome, particularly if the check often yields the same result.

Different psychological phenomena will have greater/less effect depending on the task characteristics. For example, habituation may have a stronger influence for a routine, repetitive checking task, in which parameter’s rarely change. Similarly, diffusion of responsibility may have a stronger influence when a task is checked by multiple individuals.

### What performance influencing factors positively or negatively impact the effectiveness of peer checking?

The literature review identified multiple examples of factors that affect checking efficacy. These were broadly representative of the person, task, equipment and organisational factors that are generally considered to influence human performance. Consequently, these are arguably not unique to checking tasks.

Whilst noting this, the most frequently cited factors were:

* 1. Workload
  2. Competency
  3. Divided attention.

Workload was also the most frequently cited factor in the industry engagement sessions. This indicates a strong, general recognition of the negative impact that high workload can have on checking task performance. The workload implications associated with the checking task itself were also noted, and recognised in a number of instances as an important contributory factor.

The salience of ‘workload’, and the associated factor of ‘divided attention’, may be particularly strong in the healthcare sector, which faces chronic resource challenges. As such, examining these factors may be a strong interest and concern for healthcare researchers. For example, peer-to-peer checking in a healthcare setting typically results in a checker being pulled away from their own task to check another. This practice can increase workload, which if mismanaged, may introduce distraction and interruptions, which can affect both the checker’s and task performer’s reliability.

It is noted that other standard HF aspects, such as the design of human machine interfaces (and other equipment), and the consideration of environmental factors, received comparably less recognition in the literature (albeit these factors received some mention in the industry focus groups). This may be symptomatic of the commonly used self- report/qualitative study design in the studies reviewed, as factors that require a greater degree of HF subject matter expertise to understand and identify may not have been so readily given as responses using these methods.

A particularly notable factor not identified or discussed as salient in the review was the level of independence of a check. Dependency was not directly examined, nor explicitly identified as a strong factor in the majority of literature reviewed. This is noteworthy, given the importance accorded to ‘independence’ in the nuclear industry. Consequently, based on the present review, it was not possible to confirm or refute whether the higher the independence of a check, the more likely it is that errors will be correctly identified and recovered. This was therefore identified as an important avenue for future research.

### What are the characteristics of a high reliability checking system?

As part of the research, the review sought to identify the characteristics of a high reliability checking system. However, noting the relative immaturity of the academic literature in this area, ONR does not consider the literature findings to date to be comprehensive or sufficiently robust to form the basis of guidance in this area. Consequently, the following section takes cognisance of the review findings, but is supplemented with HF principles that are grounded in extensive practical application, operational experience and learning from events. These principles also mirror the insights gathered through the industry engagement sessions.

The reviewers also found that the standards for designing and implementing high reliability checking systems are not clearly defined or consistently available within the nuclear industry, which indicates a need for additional work in this area. In the interim, ONR notes the following principles, which it considers are central to defining, implementing and maintaining high quality checking systems.

###### Apply a Graded Approach

A graded approach to applying checks should be employed, whereby the use of checks and their requirements are proportionate to the risk, and due consideration is given to the task characteristics (e.g. novelty, complexity) and the context in which the task is performed. For example, independent checks should be reserved for safety important tasks, and not implemented indiscriminately. This is because the overuse of checks can reduce their perceived safety significance, as well as increasing workload, which in turn can reduce checking rigour.

###### Use Task Analysis to Inform Check Design

Where a check is claimed as part of a safety argument, ONR considers the check to be a safety important task, and as such should be designed and substantiated in the usual way. Consequently, the factors that influence efficacy will be largely dependent upon the task context, with individual factors more (or less) influential in different scenarios.

As with any safety important human action, the performance influencing factors should be assessed through proportionate task analysis, which should in turn be used to inform a check’s design. Understanding the potential psychological phenomena at play can help to manage the effects of negative influencing factors, and harness the positive factors.

Checking tasks should be considered in the context in which they are applied, with the supporting analysis considering the impact of concurrent tasks and/or supervision responsibilities, and the associated workload implications of this.

###### Periodically Review & Monitor Implementation

Maintaining discipline in how the full intent of a check is implemented is important. Whilst a high-quality check may be initially designed and developed, the standard with which it is implemented can erode over time. Consequently, monitoring is important to ensure that checks continue to deliver their intended purpose and any related safety functional requirements. Useful ways in which the ongoing effectiveness of checks can be confirmed include:

* Compliance checks
* Conducting periodic reviews to ensure checks remain effective
* Applying learning from end user engagement, events/near misses/OPEX to improve check design.

# Limitations & Future Research

The review provided some valuable insights into the factors that affect the reliability of checking tasks. However, it is evident that the literature base is immature at present, which limits the confidence that can be drawn from the findings.

The following limitations were identified:

* **Cross sector evidence** -The available evidence predominantly originates from the medical sector with only limited cross sector evidence. Whilst parallels are undoubtedly present across industries, checking tasks can differ greatly in design and application across industries, which reduces the generalisability of the findings.
* **Quantitative and/or experimental studies** - The majority of papers utilised qualitative and/or self-report data rather than quantitative and/ or experimental approaches. Consequently, limited findings derived from objective data were available for review.

**Reference to psychological theory** - Studies were rarely grounded in psychological theory, which reduces the confidence that can be taken in the robustness and rigour of the literature reviewed.

* **Investigation of Independent Verification (IV)** - Independent checking is arguably the most prevalent check type for safety critical tasks in the nuclear sector. However very few papers reviewed investigated independent supervisory checks. The dominant check types in the literature (i.e. peer-to-peer and self-checking) are less commonly applied in safety critical nuclear operations, which reduces the direct applicability of the findings.
* **Reference to factors widely accepted in HF Relevant Good Practice (RGP)** - Factors such as independence and diversity received relatively little attention in the literature. Whilst these were sometimes indirectly referred to, they were rarely the central focus of studies.

The following avenues for future research were identified:

* Further studies to define different check types, what characterises them, and to test their relative efficacy
* Further studies that apply quantitative methods and/or experimental design, to empirically test influencing factors in controlled conditions
* Studies that empirically test the relative importance of the difference aspects of independence (e.g. person, time, location, parameter)
* Further studies in other high hazard industries (including nuclear), to better facilitate cross-industry comparisons.

# Conclusion

In conclusion, the literature review found limited evidence to provide strong support (or otherwise) for peer checking as an error identification and recovery practice. However, ONR notes that the available literature base is currently immature, with limited, empirically designed studies that are grounded in theory, being identified through the review.

The majority of the literature originated from the healthcare sector, and predominantly targeted peer-to-peer and self-checking. Few studies examined the efficacy of independent supervisory checks, which are more typical of the type of checks used for safety critical nuclear operations. It has therefore not been possible to draw firm conclusions based on the available academic evidence of what constitutes an effective check, and the extent to which the independence of a check influences its reliability.

Given this, and the immaturity of the literature, the review has not identified evidence to cause immediate regulatory concern that proportionately applied, well designed, and well implemented checks are ineffective error identification and recovery measures. However, ONR notes the importance of applying a graded approach; using task analysis to inform check design and periodically reviewing and monitoring check implementation, in designing and implementing high reliability checking systems. Factors such as high workload and distraction are likely to impair efficacy (as seen in the healthcare industry), and the influence of these (and other systemic factors) should be considered when designing checks.

Moving forward, there would be value in further research to better understand this important topic. It is also recognised that at the current time, there is lack of good practice guidance that is available and applied consistently across industry, that defines what good looks like, highlights any inherent weaknesses (e.g. diffusion of responsibility) and how these may be effectively protected against.

Given this, ONR notes the requirement for duty holders to consider the importance of checking within their safety cases, and to undertake work to address any associated uncertainties. In addition, ONR will consider whether further research in this area will be included as part of ONR’s Regulatory Research Register, and will engage with industry with

regard to progressing this.

# Further Information

This research was carried out on behalf of ONR by Abbott Risk Consulting. The full report and further details are available on request, noting the limitations identified regarding the immaturity of the literature in this area.

© Office for Nuclear Regulation, 2023

Any enquiries related to this document should be sent to [contact@onr.gov.uk](mailto:contact@onr.gov.uk) Published June 2023.