International Journal of Aviation Research

Volume 13 | Issue 01 Peer-Reviewed Article

Application of Dupont's Dirty Dozen Framework to Commercial Aviation Maintenance Incidents

Greg M. Mellema
Embry-Riddle Aeronautical University

Haydee M. Cuevas
Embry-Riddle Aeronautical University

David A. Esser
Embry-Riddle Aeronautical University

Bruce Conway
Embry-Riddle Aeronautical University

Shari L. Frisinger
Embry-Riddle Aeronautical University

In this quantitative study, twelve preconditions for maintenance errors, commonly known as the Dirty Dozen, were applied to actual incident and accident data provided by a participating airline (PA). Specifically, 25 Maintenance Event Reports (MERs) (reactive) and 60 Maintenance Operations Safety Assessment (MOSA) reports (proactive) were coded by aviation maintenance subject matter experts (SMEs) using the 12 Dirty Dozen categories as the coding scheme. Results revealed not only the presence of each Dirty Dozen category to some degree, but also the difference in sensitivity of the MER (reactive) and MOSA (proactive) to the 12 Dirty Dozen categories. Recommendations for practice and future research are discussed.

Keywords: aviation maintenance; Dupont's Dirty Dozen; human error; safety culture

Recommended Citation:
Mellema, G. M., Cuevas, H. M., Esser, D. A., Conway, B., & Frisinger, S. L. (2021). Application of Dupont's Dirty Dozen Framework to Commercial Aviation Maintenance Incidents. *International Journal of Aviation Research, 13*(1), 43-62.

According to the Federal Aviation Administration (FAA), while pilot error continues to be the leading cause of hull-loss accidents in the commercial aviation industry, maintenance errors are the second leading cause (FAA, 2014). The work of Marx and Graeber (1994), Patankar and Taylor (2004), the FAA (2018) and, most recently, Zimmerman and Mendonca (2021) estimate the maintenance error contribution to commercial aircraft accidents worldwide is between 12% and 15%. Examining the problem in an even broader scope, the International Air Transportation Association (IATA) examined safety reports filed between 2003 and 2008 and found that improper maintenance was linked to aircraft accidents worldwide as much as 40% of the time (IATA, 2008).

In and of themselves, these numbers are cause for concern. However, since each flight-hour results in an average of 12 maintenance man-hours (Hobbs, 2008), it is not unreasonable to suggest that a maintenance error may be up to 12 times more likely to occur and manifest during any given flight-hour when compared to a pilot error. Marais and Robichaud (2012) found that the likelihood of a maintenance-related accident to result in fatalities is approximately 6.5 times greater than non-maintenance-related accidents. They also found that, in accidents resulting in fatalities, those accidents related to maintenance errors generated an average of 3.6 times more fatalities, giving rise to the theory of a “fatality risk magnifier” (Marais & Robichaud, 2012, p. 111) associated with maintenance-related accidents. Regardless of the specific calculations used, it seems clear that maintenance errors play a significant role in commercial aviation safety, making any efforts to reduce them worthwhile.

The examination of human factors research in terms of aviation maintenance surged around 1990, presumably from a series of high-profile air disasters in the 70s and 80s in which aircraft maintenance was implicated (Chang & Wang, 2010; Dorn, 1996; Gramopadhye & Drury, 2000). Certain human error models and theories developed during this timeframe have become widely if not universally accepted. Examples include Reason’s (1990) Swiss Cheese model which illustrates how ‘holes’ in an organization’s systematic defenses can line up, allowing an accident to occur. Also, Dorn’s (1996) adaptation of Edward’s (1988) Software, Hardware, Environment, Liveware, or SHEL model characterized the interaction of these elements within a system such as aircraft maintenance operations. Additional contemporary works by Shappell and Weigmann (2000), Merritt and Klinect, (2006), Maurino and Seminar (2005), and others have developed models or reactive systems and taxonomies designed to help accident investigators determine what maintenance error occurred, but these systems do not necessarily offer any insight as to why it occurred.

The aviation industry is perpetually looking for new means to enhance safety and reduce costs. Proactive means (e.g., preventative measures) are preferred over reactive means (e.g., post-mishap analysis), as they do not require an incident or accident has already occurred along with all the attendant damage, cost, and potential loss of life. However, most proactive means lack sufficient prognosticative power and are therefore of limited value. To decrease maintenance errors, it is important to evaluate both reactive and proactive data to expose existing preconditions for error. One approach is to examine an organization’s maintenance culture through the construct of Gordon Dupont’s Dirty Dozen, which could yield useful information identifying the presence of preconditions for maintenance errors. Once identified, a mitigating strategy can be devised to address specific preconditions present, thereby reducing the total number of incidents and accidents that are able to manifest as a result.

Preconditions for Maintenance Errors: The Dirty Dozen

In reviewing the extant literature concerning aviation maintenance errors, a substantial amount of research supports models and theories of human behavior as it relates to maintenance errors (Gramopadhye & Drury, 2000; Langer & Braithwaite, 2016; Reason, 1990; Schmidt, Lawson, & Figlock, 2001). Additionally, significant effort has been applied to generating taxonomies to accompany these models and theories to help researchers understand what happened in terms of a given maintenance error. However, little research exists to explain why it happened. Historically, accident investigators have applied one or more of the aforementioned models and taxonomies to their investigations to essentially reverse-engineer the sequence of events that made the accident manifest physically. Boeing's Maintenance Error Decision Aid (MEDA), introduced in the mid-1990s, took a systems approach to merge accepted theories of accident causation (Reason, 1990; Schmidt et al., 2001) with a host of contributing factors, some of which are also Dirty Dozen categories (Boeing, 2013). MEDA's novel approach allowed it to perform reasonably well as a reactive investigation tool. However, much as the scientific axiom states - correlation does not equal causation, revealing what failed in a system does not necessarily reveal the underlying reason that it failed and, may even belie it to some degree. For this and other reasons, it is worthwhile to examine the contributing factors or preconditions for maintenance errors.

In 1993, Gordon Dupont was working for the Canadian airworthiness authority, Transport Canada. Dupont, along with an industry liaison committee and members of the Canadian Department of National Defence [*sic*] examined between 1,500 and 2,000 aviation maintenance incident and accident reports simply attributed to some form of human error. After approximately seven months of careful examination and discussion, the team determined the majority of these maintenance-related human errors could be attributed to one or more of 12 basic preconditions for error (see Table 1). These preconditions quickly became known as the *Dirty Dozen*. Shortly after the run of aircraft accidents in the 1980s and 1990s, the FAA's Dr. Bill Shepherd initiated a series of meetings aimed at investigating the issue of human factors as it relates to aircraft maintenance operations (FAA, n.d.). At these meetings between 1993 and 1997, Dupont first presented the Dirty Dozen to the international consortium co-sponsored by the United States, Canada, and the United Kingdom (Dupont, 1997).

Each Dirty Dozen element has a set of *safety nets* associated with it. Safety nets are regulations, policies, and practices or procedures thought to reduce the possibility that any given precondition will actually manifest as an incident or accident. These preconditions for maintenance error seemed to resonate with the personnel in the aviation industry as they offered some explanation as to why incidents and accidents occurred. By 1997, the proliferation of the Dirty Dozen framework was well underway. To date, the Dirty Dozen framework has been widely accepted by airworthiness authorities comprising 11 countries: Canada, Australia, Singapore, China, Sweden, Holland, Hungary, the United Kingdom, Ireland, Portugal, and the United States (CASA, 2013). However, despite this industry acceptance, limited scientific research exists to support the framework. Since this standard has been embraced so thoroughly across the aviation maintenance and safety culture, it would be useful to have some assurance that it is both complete and effective. Ma and Grower (2016), and even Dupont himself, have suggested that the Dirty Dozen may or may not be suitably complete as is. Therefore, evidence suggesting the completeness, or lack thereof, of the Dirty Dozen construct will be important and useful to any organization seeking to reduce its maintenance errors by identifying and reducing its preconditions for maintenance error.

Table 1

Dupont's Dirty Dozen

Precondition for Error	Definition
Lack of Communication	Failure to transmit, receive, or provide enough information to complete a task.
Complacency	Overconfidence from repeated experience performing a task.
Lack of Knowledge	Shortage of the training, information, and/or ability to successfully perform.
Distractions	Anything that draws your attention away from the task at hand.
Lack of Teamwork	Failure to work together to complete a shared goal.
Fatigue	Physical or mental exhaustion threatening work performance.
Lack of Resources	Not having enough people, equipment, documentation, time, parts, etc., to complete a task.
Pressure	Real or perceived forces demanding high-level job performance.
Lack of Assertiveness	Failure to speak up or document concerns about instructions, orders, or the actions of others.
Stress	A physical, chemical, or emotional factor causing physical or mental tension.
Lack of Awareness	Failure to recognize a situation, understand what it is, and predict the possible results.
Norms	Expected, yet unwritten, rules of behavior.

Note. Adapted from Dupont (1997).

Present Study

The purpose of this quantitative study was to apply DuPont's Dirty Dozen to examine two distinct types of reports from a participating airline (PA) for evidence suggesting the presence of one or more preconditions for error. The types and titles of the reports made available by the PA were reactive - maintenance event reports (MER) and proactive - maintenance operations safety assessments (MOSA). Such an examination of any one of these reports would yield useful information about the PA's maintenance culture. However, since proactive and reactive data each have their own strengths and weaknesses, the examination of both types of reports was posited to illustrate the PA's maintenance culture in a more holistic and complete manner. The following research question was examined: How does the reactive data (MER) analysis compare to the proactive (MOSA) analysis in terms of the Dirty Dozen? Do they echo similar Dirty Dozen categories, or do they seem to reflect different aspects of the Dirty Dozen?

Delimitations. The PA operates a fleet of over 100 Boeing 737 aircraft to destinations in eight different countries. The PA also employs approximately 3,000 maintenance-related personnel full-time, so larger organizations with a variety of aircraft makes and models were not addressed. Despite the limitations described below, the process is generalizable since the Dirty Dozen framework is largely agnostic in terms of its application across the aviation operational spectrum, be it maintenance personnel involved in commercial, cargo, or agricultural aviation. However, specific results of the application would be expected to vary from one airline to another due to the host of ethnographic variables in play at any given organization (i.e., airlines in different countries).

Limitations and Assumptions. One limitation of the current study was the finite number of reports that could be provided by the PA within a reasonable timeframe. While more reports would certainly enhance the overall validity of the research, the impact to cost and schedule was deemed too great by the airline. However, the rich variety of reports (proactive and reactive) would help mitigate any issues concerning validity that might arise from the reduced data set. A second limitation was the timeframes in which the data from the different reports were collected. The MOSA observations were conducted over four months after the MER data was collected. The PA was asked about significant turnover of personnel or major training events that may have influenced respondent's behavior or perceptions during that four-month period. The PA stated no such events had occurred. While data collected from the exact same timeframe would have been preferable, there did not appear to be any obvious reason to suspect that MOSA data collected in the last quarter of the calendar year would have been appreciably different than data collected in the first quarter of the same year. Possible additional limitations created by using aircraft maintenance SMEs were precluded by requiring them to have at least 20 years' aircraft maintenance experience, plus teaching experience specifically related to the Dirty Dozen. While individuals with such qualifications can be difficult to find, it was deemed critical to the credibility of the research.

Two assumptions for the current research were: (1) personnel filing accident and incident reports (MERs) were skilled, knowledgeable, and honest, and (2) no malice was associated with their reporting. The primary assumption made for MOSA reports was that the observers were skilled and knowledgeable personnel making sincere efforts to proactively identify potential errors or preconditions for errors. The PA had specific requirements for being a MOSA observer. While any concerned personnel in the maintenance or operations departments can file a MER anonymously, MOSA observers must have: more than four years of experience as a mechanic; qualifications in the tasks observed; knowledge of the PA's procedures; knowledge of technical English; taken the required safety course; personal characteristics that reveal ethics, neutrality, and good interpersonal relationships; and ability to generate a report with clarity and objectivity.

Method

The current study used subject matter experts (SMEs) to examine reports from the PA for evidence suggesting the presence of one or more of the Dirty Dozen preconditions for maintenance error as described by Gordon Dupont. Examined as a whole, the two distinct types of report (proactive and reactive) were expected to illustrate the PA's maintenance culture in terms of the Dirty Dozen, revealing the presence and frequency of the various preconditions for error via descriptive statistics and analysis. Once revealed, the frequency of the noted preconditions was calculated to assist the PA in targeting the most prevalent preconditions in its ongoing effort to enhance organizational safety.

Due to the narrative nature of much of the data provided by the PA, a quantitative research approach was used to answer the study's research question, examining archival data for the presence of the preconditions for maintenance error known as the Dirty Dozen. The SMEs coded two different types of reports (MER and MOSA) from the same airline within the construct of the Dirty Dozen. The results illustrated the overall presence (frequency) of Dirty Dozen elements as well as measured their prevalence (intensity) within the maintenance culture of the PA. This allowed for recommendations to the PA to focus their safety efforts on the most prevalent preconditions for maintenance error.

Design

MER and MOSA Reports. The PA provided 25 MERs collected between January and May 2017 and 60 MOSA reports collected between September and November 2017. Most of the MERs received were short narratives of maintenance-related events that did or could have resulted in injury or damage to an aircraft (see Figure 1). However, some reports simply indicated an overarching issue (e.g., fatigue) the submitter felt could precipitate an event resulting in injury or damage to an aircraft.

Occurrence ID	Type	Risk	Date and Time UTC	Occurrence Title	Registered On/By	Tail Number	Aircraft Type	Location
O212-17	MNT [redacted]	Medium	Jan 10 2017 00:30	Refueling	Jan 11 2017 10:00 [redacted]			[redacted]

DESCRIPTION OF OCCURRENCE

Description: Flight [redacted] on Jan 10th 2016, tail number [redacted], originally going to [redacted] had to divert to [redacted] due to weather. The flight proceeded with no abnormalities whatsoever. When arriving at [redacted] the aircraft was instructed to stop at a GA parking position, and a mechanic greeted us saying he was the only one working that shift; he said he had to go dispatch another aircraft bound to [redacted] and would be right back. As soon as I got the paperwork for the flight I called [redacted] Dispatch, advising them we had to hurry up with refueling because we were almost reaching our duty time limit; the person on the radio also informed us there was only one mechanic available that night and we would have to wait for him to dispatch that other flight. After a few minutes, however, I noticed the fuel quantity on the displays going up. I tried contacting the mechanic over the interphone system, but there was no response. I came downstairs to see who was refueling the plane and it shook me to see the fuel truck driver performing the refuel all by himself; I instantly told him to stop. I asked the mechanic what was going on and he answered he did not request refueling the aircraft. I then questioned the [redacted], who said he only told the driver what fuel quantity they needed. I left [redacted] without knowing who was responsible for this mishap. I am writing this report with the intents of making people aware of what happened and to help such violations from taking place again.

Figure 1. Illustrative example of MER report with identifying information redacted.

The MOSAs are proactive reports generated by periodic internal MOSA team assessments during which trained MOSA auditors observe a variety of maintenance tasks. The 60 MOSA reports conformed essentially to the format presented in AC 120-90 (FAA, 2006) for MOSA reports. Like MERs, MOSA reports also contained narratives. However, MOSA reports were the result of a proactive surveillance program aimed at identifying potentially hazardous behavior before it could manifest as an incident or accident. Nevertheless, the MOSA reports were thought to be similarly illustrative of the airline's maintenance climate, once coded.

The MERs and MOSA reports were coded by two qualified aircraft maintenance professionals. The coding scheme called for each rater to examine each event report within the context of the Dirty Dozen. An initial training session of approximately 90 minutes was held for the raters. Since some preconditions for maintenance error may not be represented in the current Dirty Dozen framework, an additional category labeled *other* was added as a category. The raters then allocated percentages (0% to 100%) to each of the categories based on their assessment of how much each precondition for error contributed to the event (see Figure 2).

Event	**Occurrence ID**												
8	01088-17												
Description of Occurance: I asked the maintenance technician over the radio if everything was ready and set on the ground for the pushback and engine startup; the technician replied saying everything was OK around the aircraft and the hazard zones were clear and ready for startup. I then turned on the strobe lights and lit the ignition on engine #2. I noticed the ground crew was taking too long to push the plane and still had not linked the tow truck to the nose landing gear. Oddly enough, I was able to see the reflection of ground crew walking away from the aircraft with wheel chocks in hands on the glass of terminal building around the plane; in other words, not only there was personnel in hazard zones but there were people around the engine that was spooling up. I inquired the maintenance technician why he stated things were good to go and hazard zones were clear when that was clearly not the case. He replied it was just a matter of connecting the tow bar. After the ground crew left the surroundings of the aircraft the personnel continued the pushback and we proceeded with the startup.													
50	**10**			**20**						**20**			**100**
Lack of Communication	Complacency	Lack of Knowledge	Distraction	Lack of Teamwork	Fatigue	Lack of Resources	Pressure	Lack of Assertiveness	Stress	Lack of Awareness	Norms	Other	Total

Figure 2. Sample coding form. Shows the event number, description of the event, and Dirty Dozen. Notional scores for Rater A shown in yellow.

When complete, the rater's scores for each event totaled 100, representing a characterization of 100% of the event expressed in terms of the Dirty Dozen. See Figure 3 for an example of this comparison. Since the minimum and maximum rating values were known (0-100), and the distance between each value is equal, the scores were considered interval measures for the purpose of determining inter-rater reliability (Hayes & Krippendorff, 2007). Following the first evaluation of the reports by the raters, inter-rater reliability was calculated using Krippendorff's alpha protocol for interval measures. A Krippendorff's alpha value of less than .80 would prompt further training on the application of the categorization scheme. Another inter-rater reliability check was conducted after the re-evaluation of the reports. This process was to continue until the minimum desired inter-rater reliability level (> .80) had been achieved.

In the event the *other* category was used, the raters were instructed to highlight the text they felt represented a precondition for error not listed in the Dirty Dozen. If present, these would be examined later for any common themes that could suggest the presence of a definable precondition for maintenance error such as those suggested by Ma and Grower (2016).

8		RATER A	RATER B
DEFICIT	Lack of Communication	30	25
	Lack of Knowledge	20	15
	Lack of Teamwork	10	20
	Lack of Resources	0	0
	Lack of Assertiveness	0	0
	Lack of Awareness	10	10
SURPLUS	Complacency	20	20
	Distraction	0	0
	Fatigue	0	0
	Pressure	0	0
	Stress	0	0
	Norms	10	10
	Other	0	0

Figure 3. Example rater scores. Consolidated and reordered according to their influence on the event as a surplus or deficit (left). Rater agreement is characterized by the 3-D chart on the right and the actual Krippendorff's alpha value (center).

Rater Selection. The two raters were selected based on two criteria established a priori. The candidates must be FAA certificated Airframe & Power Plant (A&P) mechanics, preferably certificated for 20 years or more, and the candidates must have knowledge and experience with the Dirty Dozen, preferably in a training and/or human factors environment. The use of raters with such specific qualifications was thought to enhance overall reliability and validity of the study.

Results

A Krippendorff's alpha protocol was used throughout the coding process of the MER reports to assure inter-rater reliability remained at or above .80. Once the 25 MER reports were coded by the SMEs and inter-rater reliability was assured, a single set of Dirty Dozen scores were needed to enable a comparison to the MOSA reports. Considering the lowest inter-rater reliability value was .918, any disparate scores between raters were averaged to create a single score for that Dirty Dozen category in that particular case. The raters then each examined 30 proactive MOSA reports in a similar fashion (60 total), with Krippendorff's alpha employed to assure inter-rater reliability.

Maintenance Event Reports (MERs). Each category of the Dirty Dozen can be thought of as either a surplus of an undesirable trait, such as *distraction* or *fatigue*, or a deficit of a desirable trait, such as *knowledge* or *resources*. To characterize the relationship between these two factors, the categories were re-ordered by frequency, as shown in Figure 4.

Category	Freq.
Lack of Communication	428
Lack of Resources	338
Complacency	293
Norms	278
Lack of Teamwork	260
Lack of Awareness	233
Lack of Knowledge	223
Fatigue	158
Pressure	133
Lack of Assertiveness	78
Distraction	45
Stress	38
Other	0

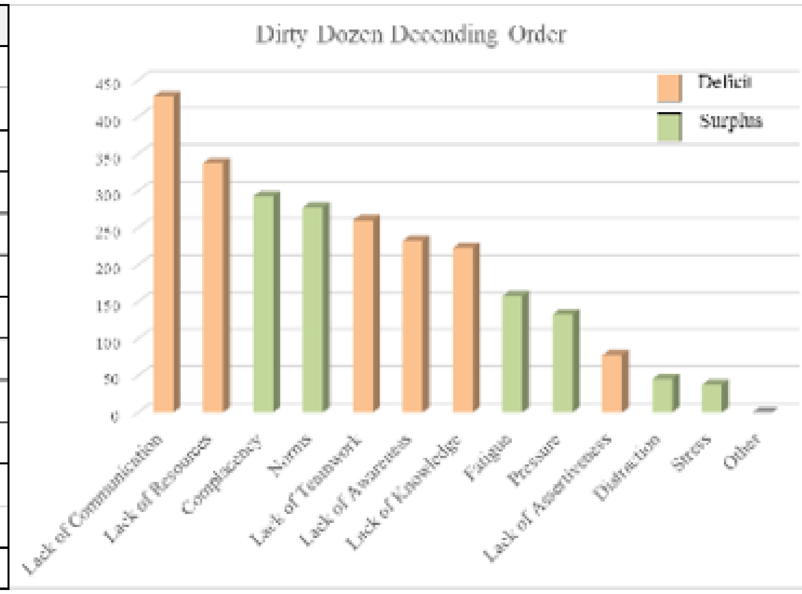

Figure 4. Dirty Dozen categories ordered by frequency and identified as a surplus of an undesirable trait or the deficit of a desirable trait.

It should be noted here that, while not shown in Figure 4, the raters expressed an interest in employing the *other* category for MER case #18. Although they were reluctant to actually score it as such, the raters felt that an argument could potentially be made for a *lack of operational integrity* as described in the works of Ma and Grower (2016) in this one case.

Maintenance Operations Safety Assessment (MOSA) Reports. The MOSA reports are a comprehensive form filled out by the assigned observer. The form contains five areas that apply specifically to Dirty Dozen categories such as communication, fatigue, knowledge, pressure, and norms. Other areas address Dirty Dozen categories in a less direct fashion. For example, comments and indications made by the observer regarding tools, calibration, and technical manuals all relate to the Dirty Dozen category of *Lack of Resources*. Thus, the raters were able to apply scores for Dirty Dozen categories not specifically mentioned by inferring from context.

In coding the MOSA reports, the raters scored the Dirty Dozen category *Lack of Resources* far more often than any other category. As a result, graphing the combined raters results became problematic in that the relatively high frequency of *Lack of Resources* modulated-down the apparent distribution of the other categories. This was rectified by applying a base-10 logarithmic scale to the results, shown in Figure 5.

MOSA Values Descending	
Lack of Resources	3148
Complacency	930
Norms	522
Pressure	355
Fatigue	325
Lack of Knowledge	295
Lack of Awareness	160
Distraction	150
Stress	100
Lack of Communication	80
Lack of Assertiveness	35
Lack of Teamwork	20
Other	0

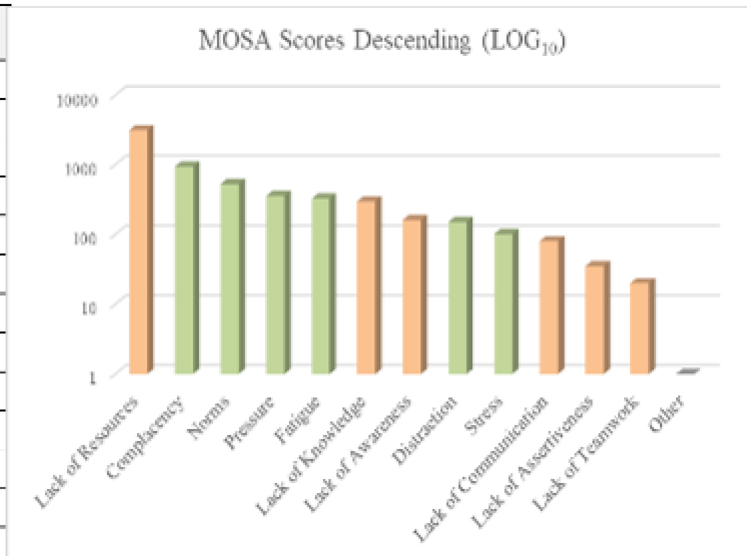

Figure 5. Combined scores showing frequency for Dirty Dozen categories, with values arranged in descending order and identified as a surplus of an undesirable trait, or the deficit of a desirable trait.

MER – MOSA SME Ratings Comparison. Since the sample sizes were not equal (25 vs. 60) and the variances were significantly different between the groups, the two data sets were analyzed by conducting a one-way, between-groups multiple analysis of variance (MANOVA), and the Brown-Forsythe test was used for the univariate analysis. For this analysis, the report type (MER vs. MOSA) was the independent variable, and the percentages reported for each of the Dirty Dozen categories by the SMEs comprised the dependent variable. Multivariate analysis statistics are reported using Wilks' Lambda and an alpha level of .05 was used for all statistical analyses. Means and standard deviations are shown in Table 2.

Table 2

Descriptive Statistics for Ratings on Dirty Dozen Categories for MER and MOSA

	MER		MOSA	
Dependent Variable	Mean	Std Dev	Mean	Std Dev
Lack of Communication	17.16	22.33	1.33	5.03
Complacency	8.92	16.41	15.50	21.93
Lack of Knowledge	10.44	10.84	4.92	12.13
Distraction	13.52	21.12	2.50	6.54
Lack of Teamwork	3.12	9.58	0.33	2.58
Fatigue	9.32	15.60	5.42	15.38
Lack of Resources	11.76	17.60	51.05	32.23
Pressure	1.80	6.27	5.92	11.52
Lack of Assertiveness	6.32	11.82	0.58	4.52
Stress	5.32	8.47	1.67	6.42
Lack of Awareness	1.52	4.41	2.67	9.13
Norms	11.12	16.26	8.45	11.81

Note. Std Dev = standard deviation

Multivariate analysis revealed a significant effect of report type on the SME ratings for the Dirty Dozen categories, F (12, 72) = 9.10, p = .0001, η_p^2 = .603. See Table 3 for estimated marginal means and standard errors along with the test statistics.

Table 3

Estimated Marginal Means, Standard Errors, and Brown-Forsythe Test Results for Ratings on Dirty Dozen Categories by Report Type

Dependent Variable	MER	MOSA	Statistic	df	Sig.
Lack of Communication	17.16 (2.55)	1.33 (1.64)	12.30	1, 25.02	.002
Complacency	8.92 (4.10)	15.50 (2.65)	2.31	1, 59.61	.134
Lack of Knowledge	10.44 (2.35)	4.92 (1.52)	4.27	1, 50.04	.044
Distraction	13.52 (2.53)	2.50 (1.63)	6.55	1, 25.94	.017
Lack of Teamwork	3.12 (1.12)	0.33 (0.72)	2.06	1, 25.47	.164
Fatigue	9.32 (3.09)	5.42 (1.99)	1.11	1, 44.44	.297
Lack of Resources	11.76 (5.76)	51.05 (3.72)	51.96	1, 76.88	.001
Pressure	1.80 (2.06)	5.92 (1.33)	4.48	1, 76.99	.038
Lack of Assertiveness	6.32 (1.48)	0.58 (0.96)	5.55	1, 26.97	.026
Stress	5.32 (1.42)	1.67 (0.91)	3.75	1, 36.04	.061
Lack of Awareness	1.52 (1.61)	2.67 (1.04)	0.61	1, 81.11	.438
Norms	11.12 (2.65)	8.45 (1.71)	0.55	1, 35.04	.462

Note. MER = Maintenance Event Report. MOSA = Maintenance Operations Safety Assessment. Standard errors presented in parentheses following means. df = degrees of freedom. Sig. = significance (p value).

Univariate tests revealed a significant effect of report type on six Dirty Dozen categories: *Lack of Communication*, *Lack of Knowledge*, *Distraction*, *Lack of Resources*, *Pressure,* and *Lack of Assertiveness*. *Lack of Communication*, *Lack of Knowledge*, *Distraction,* and *Lack of Assertiveness* were rated significantly higher on the MER than the MOSA. *Lack of Resources* and *Pressure* were rated significantly higher on the MOSA than the MER. No significant differences were found on the other six Dirty Dozen categories: *Complacency*, *Lack of Teamwork, Fatigue*, *Stress*, *Lack of Awareness,* and *Norms*.

Discussion

Despite the aviation industry's broad acceptance of the Dirty Dozen as the 12 primary preconditions for aircraft maintenance errors, a rigorous literature review did not identify any research conducted that leverages this broadly accepted framework for its potential analytical value. Yet, a systematic examination of an organization's maintenance culture through the construct of the Dirty Dozen may yield useful information identifying the presence of preconditions for maintenance errors. As illustrated in this study, maintenance-related reports can be coded and analyzed using SMEs in such a way as to illustrate the organization's maintenance culture and reveal the presence of these preconditions for maintenance errors. Once revealed, a mitigating strategy can be devised to address the specific preconditions that are present, thereby reducing the total number of incidents and accidents that are able to manifest as a result.

MER – MOSA Comparison. The study's research question asked - How does the reactive data (MER) analysis compare to the proactive (MOSA) analysis in terms of the Dirty Dozen? Do they echo similar Dirty Dozen categories, or do they seem to reflect different aspects of the Dirty Dozen? As further discussed next, the results from the analysis showed the difference between the MER and MOSA reports is complex, with the MER reports detecting certain Dirty Dozen categories better than the MOSA and vice-versa. There also seems to be a subset of categories that the MER and MOSA reports detect equally well.

To begin, the categories of *Complacency*, *Lack of Teamwork*, *Fatigue*, *Stress*, *Lack of Awareness*, and *Norms* were not significantly different between the two types of report. However, *Lack of Communication*, *Lack of Knowledge*, *Distraction*, *Lack of Resources*, *Pressure*, and *Lack of Assertiven*ess were rated significantly higher on the MER than the MOSA, while *Lack of Resources* and *Pressure* were rated significantly higher on the MOSA than the MER. This suggests the two types of report seem to echo each other in terms of some Dirty Dozen categories (*Complacency*, *Lack of Teamwork*, *Fatigue*, *Stress*, *Lack of Awareness*, and *Norms*). In contrast, the data suggest MERs appear to have somewhat greater sensitivity when applied to situations in which the categories of *Lack of Communication*, *Lack of Knowledge*, *Distraction*, *Lack of Resources*, *Pressure*, and *Lack of Assertiven*ess are prevalent. Whereas, the MOSAs appear to be more sensitive when applied to situations in which *Lack of Resources* or *Pressure* are in evidence.

As stated previously, six of the Dirty Dozen represent the deficit of a desirable characteristic (e.g., teamwork), and the other six represent a surplus of an undesirable feature (e.g., fatigue). The Dirty Dozen totals for both the MERs and MOSA reports were color coded for deficit (orange) and surplus (green), rearranged in descending order, and compared in Figure 6. A visual inspection of the two graphs shows the prevalence of *Lack of Resources* across both reports, followed by *Complacency* and *Norms*.

Two notable issues became apparent in the examination of the MOSA reports. First, the *Lack of Resources* category was used so frequently by the raters that its total for the 60 MOSA cases was 3148. To put this into perspective, the next highest value was *Complacency* at 930, hence the use of a LOG_{10} algorithm to keep the other categories meaningful on the graph. Second, while *Lack of Communication* was the number one precondition revealed in the MER reports, it ranked tenth (of thirteen) on the MOSA reports. Being reactive in nature, the MER reports document events that have actually come to fruition, unlike the MOSA reports which speculate to a large degree what forces are at work during an observation by a third party and are heavily influenced by the specific items on the MOSA checklist. Therefore, for the purposes of the current research, the MERs are considered more directly grounded in reality and more representative of actual circumstances in the maintenance department. If this is indeed the case, it prompts the questions – why is *Lack of Communication* ranked so low on the MOSA analysis? and, why was *Lack of Resources* disproportionately high in the MOSA analysis?

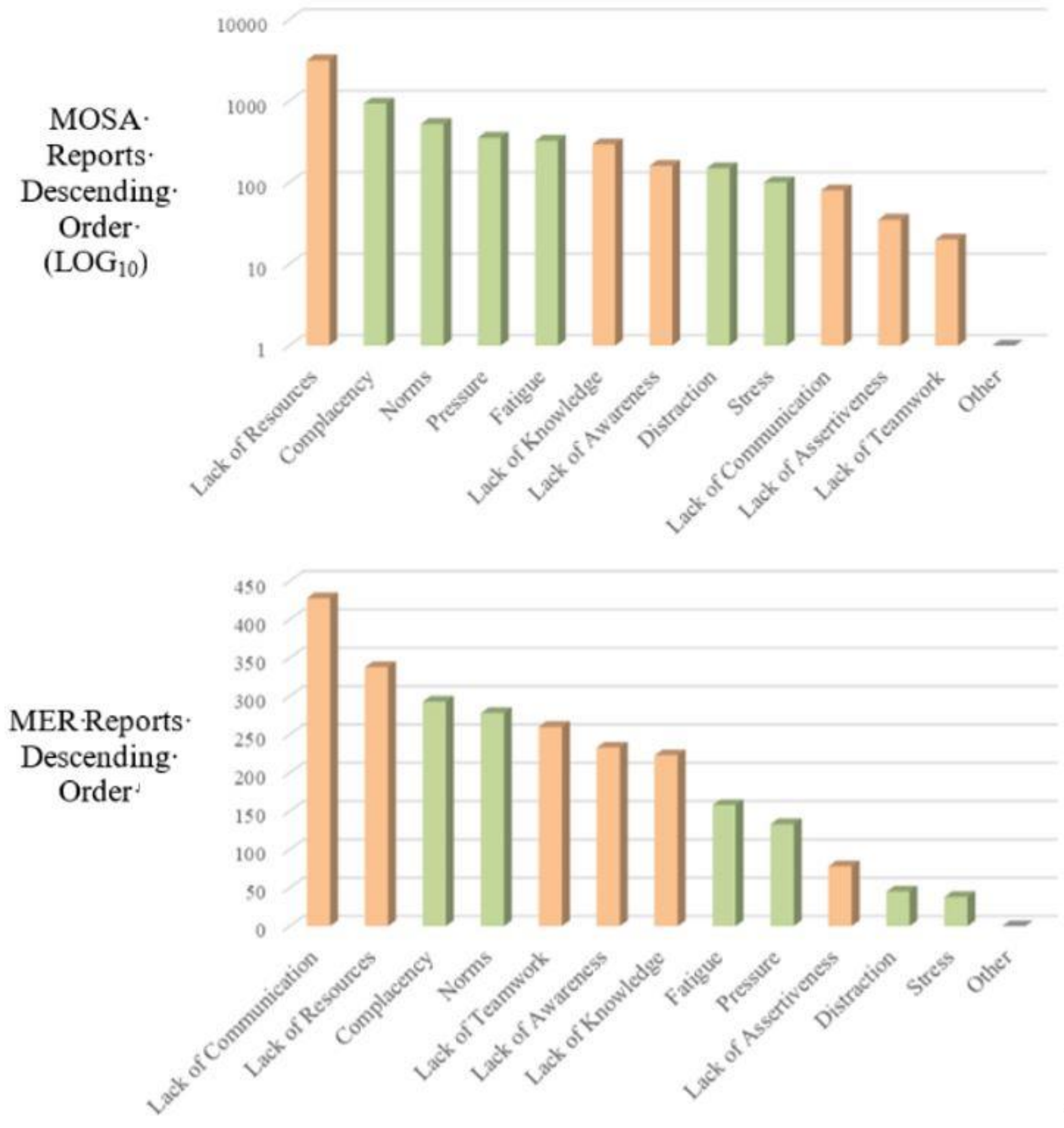

Figure 6. MOSA and MER reports arranged in descending order.

A possible answer to the first question can be found in the MOSA form itself. The last section of the MOSA form deals exclusively with communication. It asks the observer to assess six types of communication: communication between departments, between shifts, among technicians, between technicians and supervisors, technicians and inspectors, and between supervision and management. The observer is instructed to check one of three boxes next to each of the six types of communication labeled 'Yes', 'No', or 'N/A'. *Lack of Communication* ranked low on the MOSA analysis since, more often than not, the 'Yes' box was checked for all six types of communication. However, even if it is assumed that communication is observed at all six levels, which seems unlikely given the MER analysis, there is no mechanism on the MOSA form for an observer to indicate the effectiveness of said communication. Thus, it can be seen that three possibilities exist to explain the disparity between the MER and MOSA report analyses: first, communication is not, in fact, being observed at all six levels; second, communication is being observed at all six levels, but the effectiveness of the communication is often poor; or third, some combination of these two possibilities. Since few matters involving human behavior are purely binary, odds favor the third possibility as the more likely explanation. Therefore, the way in which ‘communication’ is handled in terms of both the construct of the MOSA form as well as training of the observers should be examined further by the PA.

An answer to the *Lack of Resources* question can also be found in the MOSA form itself. For maintenance personnel, the term *resources* is broadly defined. Maintenance manuals, tools, materials, parts, consumables, and more comprise a mechanic’s resources. With this in mind, an examination of the MOSA form shows that many of the headings (orange bars) contain several questions that can fairly be said to reflect resources; notably, 18 of the 60 questions (30%) on the form relate to resources in some way. Since no other precondition for maintenance error is so well represented, this sets up any analysis of the MOSA reports to be more sensitive to resources in general, and therefore creates a certain degree of bias in the results. However, given that *Lack of Resources* was coded by the raters more than the next highest category (complacency) by a ratio greater than 3:1, it seems likely that *Lack of Resources* would still rank very high in the MOSA analysis even if the bias were somehow accounted for.

Prevalence of Dirty Dozen Categories. Coding and subsequent analyses of the MERs and the MOSA reports showed the presence of all twelve Dirty Dozen preconditions for maintenance error to one degree or another. It also demonstrated that while some Dirty Dozen categories were revealed equally by both types of reports, MERs were more sensitive to some categories and MOSA to others. *Lack of Resources* ranked second most frequent in the MER analysis and the most frequent in the MOSA analysis. Notably, *Lack of Resources* was disproportionately high in the MOSA analysis for reasons already discussed. While the collective analyses of these proactive and reactive reports suggest a notable lack of resources, that should not be construed to mean the PA is knowingly under-resourcing the maintenance department. In a recent article for *Director of Maintenance* magazine, Gordon Dupont described traits exhibited consistently by maintenance personnel, including “doesn’t like to ask for help, tends to be self-sufficient, tends to think things through on their own and not share thoughts too frequently or thoroughly” (Dupont, 2018, p. 14). Dupont goes on to say that, because of these and other traits, mechanics often do not ask for the resources they need. Therefore, it is entirely possible the PA is unaware of much of the under-resourcing experienced in the maintenance department.

Being reactive in nature, the MERs document events that have already occurred and thus are considered somewhat more reliable than their proactive (MOSA) counterpart. As such, it is difficult to ignore the most frequent Dirty Dozen category found in the MER analysis. In first place, *Lack of Communication* ranked approximately 25% higher than the second-place category (*Lack of Resources*). However, the MOSA analysis did not confirm a lack of communication was present. Given the issue in documenting the quality of communication present in the MOSA forms, a distinct issue with communication could well exist, but would be difficult to detect given these limitations.

Norms and *Complacency* were the last categories prevalent in the top of the MER and MOSA analyses. It seems worth noting that *Norms* was the fourth most prevalent category in the MER analysis and third in the MOSA, while *Complacency* ranked third in the MER analysis and second in the MOSA analysis. This suggests both *Norms* and *Complacency* as preconditions for maintenance error are present and active in the PA's maintenance department as well.

Study Limitations

Although the findings are promising, conclusions drawn from these results are limited by three notable issues identified in this study. First, the data derived from the two reports were collected during different time frames, which could introduce the possibility of events occurring that might have influenced one of the reports. Ideally, the data should be collected during the exact same timeframe. A second constraint was the limited number of reports provided by the PA. A larger data set would enable a more robust evaluation of the prevalence of the Dirty Dozen categories. Finally, this study focused on reports provided by one specific airline, limiting the generalizability of the findings to other airlines, that is, each airline has a unique maintenance climate, influenced by a broad range of organizational and ethnographic variables. Nevertheless, the methodology employed in this study would be applicable to other airlines.

Recommendations for Practice

Given the findings above, the basic concept of using the Dirty Dozen as a diagnostic tool for maintenance organizations seems to have merit. Although, more work needs to be done in terms of coordinating these two differing views of a maintenance organization and maintaining better control over the data source and other noted variables. Since the MER and MOSA reports seem to have a sensitivity to certain Dirty Dozen categories, how would the use of more controlled data affect the MER - MOSA relationship? It seems intuitive that more data collected (25 MERs was a rather modest quantity) and data gathered from identical timeframes might well impact this relationship. To this end, more research should be directed.

The PA should investigate further the suggested *Lack of Resources* that seems to be present in regards to maintenance. While the original data presented a variety of challenges, the combined results indicating a lack of resources is particularly compelling and warrants further investigation to develop a mitigation strategy. Although not as strong as the evidence supporting *Lack of Resources*, a case can be made for the presence of *Lack of Communication*, *Norms*, and *Complacency* as well. Therefore, a mitigation strategy for these preconditions should also be examined.

The majority of MOSA forms had checked boxes indicating that communication between various personnel was occuring. However, if the communication being observed does not relate to the task at hand, or if it is not interpreted correctly or not received at all by the recipient, this tends to confound the performance of tasks such as noted in the MER analysis. Since the MER analysis suggested *Lack of Communication* was prevalent in events that had come to fruition, it would be worthwhile for the PA to revisit this section of the MOSA form to see how it can be improved.

Recommendations for Future Research

The apparent relationship of proactive reports (MOSA) to reactive reports (MER) in terms of their sensitivity to certain preconditions for maintenance error is intriguing and lends itself to a host of additional questions. For example, are the results found here typical, or do they tend to vary from one organization to another based on variables not considered in this study? If these results are typical, could reactive and proactive reports or their supporting documentation be improved in a manner that enhances their sensitivity to certain Dirty Dozen categories?

Conclusion

The Dirty Dozen are widely accepted to be the 12 most common preconditions for maintenance error in the aviation industry. The assumption being that if preconditions for maintenance error are found to exist, the errors themselves are likely not far behind. The reports documented the PA's maintenance activities from two points-of-view: reactive (MERs) and proactive (MOSA). It was posited that a detailed examination of these maintenance-related reports through the framework of the Dirty Dozen would illustrate and highlight these preconditions, thus helping the PA understand where best to allocate resources to reduce these preconditons, thereby reducing the chance for errors to come to fruition. This line of research could potentially enhance aviation safety through a better understanding of human error in aviation maintenance and other related areas such as ground operations and ramp operations.

Acknowledgements

The views herein are those of the authors and do not necessarily reflect those of the organizations with which the authors are affiliated. The research reported in this paper is based upon the doctoral dissertation of Greg M. Mellema, Embry-Riddle Aeronautical University.

References

Boeing (2013). *Maintenance error decision aid (MEDA)© user's guide*. Maintenance Human Factors, Boeing Commercial Aviation Services. Retrieved from https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/media/MEDA_Users_Guide_Updated_09-25-13.pdf

Chang, Y., & Wang, Y. (2010). Significant human risk factors in aircraft maintenance technicians. *Safety Science, 48* (1), 54-62. Retrieved from https://doi.org/10.1016/j.ssci.2009.05.004Civil Aviation Safety Authority (CASA) (2013). *Safety behaviours: Human factors for engineers.* Canberra ACT, Australia: Civil Aviation Safety Authority. www.casa.gov.au.

Civil Aviation Safety Authority (CASA) (2013). *Safety behaviours: Human factors for engineers.* Canberra ACT, Australia: Civil Aviation Safety Authority. www.casa.gov.au.

Dorn, M. D. (1996). Effects of maintenance human factors in maintenance-related aircraft accidents. *Transportation Research Record*, 1517, 17- 28.

Dupont, G. (1997). The Dirty Dozen errors in aviation maintenance. In meeting proceedings of *11th Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection: Human error in aviation maintenance* (pp. 45-49). Washington, D.C.: Federal Aviation Administration/Office of Aviation Medicine.

Dupont, G. (June, 2018). I'm only an AME. *Director of Maintenance (DOM)*, 12-16.

Edwards, E. (1988). Introductory Overview. In E. L. Wiener & D. C. Nagel (Eds.), *Human factors in aviation* (pp. 3-25). San Diego, CA: Academic Press.

Federal Aviation Administration (FAA). (n.d.). *Avoid the dirty dozen: 12 common causes of human factors errors*. https://www.faasafety.gov/files/gslac/library/documents/2012/Nov/71574/DirtyDozenWeb3.pdf

Federal Aviation Administration (FAA) (2006). *Advisory Circular: Line Operations Safety Audits* (AC No: 120-90): Federal Aviation Administration.

Federal Aviation Administration (FAA) (2014). *Operator's manual: Human factors in aviation maintenance*. Washington D.C.: U.S. Government printing office.

Federal Aviation Administration (FAA). (2018). *Aviation maintenance technician handbook – general*. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aircraft/media/amt_general_handbook.pdf

Gramopadhye, A. K., Drury, C. G. (2000). Human factors in aviation maintenance: How we got to where we are, *International Journal of Industrial Ergonomics*, *26*(2), 125-131, ISSN 0169-8141. Retrieved from http://dx.doi.org/10.1016/S0169-8141(99)00062-1.

Hayes, A. F. & Krippendorff, K. (2007) Answering the call for a standard reliability measure for coding data, *Communication Methods and Measures*, 1:1, 77-89. DOI:10.1080/19312450709336664

Hobbs, A. (2008). *An overview of human factors in aviation maintenance: Aviation research and analysis report – AR-2008-055*. Canberra, Australia: Australian Transport Safety Bureau.

International Air Transportation Association (IATA) (2008). *Safety report 2008.* http://www.iata.org/docx/IATA-Safety-Report-2008.pdf

Langer, M., & Braithwaite, G. R. (2016). The development and deployment of a Maintenance Operations Safety Survey. *Human Factors*, *58*(7), 986-1006. http://doi.org/10.1177/0018720816656085

Ma, M. J., & Grower, J. (2016). From "dirty dozen" to "filthy fifteen" -- Professionalism in aircraft maintenance. *Aviation MX Human Factors Quarterly*, *4*(2), 7-9. https://www.faa.gov/about/initiatives/maintenance _hf/fatigue/publications/media/december_2016_newsletter.pdf

Marais, K. B., & Robichaud, M. R. (2012). Analysis of trends in aviation maintenance risk: An empirical approach. *Reliability Engineering & System Safety*, *106*, 104-118. DOI:10.1016/j.ress.2012.06.003

Maurino, D., & Seminar, C. A. S. (2005, April). Threat and error management (TEM). *Canadian Aviation Safety Seminar (CASS)*. Vancouver, BC, 18-20 April 2005.

Marx, D. A., & Graeber, R. C. (1994). Human error in aircraft maintenance. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation psychology in practice* (pp. 87-104). Aldershot, UK: Avebury.

Merritt, A. C., & Klinect, J. R. (2006). *Defensive flying for pilots: An introduction to threat and error management*. University of Texas Human Factors Research Project, The LOSA Collaborative.

Patankar, M. S., & Taylor, J. C. (2008). MRM training, evaluation, and safety management. *International Journal of Aviation Psychology*, *18*(1), 61-71. DOI:10.1080/10508410701749449

Reason, J. (1990). *Human error*. New York, NY: Cambridge University Press.

Schmidt, J., Lawson, D., & Figlock, R. (2001). *Human factors analysis and classification system – maintenance extension (HFACS-ME).* Department of the Navy. Prepared for the U.S. Department of Transportation, Federal Aviation Administration.

Shappell, S. A., & Wiegmann, D. A. (2000). *The human factors analysis and classification system--HFACS* (No. DOT/FAA/AM-00/7). U.S. Federal Aviation Administration, Office of Aviation Medicine

Zimmermann, N. & Mendonca, F.A.C. (2021). The impact of human factors and maintenance documentation on aviation safety: An analysis of 15 years of accident data through the PEAR framework. *Collegiate Aviation Review International*, 39(2), 1-25. Retrieved from https://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8218/7624