# We know why people fail to follow procedures: now on to interventions.

**Catherine Drury Barnes, Colin G. Drury**

Applied Ergonomics Group Inc., USA

## ABSTRACT

Procedure following is used in many safety-critical enterprises to help ensure that technically correct methods are used for many tasks. Despite this, Failure to Follow Procedures is cited as one of the prime causal or contributing factors in many incidents and accidents. This paper follows an earlier study of failure to follow procedures in aviation maintenance, by going beyond the literature and pre-existing accident/incident data bases to collect data from system participants in eight different maintenance sites. Interviews were conducted with 63 users of procedures, mainly aviation mechanics, who had experience with such incidents, and with 92 managers, supervisors and procedure writes whose job is to control the maintenance process. The users provided details of an incident, then noted whether each of 90 contributing factors played a part. The managers rated the 15 “good practices” from the literature and provided detailed narrative comments. Analysis of the data confirmed the earlier analyses, although with some dissociation. Based on the data a set of interventions was developed, with training programs and audit procedures for each human function within the maintenance system. While the interventions are specific to aviation maintenance, the general good practices and contributing factors validated here have obvious application in other domains.

## KEYWORDS

Procedures, Procedure following, maintenance

## Introduction

Safety-critical organizations continue to rely on procedures for performance of routine activities, despite the fact that many in the Ergonomics/Human Factors (EHF) field have questioned whether this is an appropriate strategy, e.g. Reiman (2010). A major issue is that of failure to follow procedures (FFP) also known as Employee’s Procedural Non-Compliance or employee non-adherence to procedures (NATP) e.g. Mitchell (2005), and also as Practical Drift (Snook, 2000) or Procedural Drift (Johnson, 2003). In this paper we prefer Failure to Follow Procedures as it appears rather more neutral as to underlying causes.

In this paper we address the issue in the domain of aviation maintenance, which has had its share of FFP incidents as well as research in journals and prognostications and exhortations in trade and government publications, e.g. NTSB (2015). For data on the prevalence of FFP in aviation maintenance the publications of Boeing company (Rankin, 2008) and the FAA (Drury and Johnson, 2013) are typical. The latter source quotes Rankin’s analysis of Boeing incident analysis data as showing FFP as the primary causal factor in maintenance incidents. A further quote is that for analysis of major malfunctions following maintenance, the number one cause was failure to follow documentation (Johnson and Watson, 2001). Aviation maintenance is not a unique domain: in aviation flight operations Landry, Jacko, and Coulter (2006) found 31% of accidents involved maintenance of 55% of these were FFP accidents. Beyond aviation, the Health and Safety Executive (HSE, 2012) found strong evidence for FFP in containment level 3 laboratories, while an earlier report (HSE, 2000) reached similar findings for offshore operations.

Our first aviation work on this topic was a literature survey to find potential causal factors and good practices in a widely variety of industries (e.g. nuclear power, manufacturing, chemical plants, military aviation) as reported in Drury, Drury Barnes and Bryant (2017). Approximately 100 reference sources, ranging from quantitative analyses in research papers to internet sites with recommendations for reducing FFP incidence, were compiled and analysed to produce a list of 90 contributing factors and 15 good practices. These were further classified into five categories with many sub-categories in a hierarchical structure. The top level comprised Task, Actor, Procedure Document, Environment and Social ("TAPES"), mirroring many EHF classifications but bringing out the procedure document as a distinct category as this had the most literature counts. These lists were then validated against two data bases, one of NTSB accidents and one of self-reported incidents, largely confirming the lists with a few additions, although counts of individual factors and good practices also reflected inherent biases in the data sources (p 1116). The 15 Good Practices are given with current data later in the paper, but their TAPES counts can be summarized as follows:

| | |
|---|---|
| Task | 1 |
| Actor | 1 |
| Procedure Document | 7 |
| Environment | 1 |
| Social | 5 |

The Drury et al (2017) paper also found 90 contributing factors, which were also grouped and classified using the TAPES classification. Obviously, many of these contributing factors were the negatives of the 15 good practices, although not all were directly addressed by the good practices. The current paper follows directly from this effort by attempting to validate the lists further by detailed interviews with different "Actors" at selected aviation maintenance sites in the USA.

## Methodology

A total of 155 participants was interviewed (lasting about 45 min) across seven maintenance organizations and one aircraft manufacturer: Two other participants had unusable data. Three types of maintenance personnel were recruited; 63 aviation maintenance technicians (AMTs), as well as company supervisors, inspectors, managers, and procedure writers (N = 92). Overall, the sample was age-representative of the population of AMTs (BLS, 2016). All questionnaires were presented to participants in semi-structured private interview format. All interviews were transcribed by researchers by hand, and transferred to a computer database at the conclusion of interviews. All participants confidentiality was covered by the Institutional Review Board.

An Incident Questionnaire was developed to interview AMTs regarding their personal experiences with FFP events. Each participant was asked to recount a time when an FFP event occurred and their unprompted opinions regarding the Contributing Factors (CFs) for the event. Participants were then presented with 90 CFs, and asked to determine whether each CF contributed to their event. Finally, Respondents were asked to discuss what mitigation strategies had been put into place to mitigate FFP events, as well as any strategies which might mitigate FFP events if they were enacted. The managers etc. did not rate the 90 CFs as their questionnaires covered good practices.

A Good Practices questionnaire was used to interview company inspectors/supervisors, procedure writers and managers to discuss the FFP mitigation strategies in place at each of their sites, their personal mitigation strategies, and what strategies they would like to see enacted. At the conclusion of shared mitigation strategies, researchers asked participants to rate personnel effectiveness in implementing each of the 15 good practices. [These rating results are too extensive to present here, and all analysis will be presented on the accompanying narrative data.]

Note that interview questionnaires produced positive and negative factors related to FFP events. The Incident Questionnaire was given to participants who had direct experience with FFP incidents as this was closer to their experience than consideration of good practices, although some good practices data did emerge in their narratives. Conversely, managers etc. were asked directly about good practices as their experience was closer to implementing practices than to having FFP incidents. However, those receiving the Good Practices Questionnaire did provide contributing factors, i.e. the negative good practices.

## Results

The salient results from both types of interview are presented in turn, and finally brought together to provide a comprehensive and quantitative picture of the causes of FFP incidents.

### *FFP Incident Interview Results*

Because the basic data from the Incident interviews was a Yes or No to each contributing factor, the main analysis is on counts of the number of responses to each CF. First, all 90 CF's were cited by participants in at least one FFP incident, and no new CFs were found in narrative responses. Thus, the list of 90 CFs was found to be valid. Second, there was a small negative correlation between total CF Yes responses of each participant and their age ($r = -0.27$, $p = 0.029$).

In terms of a ranking of cited CF's, we can consider the totals by the TAPES category, shown in Table 1. The major change from earlier work is the low number and percentage for Procedure Document. This is discussed later.

Table 1: Counts and percent of Contributing Factors

| TAPES Class | Count of CFs | % of Respondents |
|---|---|---|
| Task | 304 | 32.2 |
| Actor | 255 | 28.8 |
| Social | 460 | 27.0 |
| Environment | 191 | 18.9 |
| Procedure Document | 137 | 10.9 |

The top ten (an arbitrary cut-off) individual contributing factors are shown in Table 2, again with Procedure Document notably absent. Additionally, only a single example of either Environment or Actor made the top ten list: Social and Task together comprised seven of the ten.

Table 2: The ten most cited Contributing Factors

| Contributing Factor | TAPES | % Responses |
|---|---|---|
| Task familiarity (Too low OR too high) | Task | 66.7 |
| Time / production pressure | Social | 65.1 |
| Task interruptions tolerated | Social | 61.9 |
| Task distractions tolerated | Social | 60.3 |
| Users believe they know what they are doing | Social | 57.1 |
| Did you miss a task step? | Task | 50.8 |
| Did you forget to perform a task? | Task | 49.2 |
| Was this on other than a day shift? | Environment | 49.2 |
| Was the task performed from memory | Actor | 46.0 |

***FFP Good Practices Interview Results***

The Good Practices interview was given to those in a more managerial position, e.g. supervisors, managers, document writers and was not focussed on a single FFP incident. Also, there was much more narrative solicited and provided to give depth to the findings beyond effectiveness ratings of the 15 Good Practices. Two findings need to be presented from the effectiveness ratings however. First, the mean ratings were positive, close to “Very Effective” on the 5-point scale. Second, the mean rating of good practice effectiveness for each participant did not have a significant raw correlation with age ($r = -0.15$, $p = 0.162$), although age was a highly significant co-variate in the overall GLM ANOVA ($p < 0.001$). Note that a parametric ANOVA was used on ordinal date after checking that the residuals were indeed normally distributed. The other data presented here is of counts of various responses, for which a non-parametric Chi-Square test is appropriate.

Narrative data came from three sources. An initial unprompted question asked for their, and their organization’s good practices. Narratives were next encouraged concerning each of the 15 good practices. Finally, a question was asked about any ideas participants had concerning good practices they would like to see implemented. A sample narrative data was coded independently by the authors to align criteria, then all data was coded, e.g. into positive and negative comments.

The main narrative data analysis was counting positive comments (e.g. “Safety First here”) and negative comments (e.g. “We DO feel pressure!”) for each Good Practice. A Chi-square test of whether the percentage positive differed between Good Practice number gave a highly significant result: Chi-square (14) = 49.0, $p < 0.001$). This positive percentage showed a high correlation with the mean ratings presented earlier, $r = 0.622$, $p = 0.013$. Table 3 summarizes the positive comments for each good practice.

Table 3: The 15 Good Practices with their percentage of positive comments

| TAPES class | Good Practices for each TAPES class | % Positive Comments |
|---|---|---|
| **Task** | 1. There is a known policy to deal with incorrect or incorrectly installed parts on the aircraft. | 65.2 |
| **Actor** | 2. Users are trained appropriately, experienced and knowledgeable. | 48.8 |
| **Procedure Document** | 3. Procedures are technically accurate. | 50.0 |
| | 4. Procedures are designed to conform to Human Factors guidelines for content, organization, readability and graphics. | 39.5 |
| | 5. Procedures are used only as needed, and at a suitable level for professional users. | 43.2 |
| | 6. Procedures are in a medium suitable for use at the working point. | 47.4 |
| | 7. Procedures are kept up-to-date. | 57.1 |
| | 8. Procedures have been validated by observing their use in detail. | 75.0 |
| | 9. Procedures incorporate explicit input from users, i.e. AMT's and inspectors with direct knowledge of the tasks. | 60.0 |
| **Environment** | 10. A high-quality visual environment is provided, including aids for seeing inaccessible work points. | 84.0 |
| **Social** | 11. Organizational policy on use of procedures is in place, | 50.0 |
| | 12. Organizational policy is enforced by all levels of management as well as by peers. | 42.9 |
| | 13. Procedures are available when needed and users can always find the correct procedure. | 77.8 |
| | 14. Users are insulated from time / production pressures. | |
| | 15. Users have an appropriate and known plan to improve or optimize procedures | 21.9<br>71.4 |

For the three questions *not* prompted by a good practice, Table 4 provides data on counts of comments, of which all were necessarily positive due to the question wording. The first two questions were given before the good practices questions, while the third was given after these questions. This table is given here to provide examples of the detailed comments given by system participants, rather than wording (e.g. Table 3) distilled from many sources. The total narrative frequencies of comments pertaining to each category of the TAPES classification did not differ between the three questions (Chi-square (8) = 10.02, p = 0.264). Overall, Social was the classification with the most comments (122) with Actor (74) and Procedure Document (61) next. Task (21) and Environment (18) had by far the fewest positive comments. Note also that the highest individual item totals for each question were for "Improve task cards/process".

Table 4: Un-prompted comments before and after the good practices questions

| TAPES | Detail | Pre-good practices: Org. | Pre-good practices: Personal | Post-good practices |
|---|---|---|---|---|
| **Overall** | Generally positive comments | | | 13 |
| **Task** | Support the Task, e.g. Tools/workspace | 1 | | 2 |
| | Second set of eyes, buy-back | 9 | 4 | 3 |
| | Check each other's work | 1 | | 1 |
| | Better process planning | 1 | | |
| **Actor** | Initial training, OJT, CBT, tracking | 8 | 8 | 7 |
| | Mentoring, coaching | 4 | 7 | 2 |
| | Recurrent training | 18 | 3 | 2 |
| | Reactive training after event | 5 | 1 | |
| | Hiring good AMTs | | 1 | 4 |
| | FFP training, "dirty dozen" | | | 3 |
| | Reduce turnover | | | 1 |
| **Procedure** | Improve Task Cards/Process | 19 | 13 | 10 |
| | Improve manuals, e.g. GMM, CMM, AMM | 7 | 2 | |
| | Eliminate references to other documents | 1 | | |
| | Make checklist | 3 | 2 | |
| | Validate procedures | 2 | 1 | |
| | Standardize procedures | | | |
| | Easier e-access for AMTs | | | 1 |
| **Environment** | Work time limits | 2 | 2 | 3 |
| | Shift rotation | 1 | | 1 |
| | Better hangars, less constricted | 1 | | 2 |
| | Help AMTs slow down | 3 | 2 | |
| | Less workload/schedule pressure | | | 1 |
| **Social** | Personal communication | 5 | 18 | |
| | Group communication, e.g. briefings | 10 | 8 | 1 |
| | Bulletin boards / TV monitors | 7 | 4 | 1 |
| | Use SMS / CASS (proactively) | 3 | | 1 |
| | QA Audits | 2 | 2 | |
| | STOP policy if uncertain | 5 | 5 | 1 |
| | Use Investigation of events, RCA ASAP | 12 | 4 | |
| | Stand Down day | 4 | 1 | 1 |
| | Change policy manuals | 3 | 2 | 1 |
| | Brainstorm improvements | | 1 | |
| | Assign tasks by AMT experience | | 1 | |
| | Leadership, accountability | | 2 | |
| | Speak up if something is wrong | | | 6 |
| | Better supervision of AMTs | | | 1 |
| | Enforce procedure policy | | | 7 |
| | Better leadership on FFPs | | | 3 |

**Discussion**

The results across both interview types were largely mutually supportive, and also supportive of the prior literature findings from Drury et al (2017). All of the contributing factors and good practices found some support in the data, although there was some dissociation of relative importance between the different measures and samples. The main dissociation was the low frequency of Contributing Factors for the incident interviews compared with the prior research (e.g. 15 good practices) and the Good Practices interviews conducted in this study. Some of this arises from the literature sources on FFP procedures used to develop the 15 good practices. These contained many studies and recommendations from people who were either EHF professionals or those concerned more generally with writing procedures, rather than the practicing mechanics (such as AMTs) interviewed here. Note however that AMTs who reported incidents to the Aviation Safety Reporting System (ASRS) analysed earlier *did* cite the procedure frequently (Drury et al, 2017). It appears that the main cause of the dissociation was that our interviews asked about an incident the AMT *remembered*, and most of these were from some time in the past, or were incidents the participant observed but was not a direct participant in. Indeed, many of our AMT participants said they did not remember anything about the procedure document at this later time, or had never seen it when they observed the incident. There seems to be a memory or salience issue here, rather than procedure documents not in fact contributing to the incidents. Studies that examine the design of the procedure document are still on-going in the EHF community, e.g. Hendricks, Peres and Neville (2018); Mehta and Thomas (2018), so there is still belief in the EHF community that the procedure document is important for FFP reduction.

Apart from this one finding on Procedure Document, the results were mainly consistent across the various studies in that Social variables (i.e. human interaction with other humans) represented a major factor in FFP incidents. In fact, the results should be familiar to EHF professionals who study procedural compliance, with major contributing factors such as:

- Time pressure on end user (e.g. Mehta and Thomas, 2018).
- Lack of management commitment to doing things right every time (e.g. use of "second set of eyes" or commitment to "stop & ask").
- Lack of consistent procedure policy enforcement: Most organizations were better at *having* written policies than at consistently enforcing them.
- Procedures that do not use EHF input (although such input has been available for many years, e.g. Drury, 2006).
- Inadequate policies for timely improvement of procedures, leading to AMTs not even suggesting changes after perceiving lack of management action on prior reports.

The conclusion for aviation maintenance is that interventions are needed to ensure that what is known, and has been known for some time, is addressed by an industry with time pressures typical of most enterprises. What has been encouraging from the current study is that every one of the good practices we evaluated had been implemented by at least one of the eight sites visited. Companies can and do follow selected good practices, although nowhere were all of them found. The companies following each good practice did not go out of business, which is what many companies fear if they implement policies that may have on-going costs, such as consistently enforcing policies despite schedule pressures. Perhaps a way to persuade more to implement our findings is to develop prototype materials for companies to follow (see below). Also we should emphasise that following good practices has the identical cost structure to other forms of insurance regularly bought by industry: Pay a small annual cost to prevent a major, although rare, catastrophe.

## Interventions and Conclusions

As a final output from the study, a set of interventions was developed for each of the major actors in the aviation maintenance system:

- Procedure users, mainly AMTs, who carry out the prescribed tasks.
- Inspectors, who both use procedures and check on the work of AMTs.
- Supervisors/ Managers, who control the maintenance process at a number of levels.
- Procedure writers, often engineers, who produce the documents and also help modify them.

For each group, three types of intervention were developed, each based on the same data set:

- A training program, covering the overall findings of these studies and the good practices applicable to each group.
- A small checklist on which the good practices were listed, for use before and after each task as a reminder of the training. These can also be displayed in break rooms printed as posters or as TV displays,
- An audit system to be used in regular evaluations of the extent to which good practices are being followed.

These types of interventions are familiar to all groups of actors from the many forms of training, reminder and audit materials used already in aviation maintenance.

A unique feature of all materials is that every item included was referenced back to the data from this study. Good practices from the original 15 were included if they were rated “High Effectiveness” (so that users are encouraged to continue effective practices) or “Low Effectiveness” (to signify areas in need of improvement). Items from the contributing factors list were included if cited in more than 25% of responses. Items from narratives cited more than 5 times were also included. There was of course considerable overlap between these measures, giving final counts as shown in Table 5. Some more general items were available for inclusion (e.g. group communications, motivation to follow procedures, not performing tasks from memory) but did not receive quite as much support from the data collected.

Table 5: Items included in intervention materials

| Actor | No. of Items |
|---|---|
| Aviation Maintenance Technicians | 17 |
| Inspectors | 12 |
| Supervisors/Managers | 9 |
| Procedure Writers | 6 |
| Other more general items | 15 |

The materials developed are being used in a modified form (e.g. using computer-based training as the instruction medium instead of presentations by professionals) by the agency that requested the study. As these have only just been made available, no information has yet been collected on their effectiveness in reducing failure to follow procedures incidents.

The conclusion is that failure to follow procedures has many known causes and that for one domain, aviation maintenance, it has been possible to move from literature and data base analysis, through on-site data collection to the development of interventions keyed to the findings.

## References

Baron, R. I. (2017) Procedural Drift: Causes and Consequences, The Aviation Consulting Group, https://www.tacgworldwide.com/Portals/23/pdf/Procedural%20Drift%20Causes%20and%20Consequences.pdf?ver=2018-03-03-071026-287

BLS (2016). Bureau of Labor Statistics Data on Aircraft mechanics and service technicians 2016, Document cpsaa11b.xlsx at https://www.bls.gov/cps/demographics.htm

C.G. Drury, (2006) "Procedures and Technical Documentation", chapter 6 of *Human Factors Guide for Aviation Maintenance and Inspection*. https://www.faa.gov/about/initiatives/maintenance_hf/training_tools/media/HF_Guide.pdf

Drury, C. G. and Johnson, W. B. (2013) Writing Aviation Maintenance Procedures That People Can/Will Follow, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2013,* 57: 997-1001

Drury, C. G., Drury Barnes, C. E. and Bryant, M. R. (2017) Why Do We STILL Not Follow Procedures? *Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting,* 1664-1668

Hendricks, J., Peres, S. C. and Neville, T. (2018) The impact of hazard statement design characteristics in procedures on compliance, *Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting,* 1616-1618

HSE (2000). *Techniques for addressing rule violations in the offshore industries*. Offshore Technology Report 2000/096. ISBN: 0 7176 2095 6

HSE. (2012). *Human factors that lead to non-compliance with standard operating procedures*. Retrieved from http://www.hse.gov.uk/research/rrpdf/rr919.pdf

Johnston, N. (2003). The Paradox of Rules: Procedural Drift in Commercial Aviation. In R. Jensen, (Ed), *Proceedings of the Twelfth International Symposium on Aviation Psychology*, April 14-17, 2003, Dayton, Ohio [CD-ROM].

Johnson, W. B. & Watson, J. (2001). Installation Error in Airline Maintenance. Washington, DC: *Federal Aviation Administration Office of Aviation Medicine*. http://hfskyway.faa.gov.

Landry, S. J., Jacko, J. A., & Coulter, W. H. (2006). Impact of the use of techniques and situation awareness on pilots' procedure compliance, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* Santa Barbara, CA: Sage Publications. 50.1, 40-44.

Mehta, R. and Thomas, S. (2018) Effects of Time Pressure and Experience Level on Worker Perceived Workload: Implications for procedural designs in high-risk industrial tasks. *Proceedings of the Human Factors and Ergonomics Society 2018 Annual Meeting*, 1610-1615.

Mitchell, E. (2005), Strategies to Reduce Aviation Employees' Procedural Non-Compliance, Unpublished MSc Thesis, City University London.

NTSB (2015) *Mechanics: Manage Risks to Ensure Safety*, National Transportation Board Safety Alert SA-022, March 2013 Revised December 2015. https://www.ntsb.gov/safety/safety-alerts/Documents/SA_022.pdf

Rankin, W. (2008). Safety management systems and Boeing-related safety activities, presented *November 6-7 2008 at the Safety Management System (SMS) Workshop for Air Transport Industry*

Reiman, T. (2010). Understanding maintenance work in safety-critical organisations – managing the performance variability. *Theoretical Issues in Ergonomics* Science 12.4, 339-366.

Snook, S.A. (2000). *Friendly Fire: The Accidental Shootdown of U.S. Black Hawks Over Northern Iraq*. Princeton U.P. New Jersey.