University at Buffalo The State University of New York

Language Errors in Aviation Maintenance: Year 1 Interim Report

Colin G. Drury and Jiao Ma
University at Buffalo, State University of New York
Department of Industrial Engineering
342 Bell Hall
Buffalo, NY 14260

July 1, 2003

Prepared for the
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405
Dr. William K. Krebs
Research Grant #2002-G-025

Table of Contents

List of Tables and Figures

Abstract

The existence of maintenance and inspection personnel whose native language is not English suggests that language barriers may be causing performance errors. This project is intended to find out whether such errors exist, what patterns characterize these errors, what are their contributing factors and how effectively we can mitigate these errors. Any language errors would be communication errors by definition, so first we reviewed models of communication to search for characteristic error patterns. We identified two primary communication types relevant to aviation maintenance: synchronous communications (largely verbal and informal) and asynchronous communication (largely written and formal). We then analyzed several errors databases (e.g. ASRS) and found that both the contributing factors and the use of recovery mechanisms were different for the two error types. Next, we analyzed survey data from 113 aircraft operators, covering their English speaking/reading abilities and use of mitigation strategies. There were significant differences across four world regions in the incidence of these two sets of factors. Neither of these data sources emphasized maintenance, so to discover more refined patterns of error, contributing factors and mitigation strategies, we conducted a series of focus groups at maintenance organizations. The patterns grouped, as expected, into synchronous and asynchronous. We developed classified lists of contributing mitigation factors, which will be used in subsequent stages to quantify error incidences and test the effectiveness of mitigation strategies.

In addition, we collected data from 15 focus group participants as a pilot test of workcard comprehension with and without Simplified English. No significant results were found for the use of Simplified English in this pilot test, but the data will provide a useful native English-speaking sample for comparison with international data from Year 2.

1.0 INTRODUCTION

Outsourcing is a preferred corporate strategy for reducing nonessential costs and focusing an organization on its core business (Cant and Jeynes, 1998). In aviation maintenance, outsourcing has been advocated and widely used, as it avoids tying up capital in maintenance facilities, and can reduce costs by opening the airline's maintenance operation to outside competition. One potential impact of such outsourcing is that there are more interfaces within the system, each of which represents an opportunity for error. The "system" without outsourcing includes the aircraft itself, the airline and the regulatory agency (e.g. the FAA). However, with outsourcing, a fourth organization is added to the system: the Maintenance/ Repair Organization (MRO). Drury, Wenner and Kritkausky (2000) provided models of these interactions and examined potential and actual error sources from using MROs. Data collection at a number of domestic and foreign MROs did indeed show a potential for increased errors, but little evidence of errors in practice.

Sparaco (2002) sees the formation of global MRO networks involving US and foreign airlines, as well as repair stations. In addition to offshore MROs, there are many within the USA where non-native English speakers form part of the labor pool. The difficulty of moving between languages creates an additional potential for error. The language of aviation is primarily English, both in operations and in maintenance. Aviation Maintenance Technicians (AMTs) must pass their examinations in English, and maintenance documentation in use at the Federal Aviation Administration (FAA) approved facilities is in English. This poses a second-language or translation burden for Non-Native English Speakers (NNESs) that can potentially increase their workload, their performance time or their error rate, or even all three measures.

In a 2001 report to the Secretary of Transportation by the Aircraft Repair and Maintenance Advisory Committee, many of these issues were raised in considering changes to the domestic and foreign FAR Part 145. They recommended that:

"The FAA should establish a method for determining whether language barriers result in maintenance deficiencies."

This project is a direct response to these concerns that NNES, in repair stations in the USA and abroad, may be prone to an increased error rate that could potentially affect airworthiness.

In the first year of this project we used existing data sources and site visits to provide material for our detailed data collection efforts in subsequent years. First, we reviewed communication models in general, and aviation-related communication in particular, to provide a framework for the study. Next, we analyzed several existing databases to find patterns of language-related errors and potential causal factors. This was supplemented by analysis of a current survey of language practices in civil aviation by a major company. Finally, our site visits used focus groups to directly elicit examples and patterns of language error in maintenance. These form the basis for extensive on-site survey questions in subsequent years. While at each work site, we pilot tested our methodology for our workcard comprehension studies that will test the efficacy of various interventions.

This report presents these analyses, and detailed plans for our next phase of the project.

2.0 COMMUNICATION MODELS

2.1 Communication and Aviation Safety

Communication is defined as a dynamic and irreversible process by which we engage and interpret messages within a given situation or context, and it reveals the dynamic nature of relationships and organizations (Rifkind, 1996a&b). In the context of aviation maintenance and inspection, communication has been the most frequent aspect studied since the human factors movement began in the early 1990's. Taylor and Patankar (2001) provide a historical perspective of the time since early human factors programs, showing that interpersonal communication was a major emphasis, and that training in improving communications skills was seen as the essence of applying Human Factors to aviation maintenance. In this report, we will review the literature on communications, and in particular on communications in an aviation context, to show that it is indeed an important aspect of ensuring flight safely. We will look at more general communications models as a background for an analysis of communications errors from a number of existing databases.

Communication can be formal, i.e. written documents, or informal. Most on-the-job communication is informal, unwritten, and sometimes even unspoken. Davidmann (1998) made a distinction between formal and informal communication, where formal communication implies that a record is kept of what has been said or written, so that it can be attributed to its originator. On the whole, written communications are formal. Oral (spoken) communication consists of direct or transmitted speech between two or more people. Oral communications are more likely to be misinterpreted than written ones, a nd were originally regarded as informal, but are now often recorded and treated as formal. The defining characteristic of many formal oral communications, such as selection, grievance or appraisal interviews, or negotiation, is that those participants keep a record, and hence provide an audit trail.

Formal communication within the aviation maintenance domain is defined and regulated. A hierarchy of written correspondence is defined in the Federal Aviation Regulations (FARs), which includes airworthiness directives (ADs), notices to airmen (NOTAMs), maintenance manuals, work cards, and other types of information that are routinely passed among manufacturers, regulators, and maintenance organizations. The international aviation maintenance community adopted a restricted and highly structured subset of the English language to improve written communication, such as ATA-100 and AECMA Simplified English. However, verbal communication among aircrews and air and ground controllers has significant safety implications. Communication is based on the use of language. In order to eliminate or at least minimize potential ambiguities and other variances, people establish rules regarding which words, phrases, or other elements will be used for communication, their meaning, and the way they will be connected with one another. The aggregation of these rules is known as a “protocol”. There are four types of protocol related to flight and aircraft safety (Rifkind, 1996a&b): verbal protocols, written protocols, graphical protocols, and gestural protocols. Verbal protocols have been used for many years, primarily in two-way radio communication. A number of aviation accidents have been caused by the failure to use established verbal protocols. Verbal protocols are not generally seen as applicable to aviation maintenance, although establishing verbal protocols can reduce ambiguity and uncertainty in critical maintenance tasks such as ground movement and shift turnover. According to Rifkind (1996a&b), the only verbal protocol that has been established throughout aviation, including maintenance, is the use of English as the standard language. This was done when the International Civil Aviation Organization (ICAO) was established in 1944.

About 70% of the first 28,000 reports made to NASA’s ASRS were found to be related to communication problems (Sexton and Helmreich, 1999; Connell, 1995). The importance of communication in aviation cannot be overemphasized. A full-mission simulation study conducted on pilots discovered that crew performance was more closely associated with the quality of crew communication than with the technical proficiency of individual pilots or increased physiological arousal as a result of higher environmental workload (Smith, 1979; quoted in Sexton and Helmreich, 1999). Based on examination of accident investigations and incident reports, Orasanu, Davision and Fischer (1997) summarized how ineffective communication can compromise aviation safety in three basic ways:

1. Wrong information may be used.
2. Situation awareness may be lost.
3. Participants may fail to build a shared model of the present situation at a team level.

Along with the increasing volume of international traffic, the risk of communication errors escalates even further because of participants’ culture and native language difference (Orasanu, Davision and Fischer, 1997).

Although aviation communication is extremely important to air safety, Kanki and Smith (2001) pointed out that “besides some acronyms and jargon, the essence of aviation communication is not exceedingly unique; it encompasses all of the nuances, subtleties, and complexities of human interaction.”

After analyzing a set of reports submitted to the Aviation Safety Reporting System (ASRS) and to the International Air Transport Association (IATA) on communication problems encountered by pilots flying in foreign airspace, previous studies (Orasanu, Davision and Fischer, 1997; Cushing, 1994) categorized communication failures as shown in Table 1.

Besides types of communication failures, Orasanu, Davision and Fischer (1997) also proposed levels of miscommunication:

1. A message may not get through due to transmission problems.
2. When transmission is adequate but the message is misunderstood.

The message may be accurately transmitted and understood, but may not adequate to convey the speaker’s intent.

Several different approaches may be applied to reduce these three types of failure. Transmission problems are most amenable to prevention through use of technology, such as data link or electronic transmission of text message. For reducing comprehension errors, standardized vocabulary and phraseology have been designed to eliminate problems associated with unfamiliar terms, local jargon, or ambiguous phrases. Communication failures are more likely to occur in non-routine circumstances, when non-standard language is being used. Everyday speech patterns, which may differ enormously across cultures and be exacerbated by language barriers, open the door to misunderstanding. Speakers are recommend to use their knowledge of the addressee, the situation, and social norms to formulate what they believe will be an effective message that elicits the desired response from the addressee rather than rely on assumption.

Language Category	ASRS	IATA
Language/Accent	47	5
Partial or Improper Readback	24	8
Dual Language Switching	23	2
Unfamiliar Terminology	17	4
Speech Acts	9	0
False Assumptions or Inference	7	23
Homophony	5	1
Unclear Hand-off	4	3
Repetition across Languages	3	2
Uncertain Addressee	1	13
Lexical Inference		0
Lexical Confusion (speed/heading/runway/altitude)		4
Mistakes (unexplained)		3
Total	152	68

Table 1. Categorization of communications errors (Orasanu, Davision and Fischer, 1997)

The communication concept is two-fold: communication as a tool, and communication as a skill (Kanki and Smith, 2001). The fundamental function of communication as the skill is to deliver a message from one human being to another. In almost every aspect of aviation work, communication also fulfills a secondary role as an enabler (or tool) that makes it possible to accomplish a piece of work.

Fegyveresi (1997) summarized many variables that influence communication, such as workload, fatigue, personality traits, gender bias, standard phraseology, experience level, and vocal cues, etc. An important part of aviation communication uses the radio, which eliminates some visual components (e.g., body language, lip reading) that people rely on in day-to-day communication.

Saffley (1984) stated that all poor communication involves human factors of one kind or another, and can be divided into two categories “stemming from people misusing language” and “stemming from people interacting”. Several things can go wrong when people use language:

1. The words and sentences we use are too difficult.

2. The words are so general and abstract that they mean one thing to us but something entirely different to someone else.
3. The language sometimes has such an abrasive tone that audience reaction is negative.
4. Some other contributing reasons, such as long-windedness, ambiguity, poor grammar, incoherent expression and improper logic, etc.

Analysis of business and technical communication shows that the first three are the most frequently cited weakness (Saffley, 1984).

In previous research, the role of language use in communication processes has been relatively neglected; a deeper understanding of language, its basic characteristics, and how it works should be beneficial as we move towards an era of globalization of all aspects of aviation.

Language and cultural diversity can intensify differences and confusions in communication, but a language barrier does not necessarily result in unsafe operations. Merritt and Ratwatte (1997) conducted a study to compare safety performance between mono- versus multi-cultural cockpits. They found that although language barriers and cultural differences are inhibiting the open communication and team fellowship, multi cultural crews, especially crewmembers with English as a second language had to concisely verbalize their intent and requirements and perform “by the book”. This led to rule-based behavior, with a high degree of Standard Operation Procedures (SOPs) being used. In addition, greater reliance on crew resource management principles, such as more precise communication and more crosschecking, also support the assertion that mix-cultural cockpits may actually be safer. Although English is the official language of aviation and its practice should be mandated, language training should be intensified and standardized for the non-native speakers of English. Instead of being arbitrarily granted the linguistic advantage, native English speakers should be taught how to communicate simply, slowly and precisely with their non-native English speaking colleagues.

In the ASRS database, verbal information transfer problems account for roughly 85% of reported information transfer incidents (Nagel, 1988). Matthews and Hahn (1987) identified four major contributing factors to voice communication errors in the ATC environment:

1. Quality of the Very High Frequency (VHF) radios
2. Phraseology
3. Fatigue
4. Workload

Solutions to verbal communication errors generally fall into one to two categories: those that transfer some or all of the voice communication to another communication medium (e.g., Datalink), and those that attempt to eliminate some of the current volume of voice communication (e.g., Mode S transponder, TCAS).

2.2 Communication Principles and Models

Many models have been proposed by psychologist, linguists, and engineers to study communication in the 20th century. Generally, they fall into three categories:

1. Mechanical models
2. Psychological models
3. Integrationist models of communication

Based on basic communication theories, a communication process is composed of the sender/receiver (e.g., people, manuals, procedures, instruments, computers, etc.), the message (e.g., information, facts, emotions, feelings, questions, etc.), the medium (e.g., speech, text, video, audio, sensory, etc.), filters and barriers, feedback, and so on (Kanki and Smith, 2001; Griffith, 1999).

Kanki and Smith (2001) state that human communication always takes place with a set of contexts, such as a social context, a physical context and an operational context. The social context refers whether the receiver appropriately understands the message intended by the speaker, beyond merely using the correct words and grammar. The physical context for communication refers to aspects of the location of the communication event: co-located and speaking face-to-face, or remotely located and speaking via interphone or radio. Compared to some other working settings, the aviation operational context is relatively structured by flight phase and standard operating procedures that organize task performance.

Operational aviation communications are unique in several ways as summarized by Kanki and Smith (2001):

1. Most aviation communication is confined to small audience.
2. It is usually time-sensitive and expeditious.
3. It is constrained or limited in some way by the physical environment.
4. Circumstantial factors (noise, static, vibration, weather, etc.) are combined with barriers (cockpit doors, workstations, distances, etc.) to limit, restrict, and confound the channels used in everyday communication.

In studying communication, we are naturally interested in communication errors. Nagel (1988) categorized methods of studying errors into four categories:

1. Direct observation (which can yield a wealth of information concerning the type, frequency, and causes of errors in airline operations in a natural setting)
2. Accident data and post accident analysis, such as NTSB data base
3. Self report
4. Error studies conducted in laboratory and in simulators

2.3 Use of Languages other than English in Aviation

Language is an important element in effective and competent communication. Language usage is known to be a problem in cross-cultural communication (Rifkind, 1996a&b). As the whole of aviation, including maintenance, takes on an increasingly global dimension, we need to understand the issues involved in cross-language communication. First, we must understand the demographics of globalization in maintenance. One driver in the move towards offshore outsourcing of aviation maintenance and inspection has been the relative wage rates in various countries. The US Bureau of Labor Statistics (BLS) has relevant data in the index of hourly compensation costs. They publish overall country data on 29 countries in North America, Asia/Oceania and Europe (e.g. 2000 data) and less comprehensive data for SIC codes 372 and 376: aircraft, space vehicles and parts manufacturing (e.g. 1994 data), see Table 2.

Country	Year 2000 Overall Index	Year 1994 Aircraft, etc. Index
USA	100	100
Canada	73	81
Taiwan	49	30
France	90	83
Germany	119	121
Italy	58	74
UK	63	80

Table 2. Relative wage indices for selected countries, overall and for aviation

Many other countries have no aviation data (e.g. Mexico) but do have low compensation indexes (e.g. 12). The conclusion from these statistics is that most countries of the world have lower compensation cost. In Europe the costs are comparable to the USA or even higher, but in Asia and Latin America labor costs are considerably less.

A second useful demographic comes from the US Census data of 2000, which counts the language abilities of non-native English (NNESs) speakers who are residents of the USA. Of all US households, 13.8% speak a language other than English at home. Of these 51.6% speak Spanish with the next most common language being Chinese. There is also data on the individuals' facility with English, and the number of households where there are no English speakers. This data will be used in our project as a basis for estimating NNESs in US employment, particularly to compare with NNESs in Part 145 operators. Although international agreements have designated a particular form of English as the standard for written communication in the aviation maintenance workplace (Rifkind, 1996a&b), the fraction of the available labor force, inside and outside the USA, who speak English as their primary language will decrease slowly.

To speak and to understand a language it is not sufficient to know the words and the grammar. Bilingualism consists in the capacity of an individual to express himself in another language and to adhere faithfully to the concepts and structures of that language rather than paraphrasing his native language (Connolly, 2002).

In addition to language difference there are also variation of accent and dialect within a language. The core difference between accents and dialects is that accents indicate characteristics of speech variations in pronunciation, whereas dialects indicate language differences as well as speech differences. Accent and dialects need not be international to be considered foreign (Fallon, 1997; Hulit and Howard, 1993).

Willingness to communicate (WTC) is an emerging concept to account for individuals' first language and second language communication. Yashima (2002) studied English usage as a second language in a Japanese population and found that several factors affect WTC using English, such as general attitudes toward English, motivation, and language anxiety concerning achievement/proficiency. The model proposed in the study fits the data well, which indicates the potential for using the WTC and other constructs to account for second language communication.

Previous research has revealed that gender differences influence language behavior in vocabulary, intonation and sentence structure. Turney (1997) recognized gender bias (i.e. pitch differences, volume, and or social expectations) as a factor in controller/pilot communication in a survey study.

3.0 ANALYSIS OF INCIDENT DATABASES

Before field data is collected on language-related maintenance and inspection errors, existing databases need to be searched for relevant reports of such errors. There are three sources of potential data available.

1. Aviation Safety Report System (ASRS)

Besides the United States, some other countries such as the United Kingdom and Australia operate aviation incident reporting systems too. In the United States, the primary reporting system is the Aviation Safety Reporting System (ASRS), which was developed and operated by NASA for the Federal Aviation Administration. The ASRS has more than 60,000 reports contributed by pilots, controllers, flight attendants, ground crews and others.

According to Nagel (1988), the ASRS is an excellent resource to study errors in aviation operation. First of all, data from the ASRS have proven to be a practical and indispensable source of information for the operational community and the scientist alike. For example, in some cases, modifications to the Federal Aviation Regulations (FARs) have resulted from ASRS data and analyses. Secondly, incidents of the kind and type that are reported to the ASRS are representative of those circumstances that underlie accidents. Thirdly, as an incident reporting system, the ASRS was designed to have one major advantage relative to accident analysis database, because it is possible to query the incident reporter prior to report de-identification and it is possible to learn more about why errors are made as well as something of the circumstances in which they are made. Finally, the voluntary reporting feature of the ASRS is a drawback as well as strength. The reports are not contributed on a purely random basis, for example safety conscious people may report more often than others. In practice, ASRS reports are mainly from flight crew, although maintenance is included.

2. National Transportation Safety Board (NTSB)

The National Transportation Safety Board (NTSB) has the overall responsibility to review the facts that surround major civil aviation (and other transportation system) accidents and to issue a format finding of causality. The electronic database provides complete reports and findings on all recent NTSB investigations. A search revealed no relevant reports when searched for "language" "English" or "communications".

3. Accident/Incident Data System (AIDS)

The Accident/Incident Data System (AIDS) database contains data records for general aviation and commercial air carrier incidents since 1978. The NASDAC database for AIDS contains incidents only because NASDAC uses the National Transportation Safety Board (NTSB) accident database as the primary source for accident information. The information contained in AIDS is gathered from several sources including incident reports on FAA Form 8020-5.

The Aviation Systems Data Branch, AFS620 is the custodian of record for the FAA Accident/Incident Data System (AIDS), which contains records of aircraft accidents and incidents occurring in the United States, and those involving U.S. registered aircraft if out of the United States. The definition of an aircraft accident is an occurrence associated with the operation of an aircraft that takes place between the times any person boards an aircraft with the intention of flight until all such persons have disembarked, and in which any person suffers death or serious injury, or in which the aircraft receives substantial damage. The definition of an incident is an occurrence other than an accident, associated with the operation of an aircraft, which affects or could affect the safety of operations.

3.1 Analysis Methods

Our main interest was in maintenance and inspection errors, but few were reported in the databases studied. Hence, the objective changed to include all language-related errors, whether by flight crew, ATC, cabin crew or ground crew. This decision was in line with our literature search, which we broadened to include all communication errors. With a large enough set of aviation-related language errors, we can form more general models, of which maintenance and inspection errors will be a specific instance.

Based on a preliminary reading of about 60 incident reports, a taxonomy was developed of error manifestations, causal factors and recovery mechanisms. Some entries in this taxonomy reflect the earlier analysis by Orasanu, Davision and Fischer (1997), although we have tried to separate contributing factors from recovery mechanisms. This preliminary reading also found likely key words. Two keyword searches were made of the ASRS and AIDS databases. The first was on "English" and the second on "Language". Some uses of these words were colloquial and specific, for example, passengers using abusive "language" to cabin crew. These usages have been removed from our analysis. There remained 684 incidents that were classified as shown in Tables 3, 4 and 5. Note that outcomes were not analyzed, although we did classify them, as our interest was in the causation of errors rather than the full error propagation.

The main division of error types was between synchronous communication (real time, person to person) and asynchronous (person to document). This is a standard classification of communication systems. Within these, a relatively fine classification was made by the roles of the two communicators, e.g. flight crew with ground crew. As will be seen later, this classification was eventually collapsed into four categories. Contributing factors are those noted in the reports. They do not represent the results of detailed fault tree analysis on human factors investigation, and so are biased towards factors seen as contributing by participants reporting the incidents. Note that "language" was used to refer to both of the first two items. Language could mean the actual language used (e.g. French, Spanish, Chinese, English) on the choice of words/phrases (e.g.

expected one term but communicator used what was thought to be a synonym). Some of the communication channels themselves were poor, classified here as low signal/noise ratio. In many cases, the report mentioned that at least one of the communicators was inexperienced, for example a crew's first flight for some years into a Mexican airport.

Synchrony	Error Type	# of Reports out of 684
1. Synchronous (person to person in real time)	1.1 Flight crew/ATC miscommunication	465
	1.2 Wrong/miscommunicated action by other traffic	41
	1.3 Unable to communicate	22
	1.4 Miscommunication on flight deck	61
	1.5 Miscommunication with audio FMS (?)	12
	1.6 Miscommunication between flight deck/cabin crew	4
	1.7 Miscommunication between ground crew and operations	5
	1.8 Miscommunication between flight deck and ground crew	41
	1.9 Miscommunication with passengers	7
2.Asynchronous (person to document	2.1 Wording unclear in documentation	25
	2.2 Incorrect wording on placards	1

Table 3. Initial Classification of Error Types

Contributing Factor	# of Reports out of 684
1. Communicators not using native language (includes use of foreign language, difficulty understanding accent, unclear pronunciation)	105
2. Unclear terminology/wording	169
3. Low Signal/noise ratio on communications channel	130
4. Experience/inexperience of communicators	121

Table 4. Contributing Factors

Recovery Mechanism	# of Reports out of 684
1. No recovery attempted	340
2. Readback to other communicator	175
3. Repeated message	193
4. Ask for clarification	133

Table 5. Recovery Attempts

There are specific flight crew/ATC measures to assist in maintaining error-free communications. For example, the flight crew is expected to read the information back to the controller to confirm its accuracy. Other recovery mechanisms include repeating the message verbatim, and asking for clarification in different words. In many cases no recovery mechanism was reported.

Finally, it was found that a number of reports contained more than one error. For example, the flight crew communicated with ATC about another aircraft but this aircraft did not behave as expected. When the database was expanded to include these multiple errors, the total number of errors rose to 725 from the 684 original reports.

3.2 Results

The analysis of the database used a cross-tabulation technique developed by Wenner and Drury (2000) for drawing conclusions from an aviation accident database. The aim was to relate the contributing factors to the incident types to reach one of two conclusions:

1. No significant differences in frequency of each contributing factor across incident types using Chi-square test. The conclusion is that the factor is equally important across all types.

2. A significant difference (Chi-square, $P < 0.05$) showing that some combinations of error types and contributing factors are over-represented or under-represented. The actual factors over or under-represented are determined from the standardized residuals in each cell of the contingency table. Any standardized residual greater than 1.96 is significant at $p \leq 0.05$. These significant cells lead to a focusing of countermeasures by error type.

The first analysis cross tabulated the 11 error types with the contributing factors and the four recovery mechanisms using the classification of error types in Table 3. Because of small cell frequencies for some errors (e.g. error type 1.6 had 4 cases, error type 2.2 had 1), the Chi-square tests were unreliable. Hence, a decision was made to combine logical categories by the locus of the communication error. This produced four error locus categories as shown in Table 6. Examples of the raw ASRS narratives typical of each are reproduced in Appendix 2. These give an indication of both the detail and the contractions typical of ASRS reports. They also help illustrate the multi-causal nature of most incidents.

Synchrony	Error Locus	Error Types from Table 1	# of Reports out of 684
1. Synchronous	Traffic-related	1.1, 1.2, 1.3	528
2. Synchronous	Intra-cockpit related	1.4, 1.5	73
3. Synchronous	Other Groups (ground crew, cabin crew, operations)	1.6, 1.7, 1.8, 1.9	57
4. Asynchronous	Written Communications	2.1, 2.2	26

Table 6. Final Error Classifications

The second analysis was performed to determine whether the separation of multiple errors in reports produced different patterns of analysis when moving from N = 684 to N = 725. Tables 7 and 8 show these analyses for contributing factors and recovery mechanisms, respectively. As can be seen, there were no difference in pattern and only minor differences in significance level between the two databases. Hence, all further work used the expended database of N = 725 where multiple errors per report were permitted. The overall pattern of percentages of contributing factors by error locus is shown in Figure 1. Similarly, Figure 2 shows the overall patterns by recovery attempts.

	N = 684			**N = 725**		
	Signif	**Over**	**Under**	**Signif**	**Over**	**Under**
Native language	P=0.061		(Asynch)	P=0.076		(Asynch)
Language/ terminology	P<0.001	Asynch		P<0.001	Asynch	
Low S/N Ratio	P<0.001		Asynch Other Gps	P<0.001		Asynch Other Gps
Inexperience	P<0.001	Cockpit		P<0.001	Cockpit	Traffic

Table 7. Pattern of Significance from Chi-Square Tests of Contributing Factors

	N = 684			**N = 725**		
	Signif	**Over**	**Under**	**Signif**	**Over**	**Under**
No Recovery	P<0.001	Asynch Other Gps		P<0.001	Asynch Other Gps	
Readback	P<0.001	Traffic	Cockpit Other Gps	P<0.001	Traffic	Cockpit Other Gps
Repeat	P<0.001		Asynch Other Gps	P=0.002		Asynch Other Gps
Ask Clarification	P=0.491			P=0.514		

Table 8. Pattern of Significance from Chi-Square tests of Recovery Attempts

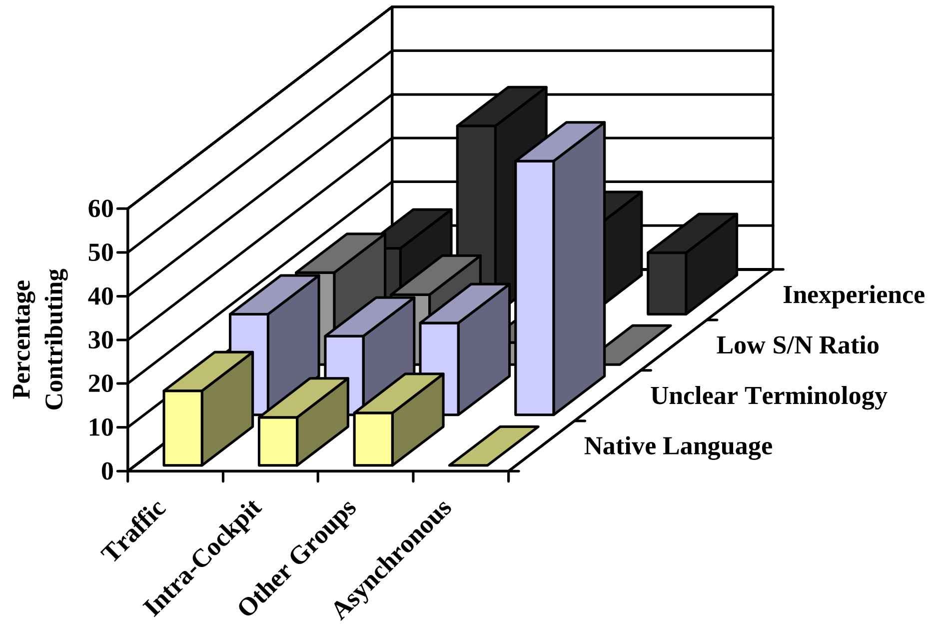


Figure 1. Pattern of Contributing Factors across Error Locii

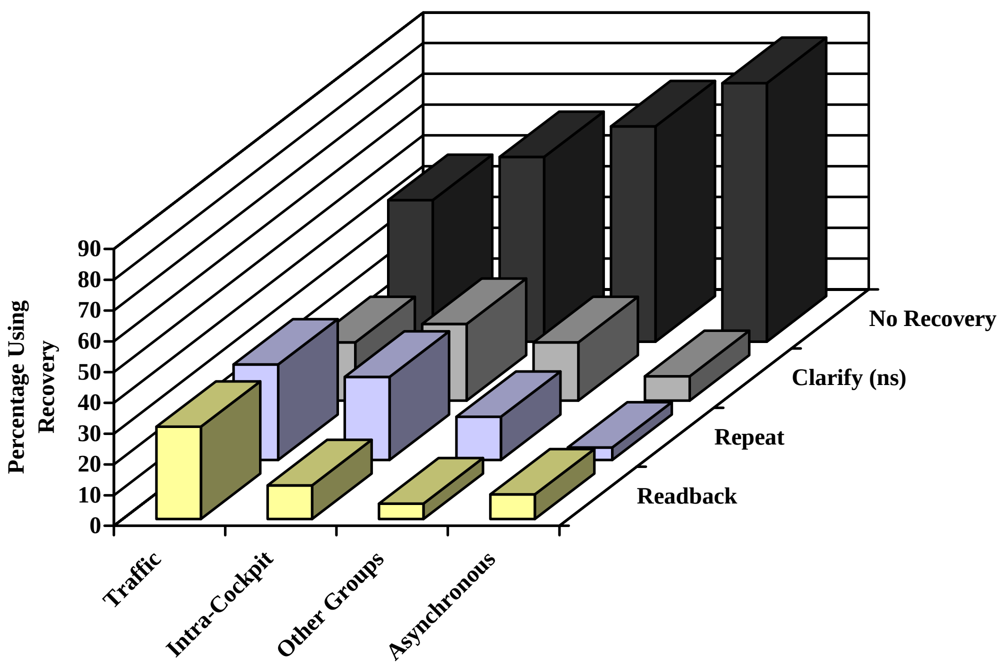


Figure 2. Pattern of Recovery Attempts across Error Locii

The first finding was that communication using native language was not significantly different by error locus ($\chi^2(3) = 6.87$, $p = 0.076$), although the standardized residual for asynchronous errors (-1.98) showed that this was significantly under-represented. In fact, for written communications, much less language difficulty would be expected, as the communication does not take place in real time. When the Asynchronous data was removed from the analysis, there was a much reduced Chi-square (χ^2 (2) = 2.09, p = 0.361) and none of the standardized residuals reached significance. The conclusion is that, apart from Asynchronous communication, difficulties with native languages are equally common for all error locii. This appears to be a general factor.

Unclear terminology showed significant effect of error locus (χ^2 (3) = 18.2, $p < 0.001$) with the asynchronous error type being significantly over-represented. For Asynchronous errors, 58% had unclear terminology as a contributing factor, compared with 23% for all other error locii. Thus, while unclear terminology was a relatively large contributor to all communications errors (23%), over twice that rate was found for communications with documents. As many have noted in aviation (Kanki and Walters, 1997; Drury, 1998) written communications need to be better designed for human use, particularly in terms of layout, wording and standardization. These ASRS findings re-emphasize the same point.

Low signal/noise ratio would logically not be expected to be an issue in Asynchronous communications, and indeed the significant Chi-square (χ^2 (3) = 15.7, $p < 0.001$) showed Asynchronous significantly under-represented with zero errors. However, Other Groups were also under-represented compared with the remaining error categories when the analysis was repeated without Asynchronous communications. This result still held (χ^2 (2) = 9.1, p = 0.010), with the standardized residual for Other Groups was –2.44. As with terminology errors, low signal/noise ratio had a relatively high incidence (21%), but for Other Groups this was only 5%. Re-reading the relevant reports showed that communications with Other Groups are often face-to-face, so this low error incidence is expected.

The contributing factor of inexperience was significantly different for the error locii (χ^2 (3) = 38.7 $p < 0.001$), with intra-cockpit errors over-represented (43%) and traffic errors under-represented (14%), compared to an overall rate of 17%. The major contribution to inexperience in the cockpit was trainee pilots. As ASRS includes general aviation and training flights, this is to be expected.

Recovery from the initial error was not attempted on 50% of occasions. Multiple recovery strategies were used at times, with rates of:

Readback	25%
Repeat Message	28%
Ask for clarification	19%

Note: For all four analyses (no recovery plus three recovery records) analyses were also undertaken with Asynchronous communications removed, and all showed the same pattern.

Where there was no recovery recorded, the over-represented error locii were Asynchronous and Other Groups, with rates of 84% and 70%, respectively. Asynchronous communication was expected to have a relatively high non-recovery rate as there are few strategies available except re-reading. Similarly, for Other Groups, less recovery strategies were available than for flight crew/ATC communications.

Readback to the other communicator was largely confined to traffic communications, where 93% of all instances were found. Thus, the other categories are relatively under-represented, averaging only 8% use of this strategy compared to 30% for traffic related communications. These differences show the power of a standard and well-practiced strategy: those trained often use it for error recovery while those who have had different training do not.

The recovery strategy of repeating a communication message verbatim was commonly used, particularly for traffic communications where 86% of all usage occurred (χ^2 (3) = 15.2, $p < 0.002$). Results were very similar to the readback strategy. For traffic, 31% used the repeat strategy, but it was only used on 14% of reports by Other Groups and only 4% for Asynchronous. Again, the results reflect the database itself, which is mainly reported by flight crew and mainly traffic-related.

Asking for clarification was statistically evenly distributed across error types (χ^2 (3) = 2.3, $p = 0.514$). This recovery strategy, like the causal factor of different languages, appears to apply to all communications errors represented in the database, with no differentiation between different types of communication.

3.3 Conclusions on Databases

From the literature on communications, particularly in aviation, we have been able to classify the communications process in context. This has led to listings of error types, difficulties and contextual factors potentially affecting communication performance. First, communications was shown to be an important aspect of human and system performance in all aspects of aviation, from maintenance to flight operations. It has been emphasized in training programs for cockpit crews via CRM training programs, and for maintenance via MRM programs. These began as close relatives of each other, but have gradually diverged, without losing their communications emphasis.

More general communications models list the tasks to be performed, attributes of the personnel communicating, and possible error pathways. For our purposes, we have been most concerned with the causation of error and potential recovery actions, rather than with relating error antecedents to outcome severity.

Analysis of the ARSR, NTSB and AIMS databases showed significant and often interesting conclusions. When the error locus was classified by the roles of the communicators, differences in contributing factors and recovery mechanisms were seen.

Our four categories of causal factors gave roughly equal counts in the databases, showing that the use of other than a native language was an important causal factor in these errors. This contributing factor appeared to be distributed across error loci, except for asynchronous (largely written) communication, where it was under represented. In fact, for asynchronous communication as a whole, native language and low signal/noise ratio were under represented factors while unclear terminology was over represented. For recovery, asynchronous had the least opportunity for recovery mechanisms, in particular the repetition so useful in synchronous communications was not usually fruitful.

Inexperience was cited as a contributing factor for many of the incidents, but primarily for traffic-related errors. Readback of the message was used mainly by flight crew for traffic-related errors. Communications with other groups, such as ground crew, had few instances of recovery.

From such patterns, the potential errors in maintenance environments can be seen more clearly. Although ASRS has few reports from this field. The characteristics of maintenance communications errors found here (asynchronous, terminology-related, few recovery mechanisms) helps set the stage for our direct measurement of these errors from maintenance participant interviews and questionnaires.

The analysis of the databases available was useful in putting language errors into context, but necessarily contains the known limitations of the databases themselves. The raw data consisted of self-reports, largely by flight crew, with some facility for further questioning (ASRS) but largely reflecting the thought and feelings of those on the flight deck. Thus their relevance to maintenance and inspection was indirect, although they did afford the opportunity to access a wide range of language related incidents.

A final quote on language is worth repeating here (Turney, 1997 quoting Brightman, 1988:

> "In order to transmit proper meaning, the encoder and decoder must be on the same wavelength. They must speak the same language. We do not hear

with our ears, we hear with our minds. And we are different from one another. All of us suffer from selective perception. What we hear depends on who we are." (Turney, 1997; Brightman, 1988).

4.0 ANALYSIS OF LANGUAGE ISSUES SURVEY DATA

From September 2002 to January 2003, a major US corporation surveyed a large number of airlines throughout the world concerning their use of English and other languages in flight operations and maintenance operations. The University at Buffalo team were given access to this data as part of their project on Language-related Errors in Aviation Maintenance and Inspection. This paper provides the University at Buffalo team's analysis of the maintenance related issues in the company-supplied database, interpreting these issues in terms of previous research on language errors. Please note that thissurvey was not designed to assist in the University at Buffalo team's effort, and that the team is presenting here only a part of the overall data.

4.1 The Survey Database

At this time, there are 113 airlines in the database. The survey was about 10 pages long, and was designed to help with the production of a new Quick Reference Handbook (QRH). The survey was confidential so that University at Buffalo team did not have access to individual responses, but worked from a summary of the data in an Excel file. The first two questions asked for the number of pilots in the company and the number of daily flights to help determine the influence of airline size on language issues. The name of the company (not in the Excel database) was used to classify the company's region of operations into:

Original Region	Number Responding
Africa	6
Asia	30
Central America	5
North America	16
South America	4
Europe	35
Russia and Eastern Europe	7
Oceania	4
Middle East	6

As we wished to perform statistical analyses testing for differences between regions, we needed to have enough responses in each region, so we combined the smaller regions as follows, based on different challenges in language they are a priori likely to face:

Region	Number Responding
Europe	35
North America	16
Asia	30
Other	32

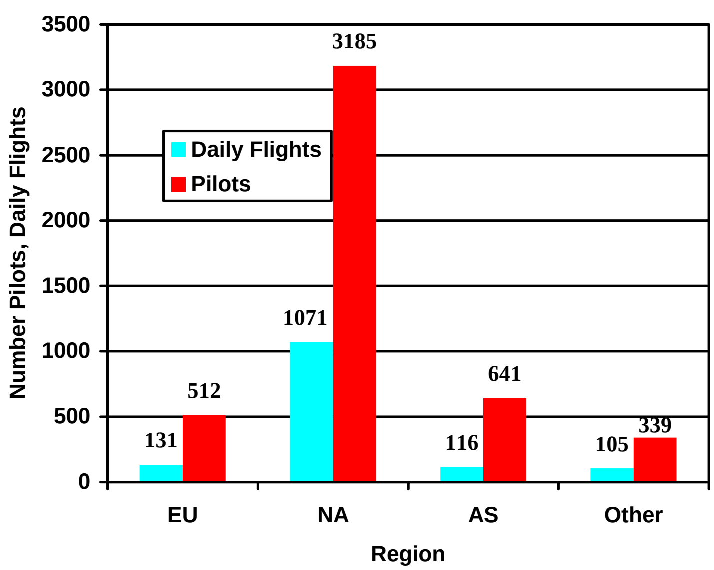
The questions analyzed in this paper were all completed by Maintenance personnel except for airline size statistics and QRH issues which were completed by Flight Operations personnel. The analysis began with determining whether size statistics were useful predictors of answers to other questions of interest. Size was different between the four Regions, and there was a high correlation (0.864) between the two measures of airline size. The size differences were tested by Analysis of Variance (ANOVA) giving the following results:

Measure	$F(3, 100)$	p
Number of Pilots	13.64	< 0.001
Number of Daily Flights	21.87	< 0.001

(Note that in our analyses, the degrees of freedom in the various tests may not match exactly as airlines occasionally did not complete all the responses.)

Correlations between size measures and other responses were low, and when used as covariates in ANOVAs, size measures proved insignificant. Additionally, as can be seen in Figure 3, average size was much larger for North America, where English is the majority national language, so that any size effects are likely to be confounded with language use. Thus we decided not to use size further in our analyses. However, in the eventual development of strategies for language intervention (e.g. changing QRH format or translation), size may indeed be a key determinant of the likelihood of strategies succeeding.

Figure 3. Average Size differences between Regions



4.2 Results

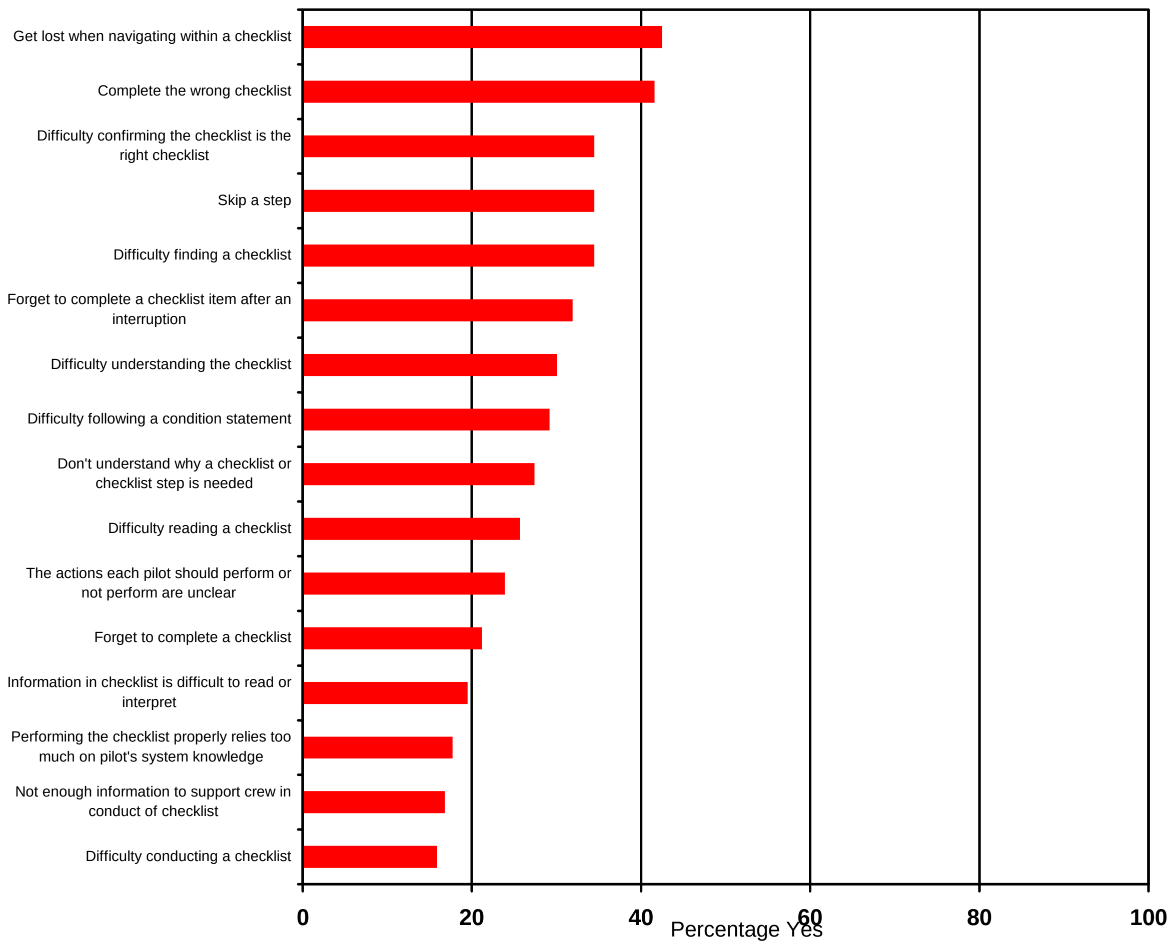
The first analysis was for the question on "Has your company identified any of the following issues with using the checklists in the QRH? Check all that apply." This question (Q12) was the only one that addressed the potential errors in checklist use, so that even though it specifically referenced Flight Operations, it was deemed relevant to any study of patterns of error in communication tasks. Each response type (e.g. Difficulty finding a checklist) was tested with a Chi-square test of equality of proportions across the four Regions. Table 9 summarizes the results:

Q #	QRH Issue	Overall % Yes	Chi-Square	Significance
1	Difficulty finding a checklist	38	10.7	0.014
2	Difficulty reading a checklist	28	13.7	0.003
3	Difficulty understanding the checklist	33	4.9	Ns
4	Don't understand why a checklist or checklist step is needed	30	4.7	Ns
5	Difficulty conducting a checklist	17	2.7	Ns
6	Difficulty following a condition statement	32	6.1	Ns
7	Get lost when navigating within a checklist	46	8.2	0.042
8	Forget to complete a checklist item after an interruption	35	1.6	Ns
9	Skip a step	38	4.7	Ns
10	Forget to complete a checklist	23	3.0	Ns
11	Complete the wrong checklist	45	1.0	Ns
12	Difficulty confirming the checklist is the right checklist	38	3.3	Ns
13	Performing the checklist properly relies too much on pilot's system knowledge	19	3.4	Ns
14	The actions each pilot should perform or not perform are unclear	26	1.4	Ns
15	Not enough information to support crew in conduct of checklist	18	2.2	Ns
16	Information in checklist is difficult to read or interpret	21	11.8	0.008

Table 9. Responses tested with a Chi-square test of equality of proportions across the four Regions

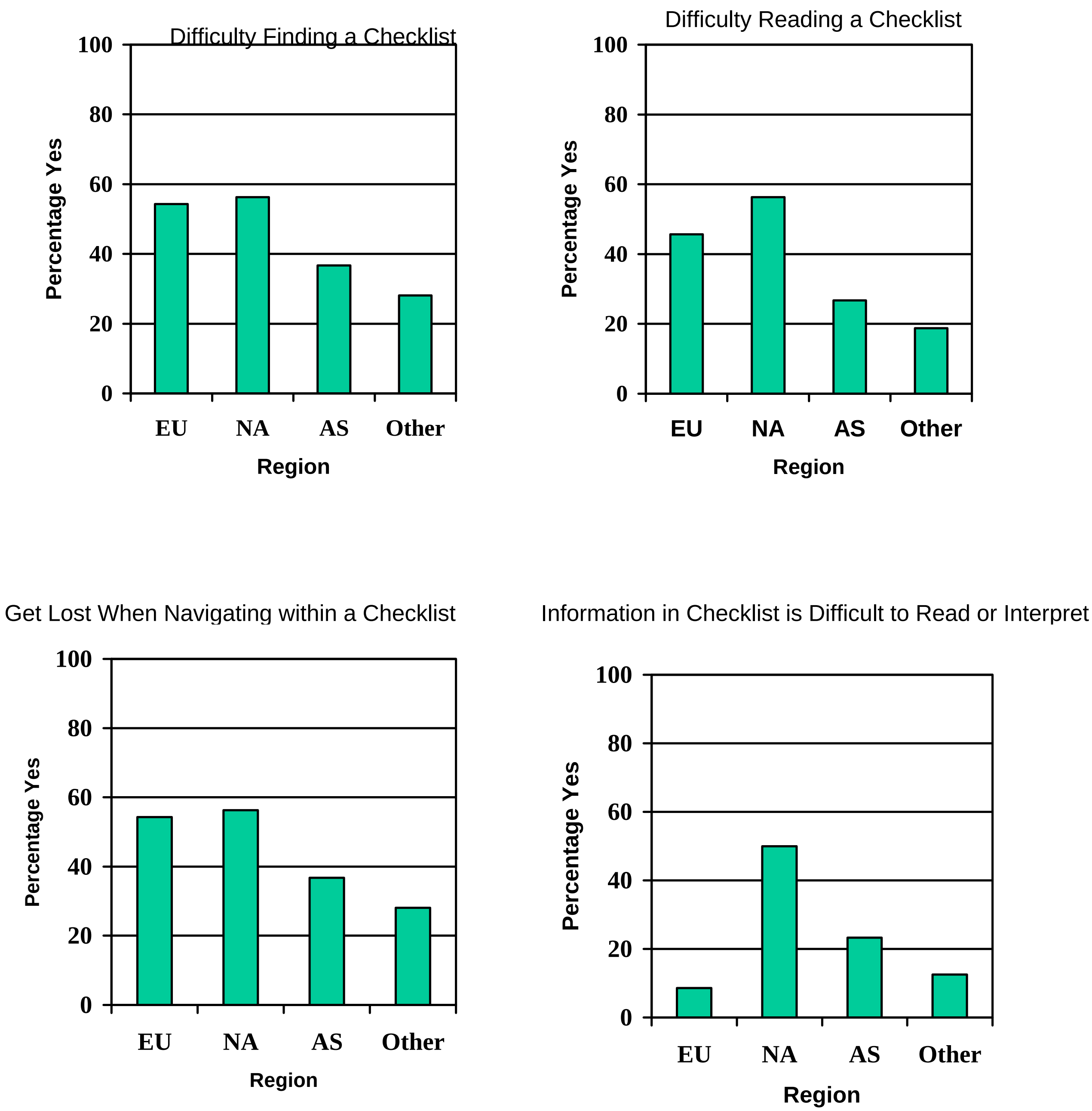
Note first that there was a considerable difference between the sixteen error patterns, with some having over twice the frequency of others. In fact, when the sixteen are ranked, we get the pattern shown in Figure 4. As can be seen, the higher frequency issues are to do with physically locating the correct checklist and performing it despite interruptions. In contrast, the lower frequency items have to do with the design, formatting and wording of the checklist itself. Clearly, the airlines in this sample reported operational rather than design difficulties with checklists.

Figure 4. Ranking of QRH Issues



For the four issues where there was a significant difference between regions, Figure 5 shows these differences. For the first three issues (finding, reading, and navigating) Europe and North America reported the most instances, with Asia and Other progressively less. There should be no regional differences in at least finding a checklist, so that differential willingness to report may be an issue here. Similarly, the final graph of Figure 5 on difficulty of reading and interpreting information on the checklist, has North America showing almost twice as many responses as any other region. Again, one would expect *less* difficulty understanding checklist information in a predominantly

Figure 5. Significant Differences between Regions of Checklist Usage



English-speaking population, so perhaps the North American users are less tolerant of sub-optimal conditions than those regions where an implicit translation from English to a native language is often required. This may also explain to some extent the findings of

Figure 4, that checklist design problems are less likely to be reported than operational problems.

Our next analysis was of the reported English language ability of mechanics. Question 6 asked "Estimate the percentage of your mechanics who are described by each of the following levels of English speaking ability". Four levels of ability were presented:

- Can speak or understand very little English
- Can speak or understand a few English words or phrases
- Can speak and understand English for simple conversations
- Can speak and understand English in long, complex conversations

Similarly, Question 7 asked "Estimate the percentage of your mechanics who are described by each of the following levels of English reading ability". Three levels of ability were presented:

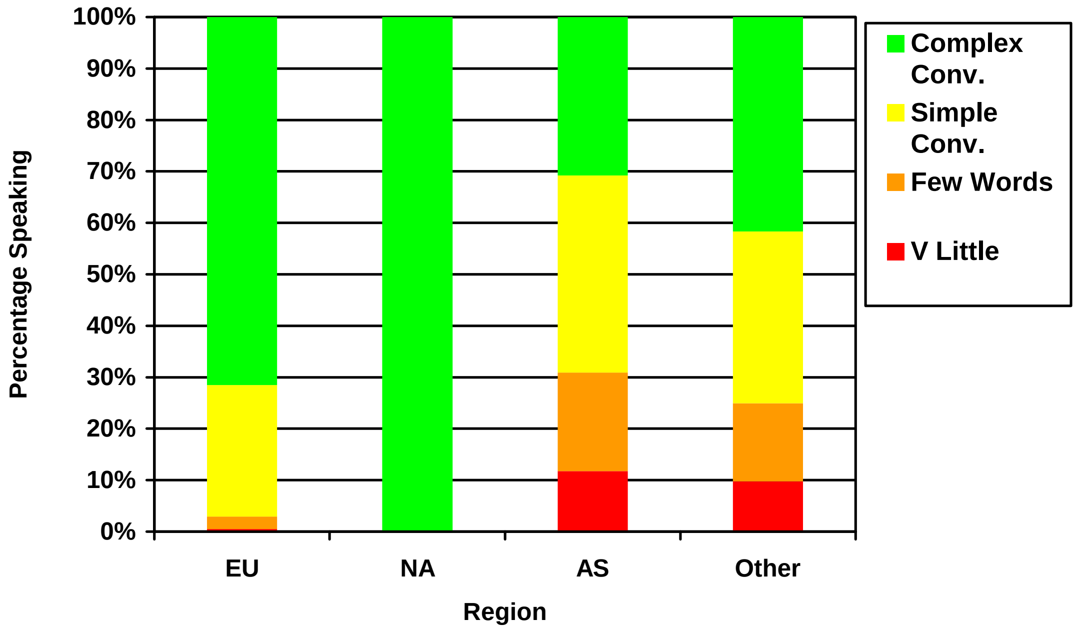
- Can read very little English
- Can read English for maintenance documents
- Can read any English document

The data from each level was analyzed separately using one-way ANOVAs between the four Regions. For Speaking ability (Q6) all levels showed significant differences between regions:

English Speaking Level	**F(3,84)**	**p**
Can speak or understand very little English	4.53	0.005
Can speak or understand a few English words or phrases	10.12	<0.001
Can speak and understand English for simple conversations	7.06	<0.001
Can speak and understand English in long, complex conversations	16.73	<0.001

Figure 6 shows graphically how the speaking ability varied between Regions. Note the contrast between Europe and North America, where most of the mechanics speak English, and Asia and Other, where there is less reported speaking ability.

Figure 6. English speaking ability reported by Region

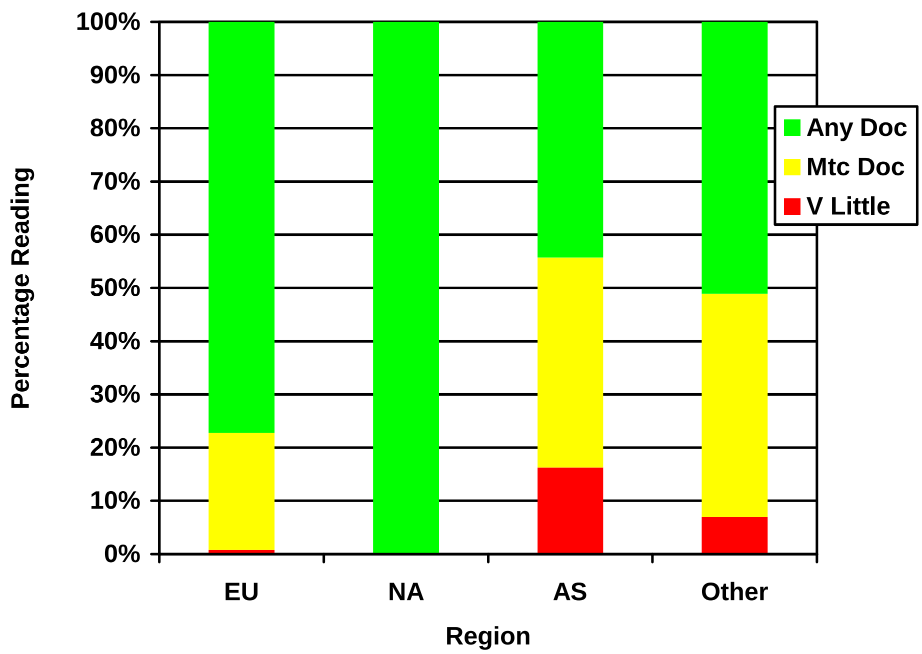


Exactly similar analyses were performed for English reading ability (Q7). The data from each level was analyzed separately using one-way ANOVAs between the three Regions. For Speaking ability (Q7) all levels again showed significant differences between regions:

English Reading Level	F(2, 82)	p
Can read very little English	7.08	<0.001
Can read English for maintenance documents	7.42	<0.001
Can read any English document	10.81	<0.001

These differences are shown in Figure 7 with common scales to provide comparisons. Note again that there is a difference in English abilities between the two groups [Europe, North America] and [Asia, Other].

Figure 7. English reading ability reported by Region



The next analysis was for the techniques used in each airline to deal with consequences of any differences between the English language of documents and the native language of mechanics. Questions 1A, 1B and 1C asked about translation of four types of document: Maintenance Manual, Maintenance Task Cards, and Structural Repair Manual respectively. The wording of Q1 was: “For each document listed here, indicate whether you use the manufacturer’s document as delivered or translate it to a different language.” First, data from Q1A to Q1C were analyzed using a Chi-square test with three responses:

- manufacturer’s document; in English and not modified

- manufacturer's document; modified or rewritten in English by your company
- Translated to a different language

Where there were too few responses other than the first, the Chi-square was invalid and so data were recoded to give just two levels:

- manufacturer's document; in English and not modified
- Any modification

For the Maintenance Manual, only two of the 88 airlines reported other than the "manufacturer's document; in English and not modified", with one reporting each modification. Clearly, the Maintenance Manual is typically left as supplied. For Q1B and Q1C, there was a significant difference between Regions when the data were combined to two levels as above:

Q #	Document	Overall % Modified	Chi-Square	Significance
Q1B	Maintenance Task Cards	13.6	26.1	< 0.001
Q1C	Structural Repair Manual	4.6	12.6	Too few data

In both cases the main difference was that only Asia and Other made modifications, for the Maintenance Task Cards 9 airlines modified them in English while three (2 in Asia, 1 in Other) translated them into the native language. Note that for Asia, 33% of the Task Cards were modified. For the Structural Repair Manual, Asia was the only Region to modify the document, with 2 performing each modification for a total of 13% modifying the Boeing originals in some way.

Questions 2-5 all dealt with languages used in common situations. They were as follows:

Q2. What language is used when an Engineering Order is created (e.g. from a Service Bulletin)?

Q3. What language is used for on-site maintenance training?

Q4. What language is used for meetings (for example stand-up meetings) with maintenance mechanics?

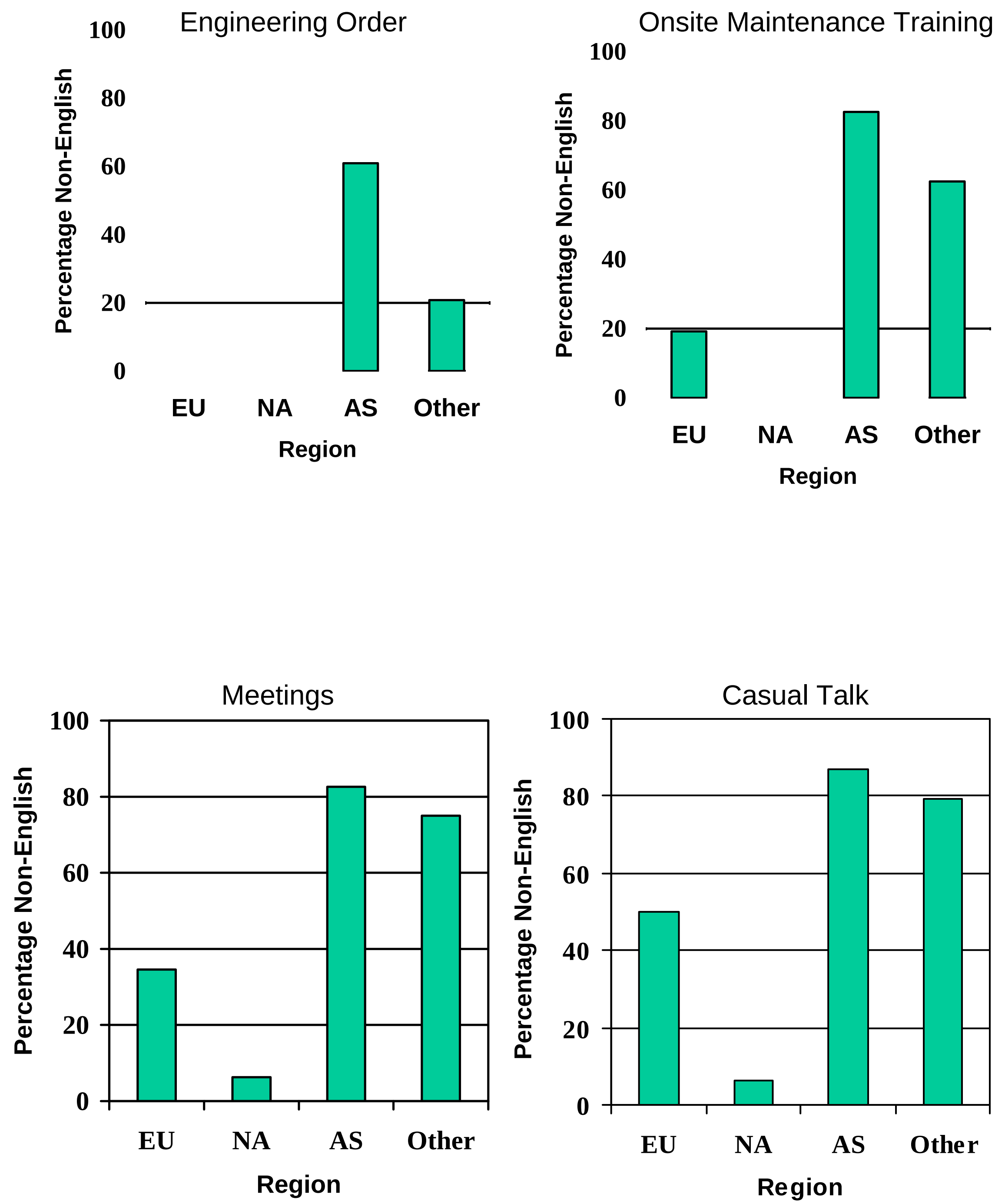
Q5. When mechanics meet casually to talk with each other, what language do they use?

Again, the languages specified were English and a language other than English, but a mixture was also allowed if respondents checked both alternatives. The following results were obtained from Chi-square tests, with either 3 categories or recoded into two categories as for Question 1:

Q #	Situation (number of categories)	Overall % Non English	Chi-Square	Significance
Q2	Engineering Order (2)	21.6	34.5	< 0.001
Q3	On-site maintenance training (3)	43.2	36.2	< 0.001
Q4	Meetings (3)	52.3	30.3	< 0.001
Q5	Casual talking (2)	59.1	30.1	< 0.001

Note that there were many uses of languages other than English, particularly for verbal communication. This is in contrast to Q1 where there were few translations of documents. It is more in line with Figures 6 and 7 in terms of English language abilities. The Region differences are shown in Figure 8, where the four graphs are again combined to show similarities. As expected, North American airlines show the least use of languages other than English, with only a single airline showing a mix of English and other languages. Europe also does not use languages other than English even half the time, presumably because of the widespread use of English in the European Union countries, as well as one whole country speaking English as the primary language. However, Asia and Other regions make considerable use of languages other than English in meetings and casual talking between mechanics, with over 79% using this strategy. Asia does translate Engineering Orders most of the time, but Other regions make less than 20% use of this strategy.

Figure 8. Regional difference of English usage



We conducted a final analysis to further investigate the relationship between an airlines' actual English ability (reading and speaking) and its strategies of handling the manufacturer's English documents and oral conversation in daily maintenance practice. For actual Reading English ability, we calculated as following from the answers to Q14.2.1-Q14.2.3:

Step 1: We took the relevant levels of Reading Ability for maintenance and calculated their expected value with 1=1, 2=2, and 3=3. This gave mean English Reading Ability for each airline, ranging from 1.0 if 100% read very little English to 3.0 if 100% can read any English documents.

English Reading Ability	Ability Value
1. Can read very little English	1
2. Can read English for maintenance documents	2
3. Can read any English document	3

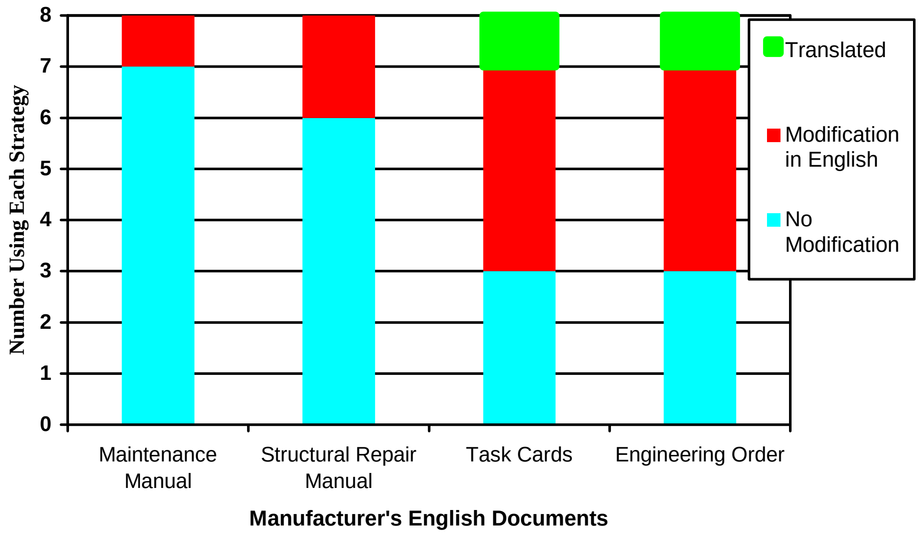
Step 2: We then categorized the Mean Reading Level into four categories as follows:

Mean Reading Level	Mean Reading Category
1.00-1.49	1. Very Low
1.50-1.99	2. Low
2.00-2.49	3. Moderate
2.50-3.00	4. High

Note: Levels 1-2 are low (worse than average) Reading English levels and Levels 3-4 are high (better than average) levels.

We expected those airlines with low level of Reading English ability would adopt some mitigating strategies in using the manufacturer's documents (i.e. modification into AECMA simplified English, translation into their native language). However, when using the Maintenance Manual, 7 out of 8 kept the original manufacturer's document in English without any modification or translation, while only one airline modified/rewrote it in English. When using the Structural Maintenance Manual, 6 out of 8 airlines did not make any modification or translation. Figure 9 demonstrates the details how these airlines deal with manufacturer's English documents.

Figure 9. The airlines with low level of English Reading ability used different strategies in handling manufacturer's English documents



In a similar way, we analyzed the relationship between an airlines' actual ability at Speaking English and its strategies of handling oral conversation in daily maintenance practice. For actual Speaking English ability, we calculated from the answers to Q14.2.4-Q14.2.7 as following:

Step 1: We took the relevant levels of Speaking Ability for maintenance and calculated their expected value with 1=1, 2=2, 3=3, and 4=4.

English Speaking Ability	Ability Value
1. Can speak or understand very little English	1
2. Can speak or understand a few English words or phrases	2
3. Can speak and understand English for simple conversations	3
4. Can speak and understand English in long, complex conversations	4

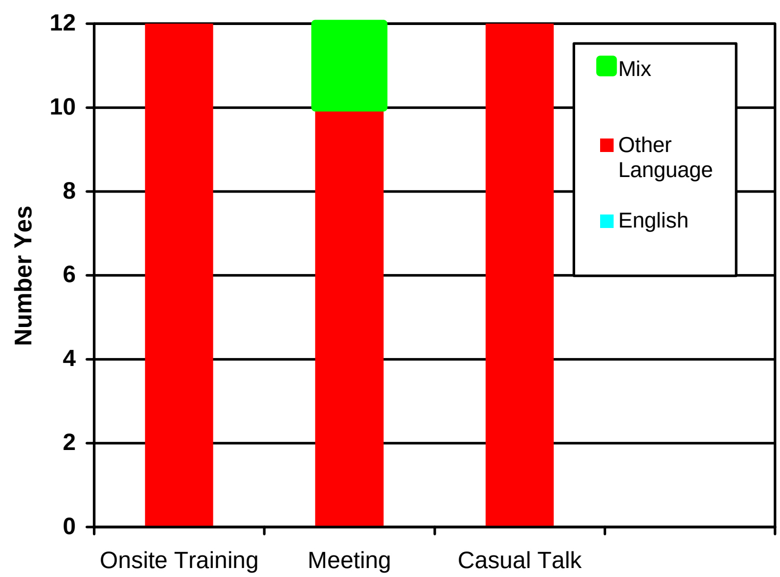
Step2: We then categorized value of Mean Speaking Ability into four levels:

Mean Speaking Ability	Mean Speaking Category
1.00-1.49	1. Very Low
1.50-1.99	2. Low
2.00-2.49	3. Low-moderate
2.50-2.99	4. Moderate-high
3.00-3.49	5. High
3.50-4.00	6. Very high

Note: Levels 1-3 are low Speaking English levels and Levels 4-6 are high levels.

For those airlines with low level of Speaking English ability (categories 1-3), when conducting Onsite Maintenance Training, 100% conducted the training in a language other than English (i.e. the native language). In Meetings, 10 out of 12 airlines used another language, with the remaining two used both English and another language. Again during Causal Talking, none of the airlines used English. Figure 10 demonstrates the details how they use different strategies in dealing with daily oral conversation.

Figure 10. The airlines with low level of English Speaking ability used different strategies in handling daily oral conversation



4.3 Discussion

This survey, devised, distributed and collected by the company, has been of considerable value in helping to focus the project on the differences between regions. There were significant differences in usage of English, the frequency of checklist problems, and the strategies to cope with a mismatch between the language of aviation and the native language used by airline personnel. The database used was based on a large sample ($n = 113$) of airlines, approximately evenly divided between North America, Europe, Asia and the rest of the world. We have largely restricted the analysis to the maintenance area, as that is the focus of our main project, but can provide similar analyses for Flight Operations data if required.

Overall Use of English: First, the data provides some evidence on the prevalence of English Language use by maintenance personnel in different regions of the world. Figures 6 and 7 show that English is spoken and read at a high level in North America, and to a large extent (75% or so) in Europe. Asia and the other countries have about 50% of users able to work with written English effectively, and about 30-40% able to work with spoken English in the same way.

Errors / Issues with Checklists: Potential problems of working in other than a native language were explored by means of the responses to issues of QRH checklist use, as this was the only document about which questions were asked on the survey. The problems reported most frequently, (30-40% of respondents) were mainly concerned with physical use of the checklists. Those reported least frequently (~20%) were typically issues of checklist design. In fact, when we classified issues as “Use” or “Design” related, the mean reporting percentage was almost statistically significant ($t = 12.57$, $p < 0.001$). This suggests that design considerations, while important and reported, are not perhaps as prevalent in operations as issues of use, e.g. being able to find the correct checklist.

QRH Use Related	QRH Design Related
12.5 Difficulty conducting a checklist	12.15 Not enough information to support crew in conduct of checklist
12.10 Forget to complete a checklist	12.13 Performing the checklist properly relies too much on pilot's system knowledge
12.4 Don't understand why a checklist or checklist step is needed	12.16 Information in checklist is difficult to read or interpret
12.8 Forget to complete a checklist item after an interruption	12.14 The actions each pilot should perform or not perform are unclear
12.1 Difficulty finding a checklist	12.2 Difficulty reading a checklist
12.9 Skip a step	12.6 Difficulty following a condition statement
12.12 Difficulty confirming the checklist is the right checklist	12.3 Difficulty understanding the checklist
12.11 Complete the wrong checklist	
12.7 Get lost when navigating within a checklist	

Airline Responses to Language Issues: The airlines cope with any potential problems by a number of means, including document translation, and conducting training and meetings in native languages. We have found that in Europe and North America, such strategies are little used, presumably because most mechanics speak English, even if that is not their native language. In contrast, Asia and the rest of the world make significant use of these strategies. Translation of documents is not a common strategy, except for Asia where 17% and 60% of airlines translated Task Cards and Engineering Orders, respectively. Otherwise, only about 4% and 20% of airlines in other parts of the world used translation for Task Cards and Engineering Orders respectively, and almost nobody translated the Maintenance Manual. The strategy of using the native for the spoken work were widely seen, with almost all Asian and most of Other airlines conducting meetings and maintenance training in languages other than English.

This is summary-based data, and thus presumably represents the respondent's best estimate of some of the variables requested, e.g. percentages of different groups with specific levels of English Ability. We have sensed some potential bias towards optimism in survey responses. This makes it all the more important that we look for ways to help these airlines that report both low levels of English ability and low use of strategies such as translation. There were only 8 airlines (5 in Asia, and 3 in Others) reporting low or very low levels of English Reading Ability, but of these, only a minority translated key documents such as the Maintenance Manual and Structural Repair Manual. This is only a small number of airlines, but the fact that they do not use translation or modification should be addressed. Note however, that of these 8 airlines most did make modification to shorter documents used as job aids, i.e. Task Cards and Engineering Orders. For oral communications, almost all airlines use their native languages for maintenance training, meetings and in casual conversation. If the oral communication is in the native language, but documentation is largely in English, there is a potential for language errors. If airlines recognize the need to match the language to the user for oral communication, then perhaps the only reason for airlines not accommodating non-English-Speakers in documents (Maintenance Manual, Engineering Orders, Task Cards, etc.) is that they lack the knowledge or ability to make effective interventions. There appears to be a need to help airlines overcome this reluctance. The current project may provide such assistance.

There are still issues of how effective strategies such as modification or translation are in error reduction, but the remainder of the project is designed to answer such questions. From this survey we have data on a larger sample than the project team could hope to visit, so that our future sample data can be seen in a broader context. Note that this survey was of airlines rather than Third Party repair stations, so that the results may be somewhat different. However many airlines are now positioning themselves to undertake third party repair business, so that the results should at least be comparable.

4.4 Conclusions on Language Issues Survey Analysis

1. The survey was based on a sample of 113 airlines, approximately equally divided between the four regions of North America, Europe, Asia and the rest of the world.

2. There was great diversity of responses to questions about use of English between regions, suggesting that solutions to any problems will require similar diversity.

3. Most mechanics in North America and Europe speak and read English at a high level. In contrast, Asia and the rest of the world have a much smaller base of English using mechanics.

4. Translation of documents was rarely used in North America and Europe, and used less than half the time in other regions. It does not appear to be the preferred strategy for response to any language mismatch issues, even where English reading and speaking ability is low.

5. Outside Europe and North America, meetings and training are often conducted in the native languages. This may be a mismatch to documentation used in the same task.

5.0 FOCUS GROUP DATA AND ANALYSIS

A focus group gathers people together to discuss the issue at hand via moderator questions and group discussions. Data are gathered through observations and conversations with participants. What the participants in the group say during their discussions are the essential data in focus groups. Typically, there are six to eight participants from similar backgrounds, and the moderator is a well-trained professional who works from a predetermined set of discussion topics. This group size allows variety in responses while still allowing all of the participants to speak. The focus group is not a collection of simultaneous individual interviews but rather a group discussion where the conversation flows because of the nurturing of the moderator (Krueger, 1994). The goal of focus groups is not to reach a consensus, but rather to obtain a variety of responses reflecting the knowledge and experience of those in the group. Focus groups have four basic uses: problem identification, planning, implementation, and assessment.

5.1 Why Use Focus Groups?

The focus group is a useful tool in gathering qualitative data in an efficient manner. Interviews and surveys are other options for gathering qualitative data. Focus groups draw on three of the fundamental strengths that are shared by all qualitative methods:

1. Exploration and discovery. Focus groups are frequently used to learn about either topics or groups of people that are poorly understood.
2. Context and depth. Focus groups encourage participants to investigate the ways they are both similar to and different from each other and help investigators understand the background behind people's thoughts and experiences.
3. Interpretation. In focus groups, people want to understand each other, and why things are the way they are and how they got to be like that. Even better, focus groups data is derived from a dynamic group process: participants influence each other, opinions change, and new insights are offered. The process allows researchers to learn more by asking participants to clarify statements and feelings, by asking follow-up questions, and by observing gestures, body languages, and tone of voice, etc.

Focus groups are particularly appropriate for use in exploratory studies when little is known about a population or phenomenon. For example, in a study on long-distance interregional commuting, Lee (1996) interwove focus groups, the qualitative approach, with quantitative data from census. Census data provided a basis for defining various types of commuters; the characteristics of these types were then elaborated via focused interviews. He found that the focused interview sessions illuminated important issues that census or conventional travel survey data alone could not.

Data collected in focus groups may be more ecologically valid than methods that assess individuals' opinions in relatively asocial setting, given that they are social events involving the interaction of participants and the interplay and modification of ideas (Albrecht et al, 1993).

Based on research by Morgan (1998), the focus group appears to be a good match for our research goal because it allows us to explore very complex behavior and motivation issues related to language errors in the aviation maintenance domain.

5.2 How to Run Focus Groups

1. Researchers' duty delegation ("Red light running study", Wissinger et al., 2000)

The moderator introduces the study to the focus group participants, asks questions, guides the discussion, and offers an oral summary to participants at the conclusion of the focus group. The notes of the moderator are not so much to capture the complete interview but rather to identify a few key ideas or future question that need to be asked. *The assistant moderator* is responsible for recording the discussion, asking any questions and addressing any issues not mentioned by the moderator. In addition to operating the tape recorder, the assistant moderator is also expected to take comprehensive notes, handle the environmental and logistic issues (e.g., refreshments, lighting, seating, etc.), and respond to unexpected interruptions. Usually, the assistant notes the participants' body language throughout the discussion.

For the moderator, a second set of eyes and ears increases both the total accumulation of information and the validity of the analysis (Krueger, 1994).

2. Discussion Questions have to:

- Address specific focus group research goals
- Be specific, and require more than a yes or no answer
    - Be small in numbers to allow sufficient time for the focus group's answers and discussions

Small Talk and Pre-session Strategy

Casual and comfortable "small talk" prior to the group discussion is essential to create a warm and friendly environment and put the participants at ease. Participants often arrive at different times; the 5-10 minute small talk maintains the warm and friendly environment until a sufficient number of participants are present to begin the session. More importantly, during this period the moderator and the assistant can observe participant interaction and learn about their individual characteristics: such as who tend to dominate the group, who are excessively shy, or who consider themselves as experts. Correspondingly, strategic positioning of participants can be applied. For example, seating the individuals who talk a lot and may later dominate the conversation at the moderator's side if at all possible. Then later on, the moderator can turn slightly away from the domineering individuals, and giving a nonverbal and diplomatic signal for others to talk when it is necessary.

Based on research conducted by Kelleher (1982; referred from Krueger, 1994), 40% of the participants are expected to be eager and open to sharing insights and that another 40% are more introspective and willing to talk if the situation present itself; the remaining 20% are apprehensive about the experience and rarely share.

A simple method to achieve strategic positioning of participants is preparing "name tents" (e.g. 5 x 8 index cards, folded in the middle and with first name printed). Based on the moderator's observation from the informal pre-session (with a quick checking with the assistant moderator's perception), the name tents will "drop" around the table in a seemingly random manner.

Opening Questions

- Greeting
- Self-introduction (first name, what do I do, etc.)
- Mentioning about video/audio taping and assure confidentiality again

Good morning and welcome to our session today. Thank you for taking the time to join our discussion of language errors in aviation maintenance domain. I am Professor Colin Drury and I represent University at Buffalo. Assisting me is Maggie Jiao Ma, also from University at Buffalo. We want to find out more about how language errors occur and therefore how to reduce them at aviation maintenance workplace. We have invited people who work in different groups to share their perceptions and ideas.

In the next hour or so, we will be discussing your experiences and your opinions about language errors occurred in your work. There are no right or wrong answers but rather differing points of view. Pleas feel free to share your point of view even if it differs from what others have said.

Before we begin, let me share some ground rules. This is strictly a research project and there are no job reviews or auditing involved. Please speak up—only one person should talk at a time. We're tape-recording the session because we don't want to miss any of your comments. If several are talking at the same time, the tape will get garbled and we'll miss your comments. We will be on a first name basis today, and in our later reports no names will be attached to comments. You may be assured of complete confidentiality. Please keep in mind that we are just as interested in negative comments as positive comments, and at times the negative comments are the most helpful.

OK, let us begin. We have placed name cards on the table in front of you to help us remember each other's names. Let's find out some more about each other by going around the room one at a time. Tell us your name and what you do.

Introductory Questions

"We are helping the FAA to reduce errors in aviation maintenance and inspection. Our aim is to find improved ways of performing maintenance and inspection jobs. One issue has been that although English is the primary language of aviation, many people do not have English as their native language."

Transition Questions

- Do you identify language errors as a problem at work?
- How much of a problem are languages errors?
- Can you give us some examples?

Key Questions

- Are there any special difficulties non-native English speakers encounters in performing their jobs?
- Do you recall any incidents where English became a barrier to performing a job?
- How do you help non-native English speakers perform their jobs without errors?
- What are the signs that a person may need your help in this way?

Ending Questions

- We have spent some time discussing language errors in maintenance. Have we missed anything?
- Give a summary on the discussion session; Get agreement from each participant.
- Thank participants for their cooperation and promise to circulate the report.

3. Participants:

Focus groups are only a viable option when the participants feel comfortable in voicing their views. Compatibility of focus group participants means not only demographic similarity but also mutual feeling of being understood and respected (Morgan, 1998). Focus groups participants are recommended to be "reasonably homogeneous and reasonably unfamiliar with each other" by Krueger (1994). In order to gain different perspectives from the focus groups, we decided to include AMTs and other technicians, QA personnel and Managers/supervisors in each discussion group. We will attempt to recruit them from different sub groups in order to avoid having a person's direct supervisor, which may inhibit the self-disclosure.

4. Communication process in focus groups

The communication that occurs in focus groups is a phenomenon sometimes overlooked by researchers seeking to gather data from these sessions. The basis of the participants' interaction in focus groups may include identification of key problems and/or solutions and enacting or informing outsiders (researchers) about cultural patterns or community (Albrecht et al., 1993). Albrecht et al (1993) stated that social interaction affects not only opinion formation but opinion articulation as well. Kelman (1961) suggested that opinions are produced through one of three processes: compliance, identification, and internalization. Focus groups interaction may include each of the three communication types, which may be deemed as a potential threat to the internal validity of the focus group data.

Compliance is the act of responding in ways one believes are expected by a questioner, in anticipation of some immediate reward. *Identification* is related to the situation in which a respondent's position on an issue is similar to the position held by someone the respondent admires or with whom he or she seeks solidarity. *Internalization* is related to the report of opinions that are deeply ingrained and personal. These opinions are potentially the most valuable, yet the most difficult data to obtain by researchers using a focus group methodology, because they are less susceptible to transient effects of material rewards or social relationships (e.g., potential artifacts of the focus group setting). These three processes interweave with one another.

We designed our study to reduce their negative effects in focus groups. First of all, participants frequently assume that a focus group moderator is professionally associated with a particular organization, product, or idea (in our case, the Federal Aviation Administration). Also the moderator's own biases and preferences may be intentionally or unintentionally signaled. The participants may wish to please (or at least not offend) the moderator; their responses may reflect what it is they think the moderator wants to hear. In order to avoid this potential bias, in our study the moderator and the assistant moderator are going to cross check the "unbiasedness" of all the questions. The cooperating relationship with the FAA for only research purpose will be emphasized. All the conversation correspondence (on video/audio tapes) collected from pilot studies will be reviewed carefully in order to find all the potentially "biased" points led by the moderator and make efforts to prevent them in the future focus groups.

Secondly, Albrecht et al (1993) pointed out one extreme case of opinion by identification, which is almost guaranteed to take place in a focus group, is involving participants from various hierarchical levels with a single reporting structure. Regardless of whether the moderator emphasizes that all responses and disagreements are legitimate, it simply may be too professionally and personally risky for a subordinate to disagree publicly with an opinion offered by a superior who holds fate control.

In order to simulate the complex working environment such as aviation maintenance context and also efficiently gain different perspectives from the focus groups, we decided to include AMTs and other technicians, QA personnel and Managers/supervisors in each discussion group. We will attempt to recruit them from different sub groups in order to avoid having a person's direct supervisor, which may inhibit the self-disclosure.

Thirdly, one classical example of difficult internalization has been seen by many experienced focus groups moderators: when participants are asked one at a time for their judgments, those responding last tend to echo the sentiments of those responding first. Albrecht et al (1993) pointed out that research on group discussion has found that the group idea-generation process benefits when it begins as a parallel, individual process.

We plan to begin with participants writing, rather than saying, their ideas. Such techniques have been examined formally as part of structured approaches in decision-making groups, and recommended as the easiest approach in the public context of the group (Albrecht et al, 1993).

5. Ethical issues:

- Participants' privacy (first name only or pseudonym names)
- Sponsor's relationship to the participants (the SCARY FAA)
- Among participants (supervisors vs. supervised, unhelpful or obstructing colleagues)
- Stressful/sensitive topics (reporting problems vs. discrimination towards NNESs)

5.3 Focus Group Check Sheet

1. Planning

________________Conceptualizing the study

________________Development of questions

________________Logistic arrangements

2. Recruiting

________________Using contacts in the companies

3. Moderating, Part II 6. Moderating Skills (P101)

There are several requirements for moderators to conduct good focus groups interviews.

No.1: Exercising a mild, unobtrusive control over the group. For example, as the discussion proceeds, participants may introduce irrelevant topics and the moderator should carefully and subtly guides the conversation back on target.

No.2: Possessing a curiosity about the topic and the participants.

No.3: Being respectful for participants. The moderator must truly believe that the participants have wisdom no matter what their level of education, experience, or background. Keeping this point in mind is especially important after four or five groups, because the moderator will have heard the topic discussed in a variety of ways and many of the concerns and key ideas have been said several times.

No.4: Having adequate background knowledge on the topic of discussion to place comments in perspective and follow up on critical areas of concerns.

No.5: Being able to communicate clearly and precisely both orally and in writing.

No.6: Keeping self-disciplined by suspending the personal views.

No.7: Possessing a friendly manner and a sense of humor.

Some techniques are proved to be very valuable in focus groups. The 5-second pause and the probe are two techniques recognized by experienced moderators in soliciting additional information from group participants. The 5-second pause is often used after a participant comment, which often prompts additional points of view or agreement with the previously mentioned position. The probe is the request for additional information. Typically, probing involves comments as the follows (Krueger, 1994):

- Would you explain it further?
- Would you give me an example of what you mean?
- Is there anything else?
- Please describe what you mean.
- I don't quite understand/ I am not very clear about what you just said.

Moderators should be attentive to how they respond to comments from participants—both in verbal and nonverbal ways. As a rule of thumb, moderators should try to restrict head nodding, which is often an unconscious response that signals agreement and, as a result, tends to elicit additional comments of the same type. Some short verbal responses, such as "OK", "Yes", or "Uh huh" are acceptable in the focus group environment, but others should be avoided if then communicate indications of accuracy or agreement. For example, "Correct", "Excellent", or "That's good" may imply judgments about the quality of the comment.

The moderator needs to pay particular attention to four types of participants, who may present some problems for the focus groups.

Types	Problems	Solutions
Self-appointed "experts"	What they say and how they say it can inhibit others in the group. They may be perceived to have more education, experience, affluence, or political/social influence.	- Underscoring everyone is an expert and all participants have important perceptions that need to be expressed. - Don't include introductory questions that would identify participants' levels of education, affluence, social influence, etc.
Dominant talker	They sometimes consider themselves to be experts. They can often be identified in pre-session small talk (e.g., talks a lot).	- Verbal: "Thank you, John. Are there others who wish to comment on the question?", "Does anyone feel differently?" - Nonverbal: avoiding eye contact
The shy	They say little and speak with soft voice, but they often have much to share but extra effort is required.	- Encouraging them open up by ensuring them that their comments are wanted and appreciated
The rambler	They use a lot of words and usually never get to the point, if they have a point.	- Discontinuing eye contacts after 20-30 seconds with the rambler.

4. Analyzing and Reporting

Part II 7. Principles of Analyzing Focus Group Results (P126)

Focus group analysis is complex and subject to human errors. The systematic approach in data collection and analysis reminds the analyst of upcoming steps but it also communicates to the user of the study that the analyst is attempting to minimize human errors (Krueger, 1994).

The Analysis Continuum

←————————————————————→

Least Time Intensive			Most Time Intensive
Least Rigorous			Most Rigorous
Memory-Based	Note-Based	Tape-Based	Transcript-Based

Part II 8. Process of Analyzing Focus Group Results (P140)

Tape-based Analysis:

1. Gather tapes and field notes by category
2. Review field notes by category
3. Enter abridged transcript on computer
4. Look for emerging themes (by question and then overall)
5. Develop coding categories and code the data
6. Sort the data into coding categories
7. Construct topologies or diagram the analysis
8. See what data are left out and consider revision
9. Prepare the draft report—begin with most important questions

The first step in the analysis of qualitative data is reducing or collapsing the typically large amount of data. A common method for data reduction is through the steps of chunking and coding (Bisantz, 2001). Chunks typically consist of text describing a single concept. Then a coding scheme, which depends on the needs of the analysts, can by applied to these chunks. The resultant coding scheme should reflect the contents of the data set, rather than predetermined theories on the part of the analyst (Strauss and Corbin, 1990). According to Bisantz (2001), coded data can then be grouped according to codes and examined for patterns and differences across variables of interest to the analysis, for example, types of users, the nature of the work, etc. For the focus groups in current project, we are going to use spreadsheet programs such as MS Excel to support coding and analysis of textual data.

There are two stages of data analysis on interview/focus group data: the preliminary analysis and the formal data analysis (Rubin and Rubin, 1995). The preliminary data analysis begins while the interviewing is still under way. By examining what we hear and pulling out the concepts and themes that describe the interviewees/focus groups' world, we can redesign our questions to focus on central themes. The formal data analysis is a more detailed and fine-grained analysis upon the completion of interviewing. The goal of this analysis stage is to integrate the themes and concepts into a theory that offers an accurate, detailed, yet subtle interpretation of your research arena.

In the formal qualitative data analysis, a routine described by Rubin and Rubin (1995) as: 1) researchers each read and reread the interviews to note core ideas and concepts, recognize emotive stories and find themes; and 2) then researchers code the material to group similar ideas together and figure out how the themes relate to each other.

<u>How to recognize concepts</u> (*concepts:* conceptual labels placed on discrete happenings, events, and other instances of phenomena.)

1. Pick out the words the interviewees frequently use that sound different from your ordinary vocabulary. Example, "It was slow night. Some drive-bys and a few carry-outs… a cold one, a couple bleeders and a blue-icer." (from conversation between paramedics). Quite different from a MacDonalds' drive-through, the paramedics use "drive-bys" for "routine stops", and "carry-outs" for "transporting people to a hospital". "A cold one" means one dead passenger, and "a blue-icer" refers to a choking victim.

2. Look for nouns/noun-phrases those are repeated frequently and seem to convey important ideas for the conversational partners (i.e., "time pressure", "planning").

3. Look for the pair, the mate, or the opposite of the concept you have just discovered. After identified a phrase as a label for an underlying concept,

looking for the discussion and definition of it (i.e., “bricks and mortar”, “bricks and clicks”).

4. Create a label for those core ideas described by the interviewees but not labeled by them.

How to hear themes (*themes*: an implicit and recurrent idea--dictionary)

Themes are those statements that offer descriptions of how people do or should behave, such as “everyone is out for himself or herself” or “good people care for their parents and help those in need”.

1. Relate themes you heard with one another, and this helps you build toward a broader description or an overall theory. For example, there are several themes can be put together as a theory about the effects of discrimination: “discrimination against African Americans in hiring”, “poor quality inner-city schools”, “residential segregation”, etc.
2. Combine information from different interviews/focus groups, the theme may emerge itself when those separate ideas are put together.
3. Tips on picking out the themes (Rubin and Rubin, 1995):
   - Some times people repeat and emphasize their own themes
   - Iconic statements and pithy summaries
   - Dramatic statements often suggest important themes (i.e., “sexy quotes” such as “shoot and fucking intermediaries)
   - Both compatible and contradictory concepts can help build themes
   - Looking for similarities in how people who are in different circumstances interpret their world

How to code interview data (*coding:* the process of analyzing data)

Based on practice, Rubin and Rubin (1995) introduced the mechanics of coding can be coding for themes, concepts, and ideas, or also the names of agencies or people, major projects, dates, stages or steps of a process, or just about anything that you think might be useful in tying things together. Coding proceeds in stages.

> Stage 1: Setting up a few main coding categories, suggested by the original reading of the interviews and the intended purposes of the report.
> Stage 2.1: While placing data into preliminary coding categories, decide whether these preliminary categories provide a good fit into the data. Make changes to the coding categories to reflect what you discover in the data.
> Stage 2.2: Add new categories when important information doesn't fit into the preliminary categories
> Stage 3: Whenever coding categories are changed or added, go back and recode the material already examined

Part II 9. Reporting Focus Group Results (P161)

The written report:

1. Cover page
2. Summary
3. Table of contents
4. Statement of the problem, key questions, and study methods
5. Results or findings
6. Summary of themes
7. Limitations and alternative explanations
8. Recommendations
9. Appendix

5.4 Focus Groups on Language Errors

While the analysis of archival data in the preceding sections could provide some insight into language errors in maintenance, such data were not collected for that purpose (c.f.

Drury 1995). More direct data collection involves the use of questionnaires and interviews specifically on the theme of language errors in maintenance. However, before we can ask sensible questions, we must have valid information on the types of errors involved. We collected such data from focus groups at MROs in different countries. So far (July 2003), we have run five such focus groups, three at US-based MROs and the other two at UK-based MROs.

We used focus groups of people at MROs drawn from AMTs, supervisors, engineers and QA specialists. Each interview lasted about 45 minutes. Our introductory statement (after introductions, ground rules and assurance of anonymity) was:

> "We are helping the FAA to reduce errors in aviation maintenance and inspection. Our aim is to find improved ways of performing maintenance and inspection jobs. One issue has been that although English is the primary language of aviation, many people do not have English as their native language."

Then, the focus groups discussed approximately ten questions with the principal investigator as moderator. When we had transcribed the data, we compared the transcripts with our notes to look for patterns of maintenance language errors or events under four headings.

1. Error types/patterns
2. Potential error detection points in the maintenance process.
3. Factors predisposing to language errors

4. Factors potentially mitigating language errors

From these lists, we were able to see the functions of aircraft maintenance and inspection (see Drury, Shepherd and Johnson 1997) and where language errors could arise. Table 10 represents our current characterization of these situations where their errors could arise, presented within a task sequence framework.

Function	Language Error Detection
Setup	• AMT may appear perplexed, or may agree with everything said.
Task Performance	• AMT may ask for assistance or clarification. • AMT may close access prematurely (i.e. before buyback)
Buyback	• Physical error may be detected. • AMT may not understand inspector's questions.

Table 10. Language Errors Arising in a Task Sequence Framework

We found the following patterns of error in both verbal (synchronous) and written (asynchronous) communication.

Verbal (Synchronous)

1. AMT unable to communicate verbally to the level required.
2. AMT and colleagues/supervisors have poorly matched models of their own and each other's English ability.
3. Native English speakers with different regional or non-US English accents (e.g. UK, India, Caribbean) prevent adequate communications.
4. AMTs unable to understand safety announcements over the PA system.

Written (Asynchronous)

5. AMT unable to understand safety placard in English.

6. AMT unable to understand written English documentation.
7. Foreign documentation poorly translated into English.

While the patterns are still being refined as further data is collected, and may eventually exhibit more of a hierarchical structure, they were reasonably consistent between the focus groups studied.

Table 11 shows the predisposing and mitigating factors identified in the focus groups. They are classified in terms of the SHELL model of human factors in aviation (Easterby, 1967).

SHELL Category	Predisposing Factors	Mitigating Factors
Software (procedures)	• Task complexity • Instruction complexity	• Document translation • Consistent terminology • Good document design
Hardware (equipment)	• Limitations of communication channel, e.g. radio, PA	• Use of aircraft as a communication device: "show me"
Environment	• Time pressure prevents AMT from querying others	
Liveware (individual)	• Inadequate written English ability • Inadequate English ability • Reversion to native language under stress	• Job familiarity • Comprehension tests for AMTs • Certify AMT for specific jobs
Liveware (inter-communication)	• Unwillingness of AMT to expose their lack of English • Time pressure	• Translator available • Assign AMTs to job based on English ability • Team AMT with native English speaker

Table 11. Predisposing and Mitigating Factors Identified in the Focus Groups

6.0 PILOT STUDY OF COMPREHENSION TEST

In Phase II of this project, we will be collecting data on both the incidence of language errors and the effectiveness of mitigation techniques. For the latter, we propose to use a comprehension test methodology, based on the one used to evaluate Simplified English by Chervak, Drury and Ouelette (1996) and further validated by Drury, Wenner and Kritkausky (1999).

One of our expected interventions will be the use of Simplified English in workcards, so we performed a pilot test of this methodology at our focus group sites with suitable participants. As workcard complexity had been shown to interact with Simplified English in previous studies, we included this factor at two levels (Easy, Difficult) as well as using workcards written in Simplified English or written by a major manufacturer in standard English.

6.1 Methodology

Materials: We selected two task cards, one "easy" and one "difficult", from four task cards used in Chervak, Drury and Ouellette (1996)'s research. Complexity of the task cards had been evaluated by the Boeing computational linguists and University of Washington technical communications researchers considering four factors in their Non-Simplified English (SE) versions (Table 12). A task difficulty rating of each task card by an experienced engineer was also used for guidance.

Task Card	Complexity	Word Count	Words per Sentence	Percentage Passive	Flesch-Kinkaid
SUNY N380	Easy	(276) 254	8 (lo)	3 (lo)	8.6 (lo)
SUNY N290	Difficult	(527) 491	17 (hi)	25 (hi)	10.4 (hi)

Table 12. Characteristics of the two non-SE versions of task cards

Both of the task cards were then prepared in SE versions, which were critiqued by Boeing, University of Washington, and the AIAA Simplified English Committee experts (Chervak, Drury and Ouellette, 1996).

Participants: We had total 15 participants, 7 from MROs in the US and 8 from MROs in the UK, comprising 5 AMTs, 4 inspectors, 1 supervisor, 2 engineers, and 3 QA specialists. We compared the age and experience distributions to the population demographics of AMTs found in a national sample compiled by the US Bureau of Labor Statistics (BLS, Washington, 1991). Our sample was significantly older with median age of 42.5 year (SD=9.6) versus a BLS median age of 36.2 years (one sample T-test, T = 2.31, $p < 0.038$), and significantly more experienced as an AMT, with median experience 16.0 years (SD=7.3) versus a BLS median of 9.4 years (one sample T-test, T = 3.95, $p < 0.002$). The age distribution of this sample shown in Figure 11 and the AMT experience distribution is shown in Figure 12.

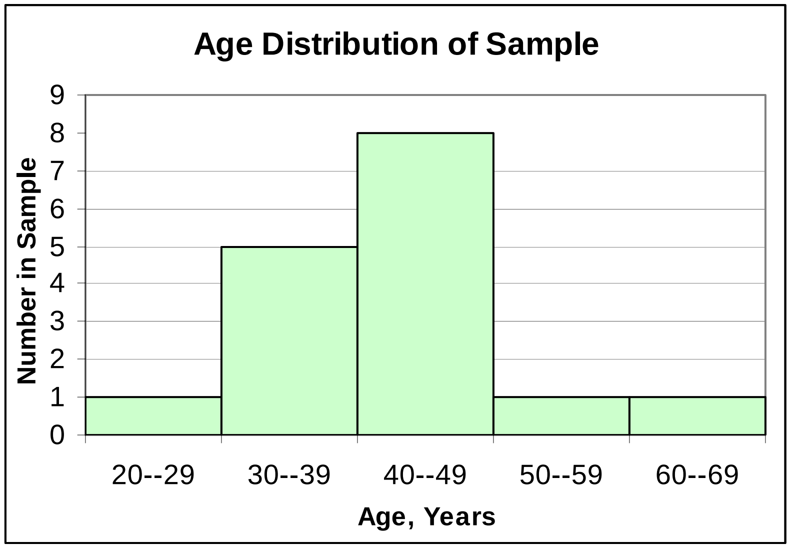


Figure 11. Age Distribution of Sample

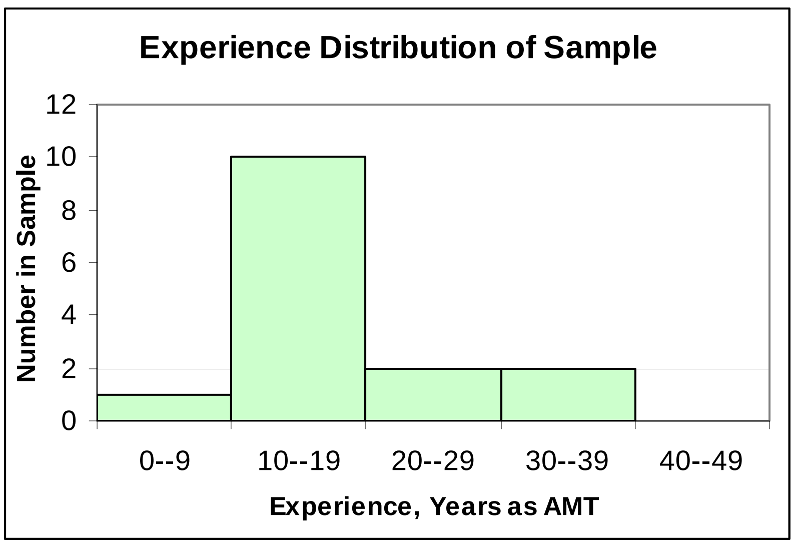


Figure 12. Experience Distribution of Sample

Figure 13 shows the distribution of scores on our chosen reading ability test, the Accuracy Level Test (Carver, 1987). This ten-minute limited vocabulary test measured the reading level of a participant as an equivalent grade level. The test has high reliability (0.91) measured on college students and has a high validity (0.77 to 0.84) when compared to a longer standard reading test (the Nelson-Denny Reading Test). Carver (1987) provides data on two appropriate comparison groups for this test: freshmen undergraduate and beginning graduate students. The mean score of our sample (14.1) was significantly higher than for college freshmen (12.5) with T = 4.35, $p < 0.001$. It was not significantly lower ($p = 0.610$) than for graduate students (14.3). Thus the reading level of our participant sample was typical of an educated adult group, i.e. above college freshmen but a little below graduate students. Interestingly, we noticed that reading level of the participants from the UK (mean = 14.9, SD = 0.45) was significantly higher than that of others from the US (mean = 13.1, SD = 1.62) with ($F = 8.52$, $p = 0.013$). Figure 14 demonstrated some variance of reading level among the different job categories. However, one-way ANOVA revealed no statistical significance between reading level by job categories.

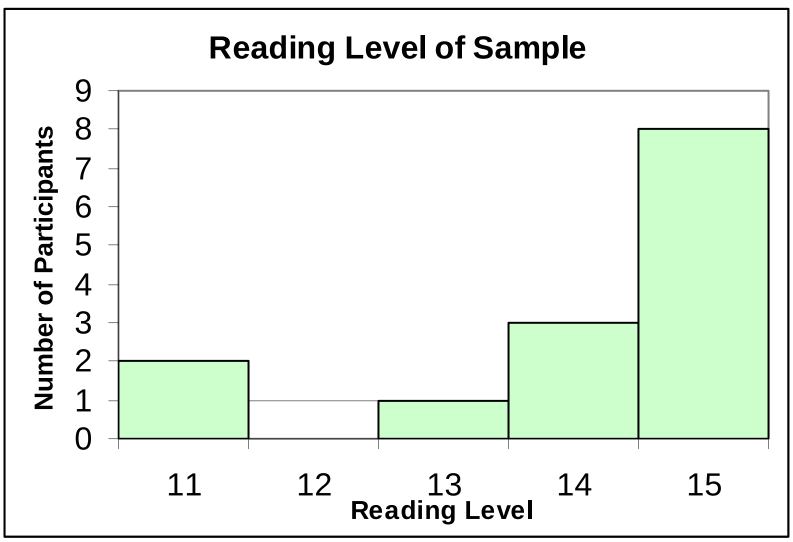


Figure 13. Reading Level Results for Participant Sample

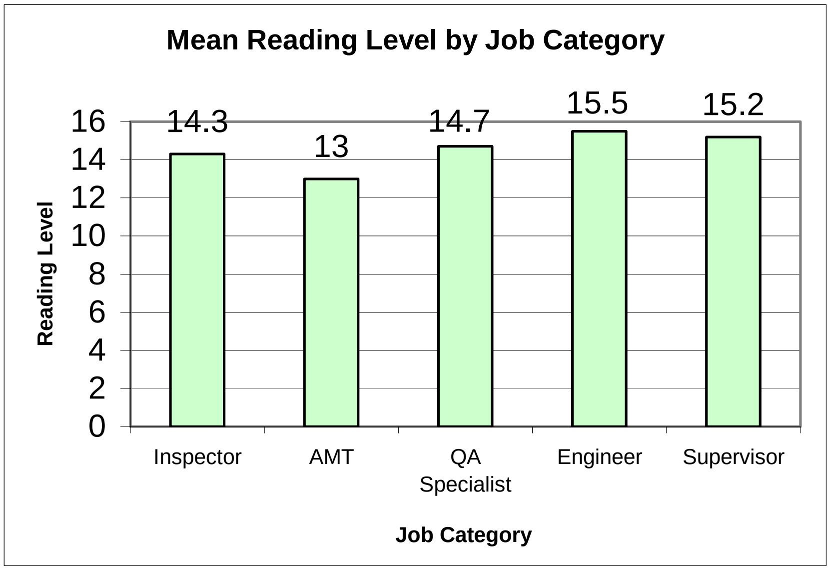


Figure 14. Reading Level Varied among Job Categories

Task Procedure: The testing took place at available conference rooms at the MROs. Each participant was given verbal instructions for completing a demographic question and the Accuracy Level Test. He/she was randomly given one of the 4 possible task cards, its comprehension questionnaire (20 questions) and a set of task card rating scales (15 scales). Distribution of task cards at each MROs was in rotating order with a new starting point at each site. The rating scales were adapted from the evaluation scales used by Patel et al (1994), covering ease of use of the task cards and attached graphics, the simplicity of the English used, and finally an overall rating on usability of the task cards. All were 9-point (0 to 8) anchored at each end with an appropriate adjective, and with their midpoints located at a scale value of 4.

In a task card comprehension questionnaire, generally a question concerning specific technical information was followed by a question asking where this information was located in the task card. The questions demanded a short answer, a “fill in the blank,” or a multiple choice. Both SE and non-SE versions of each task card had the same comprehension questionnaire. In some cases, different words were used in SE and non-SE versions of task cards to refer to the same object, so that, a neutral word with similar meaning was used in the comprehension questionnaire to prevent bias. For example, in SE version, a term "Do-Not-Operate *Tag*" was used to indicate a card that was placed on an inoperative control lever, whereas in the non-SE version the term "Do-Not-Operate *Identifier*" was used. In the questionnaire, questions regarding these cards used the term "Do-Not-Operate *Marker*."

Experimental Design: Our study was a two factor factorial design with the participants nested under both factors. The factors (Independent Variables) were:

1. Two versions of task cards: SE vs. non-SE
2. Two levels of task card complexity: Easy vs. Difficult

- Dependent Variables:
  1. Performance measures on comprehension questionnaire:
     1.1. Completion time
     1.2. Accuracy
  2. Task cards rating: 15 rates scales scores

- Possible Performance Predictors or Covariates:
  1. Age
  2. Experience as an aviation mechanic

2.1. Reading level score

2.2. Job category

2.3. Native language (NOTE: not at current stage)

6.2 Results

Performance Measures: The participants' accuracy and time in the task card comprehension test were significantly and negatively correlated (correlation coefficient = -0.692, p= 0.004): the participants tended to be either "faster--more accurate" or "slower--less accurate" (as in Figure 15). There are four possible individual variables that may affect performance: age, AMT experience, reading level score, and job category. These could be useful covariates in the analysis of main factors by reducing the expected variability between individual participants. An inter-correlation matrix of these revealed that AMT experience was highly correlated with age ($R = 0.60$, $p = 0.039$). Reading level score was moderately correlated with both performance measures (i.e. time, accuracy). Job category was not significantly related to either time or accuracy. We decided to consider two covariates: age and reading level score.

We used GLM 2-factor ANOVAs on each performance variable with the above covariates but found no statistical significance. Our finding of some differences in reading level between US and UK participants led us to explore adding this as a third factor in the analysis. There was insufficient data to perform a full factorial, so that our GLM ANOVA included only the following terms:

Difficulty (D)

SE/nonSE (S)

D X S

US/UK (U)

In three ANOVAs we analyzed Time, Accuracy and Time/Accuracy as it had already been established that time and accuracy were negatively correlated (Figure 15). The results are given in Table 13, with some results between 0.05 and 0.10 significance shown for this pilot study.

Measures	Factors			
	Difficulty	**SE/Non SE**	**D X S**	**US/UK**
Time	P = 0.010	NS	P = 0.073	P = 0.095
Accuracy	P = 0.026	NS	NS	P = 0.022
Time/Accuracy	P = 0.009	NS	P = 0.076	P = 0.024

Table 13. ANOVA Summary for Performance Measures (all with (1,10) df)

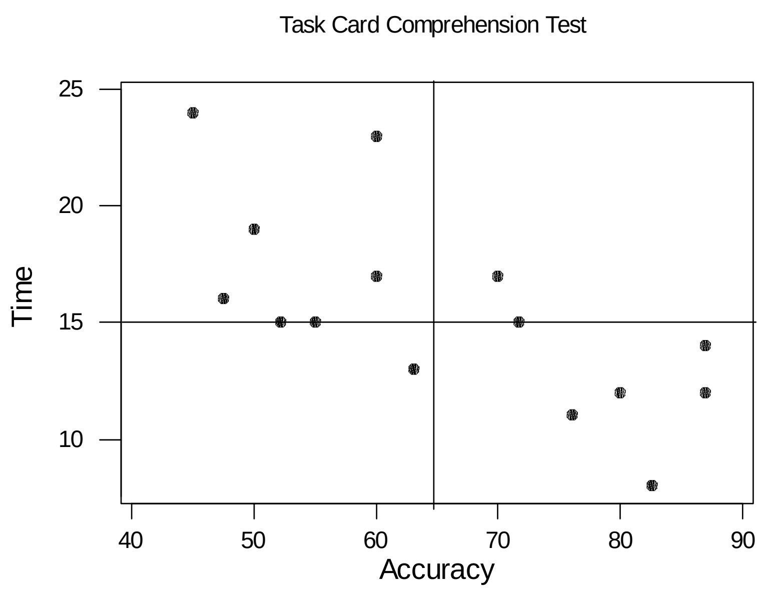


Figure 15. Time × Accuracy Plot

Generally, the UK participants outperformed their US counterparts with faster time (14 vs 17 min), higher accuracy (72% vs 58%) and better accuracy/time ratio (5.7 vs 3.8). The Simplified English intervention was never significant, but the more difficult workcard had significantly lower comprehension than the easy one, for time (18 vs 13 min), accuracy (58% vs 74% and accuracy/time ration (3.5 vs 6.3). There were suggestions of a difficulty/SE interaction, but neither reached accepted levels of significance. Accuracy/time was about the same for SE and nonSE versions for the difficult workcard (3.6 vs 3.4) but much better for SE than non-SE for the easy workcard (7.3 vs 4.8). This is somewhat contrary to preview results, where performance improved with SE more for the more difficult tasks.

The relatively low levels of significance for SE effects are probably due to our small data sample (Table 14) and the fact that all our 15 participants were native English speakers with extensive AMT experience.

	Simple Task Card	**Difficult Task Card**
Simplified English	4	3
Non-Simplified English	3	4

Table 14. Small, Unbalanced Data Sample

Rating Scale Analyses: There were few significant effects noted in the GLM ANOVAs for the rating scale scores. Table 15 summarizes significance for main factors, their interaction, and covariates.

Measure	Main Factors and Their Interaction			Covariates	
	DIFF	VER	DIFF × VER	AGE	READ
1. Readability of text	N.S.	N.S.	N.S.	N.S.	N.S.
2. Continuity of information flow	N.S.	N.S.	N.S.	N.S.	N.S.
3. Ease of information location	P=0.067	N.S.	N.S.	N.S.	P=0.030*
4. Chance of missing information	N.S.	N.S.	N.S.	N.S.	N.S.
5. Ease of understanding	P=0.066	N.S.	N.S.	N.S.	N.S.
6. Ease of location on aircraft	N.S.	P=0.017*	N.S.	N.S.	N.S.
7. Ease of relating figure numbers	N.S.	P=0.067	N.S.	P=0.045*	N.S.
8. Amount of information provided	N.S.	N.S.	N.S.	N.S.	N.S.
9. Ease of readability of attachments	N.S.	N.S.	N.S.	N.S.	N.S.
10. Relating graphics to aircraft structure	N.S.	N.S.	N.S.	N.S.	P=0.067
11. Consistency of presentation	P=0.002**	N.S.	N.S.	P=0.001**	P=0.000**
12. Compatibility with attachments	N.S.	P=0.019*	P=0.016*	P=0.025*	N.S.
13. Amount of graphics provided	P=0.029*	N.S.	N.S.	N.S.	N.S.
14. Simplicity of English used	N.S.	N.S.	N.S.	N.S.	N.S.
15. Overall ease of usability of w/c	N.S.	N.S.	N.S.	N.S.	N.S.

Table 15. Significance Levels of Main factors and Covariates for Rating Scale Data (* significant at 0.05, ** significant at 0.01)

Table 16 shows significantly different ratings for SE and non-SE versions of task cards. For rating scale No.6, “0” represents “Very Difficult” and “8” represents “Very Easy.” For rating scale No.12, “0” represents “Terrible” and “8” represents “Excellent.” Interestingly, our participants seemed to prefer non-SE versions of task cards for both scales, which was opposite to our expectation.

Measure	Non-SE	SE
6. Ease of location on aircraft	7.0	5.3
12. Compatibility with attachments	5.9	4.1

Table 16. Mean Ratings of Both Versions for Significant Measures

Table 17 shows significant ratings for easy and difficult task cards. For rating scale No. 11, “0” represents “Terrible” and “8” represents “Excellent.” For rating scale No.13, “0” represents “Too Little” and “8” represents “Too Much.” The participants felt higher consistency of presentation when the task card was easier. Although in fact the easy task card had only one graphic compared to 6 graphics of similar complexity for the difficult task card, those participants who were tested on the easy one still complained about "too much graphics."

Measure	Easy	Difficult
11. Consistency of presentation	6.2	4.9
13. Amount of graphics provided	6.4	3.3

Table 17. Mean Ratings of Both Task Cards for Significant Measures

For rating scale No.12, the interaction between two main factors was significant. The easy task card in SE had the highest preference. For the difficult task card, either in SE or non-SE didn't affect its compatibility with attachments much. The participants rated the easy task card in non-SE the least compatible with its attachments (i.e. graphic). This was consistent with No. 13 rating score in Table 18.

<table>
<tr><th colspan="2">Measure</th><th>Easy</th><th>Difficult</th></tr>
<tr><td rowspan="2">12. Compatibility with attachments</td><td>SE</td><td>6.2</td><td>5.7</td></tr>
<tr><td>Non-SE</td><td>2.7</td><td>5.6</td></tr>
</table>

Table 18. Mean Ratings of Interactions (VER × DIFF) for Significant Measures

6.2 Conclusions

This was a small study, using only 15 participants in a between-participants design, so that its power was rather low. Despite this we did find significant effects of Difficulty and hints of a Difficulty X SE interaction when a major difference between our two country data sets was included in the analysis. We did find a few significant rating scale effects, but would not read too much into them with this small and restricted sample.

What we did accomplish was to show that our methodology was simple to use, took relatively little time (30 minutes or less) and that we were able to test multiple participants at the same time. We did find that the overall experiment is likely to be sensitive, particularly with a much larger sample size. Whether the effects of interventions are large enough to be significant remains to be seen, but at least we can reliably detect differences in workcard difficulty and country of origin. In addition, this data provides a useful baseline for our tests with non-native English speakers in Phase II.

7.0 Transition to Phase II

The first phase of our project was to find the patterns of language errors, provided there is evidence that they exist. Our analysis of communication models and the company database has shown the potential for language errors by showing that responses to language differences may not always keep pace with the need for such interventions. The ASRS database analysis showed some actual errors, although these were mainly in the flight operations domain more likely to be reported to ASRS. Patterns in this data showed that maintenance language errors were largely asynchronous, while related to terminology and had few recovery mechanisms.

The five focus groups tested so far have refined our conclusions. We now have ample evidence that language errors exist, although there are recovery mechanisms and mitigating factors. The patterns found were numerous, and certainly not limited to asynchronous communication. Although documentation was an important source of difficulty, there were other patterns in verbal communication, including unexpected ones of regional accents of native English speakers. We were also able to further document the time course and propagation of errors, including error detection points and interventions. In an industry as heavily regulated as aviation maintenance, there are a number of barriers to error propagation (c.f. Reason, 1990), including the initial work assignment and inspection by a different person.

The characteristics of language errors found so far in maintenance will be refined as more focus group data is collected, but the agreement reached to date suggests that a few overall patterns may account for most of the potential errors. In subsequent years of this project, we will be collecting field data to estimate the prevalence of the patterns we have derived. This will be done using direct data collection in several regions of the world, for example those used in our analysis of the company database. We will also use our methodology of comprehension tests of workcards (e.g. Chervak, Drury and Ouellette, 1996; Drury, Wenner and Kritkausky, 1999) to test the effectiveness of intervention strategies. These include use of Simplified English, full translation, use of an English-speaking coach and provision of a local language glossary. In this way, we will be able to make recommendations to both MROs and regulatory bodies for the effective reduction of language errors.

References

Albrecht, T.L., Hohnson, G.M., and Walther, J. B. (1993). Understanding Communication Processes in Focus Groups, *Successful Focus Groups, Advancing the State of the Art, edited by D. L. Morgan*, Sage Publications, Inc., Thousand Oaks, CA. p 51-64.

Bisantz, A. M. (2001). Computer Human Interaction. In J.J. Marciniak (Ed.), *Encyclopedia of Software Engineering, Second Edition,* J. Wiley and Sons, New York.

Cant, M. and Jeynes, L. (1998). What does outsourcing bring you that innovation cannot? How outsourcing is seen – and currently marketed – as a universal panacea. *Total Quality Management*. 9(2/3), 193-202

Carver, R. P. (1987). *Technical Manual for the Accuracy Level Test.* REVTAC Publications, Inc.

Chervak, S., Drury, C. G. and Ouellette, J. L. (1996). Simplified English for Aircraft Workcards. *Proceedings of the Human Factors and Ergonomic Society39th Annual Meeting.* 303-307.

Connell, L. (1995). Pilot and Controller Communications Issues. In B.G. Kanki and O.V. Prinzo (Eds.), *Proceedings of the Methods and Metrics of Voice Communication Workshop.*

Connolly, A. (2002). To Speak in Tongues: Language, Diversity and Psychoanalysis, *Journal of Analytical Psychology*, **47**, 359-382.

Cushing, S. (1994). *Fata Words*. Chicago: University of Chicago Press.

Davidmann, M. (1998). Using Words to Communicate Effectively, online resource: http://www.solbaram.org/articles/words.html.

Drury, C. G. (1995). The use of archival data. In J. R. Wilson and E. N. Corlett (Eds), *Evaluation of Human Work* (Second Edition), Chapter 5. Taylor and Francis, New York. 101-112

Drury, C. G. (1998). Case study: error rates and paperwork design, *Applied Ergonomics,* **29(3),** 213-216.

Drury, C. G., Shepherd, W. T. and Johnson, W. B. (1997). Error reduction in aviation maintenance. *Proceedings of the 13th Triennial Congress of the International Ergonomics Association'97.* Tampere, Finland, Vol. 3, 31-33.

Drury, C. G., Wenner, C. and Kritkausky, K. (1999*). Development of process to improve work documentation of repair stations.* FAA/ Office of Aviation Medicine (AAM-240). National Technical Information Service, Springfield, VA.

Drury, C. G., Wenner, C. and Kritkausky, K. (2000). Information design issues in repair stations. *Proc. Tenth International Symposium on Aviation Psychology*. Columbus, OH.

Easterby, R.S. (1967). Ergonomics Checklists: an appraisal. *Ergonomics*. **10(5),** 549-556.

Erlandson, D.A., Harris, E.L., Skipper, B.L., and Allen, S. D., *Doing Naturalistic Inquiry, a Guide to Methods*, Sage Publications, Inc., Thousand Oaks, CA

Fallon, B. J. (1997). The Extent to Which Pilots and Controllers Consider Accents or Dialects Cause Difficulties in ATC Communications, *International Symposium on Aviation Psychology*, 76-80.

Fegyveresi, A. (1997). Vocal Cues and Pilot/ATC Communication. *International Symposium on Aviation Psychology*. 81-84.

Griffith, E. (1999). *A First Look at Communication Theory (4th edition).* McGraw-Hill, Boston.

Huit, H.M. and Howard, M.R. (1993). *Born to Talk*. MacMillian Co, New York, NY.

Johnson, D.W. (1972). *Reaching Out*. Englewood Cliffs, Prentice Hall, NJ.

Kanki, B. G., and Smith, G. M. (2001). Training Aviation Communication Skills, In E. Salas, C.A. Bowers and E. Edens. (eds). *Improving Teamwork in Organizations: Applications of Resource Management Training*, Lawrence Erlbaum Associates, Mahwah, New Jersey, 95-127.

Kanki, B. G. and D. Walter (1997). *Reduction of maintenance error through focused interventions.* Meeting Proceedings of the Eleventh Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection: Human error in aviation maintenance.

Kelleher, J. (1982). Find out What Your Customer Really Want. *Inc*, **4(1),** 88-91

Knodel, J. (1993), The Design and Analysis of Focus Group Studies: a Practical Approach, In D. L. Morgan (Ed), *Successful Focus Groups, Advancing the State of the Art,* Sage Publications, Inc., Thousand Oaks, CA, 35-50

Krueger, R. A. (1994), *Focus Groups, A Practical Guide for Applied Research (second edition)*, Sage Publications, Inc., Thousand Oaks, CA

Lee, R. M. (1993), *Doing Research on Sensitive Topics*, Sage Publications, Inc., Thousand Oaks, CA

Lee, R.W. (1996), Exploration of Long-Distance Interregional Commuting Issues: Analysis of Northern California Interregional Commuters Using Census Data and Focus Group Interviews, *Transportation Research Record 1521*, 29-36

Matthews, B. G. and Hahn, C.K.G. (1987). Voice Communication Errors in the Air Traffic Control Environment, *Proceedings of the 20th Annual Conference of the Human Factors Association of Canada*, 167-170.

Merritt, A., and Ratwatte, S. (1997). Who Are You Calling A Safety Threat? A Debate On Safety In Mono-Versus Multi-Cultural Cockpits, *International Symposium on Aviation Psychology*, 661-666.

Morgan, D. L. (1998), *The Focus Group Guidebook*, Sage Publications, Inc., Thousand Oaks, CA

Nagel, D. C. (1988). Human Error in Aviation Operations, *Human Factors in Aviation*, 263-302.

Orasanu, J., Davison, J., and Fischer, U. (1997). What Did He Say? Culture and Language Barriers to efficient Communication in Global Aviation, *International Symposium on Aviation Psychology*, 673-678.

Reason, J. (1990). *Human Error*. Cambridge University Press.

Rifkind, L. J. (1996). Communication in the maintenance work environment, *Meeting Proceedings Tenth Federal Aviation Administration Meeting on Human Factors Issues in Aircraft Maintenance and Inspection: Maintenance performance enhancement and technician resource management,* Federal Aviation

Administration/Office of Aviation, Washington, DC, 25-41. online resources: http://hfskyway.faa.gov .

Rifkind, L. (1996). Communication, *Human Factors Guide for Aviation Maintenance* Federal Aviation Administration/Office of Aviation Medicine, Washington, DC, 13/1-13/54. Online resource: http://hfskyway.faa.gov .

Rubin, H. J. and Rubin, I. S. (1995). *Qualitative Interviewing: The Art of Hearing Data*. Sage Publications, Thousand Oaks.

Saffley, J. R. (1984), Advancing Safety through Effective Communications, *29th Corporate Aviation Safety Seminar (proceedings)*, 33-39.

Sexton, J. B. and Helmreich, R. (1999). Analyzing Cockpit Communication: The Links Between Language, Performance, Error, and Workload, *International Symposium on Aviation Psychology*, 689-695.

Sparaco, P. (2002). French Foresee Broader, Multinational MRO Network. *Aviation Week and Space Technology*. 72-73.

Strauss, A. and Corbin, J. (1990). *Basics of Qualitative Research.* Sage Publications, Newbury Park.

Taylor, J. C. and Patankar, M. S. (2000). The Role of Communication in the Reduction of Human Error. *Proceedings of the 14th International Symposium on Human Factors in Aircraft Maintenance and Inspection.* Vancouver, B.C.

Turney, M.A. (1997). Analysis of Gender Bias in Controller-Pilot Communication, *International Symposium on Aviation Psychology*, 85-89.

Wenner, C. A. and Drury, C. D. (2000). Analyzing Human Error in Aircraft Ground Damage Incidents. *International Journal of Industrial Ergonomics*.

26, 177-199

Wideman, R.M. (2002). *Communication Models*. Online resource: http://www.maxwideman.com/issacons4/iac1432/

Wissinger, L.M., Hummer, J.E., and Milazzo, J.S. (2000), Using Focus Groups to Investigate Issues of Red Light Running, *Transportation Research Record,* n 173, 38-45.

Yashima, T. (2002). Willingness to Communicate in a Second Language: the Japanese EFL Context, *The Modern Language Journals*, **86**, 54-66.

U.S. Department of Labor, Bureau of Labor Statistics, September 2001.

Appendix I.

Aviation Communication Research

What	Whom/Where	Authors and Year
Voice Communication	Air Traffic Control Environment	Mattews and Hahn (1987)
Culture and Language Barriers	Communication in Global Aviation	Orasanu, Davison and Fischer (1997)
Vocal Cues	Pilot/ATC Communications	Fegyveresi (1997)
Links between Language, Performance, Error, and Workload	Cockpit Communication	Sexton and Helmreich (1999)
Gender Bias	Controller-Pilot Communication	Turney (1997)
Safety	Mono vs. Multi-cultural Cockpits	Merritt and Ratwatte (1997)
Readbacks, Volume of Information, Experience Level, Personal Problems, Standard Phraseology, and the Relationship to Safety	Controller/Pilot Communication	Wulle and Zerr (1997)
Accents, Dialects	Pilot/Controller ATC Communication	Fallon (1997)
Cockpit Data Link Technology	Flew Crew Communication	Logsdon and et al (1995)
Communication Strategies	Pilot	Fischer and Orasanu (1999)
Communication Discrepancies	Pilots and Maintenance Technicians	Mattson, Crider, and Whittington (1999)
The Impact of Automation	Flight Crew	Bowers, Deaton, Oser, Prince and Kolb (1993)
Collaboration	Pilot/Controller	Morrow, Lee, and Rodvold (1991)
Message Length, Training	ATC	Morrow and Rodvold (1993)
Routine Operation Problems	Controller-Pilot	Morrow, Lee, and Rodvold (1993)
Culture Difference	Cockpit-Cabin	Chute and Wiener (1995)
Modes of Communication	Pilot-Pilot	Zimmer and Scheuchenpflug (1995)
Aircraft Radio Communication	Radio	Weller and Wickens (1991)
Communication Strategies, Personalities, and Crew Performance	Airline Captains	Orasanu (1991)
Crewmember Communication	Astronauts and Cosmonauts	Kelly and Kanas (1992)
Bilingual ATC	ATC	Stager and et al (1980)
Data-Link Communication	Controllers and Pilots	Kerns (1991)
Mixed-media Communication (Voice, Data Link, Mixed ATC)	Flight Deck	McGann, Morrow, Rodvold and Mackintosh (1998)
Satisfaction, Information Exchange	Cockpit-Cabin Crew Interaction	Skogstad and et al. (1995)

English in Aviation

In 1995, then-Department of Transportation secretary Pena recommended requiring all commercial pilots to pass a test for proficiency in speaking English.

Famous quote about communication/ language

#1 Even in face-to-face interaction, speech is a complicated process. Language not only conveys information but also express a worldview… there is room for distortion, uncertainty and ultimate conflict.

#2 In order to transmit proper meaning, the encoder and decoder must be on the same wavelength. They must speak the same language. We do not hear with our ears, we hear with our minds. And we are different from one another. All of us suffer from selective perception. What we hear depends on who we are. (Turney, 1997; Brightman, 1988).

References

Bowers, C., Deaton, J., Oser, R., Prince, C., and Kolb M. (1993). The Impact of Automation on Crew Communication and Performance. *Proceedings of the Seventh International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp.573-577

Chute, R. D. and Wiener, E. L. (1995). Cockpit-cabin communication: I. A tale of two cultures. *The International Journal of Aviation Psychology, 5(3),* 257-276.

Fallon, B.J. (1997). The Extent to Which Pilots and Controllers Consider Accents or Dialects Cause Difficulties in ATC Communications. *Proceedings of the Ninth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp.76-80

Fegyveresi, A.E. (1997). Vocal Cues and Pilot/ATC Communications. *Proceedings of the Ninth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp. 81-84

Fischer, U. and Orasanu, J. (1999). Say it again, Sam! Effective communication strategies to mitigate pilot error. . *Proceedings of the Tenth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp.

Kelly, A.D., and Kanas, N. (1992). Crewmember Communication in Space: a Survey of Astronauts and Cosmonauts. *Aviation, Space, and Environmental Medicine*. pp. 721-726

Kerns, K. (1991). Data-link Communication Between Controllers and Pilots: a Review and Synthesis of the Simulation Literature. *The International Journal of Aviation Psychology*, 1(3), 181-204

Logsdon, E.W., Infield, S.E., Lozito, S., Mackintosh, M-A, McGann, A., and Possolo, A. (1995). Cockpit Data Link Technology and Flight Crew Communication Procedures. *Proceedings of the Eighth International symposium on Aviation Psychology*. pp. 324-329

Mattews, B.G. and Hahn, C.K.G. (1987). Voice Communication Errors in the Air Traffic Control Environment. Proceedings of the Annual Conference of the Human Factors Association of Canada. Montreal, Quebec. pp. 167-170

McGann, A., Morrow, D., Rodvold, M., and Mackintosh, M-A (1998). Mixed-media Communication on Flight Deck: a Comparison of Voice, Data Link, and Mixed ATC Environments. *The International of Aviation Psychology*, 8(2), 137-156

Merritt, A. and Ratwatte, S. (1997). Who Are You Calling a Safety Threat?! A Debate on Safety in Mono- versus Multi-cultural Cockpits. *Proceedings of the Ninth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp. 661-666

Morrow, D.G., Lee, A.T., and Rodvold, M.A. (1993). Analyzing Problems in Routine Controller-pilot Communication. *The International Journal of Aviation Psychology*, 3, 285-302

Morrow, D.G. and Rodvold, M.A. (1993). *The Influence of ATC Message Length and Timing on Pilot Communication (NASA Ames Contractor Rep. No. 177621).* Moffett Field, CA: National Aeronautics and Space Administration Ames Research Center.

Orasanu, J. (1991). Individual Differences in Airline Captains' Personalities, Communication Strategies, and Crew Performance. *Proceedings of the Human Factors Society 35th Annual Meeting.* pp.991-995

Orasanu, J., Davison, J., and Fischer, U. (1997). What Did He Say? Culture and Language Barriers to Efficient Communication in Global. *Proceedings of the Ninth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp.673-678

Sexton, J.B. and Helmreich, R.L. (1999). Analyzing Cockpit Communication: the Links between Language, Performance, Error, and Workload. *Proceedings of the Tenth International Symposium on Aviation Psychology.* Columbus, OH: The Ohio State University. pp. 689-695

Skogstad, A., Dyregrov, A., and Hellesøy, O.H. (1995). Cockpit-cabin Crew Interaction: Satisfaction with Communication and Information Exchange. *Aviation, Space, and Environmental Medicine.* Vol. 66, No.9. pp. 841-848

Stager, P., Proulx, P., Walsh, B., and Fudakowski, T. (1980). Bilingual Air Traffic Control in Canada. *Human Factors*, 22(6), 655-670

Turney, M.A. (1997). Analysis of Gender Bias in Controller-Pilot Communication. *International Symposium on Aviation Psychology*, 85-89

Wulle, B.W. and Zerr, A.R. (1997). Factors that Affect Controller/Pilot Communications: Readbacks, Volume of Information, Experience Level, Personal Problems, Standard Phraseology, and the Relationship to Safety. *Proceedings of the Ninth International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University. pp. 90-94

Appendix II.

Typical ASRS Examples of the Four Error Locii

1. Synchronous Traffic-related

172961 SYNOPSIS

ACR LGT at hold line for RWY sees SMA pass in close PROX LNDG on RWY, TWR not speaking in English.

NARRATIVE

Waiting short of hold line 05r for DEP, 05l closed, TWR is talking to numerous ACFT in Spanish. In my limited Spanish can tell TWR is CLRING numerous ACFT to land. Looking across cockpit talking to f/o when outside his window appears an early 1960's SMA y, faded red in color with Mexican registration, approx 100' to the r of our airplane parallel to us less then one wing span (30') off the GND in a 30-40 deg bank HDG around our nose for the RWY (05r) to land. ACFT passed BTWN us and RWY, approx 60' off our nose! Shocked, we contacted the TWR who said they had CLRED the ACFT to land but would not answer any other attempts by us to acquire info by radio! Why can't they speak English like most of the worlds ATC sys?

2. Synchronous Intra-cockpit related

138028 SYNOPSIS

NMAC at night. ACFT x on VFR on top descent and ACFT y opposite direction IFR.

NARRATIVE

We were descending on VFR on top from 12500 MSL. Center had advised us of IFR TFC at 9000 MSL, 1 o'clock, SBND. The TFC was an SMT jet with APCH lights on. I spotted the ACFT and advised ATC as such. My F/O was of south American background having trouble understanding English. The F/O was constantly missing radio calls and having difficulty in understanding the instrument APCH we were about to do. We were to the point the FO was a DISR rather than help. I was keeping an eye on the SMT and estimated it would pass off well to our rear. The f/o failed to call the BC loc alive and was calling out wrong MDA numbers. While shepherding the f/o along and glancing at his HSI for course interception, I had lost sight of the SMT y FLT path. Looking up I realized the SMT was closing much too close. I leveled off at about 9200' AGL and started a slight left turn. The SMT spotted us and started a left turn. The f/o looked up and gave a cry. The SMT asked about the TFC passing overhead. I apologized to the SMT. This incident was my error because I allowed cockpit DISTRS to let us get too close to the TFC. In the future I will start a new f/o earlier in the DSCNT checklist and APCH review, stay on a hard IFR CLRNC (no on top) and try to maintain continuous outside contact.

3. Synchronous Other Groups (ground crew, cabin crew, operations)

430515 SYNOPSIS

An MD super 80 was unable to taxi out after pushback from the gate due to failure of the pushback crew to remove the nose gear steering pin. FLC was unable to communicate with the GND crew due to a severe language prob.

NARRATIVE

I would like to relate to you the following story that not only happened to me this morning in ZZZ but in a couple of other cities as well. During pushback, it was just about impossible to communicate with the GNG man due to his very, very poor English. I was able to once again get around this by me asking all the questions back to him and asking for an affirmative or negative. At the completion of the push, when it was time to be shown the pin, blink the taxi light the GND man holds up his gloved r hand, and crossed wands in his l. After I blink the light I get the salute and with CLRNC from GND I try to taxi but can go nowhere. It seems that once again I am shown a bypass pin that does not have a red flag on it -- or came from his pocket. Ops was called and someone else came to remove the real pin. I am not sure what the solution is here but I believe that all pins should have long streamers attached to them and the GND personnel should not be allowed to have their own pins.

4. Asynchronous Written Communications

502081 SYNOPSIS

AN air carrier after TKOF at 1400 ft declared an EMER and diverted due to #2 eng thrust REVERSER deployed.

NARRATIVE

I was called over to ACR to placard a r eng REVERSER unlock warning LITE. Never having deactivated a 717's REVERSER I called their MAINT coordinator and asked to have the PROCS faxed. Received a fax cover sheet and 3 pages from ACR MEL manual for the 717. When finished with the deactivation called back ACR MAINT to make sure I was completing the proper signoff in the logbook and get a MAINT CTL number for the placard. The coordinator never mentioned anything about pinning the deflector doors. They were mentioned in the MEL, but the verbal language in the manual threw me off. After lift off the 717's r eng REVERSER deployed at 1400 AGL. The PLT shut down the RT eng while keeping a airspeed of 200 KTS. The ACFT returned safely. While going over the PROCS again with MAINT coordinator, found that the 4th page of the deactivation was not faxed. Without this page, missed the crucial step of pinning the deflector doors closed. This would not have happen if the MEL would state in the beginning paragraphs of the steps what had to be deactivated along with graphics and explanation. ACR XXX keeps these books on the ACFT at all times. Some of the airlines don't, you have to depend on getting info over the fax. I work at ACR XXX airlines and many times we are contracted out to work on other carriers. At ZZZ, I work on 6 carriers besides ACR XXX which has us working on 3 different kind of ACFT that we don't work on a daily basis. More in depth training would help. Callback conversation with RPTR revealed the following info: the RPTR stated the crew discovered #2 eng thrust REVERSER unlock light on during a PRE FLT CHK. The RPTR said he was contacted by the ACR to call the MAINT CTLR to get PROCS for deferring the #2 eng thrust REVERSER. The RPTR stated upon contacting the ACR MAINT CTLR was advised to follow the MEL special PROCS and render the thrust REVERSER inoperative. The RPTR said a request was made to the MAINT CTRL to fax the PROCS as the RTPRS experience was limited to a few hrs of logbook and serving training on this new airplane. The RPTR said the ACR MAINT CTLR faxed three sheets of PROCS with no page identification as 1 of 3, 2 of 3 and 3 of 3. The RPTR said the three page document was accomplished and assumed that the deflector doors were stowed meant they were in the lock pos. The RPTR said the airplane was dispatched and at 1400 ft the REVERSER deployed incurring damage to the deflector doors linkage. The RPTR stated it was then discovered a 4th page with the proc for installing lock pins locking the deflector doors in the forward thrust was never sent by the ACR MAINT CTRL. The RPTR said that, Boeing revised the proc adding pictures of the lock pins location. The RPTR stated the only work experience gained on the 717 was serving oil, hydraulics and tires. The RPTR said more training on the contract ACFT we are assigned to work would help but none of the carriers do it.

Appendix III.

Informed Consent (Focus Groups)

I, ________________________________, agree to participate in this research project on "Language-Related Errors in Aviation Maintenance and Inspection" that is being conducted by University at Buffalo from the FAA's Human Factors group.

I understand that the purpose of this study is to hold a group interview to find out about how to reduce language errors in aviation maintenance and inspection; we will discuss our general ideas about language errors in aviation maintenance and inspection.

I understand that the study involves a focus group interview that lasts about an hour, which will be audio taped (video taped).

I understand that my participation in this study is entirely voluntary, and I may withdraw from the study at any time with penalty. If I do withdraw from the study, I understand that this will have no effect on my relationship with the FAA or my company.

I understand that all the information I give will be kept confidential to the extent permitted by law, and that the names of all the people in the study will be kept confidential. Data will be logged according to participant number; data will not be associated with participants by name. Code number, identity and audiotapes will be kept in a locked file of the investigator. Only the investigator has access to the file. Audiotapes will be erased as soon as the data has been analyzed.

The members of the research team have offered to answer any questions I may have about the study and what I am expected to do.

I have read and understand information and I agree to take part in the study.

Name (print): _____________________________________ Date: ________________

Signature: ____