Factor interaction influences on human performance in air traffic control: The need for a multifactorial model

Tamsyn Edwards*, Sarah Sharples, John R. Wilson and Barry Kirwan

Abstract. In air traffic management (ATM) knowledge of the impact of human factors on performance is critical to address safety incidents. Previous research has largely focused on the effects of single factors on performance which has resulted in a comprehensive understanding of single factor effects. In current control environments however, the residual threats for incidents often result from the interaction of multiple human factors and the resulting cumulative impact on performance. This research uses a literature review, an analysis of over 400 European aviation incident reports and finally a survey of ATM professionals to assess the need for a multifactorial model of performance. Literature findings suggest that Human Factors approaches are fundamentally single-factor in nature, which is out of step with real ATM working contexts. An incident report analysis, supported by a survey of air traffic experts, suggests that multiple factor incident causation exists. This discrepancy suggests the need for a new approach to looking at how incidents occur, and their factors managed, on a day-to-day basis. The proposed solution is a multifactorial model of human performance.

Keywords: air traffic control, aviation incident reports, human factor interaction, human performance

1. Introduction

Air traffic control (ATC) is a safety critical environment [15]. Air Traffic Controllers (ATCOs) are at the sharp end of this safety critical system. To ensure flight safety, ATCOs are required to maintain a consistently high standard of performance. The potential consequences of poor performance are severe, with high costs and potential loss of life [4]. In such a safety critical environment, human performance and error are primary concerns. Although Kirwan [2] notes that ATC is a highly reliable operation, of incidents that do occur, human error has been attributed as a primary or secondary cause in 75-90% of cases [1], human factors have been repeatedly evidenced to affect human performance [10, 15] and are “major determiners of human error” [10, p330]. Knowledge of the impact of human factors on human performance and error is therefore critical in addressing safety incidents in air traffic control.

Most previous human performance and error research has focused on the effects of single factors on performance [7, 13]. This focused research has resulted in a comprehensive understanding of the impact of single factors on performance. As such, single factor issues such as fatigue, vigilance, and situation awareness problems have now largely been designed out or sufficiently mitigated by design, operational and Human Factors & Safety expertise. In current control environments, although accidents are rare, when they do occur they are often multi-causal in nature, or are seen as having no direct causes but many contributors, as highlighted by so-called ‘Swiss Cheese’ and Resilience Engineering models [e.g. 9]. Therefore, the residual threats for incidents often result from the interaction of multiple human factors and the resulting cumulative impact on performance. But if this multi-factor ‘syndrome’ is reality, are we equipped to deal with such multi-causal phenomena?
A potential solution is the development of a multifactorial model of human performance. This conceptualization would permit the modeling of the interactions between relevant human factors. Human performance limits may also be acknowledged and integrated in the model as ‘performance boundaries’. Currently, this proposal is limited by its anecdotal nature and the lack of specification of factors and thus accurate performance boundaries. However, considering the previous calls in the literature for research into multiple factor interrelations [e.g. 15] the authors believed it would be worthwhile to investigate the need for a human performance model to represent multifactor interactions and multi-factor impacts on performance, in the safety critical domain of air traffic control.

1.1. Current investigation

The current investigation aims to:
1. Assess the need for a multifactorial model of performance
2. Identify and refine key factors which impact ATCO performance for later integration into an envelope model

The aims of this investigation were addressed using three separate methodologies. A literature review was conducted to summarize the sporadic research to date on factor interaction effects on performance. An incident report analysis subsequently aimed to investigate the multifactorial nature of incidents in the field. Finally, a survey for ATM professionals (controllers and incident investigators) was utilized to refine and prioritize factors that should be considered for further investigation and potentially integrated into a multifactorial model of ATCO human performance.

2. Literature review

2.1 Method

Following a systematic search of abstract databases, a review of 83 peer-review journal and conference papers was conducted. The abstract search was focused on nine human factors [attention, communications, fatigue, mental workload, situation awareness, stress, teamwork, trust, vigilance]. These nine factors had been previously identified by subject matter experts as factors that could have a large impact on ATCO performance. The review was guided by two primary aims. The first was to provide justification that the nine human factors, previously specified for inclusion in future investigation, did impact performance. The second aim of the literature review was to identify and summarize previous research on relationships between the nine factors. In order to maintain a clear focus on these aims in the review, strict selection criteria were adopted. Papers were only included in the review if the relationship between at least one of the nine factors, and an additional human factor, also of the pre-established nine, were considered. Therefore, the main focus of this review was the identification of relationships between a set of factors, not the identification of factors known to affect performance. Due to the relative lack of research focused on the relationships between factors, the reviewed articles were not restricted to a specific domain. In addition, the papers selected for review were also not restricted by measure or method differences. This decision was made so that a larger number of relevant articles investigating the relationship between specific human factors could be included in the review.

2.2 Results

In the sample of articles reviewed, each of the previously identified nine human factors was evidenced to affect human performance. Figure 1 presents a visual representation of the results. Each factor is represented by a circle. The circle’s size is dependent on the number of articles in the review that considered that factor; the larger the circle, the greater the number of articles which considered that factor. The lines do not imply a direction of causal influence. Additionally, there is no meaning to the position of the circles or length of lines.

![Fig. 1. Frequencies of research articles relating to factor interactions](image-url)
The factors of fatigue, MWL and stress and their interactions with other factors were most frequently reported. Figure 1 suggests that the dyadic relationships for which research evidence was found include relationships between mental workload (MWL) and situation awareness (SA), MWL and fatigue, MWL and vigilance.

2.3 Discussion

Findings provided evidence that the nine factors previously specified for inclusion in this research did impact performance, therefore providing a justification for the continued examination of the these factors and the relationships between them. A number of relationships were identified between the factors. An example of an established relationship is stress and attention. This is a bi-directional relationship – stress is known to narrow the attentional field in order to reduce cognitive processing load [12]. Additionally, focused attention for long periods induces stress [14]. This may obviously have important applications in the ATC field and is one of the reasons why time on shift is strictly regulated. Significant controversies exist in the literature regarding the direction of relationships. Unfortunately, most studies were focused on cross-sectional designs which resulted in a correlation analysis. Causation cannot therefore be explored. Of the few studies that utilized a causal path analysis to investigate factor relationships [e.g. 5], only dyadic relationships were examined. No further research on the causal relationship between more than two human factors was found. Reasons for the lack of research into multifactor relationships may be surmised. The lack of focused research in this area may simply reflect the preference in research to focus on exploring one factor at a time. However, an additional reason may be that it would not be useful to examine the interactions of factors, and the combined factor impact on performance, if the components and mechanisms of single factors’ impact on performance were not known. By first developing a large body of knowledge resulting from in depth research on single factors, it is now possible to produce hypotheses on what the potential impact of multiple factor interactions on performance might be. The consideration of multiple factors may therefore be viewed as a development in research focus which will use previous knowledge of single factors to inform future research. A number of weaknesses of this review suggest that caution should be taken when interpreting these findings. As most studies were focused on cross-sectional designs which resulted in a correlation analysis, causation cannot be explored. Additionally, the majority of articles (77%) either utilized an experimental or review approach which suggests the results of this review mainly represent experimental findings. Although the experimental approach enables tight control over variables, the findings may not be ecologically valid, potentially affecting the reliability and generalization of the results. Further research is needed to investigate the nature of dyadic and triadic relationships between factors, and the impact of multifactor interactions on human performance decrements, especially within safety-critical environments such as ATC.

3. Incident reports

3.1. Method

An analysis of European incident reports was conducted. The analysis aimed to provide justification that each of the nine human factors considered in the literature review did have an impact on performance in the field. An additional aim was to establish the frequency with which each factor contributed to incidents. From these aims, and previous knowledge from the literature review, three working hypotheses were constructed:

1. Factors recorded in the incident reports correspond to factors in the literature.
2. There is not a reporting bias in incident reports.
3. Interaction effects between factors exist in the field.

Incident reports were collected from EUROCONTROL’s SAFlearn database. The database contained 420 safety occurrence reports submitted to EUROCONTROL by European air navigation service providers (ANSPs) between 2001-2003. Errors contributing to each incident were recorded according to the HEIDI (2001) classification system. Incident report selection criteria focused on ATCO contribution to incidents. Incidents in which the ATCO had no responsibility, control or contribution were excluded from analysis. After applying these criteria, a total of 275 reports out of 420 incident reports were included for analysis.
3.2. Results

Table 1 presents the frequency results from the incident report analysis, rank ordered according to frequency of incident contribution. Attention, SA, communication, teamwork were most often cited as contributing to incidents. Trust, fatigue and stress were reported least, each contributing to less than 10% of incidents.

An analysis was then conducted which examined the frequency of factor dyads present in the incident reports (Table 2). Factor dyads with lower occurrence frequencies (0 – 14) are not presented due to the large number of factor dyads with less than 14 occurrences in the incident reports. As causality relationships cannot be inferred from the reports, Table 2 displays observed common presence of factors. The frequency of occurrence of each dyadic relationship in this table ranges from 15 – 44 incident reports, out of a total of 272 reports. Attention and communications most frequently co-occurred in incident reports, with SA and attention/vigilance co-occurring with a slightly lower frequency, in 40 reports.

3.3. Discussion

All factors that were noted in the literature review as impacting performance were also recorded in the incident reports. This provides justification that the selected nine factors do impact human performance, and supports the first working hypothesis. Again, this provides further evidence for considering these factors in future research and a model of safe human performance. Two other factors were dominant in contributing to incidents: memory and perception. Although these were not included in the initial set of nine factors to be reviewed, results from the incident reports suggest that future research should include these factors in investigations of the combined effects of factors on performance and safety incidents.

Although the factors recorded in the incident reports corresponded to factors in the literature, the relative importance of each factor differed. In the literature, workload, fatigue and stress are repeatedly noted to impact human performance. This is not reflected in the incident reports. Workload was only observed to contribute to 24% of incidents, whereas fatigue was recorded to contribute to 2% of incidents and stress 1% of incidents. This represents a considerable discrepancy between literature and aviation incident reports regarding the relative impact of each factor on human performance decrements.

One explanation for this result is that it is an artificial discrepancy. The recorded lack of workload, fatigue and stress in incident causation may be due to a reporting bias. For example, [3] found that certain factors which are seen as ‘static’, i.e. normally present, tend to be under-reported, since they are always there. The rationale underlying this is that the workforce feels such items are part of the job, and never going to change, so why report them? This applies equally to investigators, since their aim is to determine measures to prevent recurrence – why document items for which there is no chance of change? In such cases only ‘dynamic’ or unusual factors are reported. In addition, other reporting biases such as recency or familiarity biases may have influenced the frequency with which certain factors were recorded. Hypothesis 2 is therefore not supported, as reporting biases are believed to impact the accuracy of incident reports. Data may therefore need to be interpreted with some caution and possibly be confirmed with additional research. However, the incident report results suggest that interaction effects between factors do exist in control rooms, and so the third working hypothesis is supported. This knowledge may be used to facilitate investigators in identifying and recording the contribution of human factor interactions in incidents.

Although this discrepancy may therefore be artificial, it was important to confirm the accuracy of the previous results. The research findings regarding the
relative prioritization of factors which most negatively impact controller performance would be used to inform future research and the factors to be included in a human performance model of performance. Therefore, it was decided that further research was needed to clarify the discrepancy and confirm which factors were indeed most negative for controller performance. The method selected to do this was an anonymous survey of air traffic professionals, who could provide expert opinion on which factors and factor combinations they saw most negatively influenced performance.

4. Survey

The aims of the survey were to refine and prioritize the factors for inclusion in a multifactorial model of human performance, and establish a relative prioritization of the factor combinations which most negatively influenced performance, to be used in future research.

4.1 Method

4.1.1. Refinement of factors for inclusion in the survey

The questionnaire was designed in collaboration with 5 ex-controllers. During the development process, several factors were found to be difficult to respond to, or not appropriate to an ATC context. For example, controllers fed back concern that high workload and underload would have differential effects on performance. Therefore, in order for the factor to be meaningful, mental workload was separated into high workload and underload. The specification of underload rather than low workload was important. Underload represented a specific phenomenon to operational experts in which workload become so low that controllers may struggle to maintain a high performance. This factor was specifically of interest as it may affect performance as negatively, or even more so, than a high workload. Therefore, high workload and underload were included as separate factors in future research. In addition, concerns were raised by the ATC professionals that the similarity between the factors of attention, memory, perception & vigilance was confusing when attempting to respond to questions. Each of these factors can be said to operate at level 1, 2 or 3 of SA. Therefore, all factors were collapsed within the broader factor of SA for future investigations. In addition, the factor of trust was found to be very challenging to respond to by ATC professionals. Controllers need to trust in their colleagues to complete their tasks safely and expeditiously, or else the system would not function. Similarly, ATCOs either trust in the systems they are using, or they do not use the system. To resolve this issue, interpersonal trust was collapsed into the factor of teamwork, whilst trust in systems was rejected from further investigation. This resulted in the inclusion of 7 of the original factors in the survey.

4.1.2. Design

The online survey consisted of a total of 28 items and aimed to investigate the frequency of the co-occurrence of dyadic factors in performance related loss of separation incidents and controller performance decrements. Each factor was paired with each of the 7 other factors once, (high workload, underload, stress, fatigue, inadequate situation awareness, inadequate communication and inadequate teamwork) to create 21 questions. The factor pairs were counterbalanced so that each factor appeared as the first factor in three pairs and the second factor in three pairs, to reduce the potential of order effects on participant responding. Each item utilized a 5-point Likert response scale, ranging from ‘Very rarely’ to ‘Very frequently’.

4.1.3. Participants and Procedure

The target population for the survey was ATCOs and incident investigators. The survey was developed online and so snowball sampling was utilized. The survey was closed two months after it was first accessible. Of 65 respondents, the majority consisted of active controllers (56.9%), 20% of respondents were incident investigators, and 23.1% were both active controllers and incident investigators. Respondents identified themselves from a total of 24 European countries. The survey confirmed the multifactorial nature of incidents, from expert opinion, and identified the factor dyads which, when co-occurring, were considered to most threaten performance.

4.2. Results

A Mann-Whitney U test confirmed significant between group differences for 12 items (U=254 – 383, p<0.05). Therefore, the results for controller respondents and investigator respondents (including respondents who were investigators and controllers) were analyzed separately.
Table 3. Ranked Means and Standard Deviations for each Item - Controllers

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWL &amp; Inadequate TW</td>
<td>2.59</td>
<td>0.86</td>
</tr>
<tr>
<td>Inadequate SA &amp; HWL</td>
<td>2.54</td>
<td>0.93</td>
</tr>
<tr>
<td>HWL &amp; Stress</td>
<td>2.46</td>
<td>0.96</td>
</tr>
<tr>
<td>Inadequate SA &amp; HWL</td>
<td>2.46</td>
<td>0.90</td>
</tr>
<tr>
<td>Inadequate comm. &amp; HWL</td>
<td>2.43</td>
<td>1.01</td>
</tr>
<tr>
<td>Inadequate SA &amp; Inadequate comm.</td>
<td>2.43</td>
<td>0.93</td>
</tr>
<tr>
<td>HWL &amp; Underload</td>
<td>2.38</td>
<td>0.98</td>
</tr>
<tr>
<td>Stress &amp; Inadequate SA</td>
<td>2.35</td>
<td>0.75</td>
</tr>
<tr>
<td>Fatigue &amp; HWL</td>
<td>2.35</td>
<td>0.82</td>
</tr>
<tr>
<td>Stress &amp; Inadequate comm.</td>
<td>2.35</td>
<td>0.82</td>
</tr>
</tbody>
</table>

4.2.1. Analysis – controllers

The range of means was calculated from 1.73 (indicating a general low response to the item, suggesting the factor pair would not often contribute to incidents) to 2.59 (indicating the item may rarely - sometimes contribute to losses of separation). Table 3 presents the means and standard deviations of the ratings of the top 10 factor pairs in rank order.

Table 3 shows that the top 6 factors pairs have mean ratings between 2.43 – 2.59. High workload and inadequate teamwork received the highest mean rating (M=2.59, SD=0.86), indicating that this pair contributes to performance decrements more often than other factor pairs presented. The standard deviations are also relatively small (0.75 – 1.06), indicating concordance between respondents. A data trend can be seen in that all of the top 4 factors contain high workload. Data were further analyzed with inferential statistics. A Levene’s test revealed the homogeneity of variance assumption was violated for 6 factor pairs (p<0.05), and a Kolmogorov-Smirnov analysis revealed that 18 out of 21 results did not have a normal distribution (p<0.05). Therefore, non-parametric statistics were utilized. A Friedman’s ANOVA suggested that participants did rate factor pairs significantly differently (X²(20) = 77.2, p<0.0005). Wilcoxon tests were used to follow up this finding. A bonferroni correction was applied and so all effects are reported at a 0.001 level of significance.

Comparisons between 13 pairs of factors were rated significantly different at the p<0.0005 level of significance. Several differences existed between high workload and underload.

Wilcoxon statistical analyses revealed each of the top rated factors containing high workload were rated as significantly more frequently contributing to incidents than those factor pairs containing underload. For example, inadequate communication and high workload was rated significantly more frequently contributing to incidents than factor pairs containing underload, including inadequate teamwork and underload (T=37.5, z=-3.59, p<0.0005) and high workload and inadequate situation awareness (T=39, z=-3.56, p<0.0005).

In addition, inadequate situation awareness and high workload was rated significantly more frequently as contributing to a reduction in safe service than four factor pairs containing underload, including stress and underload (T=37.5, z=-3.59, p<0.0005) and underload and inadequate situation awareness (T=34, z=-3.45, p=0.001).

Table 4. Ranked Means and Standard Deviations for each Item - Investigators

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate comm. &amp; HWL</td>
<td>3.43</td>
<td>0.96</td>
</tr>
<tr>
<td>HWL &amp; Inadequate TW</td>
<td>3.25</td>
<td>0.89</td>
</tr>
<tr>
<td>Inadequate SA &amp; HWL</td>
<td>3.18</td>
<td>0.77</td>
</tr>
<tr>
<td>HWL &amp; Underload</td>
<td>2.96</td>
<td>1.17</td>
</tr>
<tr>
<td>Inadequate SA &amp; Inadequate comm.</td>
<td>2.93</td>
<td>0.88</td>
</tr>
<tr>
<td>Inadequate comm. &amp; Inadequate TW</td>
<td>2.86</td>
<td>1.01</td>
</tr>
<tr>
<td>Stress &amp; Inadequate SA</td>
<td>3.82</td>
<td>1.12</td>
</tr>
<tr>
<td>Stress &amp; Inadequate comm.</td>
<td>2.82</td>
<td>0.94</td>
</tr>
</tbody>
</table>

4.2.2. Analysis – investigators

The range of means was calculated from 2.0 (indicating the factor pair would ‘rarely’ contribute to incidents) to 3.25 (indicating the item may ‘sometimes’ contribute to losses of separation). Table 4 presents the means and standard deviations of the top 10 factor pairs in rank order. Table 4 shows that high workload appeared in all the top 5 factor pairs rated as contributing most frequently to incidents. Inadequate communication and high workload received the highest ratings (M=3.43, SD=0.96). Friedman’s ANOVA suggested that participants did rate factor pairs significantly differently (X²(20) = 92.66, p<0.0005). Wilcoxon tests were used to follow up this finding. A bonferroni correction was applied and so all effects are reported at a 0.001 level of significance.

Comparisons between 13 pairs of factors were rated significantly different at the p<0.0005 level of significance. Several differences existed between high workload and underload.

Wilcoxon statistical analyses revealed each of the top rated factors containing high workload were rated as significantly more frequently contributing to incidents than those factor pairs containing underload. For example, inadequate communication and high workload was rated significantly more frequently contributing to incidents than factor pairs containing underload, including inadequate teamwork and underload (T=21, z=-3.77, p<0.0005) and stress and underload (T=21, z=-4.11, p<0.0005). In addition, high workload and inadequate teamwork was rated significantly more frequently as contributing to loss
of separation incidents than underload and fatigue (T=12, z=-3.55, p<0.0005) and stress and underload (T=16, z=-4.05, p<0.0005).

4.3. Discussion

4.3.1. High workload vs. underload

The top 4 factor pairs as rated by controllers, and the top 5 factor pairs as rated by investigators, all contain high workload. This may indicate that high mental workload is perceived by ATM experts to be one of the main contributors to loss of separation incidents. Controllers, however, are highly trained to effectively manage high workload, adopting various strategies to maintain a high standard of performance [8]. Therefore, the results suggest that in order for performance to be maintained under high workload, other factors within the social-technical system must be operating adequately. If a component is less than adequate when a high workload is present, the risk of a performance decrement or loss of separation incident may increase. Both controllers and investigators rated fair pairs containing underload as very infrequently impacting performance. The factors paired with underload were the same factors that appeared in the top four factor pairs with high workload, suggesting the level of workload is the determiner of the impact of a factor pairs on performance. For example, under a high workload, inadequate teamwork could enhance the complexity and risk of a situation. However, with few aircraft on frequency, a controller may have the spare capacity to support the deficit arising from inadequate teamwork.

4.3.2. High workload and underload

Both controllers and investigators highlighted the transition between workload extremes as frequently contributing to loss of separation incidents and performance decrements. These workload transition effects are well-known colloquially [e.g. 11]. However, very little research has investigated this effect. Future research should investigate the relationship between workload extremes, and the impact of the transition effects on performance.

5. General discussion

A review of literature and analysis of incident reports revealed a discrepancy in the relative frequency of focus of human factors impacting performance. It is difficult to form a direct comparison between these methods and the survey as the considered factors were refined. However, the survey results may contribute to resolving some of the discrepant findings. For example, the literature most often considered high workload in investigations although workload was not often classified in incident reports. High workload was found in the surveys to be a dominant factor that may influence performance, although potentially only when other aspects in the socio-technical system are not functioning adequately. This may resolve the discrepant findings between literature and incident reports. The survey also supported the incident report results that inadequate SA is a key contributor to incidents, as is inadequate communications and teamwork. Therefore, these factors should be examined in any further studies. Finally, the survey results regarding fatigue and stress partially support the results of the incident analyses. Fatigue and stress were mostly under reported in the survey, with a few notable exceptions. High workload and stress was rated as frequently occurring by both incident investigators and controllers, as was fatigue and inadequate teamwork by controllers. This may suggest that these factors, which more indirectly impact performance, may only be perceived to impact performance when combined with certain factors. Further research is needed to understanding these relationships.

The survey results also partially confirmed the dyadic factor co-occurrences revealed from the incident report analysis. The incident reports revealed that attention/vigilance and inadequate communications most often co-occurred in incidents. This factor was rated as 6th most frequently occurring by atc professionals, supporting the importance of this relationship for performance. In addition, mental workload and attention/vigilance was ranked as the 3rd most frequent co-occurrence in incident reports. The survey results supported this, as high workload and inadequate SA were rated within the top five most frequently occurring factors to negatively impact performance by both controllers and incident investigators. This concordance suggests a key focus on the factors of high workload, SA, communications and teamwork, and their associated relationships when investigating human performance in air traffic controllers.

6. Conclusion

In conclusion, literature review findings suggest that most Human Factors approaches are fundamen-
tally single-factor, or at most two-factor in nature. This has resulted in limited knowledge of the impact of multiple factors on human performance, and has restricted the study of human performance decrements to a reactive, retrospective analysis of likely causes [6]. The single factor approaches are also out of step with real ATM working contexts. An analysis of over 400 ATM incidents further supports the notion that multiple factor co-occurrences in incidents exist, which may have implications for applying supportive strategies at particular points in a causal chain. These two results suggest the need for a new approach to looking at how incidents occur, and their factors managed, ultimately on a day-to-day basis. The proposed solution is a multifactorial model of human performance. A survey was utilized to refine the factors to be investigated in future research, and identified the most frequent factor relationships to negatively impact performance, and the most negative factor pairs to negatively impact controller performance. These results will be utilized to provide hypotheses for further investigation, and suggest a core set of factors that should be integrated into a model of safe human performance in ATCOs. Such a model may encourage early detection and mitigation of situations where performance is likely to decline, and potentially allow for general performance boundaries or tolerances to be hypothesized. Additional future research should investigate effects of multifactor interactions on performance decrements, in addition, exploring the causal relationships between human performance factors. Through this knowledge, a shift towards a greater understanding of human factor interactions and the impact on performance in aviation incidents may be achieved.

References