TOOLS AND TECHNIQUES FOR EVALUATING THE EFFECTS OF MAINTENANCE RESOURCE MANAGEMENT (MRM) IN AIR SAFETY

2001 Report of Research Conducted
under NASA-Ames Cooperative Agreement No. NCC2-1156
(SCU Project # NAR004),
to NASA-Ames Research Center, Moffett Field, CA
and FAA Flight Standards Service, Washington, DC,

April 30, 2002.

James C. Taylor, Ph.D.
School of Engineering
Santa Clara University
Santa Clara, CA 95053-0590

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TOOLS AND TECHNIQUES FOR EVALUATING THE EFFECTS OF MAINTENANCE RESOURCE MANAGEMENT (MRM) IN AIR SAFETY[1]

James C. Taylor, Ph.D.
School of Engineering
Santa Clara University
Santa Clara, CA 95053-0590

SUMMARY

This research project was designed as part of a larger effort to help Human Factors (HF) implementers, and others in the aviation maintenance community, understand, evaluate and validate the impact of Maintenance Resource Management (MRM) training programs, and other MRM interventions; on participant attitudes, opinions, behaviors, and ultimately on enhanced safety performance. It includes research and development of evaluation methodology as well as examination of psychological constructs and correlates of maintainer performance.

In particular, during 2001, three issues were addressed. First a prototype process for measuring performance was developed and used. Second an automated calculator was developed to aid the HF implementer user in analyzing and evaluating local survey data. These results include being automatically compared with the experience from all MRM programs studied since 1991. Third the core survey (the Maintenance Resource Management Technical Operations Questionnaire, or "MRM/TOQ") was further developed and tested to include topics of added relevance to the industry.

BACKGROUND

MRM Evaluation Tools

Since the early 1990s research into the field of "macro" human factors in aviation maintenance indicates that many airlines have opted to improve awareness of communication, safe practices, and professionalism. But only a few of these programs have also included skill-based training in such topics as decision-making, or assertiveness (Taylor & Robertson, 1995; Taylor, 1998), and recently written communication (Taylor & Thomas, 2001a). Protocols and worksheets for capturing this last topic -- archival written communication -- were developed during 2001 and their results are reported here. Specifically written work turnover, a behavior emphasized in a particular MRM training program, was targeted for measurement in order to evaluate changes in this important

[1] The research reported here, as well as this report, benefited greatly from the help of Professor M.S. Patankar (San Jose State University) and Mr. Robert Thomas, the program's graduate research assistant during 1999-2001. Excellent guidance and encouragement by the project sponsors' technical officers, Ms. Jean Watson and Dr. Barbara Kanki, was always available and freely given. Finally, this research was supported throughout in the unstinting cooperation and assistance of our five partner companies during 2001 who remain unnamed, but not unappreciated.

behavior as a result of the training. This case provides added evidence for the effectiveness of MRM training, but perhaps more importantly it offers a model and encouragement for airlines wanting to create measures and collect data for performance targeted for improvement, but not currently measured. It also offers a caveat to managers who wish to succeed in such efforts over the long term. This case and the performance measures we developed are presented in section I below.

User-centered tools and usability

An important set of deliverables from our research program includes methods and practices to assist airline companies and other users collect psychological and behavioral data, while maintaining the conditions required for reliability and validity of those data. Over the course of this program such methods have been planned and developed. They are now documented and are ready for distribution. A shortened version of our core survey questionnaire, the Maintenance Resource Management Technical Operations Questionnaire (or "MRM/TOQ") was tested and validated during 2001 and is reported in Section III below. Such data collection methods are, however, of little use to the HF implementer without parallel methods of analysis and interpretation. Part of the ongoing work of this program since 1991 has been the collection and organization of a "benchmark" database of psychological and behavioral data from aviation maintenance personnel in the United States. The second of our three products this year are interpretive tools and algorithms, incorporating that benchmark, which form a companion to the data collection instruments described in Section III. These tools are collectively called the MRM/TOQ Evaluation Results Calculator (ERC). One part of this tool is the "MRM attitude and opinion profile." It provides the calculation of percentile scores for any maintenance work unit or site entered by the user. These profiles, in the form of standard scores ("Z"), can be used to compare the percentile rank of MRM attitudes and opinions in any given company at any stage in its MRM program with attitudes from a large database of like employees – called the "Benchmark dataset." The second part of the tool is a statistical test of attitude and opinion change between "before" and "after" MRM training. This statistic, or "t" test between pre- and post-training surveys is calculated automatically after the user has entered the individual questionnaire answers. The ERC is described in Section II below.

Measuring the Constructs of Trust & Professionalism

<u>Professionalism and trust in a fluctuating, mobile and transient maintenance workforce.</u>

Recent studies have confirmed the uncertain nature of employment security in aviation maintenance. The influence of economic conditions on maintenance employment security is strong. According to a study by the National Research Council, airlines respond to industry recession with reduced employment and lay-offs. The industry's employment levels gyrate substantially from year to year and during peak hiring periods less qualified applicants become more attractive candidates (Hansen & Oster, 1997). It is reasonable to assume that experienced mechanics' trust in companies that lay them off in bad times will be diminished; and if rehired during good times these mechanics could well resent the less-qualified applicants hired in at the same time.

With the increased use of third-party maintenance facilities by airlines, the airline industry seems to be moving toward virtual organizations which further lowers employment security. Almost all the functional units of an airline could be contracted out to third-party vendors, who specialize in such operations, and the core of the airline could focus on managing the services of these specialty vendors. This seems to be an attractive economic possibility, but the implications of such an approach could be catastrophic (NTSB, 1997). If the trend toward outsourcing continues, virtual airlines are inevitable. A likely byproduct of such an organizational structure is a highly mobile and transient workforce. Therefore, from the maintenance perspective, mechanics function as independent contractors with the repair stations and/or airlines. This could result in a workforce that is more directly dependent on the fiscal fluctuations, less loyal to employers, and more independent-minded than in the past.

The important role of the FAA in creating and supporting a maintenance safety culture has earlier been noted (Marske & Taylor, 1997). This past year we have addressed the concepts and measurements of "professionalism" and mutual trust in an aviation maintenance environment because they are postulated to be keys to building safe virtual organizations in uncertain times.

The new version of a shortened and revised version of the core survey questionnaire, the Maintenance Resource Management Technical Operations Questionnaire (or "MRM/TOQ") measures trust and professionalism – core elements of a safety culture -- developed with industry partners. The noteworthy results are reported in section III below.

I. MRM Performance Evaluation Tools

"Written Communication Practices as Impacted by a Maintenance Resource Management Training Intervention"

Written communication was examined in the context of the maintenance station of a large airline company that had implemented a Maintenance Resource Management (MRM) training program. Data were collected and analyzed from written work turnover documents to explore written turnover practices and examine training effects on such practices. Trends in archival paperwork error data were also examined throughout training periods, along with respondent recollections of training content regarding written communication. Implications for successful program management, and for future research geared to airline maintenance error reduction are discussed.

A concept of central importance to aviation safety that is covered in most Maintenance Resource Management training programs is the practice of clear and thorough communication. A number of airline accidents caused by human factors can be traced to erosion in either verbal or written exchange of critical information (Taylor and Christensen, 1998). The role communication has been shown to play in human factors error underscores its value as a research construct. More specifically, written work-turnover and other documentation represent critical aspects of high-risk organizational systems. Because complexity of such high-risk systems has been a theorized contributor to accident rates (Perrow, 1999), the clarity and accuracy of written turnover are a critical leverage point for maintenance error reduction. Essential components of accountability, information flow and quality, and safety assurance hinge on the proper and complete use of written communication.

As written communication is so vital to safety in airline maintenance, it is no surprise that efforts have preceded the present research to increase the quality of documentation. Hutchinson (1997) examined work cards in a large repair station and found that over a twelve-month period, 40% of them contained vague, ambiguous or abbreviated phrases that missed intended standards of federal aviation regulation. A feedback system was implemented on the hangar floor whereby work-record error rates were posted daily for mechanics to see. Being shown error rates with such rapid feedback had a profound impact on documentation practices, with the 40% error rate dropping to zero in eight weeks.

Taylor and Christensen (1998) highlight the importance of written communication in airline maintenance, calling it "the bedrock of all communication in maintenance." Of all modes of communication operating in such a system, these authors see the written message at the core. They cite three critical factors in improving written communication in airline maintenance. One is employee participation. Involving employees in the improvement process has shown to be a positive force in reducing paperwork errors (Taylor, 1994). A second important factor is ergonomics and forms design. Research has explored this area to maximize the clarity and usefulness of work documents in airline maintenance (Patel, Drury and Lofgren; 1994). Finally, measurement and feedback on performance is important as Hutchison (1997) has shown. Efforts to measure patterns in written communication and provide feedback to researchers, managers and mechanics about improving this skill help initiate a process geared toward safer airline maintenance departments.

The present study marks an initial attempt to measure some qualities of written communication beyond the absence or presence of discrepancies. It is also an effort to examine the effects of a Maintenance Resource Management (MRM) training program with modules on improving written communication in general and written turnovers in particular. That training took place in two phases. For the large repair hangar described here, phase one occurred from January 2000 through April 2000 (the time it took for all participating employees to go through the one day training). Phase two began for this "subject site" in June 2000 and concluded in August of 2000. Other sites in the same company (hereafter called the "subject company") have started the training, but have not yet completed it. Their interim results will also be compared with the subject site. Further comparison uses some results from MRM programs in two other companies, whose programs did not include modules on written communication and whose training was completed in one phase.

A definition of written turnover. "Turnover" in organizations employing shift work denotes passing of partial or incomplete jobs from one shift to the next. In the present case, written turnover is the documentation of work performed and passed from at least one shift to another during aircraft overhaul. Such a written account, according to most FAA-approved maintenance manuals, must be recorded for the employee attempting to complete a job on a subsequent shift. Written turnover in the airline industry serves two crucial purposes: 1) it leaves a paper trail of accountability for each step in a set of maintenance procedures, and 2) it provides the next work shift with information vital to assuming the next stage of a task, and ultimately completing the entire job. Important to conclude from this description is that the work card represents a carefully crafted centerpiece to a system of checks, re-checks, accountability and safety nets. Written turnover practices represent the critical human component to this system that ultimately determines the system's ability to attenuate maintenance error.

For the subject company, written turnover was emphasized primarily in Phase I of the training, with cursory reminders occurring during Phase II. Specifically in Phase I, Clarity, Completeness and Correctness ("the three C's") were stressed as critical to written communication. Exercises demonstrating the importance of such written communication included a task that involved following a complete set of directions, the clarity (or unclarity) of which was not apparent to participants until the very last step. A second exercise had participants write a work document entry, striving for enough clarity, completeness and correctness to enable a second, naïve participant to correctly assemble a set of objects in a particular fashion based on what was written. Additionally, considerable time was spent in discussing and examining company turnover documents and how to fill them out properly.

Based on the emphasis in Phase I toward written communication and turnover, our expectation was that turnover quality and attitudes toward written communication would be most improved immediately following this period, and that errors in written documents would be diminished. Stated more specifically, our hypotheses were that following training: 1) the subject site would show significant increase in intentions to improve written turnover, 2) performance data such as paperwork errors should show a decrease, and 3) the actual written turnovers would improve in length (completeness), in legibility (clarity) and in content (correctness), compared with appropriate baselines. Specifically, intentions could be compared with respondents in other companies not receiving the specific training modules, discrepancies in written documents could be compared before and after the training; and current written turnover length, completeness and clarity could be compared with the subject site's prior performance in the year preceding the training.

Method

Subjects and Samples

The subjects (employees of the "*subject site*") are aviation maintenance repair mechanics and quality inspectors, plus their immediate supervisors and middle managers who have completed a two phase MRM training program in a maintenance repair site belonging to a large airline. The *subject site* is unique in that *all* its employees have completed both phases of this MRM training, which emphasized improving written turnovers. Initial field interviews in the subject site during and after the training period revealed that many participants especially valued its sections on written communication and turnover. Results from this subject site are compared with other heavy maintenance facilities in the same company ("*subject company*") that had begun, but had not yet completed, the same MRM training. Survey results from the subject site and its larger company are compared with heavy maintenance operations in two other airlines (*comparison companies "A" & "B"*) whose MRM training did not include the topics of written communication or improving written turnovers. Survey respondents in the comparison companies include mechanics, inspectors, management and support personnel in similar proportions to the subject company.

Data

Assessment of Written Turnover Quality

The documents from which we assessed the quality of written turnover in the subject site consist of "non-routine work cards" that are included in the document packages resulting from aircraft heavy maintenance overhaul, or "maintenance checks." These "checks" are a set of preplanned maintenance inspections and procedures, which are conducted at required intervals for aircraft of a particular model. The "non-routine work" results from defects or damage found during the preplanned inspections. The overhaul process studied here is called "C-check" in the industry and it is a fairly extensive overhaul process. Because the set of maintenance procedures is so large for a C-check, the subject company has divided theirs into six parts that can each be performed usually in three to four days (nine to twelve eight-hour shifts).

For each non-routine job card they work on, these maintenance employees are required to sign (actually, stamp) the entries for which they accept responsibility using their own stamp issued with their employee ID number. The employee who stamps the "repaired by" section on the front of the card accepts responsibility for his/her section, as well as any entries on the card that have not been stamped. The "checked by" section of a work card is generally stamped by an inspector, meaning this individual is accepting responsibility that the completed job has been conducted properly, and that any "required inspection items" have been properly inspected.

Sampling Written Turnover Data

The subject site's data sample represents turnover data entries recorded by the mechanics, inspectors, supervisors and managers in this one heavy maintenance station. All of these people had completed both phases of the MRM training at the station during the preceding year. Turnover data were collected and coded from completed work documents during visits to the company archives. A purposeful sample of document packages was drawn. We could not review all non-routine work cards for the subject station with the time and manpower available. We therefore sampled the documentation of approximately 10% of all C-checks performed at the subject site for a two-year period. Because no grounded or theoretical reasons could be conceived to choose one phase of the check over another, our sample was selected without regard for phase of check other than gaining an adequate proportion of the total checks conducted in 1999 and 2000. The population consisted of 179 document packages in 1999 and 169 more in 2000, a total of 348. From this, a sample of 32 packages was drawn, with a roughly even distribution among each of the two years included. Sixteen packages, each from 1999 and 2000, were included in the sample. Phase I training began in January of 2000 and concluded in March of 2000. Phase II began in June of 2000 and concluded in August of 2000.

Figure 1. Total number of turnover entries for each sampled month in 1999 and 2000

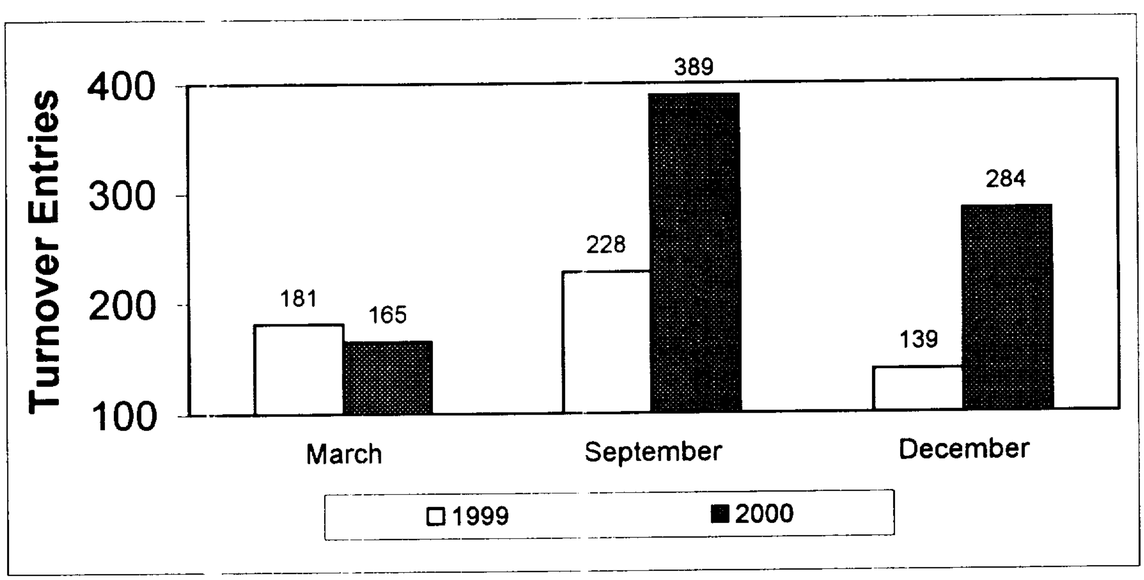


Figure 1 shows the distribution of the 1,386 separate turnover entries obtained from the 32 package sample. March, September, and December were selected as appropriate periods in each year to draw samples based on their proximity to 2000 training onset and conclusion. The sample chosen allows examination of changes in written turnover performance at critical points coincident with onset and termination of training. It also allows for comparisons to baseline from the same months in 1999, during which training had not yet been implemented.

Coding the Turnover Data

Turnover written in response to the initial inspection and defect description were assessed and coded by two raters. Turnover length (completeness) was recorded by counting the number of words included in the turnover, including reference numbers and abbreviations. Legibility (clarity) was recorded by assigning a rating from 1 (completely illegible) to 4 (completely legible) for each turnover entry. Content (correctness) was recorded by counting the number of times an entry included "what was done," "where I stopped/how I left the situation," (these are considered correct); or "what to do next," which was considered incorrect by industry standards. Raters were compared on turnover length, content and legibility for each time block separately using independent samples t-tests. Number of words (length) and content were stable across raters, with no significant differences between raters. However, comparison of raters on legibility yielded significant differences at almost all time blocks, reflecting the increased subjective judgment inherent in this measure.

Measuring Paperwork Discrepancies

The subject company's airline maintenance department, in which the new training on written communication had been implemented, has measured and reported total paperwork discrepancies for each station by month between 1995 and 2001. The subject company's monthly reports were made available to the researchers for use in identifying improvement trends coinciding with the training. In order to compare the subject site with others in the subject company, the raw data contained in these reports was corrected for station size through the use of personnel headcount. Trends for these corrected data were examined for a period prior to the onset of the training and for the available months thereafter. Viewing these trends we expected to find the most impact of the MRM training on the subject station in which all employees had completed both phases; and to a lesser degree in the other maintenance stations in the company where not all employees had yet been trained.

Survey Measurement

Employee intentions to improve their written communication following their training, and their reports of actually doing so, were collected using post-training surveys. Survey data were collected from the subject company and from two comparison companies, "A" and "B," using the Maintenance Resource Management – Technical Operations Questionnaire (MRM/TOQ), a well-tested and validated survey instrument

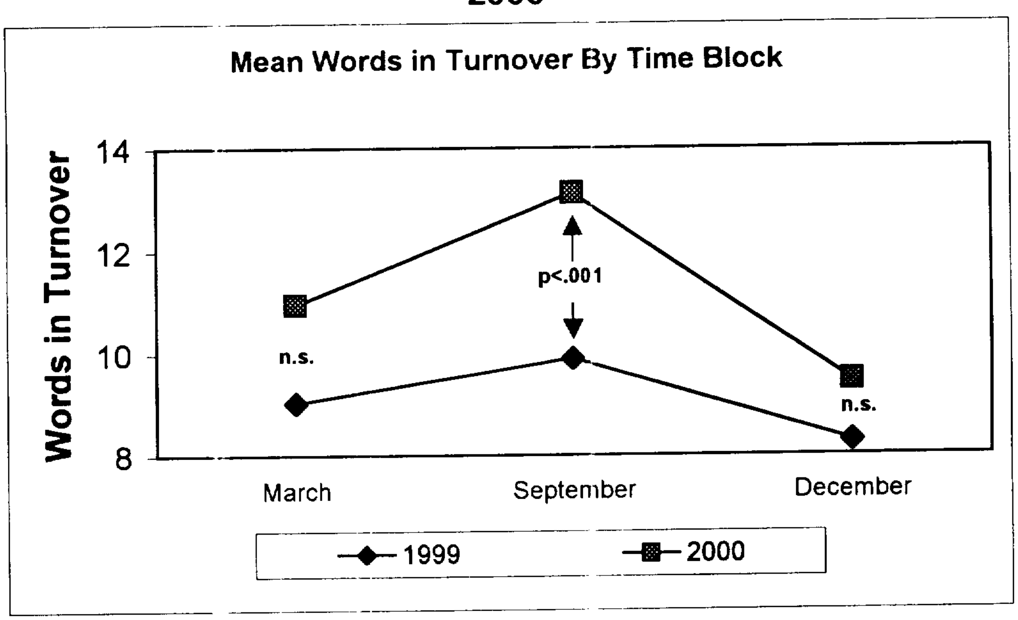
(Taylor, 2000). Training participants completed surveys immediately after their training. In the subject company's sites where training occurred in two phases, questionnaire data were collected after each phase. The MRM/TOQ data used to explore the effect of the training on written turnover come from responses to previously validated open-ended items that are subsequently coded into fixed categories (Taylor, 1998, 2000). Initial responses come from the immediate post-training questionnaire, in which participants were asked what was memorable about the training they had just received, and how they intended to use the training. Further responses were collected from participants several months after their training when these respondents received another MRM/TOQ in which they were asked to describe what changes they had actually made as a result of their training. Since the coding scheme included categories for both "writing more clearly," and "improving my turnovers," we expected to find such responses in greater proportion in the subject site, next most frequent in the remainder of the subject company, and the least in maintenance operations "A" and "B" where the MRM training curriculum didn't include written communication as a topic.

Results

Comparisons of Written Turnover Before and After MRM Training

Figure 2 shows the written turnover length for the "subject site" for 1999 (the year before MRM training) and 2000 (the year in which training occurred). As shown in Figure 2, the distribution of mean "number of words in turnover" arrayed across sampled months in each year are roughly parallel for this measure and higher for 2000.

Figure 2.
Turnover Length: Subject Site Comparison for Six Time Periods, 1999 and 2000



A one-way ANOVA was conducted for turnover length with time period as the factor, and it was significant (F=7.95, df=9, 2,083, p<.001). Tukey HSD post hoc analysis revealed the following: Turnover length remains fairly stable and free of significant variation across same months in 1999 and 2000. The exception is that in September 2000 (the month following the completion of all training), an increase is shown over the same period in 1999. The increase in length between December 1999 and March 2000 is also statistically significant, suggesting an improvement resulting from phase I training.

Figure 3.
Turnover Legibility: Subject Site Comparison of Six Time Periods, 1999 and 2000

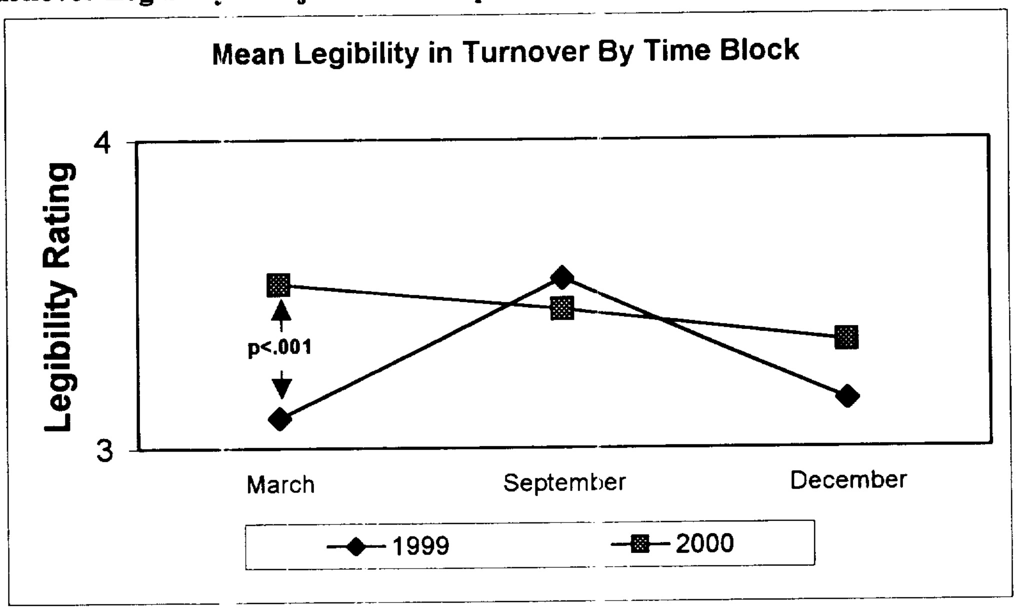


Figure 3 shows somewhat similar results for turnover legibility. The one-way ANOVA of turnover legibility is also significant (F=10.82, df=(9, 2,083), p<.001). Tukey HSD post hoc analyses revealed a significant higher level occurs in March 2000, immediately after Phase I training concludes than its counterpart a year earlier. Also, as with turnover length, a significant increase in legibility was found from December 1999 to March 2000 (suggesting an effect of phase I training). No other significant differences emerged for legibility.

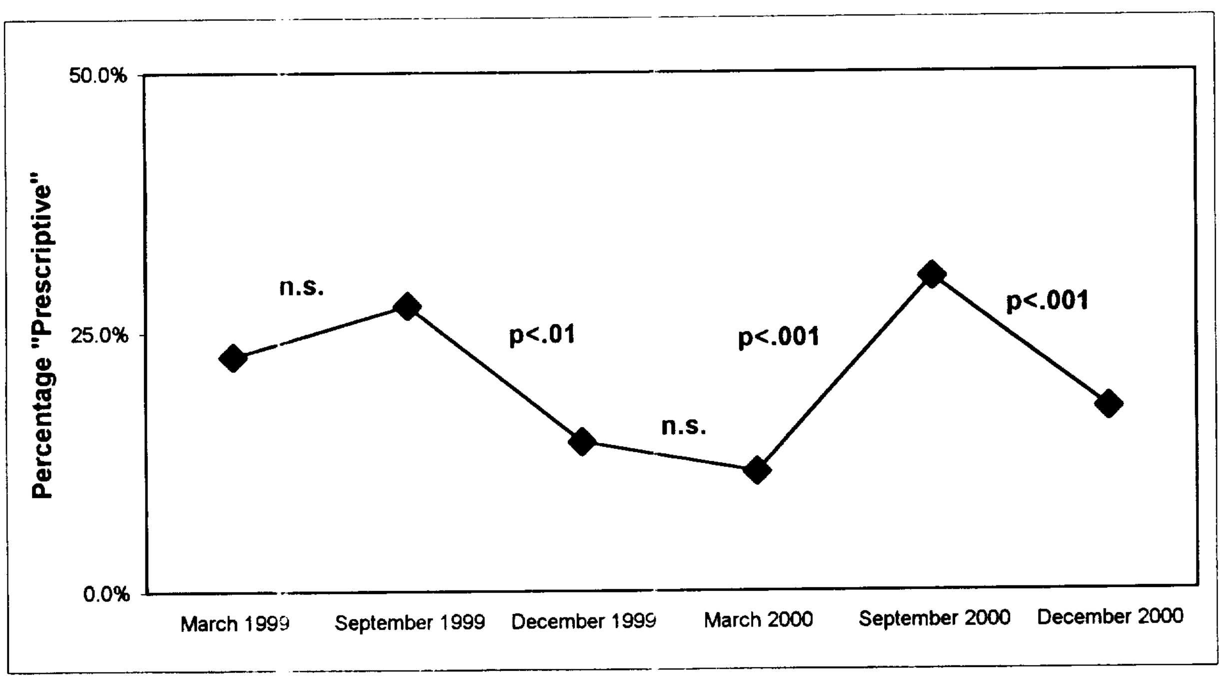
<u>"Descriptive" vs. "Prescriptive" Turnover Content</u>

Among the hypotheses tested in this research is the improvement in content and correctness of written turnover documents. As previously mentioned, policy at the subject company and elsewhere in the industry discourages maintenance employees from making statements in the turnover about what the next course of action should be for the employee receiving the turnover. This is because such statements can limit the decision making of the turnover recipient, and the suggested comment may be against authorized procedures. For this reason, we compared "descriptive" turnover (only stating "what was done" or "how the job was left") and "prescriptive" turnover (adding statements about what the next mechanic should do) on turnover length and legibility. Legibility was not different between "descriptive" and "prescriptive" turnovers (t= -1.95, df=2091, n.s.). However, for total number of words the "prescriptive" turnover entries had significantly more words than the "descriptive" turnover entries. Levene's test was significant for the t-test used for analysis (F= 32.70, p<.001), and the group sizes were unequal, necessitating a non-parametric analysis. The Mann-Whitney U test showed significant difference in mean ranks at z= -16.154, p<.001. The greater number of words in the "prescriptive" turnover is no surprise, as additional writing should be required to include direction about what should be done next. This finding reinforces a point made in the subject company's MRM training that longer turnover is not necessarily better turnover.

Unfortunately this advice did not have a measurable effect on performance. Figure 4 shows the percentage of "prescriptive" turnover entries across time blocks. An overall chi square test of the 6 time blocks by inclusion of prescriptive turnover was significant ($X^2(5)= 37.772$; $p<.001$). Post hoc chi square tests were conducted for adjacent time blocks, and significance values are shown in Figure 4. A significant decrease was shown from September 1999 to December 1999 ($X^2(1)= 8.654$; $p<.01$), a significant increase was shown from March 2000 to September 2000 ($X^2(1)= 22.044$; $p<.001$) and a decrease was found from September 2000 to December 2000 ($X^2(1)= 14.198$; $p<.001$). No clear effect of MRM training on writing "prescriptive" turnover can be discerned from the current analysis.

Figure 4.

Turnover Content: Subject Site's Percentage of "Prescriptive" Responses for Six Time Periods, 1999 and 2000



Job Title Comparisons

Because all maintenance employees do not perform the same roles and functions, researchers were interested in examining comparisons of turnover entries among job titles. One-way ANOVAs were conducted for turnover length and legibility with job title as a factor. Groups included mechanics, inspectors and managers for both dependent measures. The ANOVAs were significant for both legibility [$F(2,1825)= 29.68$, $p<.001$] and length [$F(2,1827)= 6.982$, $p<.001$]. Tukey post hoc analyses indicated that inspectors write shorter turnover than mechanics but write more legibly than both mechanics and managers.

Figure 5:
Mean number of words per turnover entry for subject site's inspectors, mechanics and managers across all time blocks

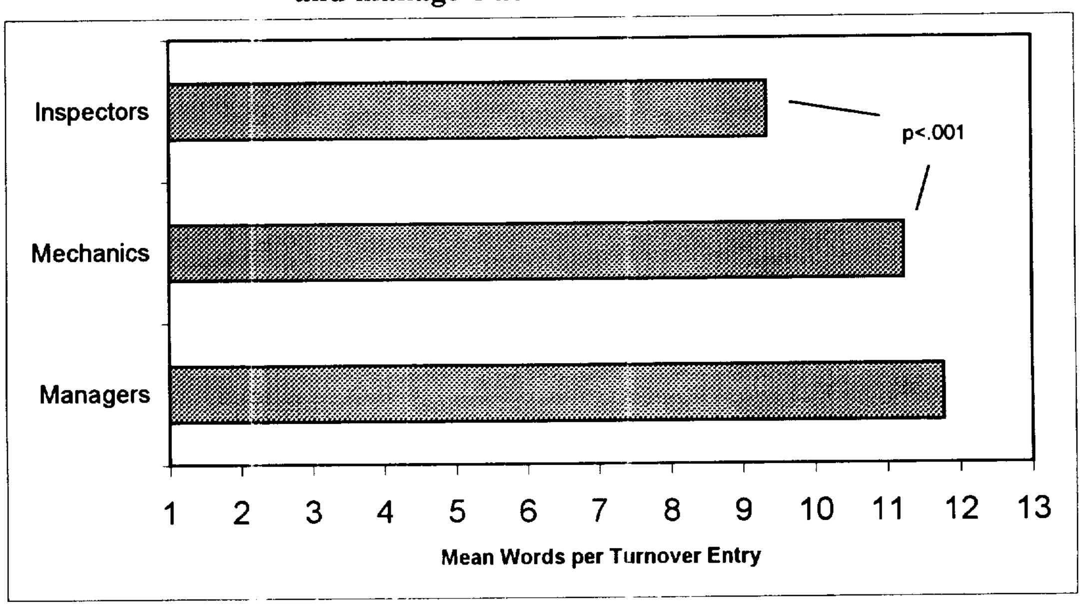
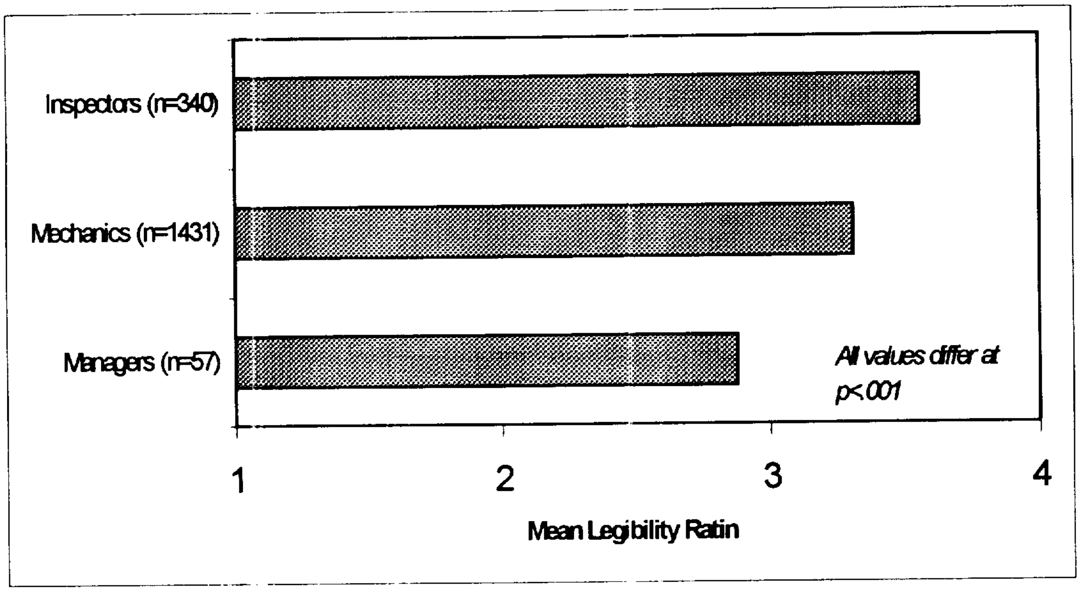


Figure 6:
Mean legibility rating per turnover entry for subject site's inspectors, mechanics and managers across all time blocks



Also recorded was the correctness of the written turnover. Each entry was dichotomously coded as having either included or not included what was done, how the situation or job was left, and what needed to be done next. Pearson's Chi-Square statistic was conducted for each of these variables in cross-tabulation with the three main job titles of mechanic, inspector and manager. Overall 2X3 cross-tabulations yielded significant chi-square statistics ($X^2(2)= 21.947$, $p<.001$), indicating a relationship between turnover content and job title. In 2X2 chi square tests, mechanics were shown to be more likely than inspectors ($X^2(1)= 32.807$, $p<.001$) and managers ($X^2(1)= 7.082$, $p<.01$) to write the prescriptive response, "What to do next". Managers and inspectors did not differ.

Paperwork Errors

Figure 7 shows the total number of errors per month from January 1995 to April 2001 for the subject site and the average errors per month for all remaining base maintenance stations in the subject company. A slight positive trend is shown in number of errors across time (the trend line for the subject site is solid and the trend line for the average of the remaining stations in the subject company is dashed), with a sharp increase occurring in 2000 and 2001. Both trend lines in Figure 7 shows a positive slope after 1998. This seems perplexing considering the ongoing training program in progress designed in large part to reduce these types of errors. However, a hiring freeze ended in the subject company at the beginning of 1998, and a number of young and less experienced mechanics began work for the company at the beginning of 1999.

Figure 7.
Paperwork Errors from January 1995 through April 2001

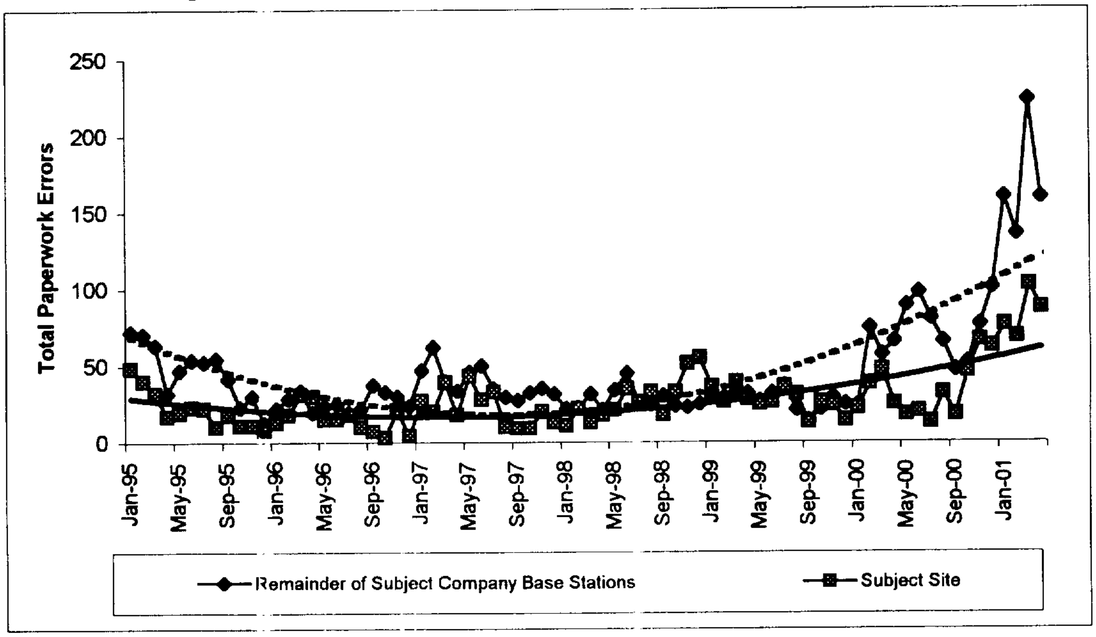
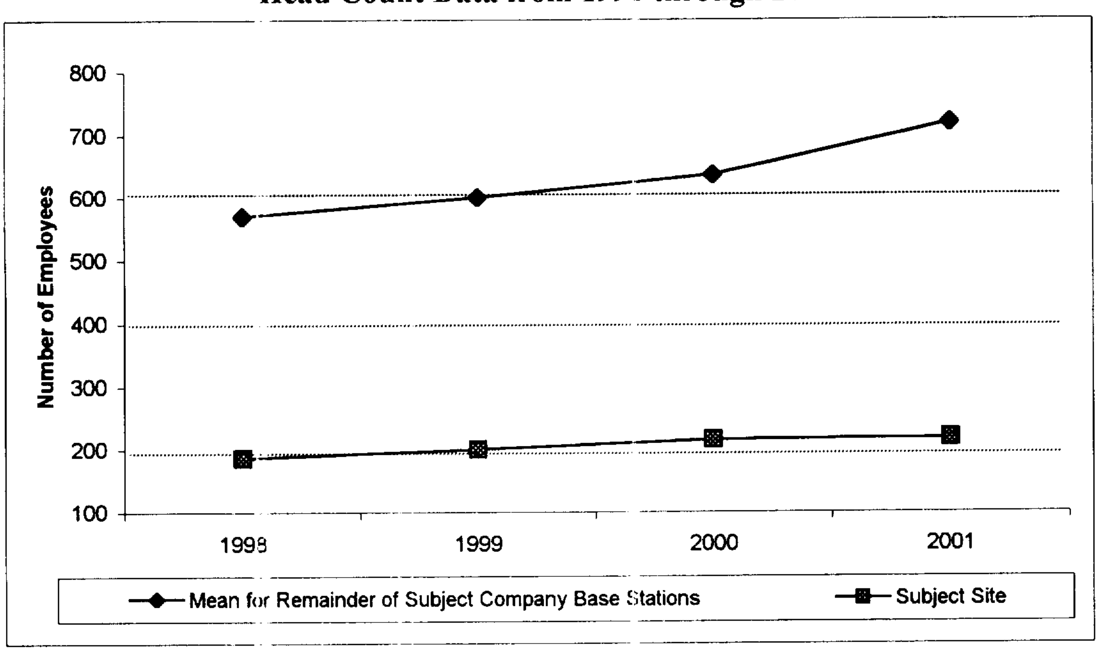


Figure 8.
Head Count Data from 1998 through 2001



Head count data is shown in Figure 8. This shows an increase in the number of employees from 1998 to 2001 in the subject station and the remainder. Head count data was not available prior to 1998.

We could easily expect that a population suddenly infused with new employees would yield an error trend with a positive slope. Any significant effects of MRM training are likely overshadowed by the propensity of a new hire to commit error. To assess the possible effects of new employees hired, we adjusted errors by head count and compared the trend line slopes before and after January 1999. Figure 9 shows the year 1998 and the different trends in paperwork errors between the subject site and the remaining heavy maintenance stations in the subject company. The subject site is less affected by new hires in 1998 and shows an error rate increasing more sharply than the head count rate over time, which shows an overall increase in errors per employee during this time.

Figure 9.
Paperwork Errors Adjusted for Head Count for 1998

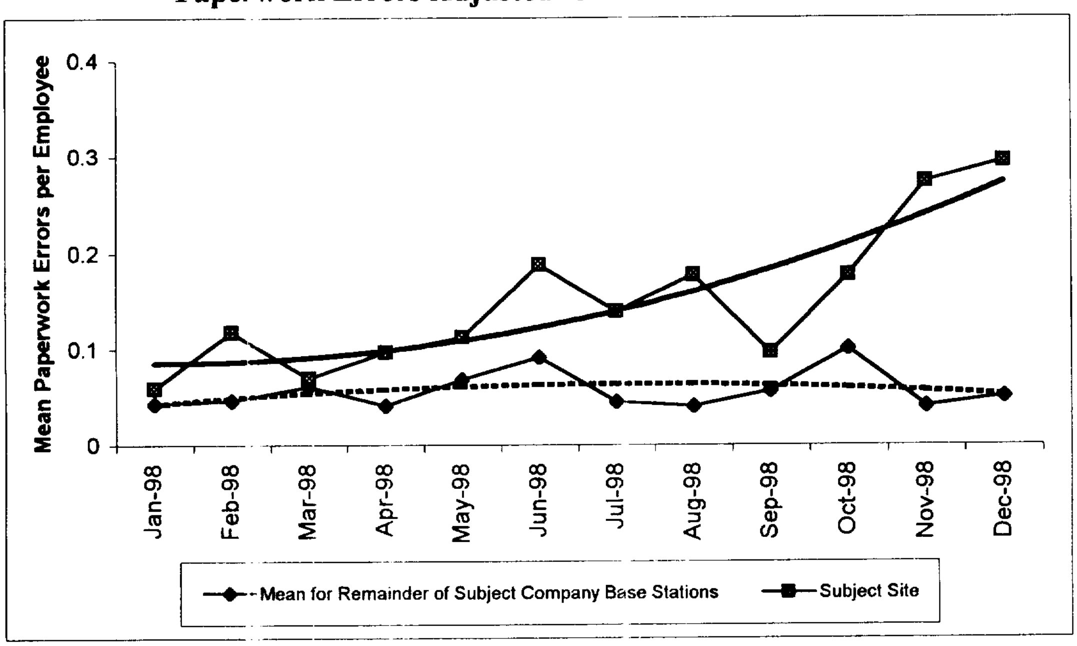
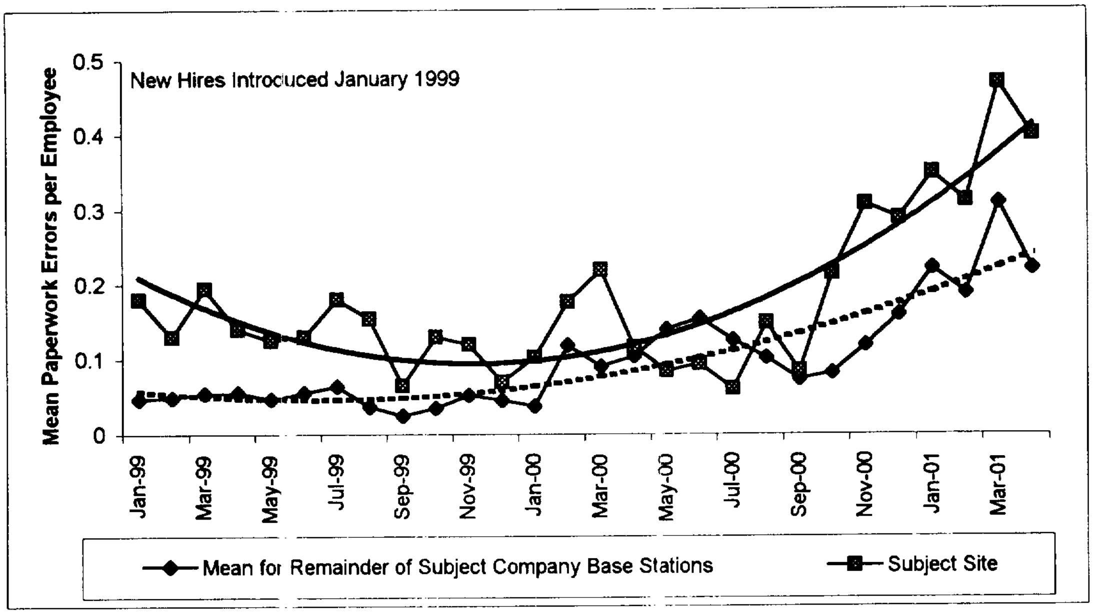


Figure 10.
Paperwork Errors Adjusted for Head Count for 1999, 2000 and 2001 (During Training and after New Employee Hiring)



For 1999 through 2001, corrected for head count, Figure 10 shows an increasing trend for both the subject site and remaining stations.. This similar shift in trend for both groups lends support to the idea that new and relatively inexperienced mechanics can be largely responsible for the diminished paperwork skills and the increase in paperwork error rates in 1999-2000.

Field Interviews and Survey Data

Recollections and Intentions

In field interviews conducted in June 2000, shortly after phase I training was completed, a sample of 46 maintenance employees from the subject site were asked what they remembered best about the training. "Turnover" tied for the highest response with "Case studies and videos" at a 15% response rate. This apparent enthusiasm and remembrance for written turnover was encouraging, since written turnover was a primary component of phase I training.

Following both phases I and II, the MRM/TOQ included the questions "what are good aspects of the training?" and "how will you use this training on the job?" Among the general themes that are coded for each of these, three bore some relationship to the topic of written turnover. Those themes were" improve turnovers," and "write more clearly," as well as "communication" (coded if the respondent wrote only the word "communication" and nothing else),. Data from the subject site are compared with the results from remaining heavy maintenance hangars in the same company; and both of those are compared with companies "A" and "B" that are engaged in similar heavy maintenance operations, but whose MRM training did not cover written communication.

Table 1 shows the degree respondents felt the three selected communication topics were memorable (or good) in the training they received.

Table 1.
Communication and Turnover Responses
"What were the good aspects of the training?"

What were the good aspects of the training?	"Improving turnovers"	"Writing more clearly"	"Communication"
Phase I Subject Site (n = 245)	7.4%	1.6%	4.2%
Phase II Subject Site (n = 263)	0	0.5%	2.1%
Phase I Remainder of Subject Company (n = 837)	7.3%	3.4%	7.3%
Phase II Remainder of Subject Company (n = 236)	0	0.4%	1.2%
Comparison Company A (n = 1,844)	0	0.3%	4.1%
Comparison Company B (n = 153)	0	0.6%	3.8%

The results in Table 1 reveal a difference among the six survey samples in their mention of memorable topics that is statistically significant (Chi Square = 41.62, df = 10, $p<.001$). These results show a substantial regard for the treatment of improving turnovers in the subject station and in the remainder of the subject company immediately following their phase I training. Improving turnovers was not mentioned at all in the two comparison companies following their MRM training and this is to be expected insofar as their training programs did not emphasize that topic. Likewise, and for the same reason, no mention of the turnover topic was made following the phase II training in the subject site and the remainder of the subject company. A smaller proportion in the subject sites mentioned clearer writing as a memorable aspect of their phase I training and this appears as a very small percentage following phase II training as well as for the two comparison companies. There appears to be little difference in the general "communication" topic among the six samples except that it seems to diminish in the subject site and remainder of the subject company after phase II training.

Table 2.
Communication and Turnover Responses
"How will you use this training on the job?"

How will you use this training on the job?	"Improving turnovers"	"Writing more clearly"	"Communication"
Phase I Subject Site (n = 245)	6.6%	8.1%	4.1%
Phase II Subject Site (n = 263)	1.1%	0.6%	3.0%
Phase I Remainder of Subject Company (n = 837)	15.6%	8.7%	6.1%
Phase II Remainder of Subject Company (n = 236)	0.1%	0.8%	3.5%
Comparison Company A (n = 1,844)	0	0.1%	7.2%
Comparison Company B (n = 153)	1.3%	0	7.8%

Table 2 shows the degree respondents expected -- as a result of their training -- to improve their turnovers, to write more clearly, or to just "communicate." It shows that participants in the subject station, and in the remaining heavy maintenance stations in that company, more frequently express intentions to improve turnover and write more clearly than in the other two companies. These respondents also most frequently expressed intentions to improve turnovers and write more clearly after phase I than after phase II. This reduction of intentions following phase II training is not a surprising finding considering these topics were not much emphasized in phase II content. The two comparison companies show minimal intentions to practice either improved turnovers or clearer writing. Once again, the general communication topic shows little difference among the six samples. The Chi Square test for difference among the six survey samples over the three response categories is statistically significant (Chi Square = 46.76, df = 10, p<.001).

Reports of Actual Behavior

Table 3 displays data collected from the subject company's MRM/TOQ following phase II, and shows and the degree to which respondents say they *did* improve their turnovers, they *did* write more clearly, or if they better communicated in general as a result of their training. These results are compared, in table 3, with data collected from respondents in the two comparison companies in a follow-up MRM/TOQ survey administered two months after their training.

Table 3.
Communication and Turnover Responses
"What changes have you made on the job?"

What changes have you made on the job?	PhaseII, Subject Site (n=180)	Phase II, Remainder of Subject Company (n=259)	Comparison Company A (n=585)	Comparison Company B (n=150)
"Wrote more clearly"	0.6%	2.3%	0	0
"Better turnovers"	1.1%	1.9%	0	1.3%
"Communication"	2.7%	1.9%	1.6%	6.0%
Chi Square = 10.66, df=6, n.s.				

These reports of behavioral change several months after the initial training cannot be said to support the prediction of respondents' actual change in written turnovers resulting from the training. Although Table 3 seems to show a slight trend in subject company respondents' reports of writing more clearly and improving their turnovers, the Chi Square test does not show a significant difference among the several samples.

Discussion

MRM Training Effects on Turnover Practices

The most direct evidence we have presented here, the analyses of written turnover length and legibility, does yield findings showing benefit of MRM training. For our subject site, which received the maximum effect of the training, turnover length increased over 1999 baseline levels after Phase II in September 2000. This is not a complete support of our hypothesis because we expected an increase in turnover length occurring after Phase I, where written communication is emphasized. The second direct, but partial support for our hypotheses lies in the legibility results -- legibility increased over baseline after Phase I, but returned to 1999 levels after Phase II. Possibly, legibility is a habit more quickly and readily improved than writing more complete descriptions.

This failure to fully support our hypothesis might be explained by participant reaction to a second training module. After a second training, participants get a reminder of Phase I content, and may hear an implicit message that management is committed to the values and ideas advocated in the training. Those results (Figure 2) do show an increasing length of written turnover from January to March and again from March to June 2000 where the difference is finally significant. It may require some time and encouragement from others to make the extra effort to increase turnover narrative.

The analysis of job titles and turnover content showed mechanics to be the most thorough in their entries, being more likely than managers or inspectors to include all three types of content recorded. These findings are consistent with job roles. Because mechanics are performing a bulk of the actual work, occupational demands may motivate them to write longer and more comprehensive turnover. Consistent with this explanation are the positive sentiment and the stronger intent to improve turnover shown after phase I than after phase II in the survey data (*cf.*, Tables 1 and 2).

Participants may have made an initial effort to write more legibly after the first training because it was not too demanding and cumbersome. Little management commitment at the subject site was dedicated to this change, and little reinforcement was reported to be received by mechanics. Thus, the efforts waned in the absence of reminders or internal incentives.

Other measures of paperwork errors provided additional means by which to assess MRM training effects. However, the introduction of a substantial number of new personnel into the subject company at the beginning of 1999 seems to have confounded those efforts to detect any training impact on paperwork error rates. Under these circumstances special technical training program in the proper use of forms would be of benefit for the new hires as well as for the more experienced mechanics who were providing them on-the-job guidance and advice. Without such technical training the influence of this diminished basic skill may outweigh any error-reducing effects the MRM training may have provided. That less experienced workforce is likely responsible for some if not much of the increase in errors following 1998. Similar data were not available from the comparison companies because they had not collected similar or comparable paperwork errors.

The effects of MRM training on measures other than written turnover quality, are also short-lived (Tables 1 & 2). An analysis of the enthusiasm data between Phase I and Phase II suggests that the enthusiasm for MRM training had decreased significantly, especially at the subject site.

Many mechanics in the subject site appear to have made an initial effort to write more legibly after the first training (Figure 3). Probably because little commitment at the subject site was dedicated to this change, and little reinforcement received by mechanics, their efforts waned in the absence of reminders or internal incentives. Anecdotal reports from the field visits suggest that local management did little to reinforce the content of the Phase I training and may actually have stymied it. This had dampening effects on mechanics' motivation to apply the training further.

This study focused on written turnover content, and measured it -- in a marked departure from earlier studies. The use of direct qualitative and quantitative variables reported here lend support to our hypothesis that training can improve written turnover. These results provide knowledge about how one might typically expect these constructs to behave in future programs. Such a framework is important for subsequent work in this important subject area.

Other data used and reported here -- the survey and interview data -- reveal the longer-term effects of management support (or its lack) on implementing the message of the MRM training. The fact that local management was not consistent and forceful in its support of this airline training program provides reinforcement for previously reported results regarding obstacles to successful organizational change in the airline industry (Taylor, 1998; Taylor & Christensen, 1998; Patankar & Taylor, 1999).

II. User-centered tools and usability

Evaluating MRM Programs: A New Method and Tool

1. The Use of Company- and Department-Level Percentile Ranks in Industry-Wide Organization Research

A common method of evaluating organizational success is by comparison to other organizations within the same industry. When data are collected from a number of companies with similar function or purpose, an organization can be placed along the distribution of all the companies and assigned a percentile rank. This ranking indicates where a particular organization ranks among its industry peers. This paper provides a basic description of percentile ranks, and discusses the practical implications of their use in organization research.

In our lab at Santa Clara University, we have collected an industry-wide MRM/TOQ survey database, numbering over 43,000 individual questionnaires, from which we can calculate the percentile ranks of any company, maintenance department, or sample we choose. We employ these percentile ranks for all companies interested in how attitudes before and after their training programs measure up to the levels that are typical in aviation maintenance. This analysis is provided in bar graphs that show each scale in relationship to the 50th %ile, which indicates participant attitudes are the same as the average in the population.

<u>Why Percentile Ranks?</u>

Percentile ranks are appropriate for industry-wide organizational research for much the same reason they are used in clinical and educational settings: The desire for a benchmarked comparison of performance. In addition to the longitudinal means comparisons, which show how much a company has changed over time, the percentile ranks calculator shows the position of a company in the industry at a particular point in time. Both pieces of information are important, but different, and provide a richer assessment of cultural change when taken together.

<u>The Nature of Percentile Ranks</u>

Percentile ranks are a descriptive measure derived from standard scores that identify the location of an individual or subgroup along a distribution of a larger population to which that individual or group belongs (see Downie & Heath, 1974). Such measures have typically been used on standardized individual achievement tests, where results are to be interpreted in the context of the population to which the test-taker belongs. Application to organizations and group scores on standardized attitude surveys presents another valid use of percentiles.

Interpretation of Percentile Ranks

A few basic rules are important to the interpretation of percentile ranks. Percentile ranks range from 0 to 100, with higher ranks indicating a larger portion of the distribution of scores falling below the individual or group in question. Brown (1991) offers cautionary advice about the interpretation of percentile ranks. First, differences in scores on the extreme ends of the percentile rank distribution carry more weight than differences toward the middle. For example, the difference between a percentile rank of 50 and 55 is less meaningful than the difference between 5 and 10, between 30 and 35, or between 90 and 95. Also, percentile ranks are not to be averaged or summed. Percentile rank, an index of individual standing among a group, should not be confused with *percentage*, an index of proportion of a total group.

Percentile ranks in organization research can act as an indicator of where a company or department resides among its industry peers, but not necessarily as an indicator of individual or group improvement. As an example, Company A might already have very high trust in it's organizational culture. Therefore, Company A scores very high on the trust scale for both pre-test and post-test with no statistically significant difference between their average scores on that scale. Despite no significant improvement, Company A would show high percentile ranks. By contrast, Company B has moderate or relatively low trust in it's culture. This company would score low on the pre-test measure of trust, and have a lower percentile rank; but it might be expected to get more training benefit than Company A and score significantly higher on the post-test measure. Alas, though Company B has made significant improvement in trust, its post-training percentile ranks could still be comparatively low.

II. A Tool for the Calculation of Percentile Ranks

A tool for the calculation of percentile ranks has been developed for use with Maintenance Resource Management training evaluation in aviation. The following section describes a tool that allows trainers on-site to enter data and get percentile ranks on five survey scales. The tool is designed to readily provide benchmarked feedback to MRM trainers using percentile ranks.

The Evaluation Results Calculator for MRM Trainers and Implementers: Including Percentile Rank and Longitudinal Means Comparison

The *MRM Evaluation Results Calculator* (ERC) introduced here is a tool for organizations to examine themselves in relation to other companies. The tool has been developed specifically for use by Maintenance Resource Management trainers and implementers using the Maintenance Resource Management / Technical Operations Questionnaire, or "MRM/TOQ" (Taylor & Thomas, 2001). This application has implications for almost any instance where data is acquired for a variety of same-industry companies. The aim is to provide a tool for self-evaluation that will assist trainers in tailoring their content and approaches to reach desired learning objectives. Trainers will be immediately able to enter survey data on-site and acquire a picture of where they stand in the industry. Because rapid and consistent feedback is such a critical part of learning and personal improvement, trainers will likely find this self-usable calculator a welcome addition to training improvement pursuits.

How the Evaluation Results Calculator Works

The ERC presented here is an MS Excel program. It operates by converting raw survey scores (entered by the user) into z-scores, and calculating the area of a normal curve **below** that z-score. This is accomplished by embedding a Standard Normal Distribution Table (found most introductory statistics textbooks) into the Excel program. The percentile rank calculation is not statistically complex, and does allow a readily available way to achieve useful information with data collected on-site. The calculation procedure is described in more detail below:

1) Scale means are calculated from survey data entered by the user. Scale formulas are shown in Appendix A.

2) The Z-score for each scale mean is then calculated using the formula:

$$\frac{\textit{(Sample Scale Mean Score – Average of all Population Scale Mean Scores)}}{\textit{Standard Deviation of all Population Scale Mean Scores}}$$

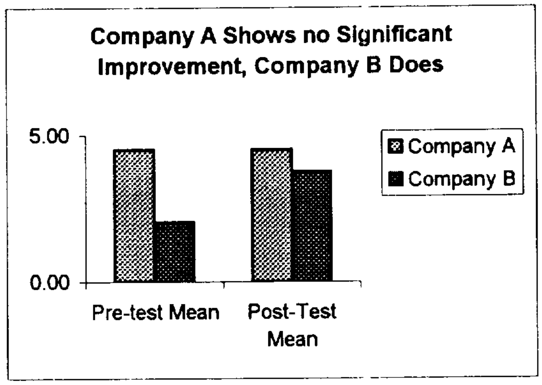
3) This produces a distribution of sample Z-scores of which the mean is taken to produce the mean Z-score for each scale.

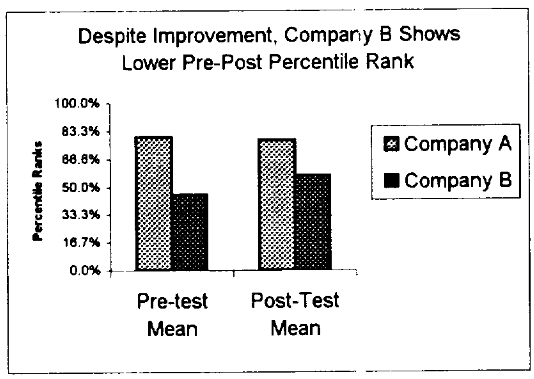
4) The mean Z-score is converted to the area under the normal curve between the sample Z-Score and the center of the distribution using a Standard Normal Distribution Table (Appendix).

5) Finally, .5 is added to the outcome of step 4 to arrive at the percentile rank of the sample being evaluated.

Hence, the ERC uses the mean and standard deviation of the industry population to calculate the benchmarked attitude ratings of training participants. This is shown in the form of pre- and post- percentile ranks. In addition to percentile ranks, the calculator also provides pre- and post-training mean scores and calculates an independent samples *t-test* to determine statistical significance. When scale means are statistically significant at the .05 alpha level, the scale and means scores are highlighted in orange. Graphs are included in the program output, which automatically update as data are entered. Samples of these graphs are shown in Figure 11. The user needs only to enter the data, and then print the graphs.

Figure 11
Samples of ERC Output





Instructions for Using the MRM Evaluation Results Calculator

The ERC has been initially designed for use with Pre- and Post- versions of the MRM/TOQ. Its operation is summarized in three simple steps: data entry, interpretation of results, and graphs:

Step 1) Data Entry

The MRM/TOQ Evaluation Results Calculator requires data entry into Excel worksheets designated for pre- and post-training data. The questions are listed across the top of each worksheet in the same order they appear on the pre- and post- survey instruments. Illegible or omitted survey responses should simply be skipped during data entry. After all the surveys at hand are entered, results are obtained by clicking on the *Scale Means and Ranks* worksheet. To summarize, data entry for the evaluation results calculator occurs in three steps:

1. Enter Pre-Training Data into *Pre-Training Data Entry* worksheet.
2. Enter Post-Training Data into *Post-Training Data Entry* worksheet.
3. Go to *Scale Means and Ranks* worksheet to view calculated results.

Step 2) Interpretation Of Results

The MRM/TOQ Evaluation Results Calculator yields Pre- and Post- mean scores, as well as Pre- and Post- percentile ranks. These calculations are made for several validated survey scales, described below in Section III. When Pre- and Post-Training mean scores bear a significant difference at the .05 level, or better, those scores and the respective scale are highlighted in orange.

An important note applies to the use of percentile rank to determine success of a training intervention as applied here. For the purposes of the MRM/TOQ pre-post surveys, an increase in percentile rank from pre-test to post-test does not mean that an actual increase took place by the group being examined. This is because the scores are being calculated against two different distributions (pre and post). Rather, the pre- and post- percentile ranks show group or individual standing against industry measures at separate points in time. If the larger population happened to increase on average at a lower rate from pre to post, then a particular group could show an increase in percentile rank by merely maintaining the same raw mean score or decreasing to a lesser extent.

Step 3) Graphs

Results are graphed at the bottom of the *Scale Means and Ranks* worksheet in two ways: Scale means and scale percentile ranks. Further, scale mean and percentile rank results are separated into pre- and post-training.

Measures used in the Evaluation Results Calculator

The following are measures used in the MRM Evaluation Results Calculator as evidence of training impact. They were developed and validated through factor analysis using the MRM/TOQ described in Section III.

<u>Trust Supervisor's Safety Practices</u> This scale reflects the quality of the relationship between the respondent and her/his supervisors or managers on safety related matters. Survey questions that comprise this scale probe for how much the respondent feels she/he can approach management without fear of punishment, backlash or inaction (especially with safety issues and suggestions).

<u>Value Trust and Communication with Coworkers</u> This scale, also a trust measure, indicates the importance of trust and quality communication among the respondent's coworkers. General importance and feeling of open communication, debriefing and shift meetings are measured by this scale.

<u>Value of Assertiveness</u> A critical component of good communication in aviation maintenance that is stressed in MRM training is the ability to speak and listen assertively when doubt arises or a situation seems unclear. This scale measures the respondent's comfort in disagreeing with or speaking out against the opinions of others in maintenance.

<u>Understand Effects of Stress</u> This scale measures the respondent's awareness of the impact and importance of individual stress factors to her/his performance. The degree

to which the respondent believes that fatigue and personal problems degrade safe performance are measured with this scale, as well as self-perceived ability to separate personal problems from work.

Enthusiasm for the Training Post-training enthusiasm measures are taken to assess trainee motivations to transfer training concepts to the work environment. Enthusiasm is measured only for post training, and is comprised of three statements for which respondents are to rate their level of agreement: *1) This training can increase safety and teamwork, 2) This training will be useful to others* and, *3) This training will change my behavior.*

III. Future Directions and Applications

The ERC introduced here has many possibilities for increasing accessibility to benchmarked training evaluation. As the evaluation process becomes more automated and user-friendly, training development efforts will improve and become based more on systematic measurement rather than trainer intuition. The instant quality of the feedback provided by the Evaluation Results Calculator allows benchmarked feedback to be used immediately for application toward improving the next training session.

Future developments of the ERC should involve two basic directions: 1) More comprehensive comparisons with other surveys, e.g., with "baseline" surveys before a program is implemented, and with "follow-up" surveys administered months after attending training. and 2) creating richer and more detailed feedback from the instrument, including analysis of write-in answers from the post-training and follow-up surveys.

Quickness and Usability

One of the fundamental purposes of the ERC is to speed-up the feedback process by putting it in the hands of those closest to the training. To this end, improvements to the tool should focus largely on this component. Currently, the greatest obstacle to speed of use with the ERC is the data entry process. Developments will need to provide a more efficient method of data entry than keyboard data entry. Two main options being considered are scanning technology and web-based data entry. With scanning technology, trainers could collect surveys and immediately scan responses into the program without having to hand enter data. With the web-based option, training participants could enter their own data via the web, and feedback results could be accessed by designated parties instantly. Each of these improvements would increase the quickness and usability of the ERC.

Increased Feedback Detail

This newly introduced first edition of the ERC provides pre- and post- scale Means with tests of significance, and pre- and post-training percentile ranks based on pre- and post-training industry databases. As indicated earlier in this paper, the percentile ranks as currently calculated say nothing of actual improvement from pre- to post-training. Percentile ranks could shed greater light on actual attitude change or improvement if company samples were ranked on *gain scores*. Warr, Allen and Birdi (1999) identify two, and only two, types of outcome data examined in publications about training. The first type is score attainment, which is merely the measure at either pre- or post- training (generally post) of a certain criteria. Score attainment is the outcome data type for which percentile ranks are being calculated in this first edition. The second type of outcome data is *gain scores* (also referred to as change scores). Gain scores are the difference between pre- and post- measurement and provides a quantification of the magnitude of training effects. This latter analysis is much preferred because it controls for pre-test difference among groups being compared. As a next step, a single industry database of gain scores could supplement the current pre- and post-training databases, and a single gain score percentile rank could be calculated for the amount of attitude change. This percentile rank would represent where the designated sample ranks in the industry on how much actual change took place.

Yet another improvement in the quality of feedback provided by this tool is in the populations used for benchmarking measured attitudes. In clinical and educational settings, an individual's score is often only ranked among members of that person's own group. As an example, members of particular cultures or ethnicities can be ranked among the population of test-takers from that same culture or ethnicity to attenuate cultural bias that may exist in the instrument.

This method, common in psychological and educational testing, can be employed with our instrument by allowing users to compare their sample group to different populations. For instance, if the evaluation of a training with only managers was desired, then the user could designate only managers be used for contrast in the total benchmark population. The same could be done for training participants with different job titles, levels of experience, age, etc. Users might also designate only to use their own company as the comparison population rather than the entire aviation industry.

Summary

The MRM Evaluation Results Calculator contains tools designed for MRM trainers and implementers to quickly and conveniently obtain feedback on the impact of their program. The Calculator shows pre-post change, as well as percentile ranks, indicating a respondent groups' standing among the industry. These calculations are performed for survey scales and enthusiasm measures from the Maintenance Resource Management / Technical Operations Questionnaire (MRM/TOQ).

III. The Constructs of Trust & Professionalism

Toward Measuring Safety Culture In Aviation Maintenance: The Structure of Trust and Professionalism

Introduction

The past decade has seen a dramatic increase in aviation maintenance safety programs incorporating principles of Human Factors and Organization Psychology (Taylor, 2000a). These programs are intended to influence the attitudes and behaviors of aircraft mechanics (following current US practice, hereafter called Aviation Maintenance Technicians, or AMTs). Additionally, these programs have also targeted those people in support of AMTs, including their supervisors and managers as well as other related occupations and professions.

Evidence is growing that AMT professionalism and interpersonal trust are key to building aviation organizations with excellent safety records. Persistent awareness of professional responsibilities is a necessary condition for maintenance safety and this element has been shown repeatedly to be a key factor in safety and human factors training (Taylor & Patankar, 2001). This professionalism however is not sufficient in itself. It is widely believed that interpersonal trust is also required for effective communication. Mutual trust among AMTs and other ground support personnel cannot be taken for granted and must be consciously supported and encouraged. This is true not only because of the historically solo nature of the AMT's occupation, but also because aviation is a multinational business, and because attitudes toward open communication and willingness to communicate have been shown to differ among national cultures (Helmreich & Merritt, 1998; Taylor & Patankar, 1999). Many airlines are trying to improve their safety culture by emphasizing communication and professionalism, together with awareness of decision-making, employee participation, and effective safety systems. To fully understand the concept of safety culture, significant research now needs to be directed toward developing the concepts and measurements of trust and professionalism.

Interpersonal Trust as Concept and Measure

The concept. Investigators have confirmed that the concept of trust is bipolar (includes "distrust" and "trust") and that trust is a generic concept that includes interpersonal trust as well as trust of technology (Jian, Bisantz & Drury, 1998). In understanding the dynamics of trust in organizations, one can variously focus on the macro level or micro-level of theory and analysis (Kramer & Tyler, 1996). From the macro level, investigators answer questions about how trust is related to organizational dynamics or management. Examples of such questions are whether trust in an industry or company has declined or whether trust can be rebuilt.

The micro-level perspective of trust considers the psychology of the individual -- why people trust, and what aspects most influence individual trust. From this micro-level, investigators posit that trust facilitates truthful communication, and leads to collaboration (Mishra, 1996). We are interested in this aspect to the degree that variables like an individual's age and experience can influence trust.

The measure. Questionnaire scales developed during the 1960's and 1970's measure micro-level trust as an attitude, or affective state ("being trustworthy is important"), or as an opinion or evaluation ("this person is trustworthy"). Reported scales are found to rate high in construct validity, and reliability usually using samples of undergraduate students. In use they emphasize the belief of trustworthiness (the degree to which others are seen as moral, honest and reliable) (Wrightsman, 1974). In the present study both measures for trust (attitudes and opinions) are considered and at both the micro and macro levels. Our purpose is to examine how the measures of levels of trust match the characteristics and conditions of the airline maintenance industry.

Method

Subjects:

During 1999-2000, 3,150 employees in five aviation maintenance organizations completed questionnaires measuring their attitudes and opinions about safety, communication, goal attainment, stress management and trust.

Respondent sample

The respondents come from samples that bracket the range of organizations and job types in the commercial aviation maintenance industry. The group includes employees in maintenance departments in major airlines, maintenance departments in small airlines as well as employees of commercial aviation repair stations. Each sample represents a US-based air transport company or a separate sample within an airline company. Participants include AMTs, maintenance managers, and maintenance support personnel. All can be considered naïve subjects in so far as they completed our survey before they were exposed to organizational change programs intended to influence their attitudes or opinions. All surveys were collected in the years 1999 and 2000.

Sample A (n = 119) is a 10% stratified random sample of the maintenance department of a large passenger airline who received the survey by company mail with a cover letter from the head of maintenance. The participation (75% return rate) was quite high for this type of mail survey.

Sample B (n = 152) consists entirely of volunteers from the maintenance department of a large airline who elected to attend a company-sponsored Human Factors and Safety Training program. Sample B's surveys were administered before the training began. This sample contains a larger number of college-educated and female respondents, and is more heavily weighted toward management respondents than sample A.

Sample C (n = 2,574) respondents are maintenance department participants in another airline's Human Factors and Safety training. Sample C's surveys were also administered just before the training began. Company C's distribution of job titles is closer to Sample A for its proportion of hourly workers in the line and base maintenance operations and its proportion of middle management.

Sample D (N = 78) respondents are all the maintenance employees in a small regional airline. Like Sample A they received their surveys by company mail with management encouragement to complete it.

Sample E (n = 227) is from a large US-based aircraft repair station. Sample E's responses are from two data collection efforts. Over forty percent (n = 96) of data set E is comprised of a 10 % random sample of AMTs who participated in a mail survey. The other 131 respondents in the company E data set are the company's entire population of maintenance managers. The managers completed the same surveys as the AMTs, but did so immediately prior to receiving company endorsed Human Factors and Safety training.

Analysis of Variance (ANOVA) was used to test differences in background characteristics among the five samples. All samples differ significantly in age ($p < 0.000$, $F = 29.2$, $df = 4, 3137$), years in present position ($p < 0.001$, $F = 28.7$, $df = 4, 3179$), years in college ($p < 0.001$, $F = 99$, $df = 4, 2593$), years in the military ($p < 0.001$, $F = 79.5$, $df = 4, 2671$,), years in trade school ($p < 0.001$, $F = 137.5$, $df = 4, 2497$), and years with other airline ($p < 0.001$, $F = 146$, $df = 4, 2578$). Chi-square tests show that the samples differ significantly in proportion of respondents who are managers, AMTs, cleaners, inspectors, clerks, and engineers ($p < 0.000$, $X^2 = 339.18$, $df = 20$); as well as the proportion of male to female respondents ($p < 0.000$, $X^2 = 34.78$, $df = 4$).

The Survey Measure: The "Maintenance Resource Management Technical Operations Questionnaire" (MRM/TOQ).

The MRM/TOQ developed for the present study is a further modification of a survey developed in 1991 (Taylor, 2000b). The MRM/TOQ questionnaire is a self-report measure of attitudes and opinions that are related (conceptually or empirically) to human factors and safety training in maintenance and maintenance support functions. Respondents are asked to express their degree of agreement in a series of statements. A five-point agreement scale is used.

The initial questionnaire in the present study begins with a core of 34 statements. Some of them were new items introduced to the MRM/TOQ to examine interpersonal trust. Others were carried over from earlier surveys such as the Cockpit Management Attitudes Questionnaire (CMAQ) (Helmreich, Foushee, Benson, & Russini, 1986; Taggart, 1990). These 34 items were successively reduced to 27, 18 and finally 15 items through a series of Factor Analyses conducted with the five unique respondent samples described above. The final 15-item survey is included as Appendix B.

Factor Analysis: Methodology for Combining Survey Items Into Scales

Several previous studies report using Factor Analysis to explore and confirm the internal structure for the core questionnaire items of the CMAQ (Gregorich, Helmreich, & Wilhelm, 1990; Sherman, 1992) and the original MRM/TOQ (Choi, 1995; Taylor, 2000b). The purpose of these analyses is to provide greater reliability and simplify interpretation of survey results by combining individual item responses into a fewer number of multi-item scales. Those studies also sought to create a valid instrument to assess the degree of change and improvement achieved by the companies' safety and human factors programs. Like those predecessors the present study seeks to use Factor Analysis (hereafter referred to as FA) to determine the smallest number of reliable measures for the revised survey of AMTs and others in aviation maintenance; but it also uses FA to determine what new internal structure emerges when using new survey items on safety practice and interpersonal trust.

Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure were conducted for each sample to test the appropriateness of the data for Factor Analysis (Norusis, 1990, pp. 316-317). The KMO ranged from .672 to .840, and the Bartlett tests were significant ($p<.001$) in all cases. For each of the analyses for each of the samples a principal components analysis was run and initial factors were extracted based on Eigenvalues. From the scree plots obtained, the appropriate numbers of the factors were determined as specified by Norusis (1990). Initially both oblique (Quartimax) and orthogonal (Varimax) rotations were tested; however, since the varimax solutions were uniformly more parsimonious than the quartimax the former technique was employed thereafter. In all cases the factor solutions offered good interpretability and simple structures.

Results

Iterative Factor Structure

Progress occurred in several steps. A first exploratory FA was conducted using Sample A data. It used 34 items and resulted in 9 factors, together accounting for 66% of the variance, with the primary factor containing 8 items with loadings greater than .40. A second exploratory 34-item FA was duplicated in sample B. For sample B, this FA resulted in a larger structure of 10 factors, with a primary factor with 18 items loading above .40. Next, the 34 item exploratory analysis was repeated using two internal sub samples (maintenance stations in separate cities), from Sample B. Seven of the 18 items of factor #1 were inconsistent in their loadings across the two sub-samples and were dropped from further analysis, which left 27 items to analyze.

Factor Analysis was then repeated with the 27 items for the total B sample, in order to confirm the preceding exploratory FA results using 34 items in samples A and B. This 27 item FA extracted nine factors accounting for 62% of the variance. The resulting structure of factors and item loadings after rotation are shown in Table 4. The first seven factors contain multiple items with loadings greater than .40. Only two of the 27 items have loadings this high or higher in two factors simultaneously. This seven factor structure is interpretable and the factor labels are shown in Table 5. Factor I, "Supervisor trust and safety," and factor II, "Value coworker trust and communication," echo the primary factors extracted in the 34 item FA computer for samples A and B. They are trust factors with different foci and meaning from one another. Factor V, "assertiveness" (a reflected factor because of negative loadings for both items), and factor VI, "effects of stress," are similar to factors derived from the earlier version of the MRM/TOQ (Taylor, 2000b). Factors III, IV, and VII although clearly interpretable are new to the 27 item FA. Of these, factor IV is of most interest in the present study, being the third trust factor in the structure, and it is different again in content and focus from either factor II or I. Factors VIII and IX contain only one item each and are thus not of significance to the present structure – except in their remoteness from its core.

Table 4: Confirming FA Using 27 Items, Sample B

	I	II	III	IV	V	VI	VII	VIII	IX
Factor I (Supervisor trust & safety)									
1. My supervisor can be trusted	.80								
2. Supervisor makes realistic promises and keeps them	.80								
3. My safety ideas would be acted on if reported to suprv.	.76								
4. My supervisor protects confidential information	.69								
5. We get feedback about the performance	.51								
6. AMTs ideas go up the line	.47								
7. I know proper channels to report safety issues	.45								.43
Factor II (Value coworker trust & communication)									
8. Having the trust of my coworkers is important		.75							
9. Debriefing after major task is important		.70							
10. AMTs contribute to customer service		.65							
11. Start of shift meetings are important		.59							
Factor III (Pride in company)									
12. Proud to work for this company			.76						
13. Others should make the effort for open communication			.65						
14. Other groups share our goals			.63						
Factor IV (Coworker personal trust)									
15. My coworkers can be trusted				.71					
16. Personal Problems can affect my performance				.66					
17. Mechanics in other departments can be trusted				.61					
Factor V (Value assertiveness)									
18. Should avoid disagreeing with others					.77				
19. Mgt effectiveness results from technical competence					.44				
Factor VI (Effects of my stress)									
20. Even when fatigued I perform effectively						.71			
21. Management should take control in emergency						.55			
22. As a professional I can leave problems behind						.53			
Factor VII (Need to speak up)									
23. Important to avoid negative comments about other's work					.51		.59		
24. Coworkers value consistency between words and action							.58		
25. We can question goals							.55		
26. I should provide written & verbal turnovers								.83	
27. My work affects passenger safety & satisfaction									.84
Eigenvalues	5.34	2.00	1.81	1.55	1.41	1.32	1.23	1.09	1.02
Percent of variance	20.1	7.4	6.7	5.8	5.2	4.9	4.6	4.0	3.8

Factor Analysis for the 18 Items Common to All Samples

The surveys collected from the three additional aviation maintenance companies (C, D, E) were available for further test. Each of these samples was missing one or more of the 27 items used in Samples A and B. In total, nine items from the original 27 were missing from at least one of samples C, D, or E. These nine items (numbers 2,10,12,15,17,19,25,26 and 27 in Table 4) had not been used either because the companies (being quite different from one another) requested they not be used, or the investigators felt some items were inappropriate for that application or sample. These final analyses to confirm Sample B results with the reduced set of 18 items were conducted in the three additional sites (C, D, and E) as well as the original two sites (A and B). The five samples were analyzed separately, but in a similar fashion.

Table 5 contains the factor loadings for the 18 items for all five samples. It shows that Varimax rotation resulted in 13 of the 18 items loading clearly and consistently into four scales over the five company samples. The item numbers used in Table 5 are the same as those used in Table 4. Factor loadings above .50 for any sample are considered strong, and those above .40 are considered at least supportive to the factor structure. Item or identifier consistency among the five samples is determined by at least four samples having a loading of .40 or greater.

Table 5. Factor Loadings Using 18 Items For Each of Five Companies

Factors & Items	Samples				
	A	B	C	D	E
Factor 1 - Supervisor Trust & Safety					
Consistent Identifiers					
1. My supervisor can be trusted	0.534	0.778	0.723	0.830	0.824
3. My safety ideas would be acted on if reported to suprv.	0.729	0.776	0.728	0.673	0.653
4. My supervisor protects confidential information	0.514	0.748	0.681	0.503	0.693
7. I know proper channels to report safety issues	0.007	0.512	0.432	0.476	0.654
Inconsistent Identifiers					
6. Mechanics' ideas go up the line	0.764	0.593	0.641	0.059	0.279
5. We get feedback about the performance	0.791	0.487	0.685	0.108	0.325
14. Other groups share our goals	0.270	0.239	0.515	0.121	0.006
Eigenvalue	3.967	3.716	4.051	2.038	3.819
Percent of Variance:	22.0%	20.6%	22.5%	11.323%	21.2%
Factor 2 - Value coworker trust & communication					
Consistent Identifiers					
8. Having the trust of my coworkers is important	0.810	0.620	0.699	0.486	0.648
9. Debriefing after major task is important	0.003	0.801	0.692	0.729	0.665
11. Start of shift meetings are important	0.161	0.601	0.628	0.757	0.655
13. Others should make the effort for open communication	0.510	0.208	0.773	0.748	0.706
24. Coworkers value consistency between words and action	0.697	0.150	0.733	0.527	0.431
Eigenvalue	2.278	1.74	2.057	1.602	1.885
Percent of Variance:	12.7%	9.7%	11.4%	8.9%	10.5%

Factor 3 - Effects of my Stress					
Consistent Identifiers					
16. Personal Problems can affect my performance	-0.809	-0.554	-0.696	-0.807	-0.776
20. Even when fatigued I perform effectively	0.742	0.683	0.664	0.235	0.599
22. As a professional I can leave problems behind	0.719	0.715	0.645	0.292	0.753
Eigenvalue	1.366	1.336	1.203	1.506	1.392
Percent of Variance:	7.6%	7.4%	6.7%	8.4%	7.7%
Factor 4 – Value Assertiveness (reflected)					
Consistent Identifiers					
18. Should avoid disagreeing with others	0.789	0.664	0.815	0.870	0.737
23. Important to avoid negative comments about other's work	0.743	0.617	0.787	0.396	0.738
Inconsistent Identifiers					
21. Managers should take control in an emergency	0.004	0.569	0.434	0.006	0.000
Eigenvalue:	1.030	1.302	1.517	1.167	1.160
Percent of Variance:	5.7%	7.2%	8.4%	6.5%	6.4%

Of the five items not loading as strongly on one factor and/or not consistently loading across the five samples, three (numbers 5, 14, 21) are dropped from further consideration. Although there were differences in detail and minor differences in the structures among the solutions extracted using the separate company samples, the same four factors were derived for all five samples. Furthermore, two of these four factors reproduces the essence of the first two trust factors from the 27 item analysis, "Supervisor trust and safety," and "Value coworker trust and communication;" as well as the "Assertiveness" and "Effects of Stress" factors extracted from previous versions of the MRM/TOQ. This 18-item replication concludes the final development of scales derived in the present study.

Factor I: "Supervisor trust and safety. As seen in Table 5, Factor I is consistently characterized by four items that suggested a trust of one's supervisor in regard to ethical behavior and safety practices involving their superior-subordinate relationship. They are "My supervisor can be trusted," " My safety ideas would be acted on if reported to my supervisor," and "My supervisor protects confidential information," and "I know proper channels to report safety issues." Three other items (5, 6, and 14) are less consistent in their loading on this factor, but also express related assessment of vertical communication. One of these less consistent identifiers, "Mechanics ideas go up the line" (#6) has reasonably strong loadings for three of the five samples. It was decided to include the 'ideas up the line' with the four more clearly consistent identifiers/items into an index of five items for this scale. Theoretically, endorsement of the five items identifying this factor implies a favorable opinion toward a superior's trustworthiness in support of safety. The remaining two items (#5 and 14) were dropped from further deliberation.

Factor II: "Value coworker trust & communication." Factor II indexes a belief in trusting one's coworkers in association with consistency in their words and deeds and their open communication in meetings and discussions. Agreement with the five items related to this factor suggests a high value for trusting coworkers in work-related discussions.

Factor III: "Effects of my stress." Three items describing the effect of stress on one's performance identified factor III. Agreement with two of these items denoted imperviousness to stress, while the third was stated as a direct effect. This item, "Personal problems can affect my performance," was consistently and negatively loaded on Factor III in all five samples, while the other two items (20 and 22) had strong positive loadings for four of the five samples. Agreement with the first item and disagreement with the second and third one can be seen as congruent with professionalism, indicated by the stress management goal of many human factors and safety training programs in maintenance (ATA, 2001).

Factor IV: "Value Assertiveness." Two items that suggested avoidance of interpersonal conflict represented factor IV. These items, "We should avoid disagreeing with others" and "It is important to avoid negative comments about other people's work," were each strongly and negatively loaded for four of the five samples. Disagreement with both items is interpreted as endorsing the professional goal of candor and openness in maintenance and safety-related communication (ATA, 2001). A third item (#21) shared less consistency than the others and was dropped from further consideration.

Creating Measures of Trust and Professionalism -- Scale Construction

Creating scales from the FA. In the present case, scales are created by averaging the raw scores of variables that consistently identified each factor across solutions.

The scale for Factor I, labeled *Supervisory trust & safety,* is created by summing each respondent's raw scores for items 1,3,4, 6 and 7, and dividing that sum by five. Scale for Factor II, *Value coworker trust & communication* contains the sum of raw scores for items 8,9,11, 13 and 24, divided by five.

Scales for factors III and IV are treated slightly differently. To facilitate discussion and scale interpretability, the scale for Factor III, *Effects of my stress,* is constructed by summing the raw score of item 16 with the reflected (or reversed) scores of items 20 and 22 and dividing that total by three. Likewise the two Factor IV items are combined into the scale called *Value Assertiveness* by summing their reflected raw scores before averaging.

Correlations among the developed scales were calculated for each sample to arrive at conclusions about the nature of the measures and the relationships among them. Given the orthogonal FA rotation solution used in the present study, we expected independence among the derived scales. We found a low, but remarkably consistent significant correlation (ranging between +.33 and +.39) across all five samples between "*Supervisor Trust & Safety*" and "*Value Coworker Trust & Communication.*" Despite this effort to retain independence, correlations between these two scales are perhaps explainable as evidence for a trust culture; in which employees who can trust their supervisors may also be more likely to value trust and communication with their coworkers. Evidence for relationships between stress and assertiveness scales and between them and the two trust scales was not found. Sample C yields a higher number of low magnitude, yet significant inter-correlations, but these likely indicate the effect of type I error due to the substantially larger number of respondents in the company C sample.

Reliability of the MRM/TOQ item and index measures

Cronbach's Coefficient Alpha was used to assess internal consistency of the scales. Alpha was calculated for all four factors for each sample used in the current study. Alpha coefficients for Supervisory Trust & Safety (a 5-item scale) range from .72-.75 for the five samples, for Value Coworker Trust & Communication (5-item scale) range between .65-.77, for Effects of My Stress (3-item scale) are .43-.67, and Value Assertiveness (2-item scale) are .42-.62. Although the two trust scales are clearly more reliable than the stress and assertiveness measures, this is at least in part a consequence of the larger number of items that comprise the trust scales. In any event, reliability as assessed here is quite good for all measures.

Validity of the MRM/TOQ index measures

Macro-level Analysis

<u>Construct Validity: Factor Analysis</u>

As Stapleton (1997) asserts, factor analysis is a useful tool with which to evaluate score validity. Construct validity can be defined as the ability of variables chosen by a researcher to represent a theoretical construct. Factor analysis can tell us the extent to which our variables are measuring the same concepts. The implication is that when a large set of variables can load neatly into a few intended factors, evidence is granted that these variables are tapping the desired constructs. Hence, the factor analyses demonstrated here serve to establish construct validity for the MRM survey.

<u>Construct Validity: Organizational and occupational differences among the scales.</u>

A benefit for including the five separate samples in the current study is to examine the sensitivity of scale scores in distinguishing among aviation maintenance organizations. Investigators' prior knowledge of these samples also provides an opportunity to validate the measures based on grounded knowledge and observation about their respective histories and organizational contexts. The macro-level model of trust in organizations suggests that differences in organizations should be expected, given conditions allowing for differences in leadership climate and company culture. Table 6 shows the mean scores for each of the four index or scale measures among the five subject samples. Analysis of Variance (ANOVA) test reveals significant differences among companies for two of the scales --Supervisory Trust & Safety (p=.000, F=7.69, df=4), and Effects of My Stress (p=.036, F=2.58, df=4).

Table 6. Index (Scale) Mean Scores by Company Sample

INDEX	Compan	N	Mean	Std. Deviation
I. Supervisor Trust & Safety	A	116	3.65	0.86
	B	129	3.93	0.75
	C	240	3.41	0.84
	D	76	4.06	0.66
	E	209	4.01	0.75
	Total	293	3.50	0.85
II. Value Coworker Trust & Communication	A	116	4.53	0.52
	B	129	4.50	0.47
	C	240	4.44	0.59
	D	76	4.39	0.50
	E	209	4.62	0.42
	Total	293	4.46	0.58
III. Effects of my Stress	A	116	2.66	1.06
	B	129	2.94	0.88
	C	240	3.11	0.83
	D	76	2.72	0.79
	E	209	3.14	.0.93
	Total	293	3.08	0.86
IV. Value Assertiveness	A	116	2.95	1.13
	B	129	2.82	1.02
	C	240	3.10	1.09
	D	76	2.86	0.93
	E	209	2.68	1.02
	Total	293	3.05	1.09

Further, examination of interpersonal trust at the macro-level would also lead us to expect to see differences among the different occupations in aviation maintenance. Table 7 contains the mean scores for the maintenance and support occupations for the five samples.

Table 7. Index (Scale) Mean Scores by Occupational Group

INDEX	Occupation	N	Mean	Standard Deviation
I. Supervisor Trust & Safety				
	Mechanics & Leads	181	3.35	0.84
	Inspectors	112	3.34	0.88
	Management & Supervisors	290	4.18	0.63
	Utility & Cleaners	160	3.48	0.82
	Engineers	92	3.49	0.93
	Clerks, Analysts, Planners	471	3.68	0.76
II. Value Coworker Trust & Communication				
	Mechanics & Leads	181	4.41	0.59
	Inspectors	112	4.38	0.63
	Management & Supervisors	290	4.70	0.44
	Utility & Cleaners	160	4.40	0.65
	Engineers	92	4.51	0.56
	Clerks, Analysts, Planners	471	4.52	0.49
III. Effects of my Stress				
	Mechanics & Leads	181	3.06	0.86
	Inspectors	112	3.21	0.83
	Management & Supervisors	290	3.30	0.80
	Utility & Cleaners	160	2.91	0.93
	Engineers	92	3.15	0.76
	Clerks, Analysts, Planners	471	3.05	0.85
IV. Value Assertiveness				
	Mechanics & Leads	181	3.12	1.07
	Inspectors	112	3.26	1.04
	Management & Supervisors	290	2.97	1.08
	Utility & Cleaners	160	2.77	1.13
	Engineers	92	3.07	0.94
	Clerks, Analysts, Planners	471	2.90	1.12

Multivariate Analysis of Variance (MANOVA) was used to test the scale differences for the six occupational categories among the five companies. Two scales, "Supervisor Trust" and "Effects of Stress," were found to significantly differentiate among the companies. These results will be discussed later. For the maintenance occupations, three of the four scales reveal statistically significant differences. They are Supervisory Trust & Safety ($p=.000$, $F=8.55$, $df=5$), Value Coworker Trust & Communication ($p=.006$, $F=3.25$, $df=5$), and Effects of My Stress ($p=.002$, $F=3.92$, $df=5$). Managers had the highest scores for all three of these scales and AMTs and

Inspectors had the lowest scores. The "Value Assertiveness" scale was the only scale not demonstrating significant differences among the occupational types or the companies.

The interaction between occupation and company for the "effects of my stress" scale was found to be significant (p=.018, F=1.80, df=19). This sole significant interaction effect reflects some modest differences on the stress scale among utility cleaners, engineers and inspectors between companies. The lack of interaction effects for any of the scales between the AMTs or managers and other occupational subgroups for the other three scales confirms that there are only minor differences among the relative ranks for the occupations over companies. This supports the assumption of validity for the scale scores for distinguishing these two occupational groups, which are the central focus of the present study.

<u>Construct Validity: Interdepartmental differences among the scales.</u>

Next we tested the main differences for the four index measures between the two different maintenance departments (Flight Line maintenance and Base Hangar maintenance) across the five subject samples using the one-way Analysis of Variance (ANOVA) test. Only one of the four indices, "value coworker trust & communication" reveals statistically significant difference (p.000, F=20.8; df=1, 1418). Apparently the other three scales are not sensitive to the differences between the departments. Despite the fact that the Line maintenance mean score for "value of coworker trust & communication" is quite high (Mean = 4.385, Standard Deviation = .622, n = 643), it is still significantly below that of Base maintenance (Mean = 4.522, Standard Deviation = .508, n = 777). AMTs in the base hangars tend to be assigned to work together on complex jobs lasting as much as a week, while AMTs in flight line tend to be assigned to work alone on much shorter jobs. These conditions may well engender greatest value for collaboration among the base-hangar AMTs and the lesser value for this attribute on the flight line.

<u>Content Validity: Effect of Training</u>

Company "C" has created a one-day human factors and safety training program, called Maintenance Resource Management (MRM) training, for all maintenance employees. The training curriculum includes modules on communication and teamwork, the effects of fatigue and pressure on stress and performance, and speaking up (assertiveness) for safety. Supervisors, managers and maintenance executives attended and participated in the program along with mechanics, inspectors, utility cleaners, and clerical employees. Previous field work had established that Co C's MRM program had succeeded in short term change, but had not sustained it due to a lack of management support (Taylor & Thomas, in press). Training participants in company C completed the MRM/TOQ immediately before their training (these "pre-training" surveys were used in the FA described earlier). Immediately after their training, company C participants completed a "post-training" survey and then completed the survey again several months later (phase two, or "two-month follow-up" surveys). The three attitude or belief scales ("Value coworker trust," "Effects of stress," and "Value assertiveness") were expected to be sensitive to the effects of this training. The "Supervisor Trust & Safety" scale, representing respondent opinions of supervisory behavior, was expected to be more sensitive to changes in the leaders' subsequent behavior than the other three scales and to show this in the follow-up survey. A one-way ANOVA comparing the scale scores over the three surveys and those results showed significant changes for all four scales. Figure 12 shows the company C mean scores for the four scales before and after the training and again several months later.

Figure 12 Comparing Scales Before and After Training

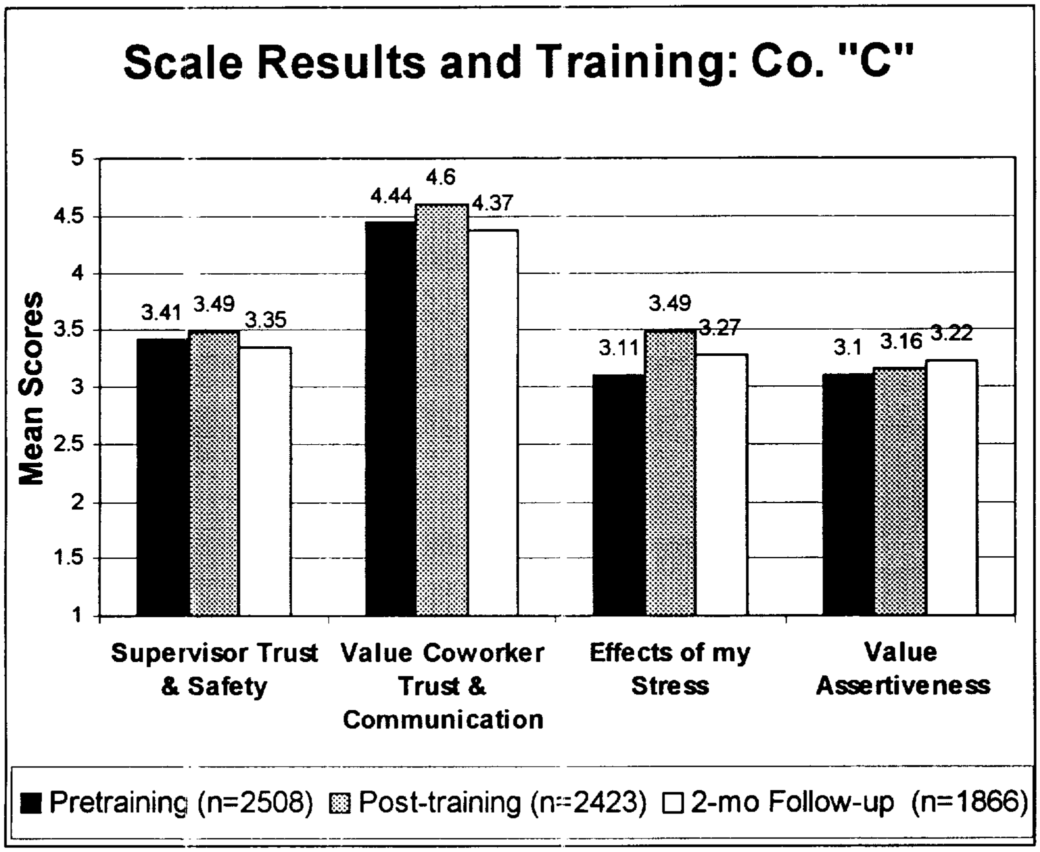


Figure12 shows that the training is accompanied by an increase in scale scores, but for three of the four scales this rise is then followed by decline two months later. Bonferroni post-hoc tests established statistical significance for the rise and fall of the supervisor trust, valuing coworker trust, and recognizing stress effects scales that are pictured in Figure 12. A post-hoc test also reveals that the rise in valuing assertiveness over time is significant only between the survey two months after training compared with the pre-training level.

Estimation of Concurrent Validity through Item Analysis

Obtaining index scores on a scale of measured intervals has important practical value for applied problems. Attitude surveys normally result in nominal or partly ordered scales, which are substantially weaker than ordinal or ordered-metric scales in their ability to describe respondent samples or be used with more stringent statistical tests and large samples. Scaling is used to overcome the problems of weak scale strength due to unsystematic combination of items or the use of single items as scales.

There are various scaling techniques to generate robust and reliable scales approaching ordinal or even ordered-metric strength. The Likert scaling method is one of these and is fairly simple to construct, although certain conditions and steps must be satisfied. Likert scales provide improvement over individual survey or test items as well as scales simply combined by intercorrelation (Selltiz, Wrightsman, & Cook, 1976). An essential component of "Likert-type" instruments is that scale items should correlate highly with total scores on the entire scale (Selltiz, et al., 1976, pp. 418-421). Also, items should show substantial disparity between those who score high and those who score low on the scale. In other words, good concurrent validity is required for a true Likert scale. The combination of FA helping to distinguish which items are identified most clearly with a common construct (Table 5), and the Alpha correlations also described earlier, which confirm the internal consistency of the scales comprising that construct, provides evidence that further testing the requirements of the Likert-type scale could be satisfied for the four scales described in the present paper. To address these requirements, item analysis was conducted for each item used in construction of the four scales generated through factor analysis. This was accomplished by conducting t-tests of item mean scores between the highest and lowest quartiles for each scale. Robust differences between the highest and lowest quartiles serve as evidence that a particular item is adequately discriminating between low and high groups on the scale construct to which it is associated. Table 8 shows the Item Analysis.

Table 8.
Item Analysis: Mean Differences Between Lowest and Highest Quartiles for Each Item

SCALES & ITEMS	LOWEST QUARTILE	HIGHEST QUARTILE	MEAN DIFFERENCE *
TRUST SUPERVISOR AND SAFETY			
My Supervisor can be trusted	1.94	4.60	-2.66
My supervisor protects confidential information	2.28	4.66	-2.38
My safety suggestions would be acted upon if I reported them	2.20	4.54	-2.35
AMTs ideas go up the line	1.87	3.97	-2.10
I know proper channels to report safety issues	3.42	4.76	-1.34
VALUE COWORKER TRUST AND COMMUNICATION			
Debriefing after a major task is important	3.50	5.00	-1.50
Start of shift meetings are important	3.51	5.00	-1.49
Having the trust and confidence of my coworkers is important	3.88	5.00	-1.12
My coworkers value consistency between words and actions	4.07	5.00	-.93
Employees should make the effort for open communication	4.11	5.00	-.89
EFFECTS OF MY STRESS			
I can leave personal problems behind (reflected)	1.67	4.13	-2.46
Even when fatigued, I perform effectively (reflected)	1.97	4.34	-2.37

Personal problems can affect my performance	3.52	4.77	-1.25
ASSERTIVENESS			
Avoid disagreeing with others	1.40	4.78	-3.38
Avoid negative comments about others' work	1.68	4.82	-3.13
**All Mean Differences Significant at p<.001*			

Results shown in Table 8 indicate that most of the items used in the present factor analysis and scale construction are able to discriminate well between the lowest and highest quartiles. Mean differences between the lowest and highest quartile for all items were significant at p<.001, and non-parametric comparisons confirmed these results.

Micro-level Analysis

Demographic characteristics were shown to differ within the set of respondents in the present study. Some of these individual characteristics such as time with the company, time in job or education are occupationally specific. On the other hand, the age and gender variables can be considered more independent of the industry and thus can be used to test the sensitivity of the four scales -- and in particular the two trust scales -- to individual differences. Several main effects of age and gender on the four scales were evident using MANOVA. There were no significant interactions found between age and gender for any of the four scales.

Three scales showed significant differences between men and women. The differences in gender showed higher "Supervisor trust" (p=.002, F=9.58, df=1), and "Value of coworker trust," (p=.028, F=4.86, df=1) for women than men; and for the "Value of assertiveness" to be greater for men than women (p=.008, F=7.07, df=1).

Three scales were significantly different for respondents of different ages as well. In the case of the "Supervisor trust" scale, a significant curvilinear effect (p=.002, F=4.13, df=4) was manifest where the level decreased with age until 45 years and then increased again. The age and "Value of assertiveness" relationship was also found to be significant and curvilinear (p=.007, F=3.51, df=4), with this attitude increasing with age until 45 when it decreased again. A significant linear relationship was seen for "Effect of my stress" (p=.027, F=2.74, df=4) where this appreciation increased from the youngest to the oldest category.

Summary

A survey of forty-eight survey questions administered to airline maintenance personnel at five qualitatively different companies and sites was factor analyzed and reduced to a valid and reliable set of scales that measure trust, assertiveness and stress. Item reduction was determined by the strength of the loadings and the availability of item data from each sample. Variables ultimately yielded 4 distinct factors after data reduction to a set of 15 items common to all samples that loaded with at least moderate strength onto one of the factors. In addition, participants answered demographic and experience related questions. The purpose of the questionnaire is to measure attitudes, opinions and skills that subsequent human factors training aims to influence. An impetus for including five distinctive samples in the current study was to examine the stability in factor structure across differing organizational environments within the same industry.

The four factors produced after data reduction were: Supervisor Trust and Safety, Value Coworker Trust and Communication, Value Assertiveness and Awareness of Stress Effects. Little inter-correlation was found among the scales, with exception to the two trust measures. These showed a consistent positive relationship across company samples. Reliability of the scales was shown to be high. Validation at the macro and micro level of analysis was established. Training effects on the scales were also examined. These results -- as well as comparisons among the companies; between departments, among job titles, and among differences in demographic data across the companies -- show the scales to be good measures that are accurately conveying information about their intended constructs. Additionally good strength as "Likert Scales" is indicated by an item analysis, which showed ability of constituent items to discriminate quite well between high and low groups for each scale.

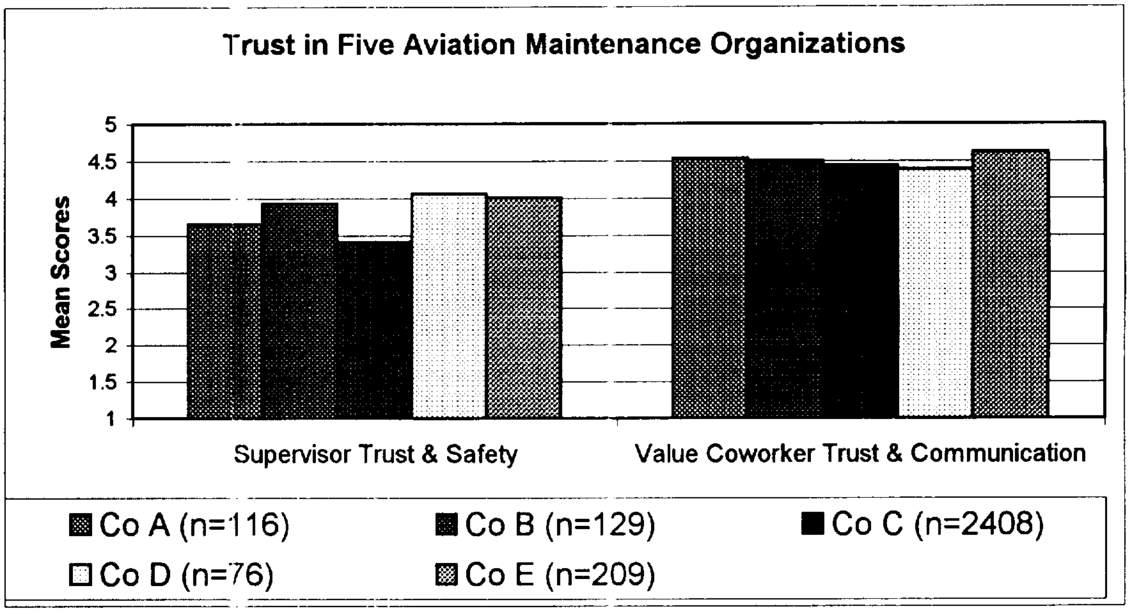
How Much Trust and Professionalism is there?

As already reported, the employees in five very different aviation organizations were found to differ in the degree of trust they have in their superiors' safety practices.

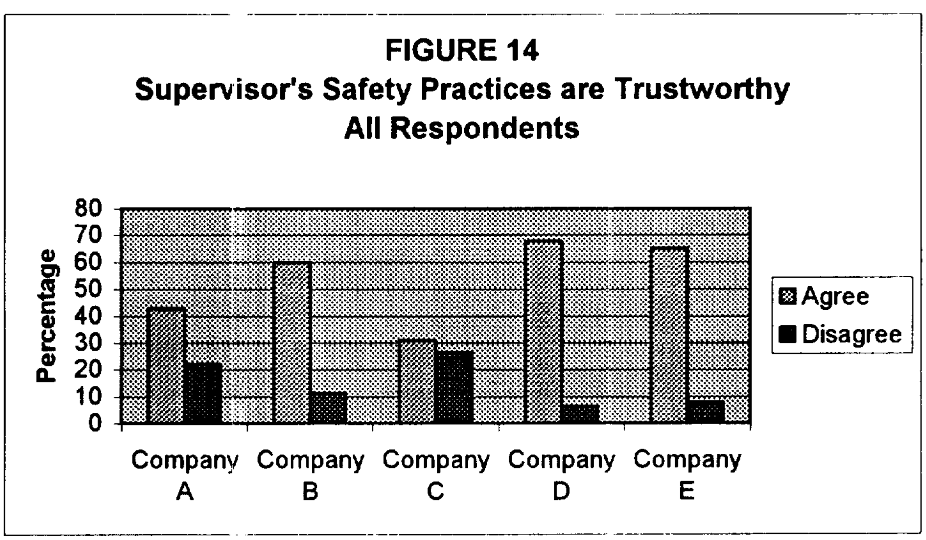
Multivariate Analyses of Variance (MANOVA) were used to test trust scale differences among the five companies, among occupational categories, between gender, and among age categories.

Intercompany Differences: Significant differences found for "supervisor trust & safety" ($F=7.69$, $p<.000$), but not for "value of trusting coworkers." Figure 13 shows mean scores for the two trust scales among the five companies. Post hoc tests show that company C has significantly lower "Trust Supervisor" scores than each of the other four company samples.

Figure 13



Across the five companies, we find a high of 68% and low of 31% of all respondents who say they agree or strongly agree that their supervisor is trustworthy regarding safety issues. Conversely, 6% to 26% respondents either say they disagree or strongly disagree with this (see Figure 14). The remaining proportion in each company report neither agreement or disagreement.



Occupational Differences. In general there is a perceived difference between mechanics and managers in their interpretation of their supervisor's safety practices. As a probable consequence mechanics tend not to trust their managers as much as we might want in this high-risk industry. Figure 15 shows the mean scores among occupations for all five companies combined for the scale "Supervisor's Safety Practices are Trustworthy." The MANOVA "F" score of 8.55 is significant p<.00).

FIGURE 15
"Supervisor's Safety Practices are Trustworthy":
By Occupation

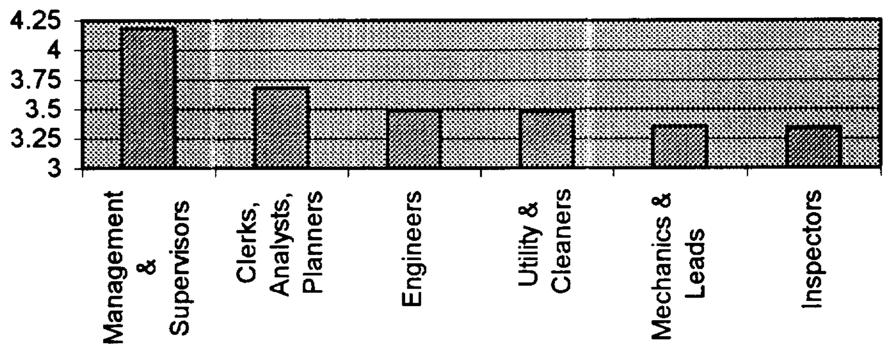
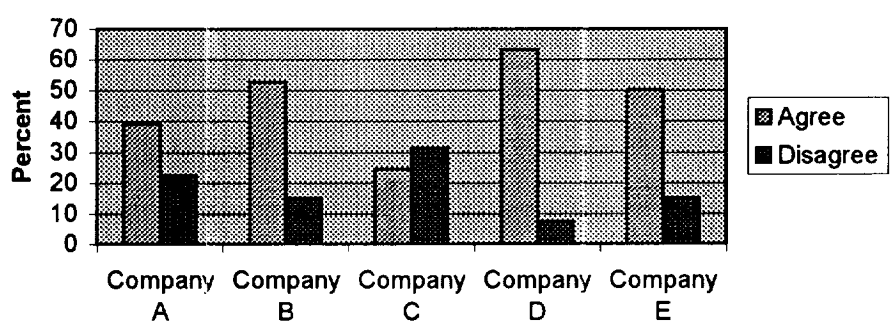


Figure 16 shows that across the five companies, a high of 63% and low of 24% mechanics say they agree or strongly agree that their supervisor is trustworthy regarding safety issues – Stated as the converse, 7% to 31% mechanics say they disagree or strongly disagree with this.

FIGURE 16
Supervisor's Safety Practices are Trustworthy
Mechanics, Inspectors & Leads only

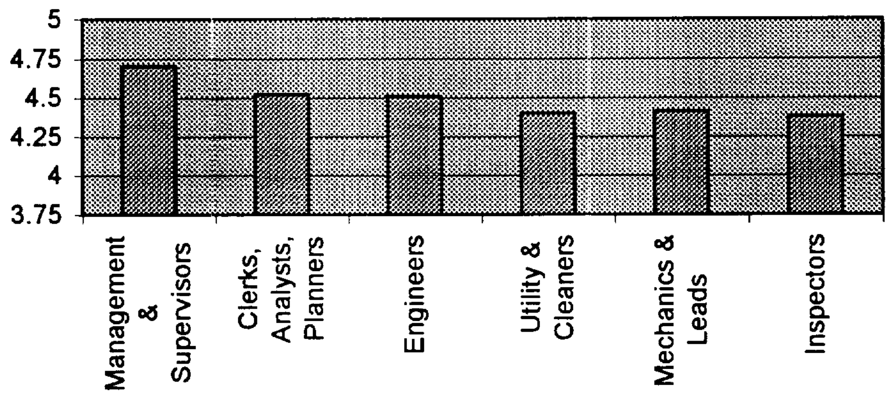


These results show that there substantial differences among companies in the de3gree of AMTs' thrust in their management. Such differences illustrate an important aspect of safety culture.

Trusting One's Coworkers

Figure 17 displays means among occupations for the scale "Value of Coworker Trust and Communication." Substantially more respondents from all companies "value open, trustworthy communication with coworkers," but managers are still higher than mechanics. The "F" score for these results is 3.25, p<.00.

FIGURE 17
Value of Coworker Trust & Communication: By Occupation

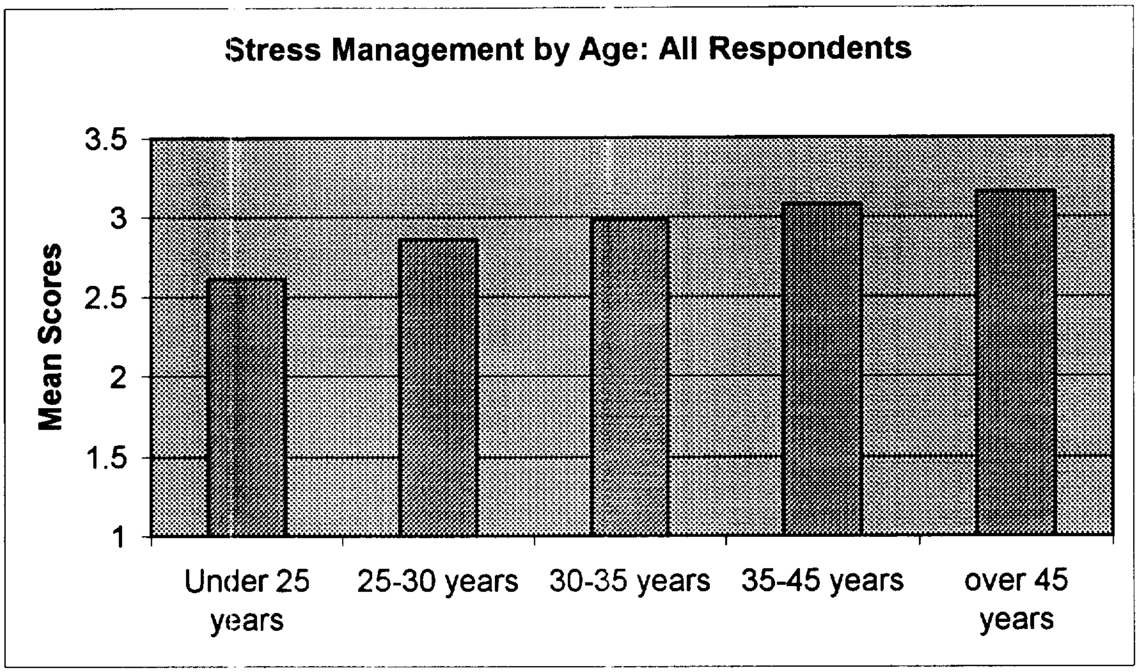


Professionalism

This study found two other scales dealing with support of professional issues: Importance of Stress on Decision Making; and Importance of Assertiveness. Like the two trust scales, these professionalism scales revealed a high reliability and validity across the five samples and showed an ability to differentiate among different occupations, gender, age categories and /or organizations. Historically these two professionalism scales have shown a sensitivity to MRM training – they both increase after training (Taylor & Robertson, 1995; Taylor & Patankar, 2001).

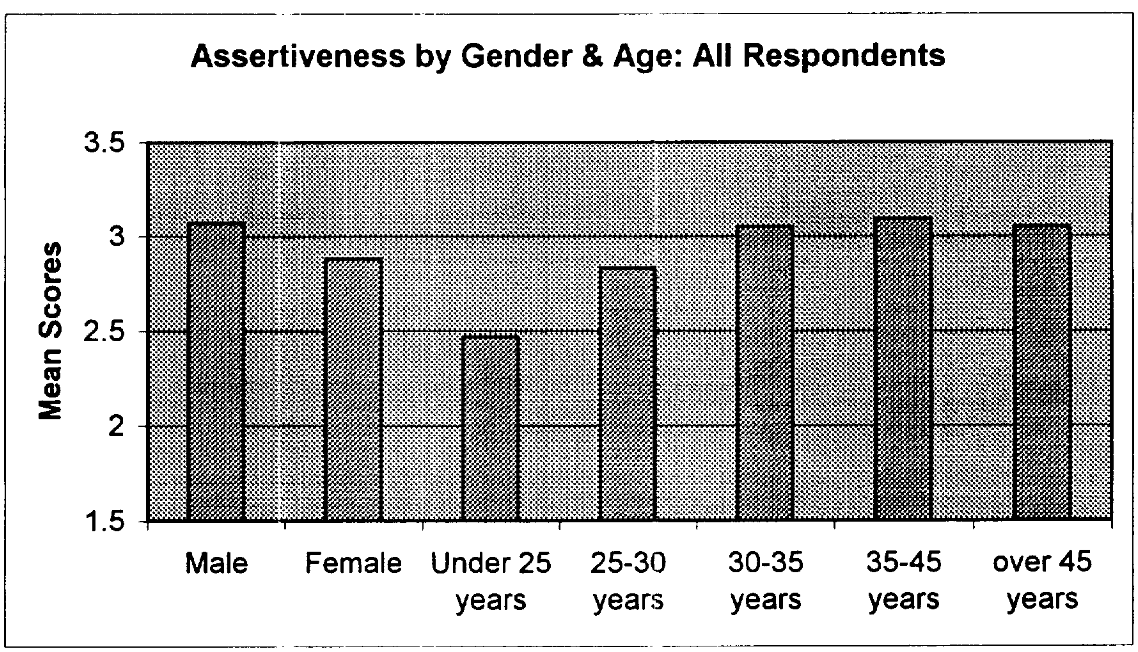
Significant differences among companies and among occupations were found for "stress management," but not for "value assertiveness." A significant and linear relationship was found between "stress management" and age, where this appreciation increased from the youngest to the oldest category. Figure 18 shows this comparison. A One-Way Analysis of Variance (ANOVA) was statistically significant (F = 10.22, p<.000)

Figure 18



Assertiveness was not significantly related to Company or Occupation for the five aviation samples reported here, but it was significantly related to gender and age. Figure 19 shows these relationships (F gender = 7.41, p<.006; F age = 5.61, p<.000)

Figure 19



Discussion

The present factor-analytic approach provides a useful and parsimonious solution for a survey assessment of maintenance human factors training and its subsequent diffusion and implementation. The data support the reduction of 18 variables into 15, clustered into four stable factors. Of the 15 surviving variables, 10 of these items date back to the original 1986-1990 CMAQ (Gregorich, et al., 1990) and successor surveys, and five are newly-created items measuring interpersonal trust. The two trust scales exhibit reasonable independence from the other professionalism scales across samples and show good reliabilities. Construct validity and discriminant validity among companies, departments, and individual differences were also demonstrated.

Factor I, "Supervisor trust and safety incorporates a trust of one's supervisor in regard to ethical behavior and safety practices involving their superior-subordinate relationship. Agreement with the five items identifying this factor implies a favorable opinion toward a superior's trustworthiness in support of safety.

Factor II, "Value coworker trust & communication" expresses a high value for trusting one's coworkers' communication in meetings and discussions. These two factors do support the expectation that aviation maintenance people find interpersonal trust to be a central concept in human factors.

Factor III, "Effects of my stress" emphasizes the consideration of stressors at work and the possibility of compensating for them. Though not related to the theme of human communication or interpersonal relations this factor proves to be an important concept for maintenance professionalism and is central to the curriculum of most human factors training programs.

Factor IV, "Value Assertiveness" emphasizes the goal of candor and openness in maintenance and safety-related communication. It is apparent from the present data that valuing assertiveness is independent of trusting others or their trustworthiness. Despite this, candor and honesty are also central to maintenance personnel and it is also an important part of many human factors programs.

Both factors III and IV reflect professionalism of the maintenance occupation. Stress management shows professional awareness by granting importance to conditions that may degrade decision making. Likewise, being willing to speak candidly can show a professional concern for safety and quality.

This new version of the MRM/TOQ has several uses as an investigative tool. Evaluation of the current status of maintainer attitudes within or across organizations and historical time frames is made possible. This includes assessment of the effects of particular human factors training when pretraining and posttraining and follow-up measures are obtained As more data on trust and professionalism is collected, the opportunity to compare even small samples to an accumulated benchmark increases. As more self-disclosure safety processes are introduced into aviation maintenance operations the more important will interpersonal trust become. Continued use of the MRM/TOQ to explore linkages to safety performance should benefit from the use of the two new trust measures introduced here.

This study demonstrates that aviation safety culture, although influenced by other cultures (national, organizational and professional), can be organized and studied in terms of two parameters: professionalism and trust. These two parameters can now be measured using a simplified 15-item MRM/TOQ presented here.

IV. Conclusions

State of MRM Measurement

This year we have attained several milestone achievements. First we have created performance measures of particular relevance to a specific MRM program – providing results that would have otherwise remained uncounted -- but with ready transferability to other programs as well. The measures – length, readability, and descriptiveness of written turnovers – were developed to show accurate and realistic testing of a particular program, but are here described to allow other to duplicate these or similar measures in other settings. They were shown to be sensitive to the effects of a specific MRM training course designed to improve written communication.

Second we have continued to show the usefulness of self-reported behavior measures. The turnover qualities, described above, were shown to be related to the open-ended questions, "How do you expect to use the training?" and "how have you used the training?"

Third we have updated and streamlined our basic survey instrument, the MRM/TOQ. It is now shortened, yet it contains questions that are summed to provide valid and reliable measures of aspects of professionalism (assertiveness, and stress management), and two aspects of interpersonal trust (trust of one's supervisor's safety practices, and importance of trust and communication with coworkers). The two professionalism scales, and three enthusiasm items from the post-training survey, can be compared back with our earlier MRM/TOQ surveys collected since 1991 (n>43,000). Yet even the new trust items already have an experience base of over 3,000 cases, and this number continues to grow. This means that a sizable and usable benchmark database is now available for use.

Fourth we have developed a tool that helps trainers and human factors professionals in the field to measure their organizations' survey responses over time and to compare these responses with the larger industry benchmark. This tool, the Evaluation Results Calculator (ERC) automatically computes the user's organizational mean scores pre- and post-training and computes its percentile rank compared with the overall maintenance benchmark.

Fifth, examination of results from the new trust scales suggests real differences in safety culture among companies. This extrapolation awaits further development and test.

When this year's achievements are added to our program's accomplishments of past years (Taylor & Robertson, 1995; Taylor, 1998, 2000c), a comprehensive and well-tested measurement plan for assessing MRM programs at all four levels of evaluating training interventions (Kirkpatrick, 1983) has been attained.

V. Recommendations

Success in improving safety performance over the long run is a complex of several efforts. All of them are necessary for success, but none are sufficient alone. With this year's results even more evidence has accumulated to bolster the following recommendations. These are the complex of key variables that must be controlled for long term safety improvement in aviation maintenance.

1. Start with the end in mind. We have previously discussed the importance of targeting outcomes (Patankar & Taylor, 2000) and our results this year show that a program to improve written turnover between shifts did improve that behavior for a short while – despite a lack of management support and guidance. The newly created measures of written turnover quality illustrate a practical approach to assessing performance previously targeted for improvement.
   No program in aviation maintenance is known to have consciously planned to increase trust of supervisors by AMTs, but if the wide variation among companies we have documented is to be reduced such a target must be consciously set.
2. Create high quality instructional programs. Building awareness of safety hazards and the positive effects of stress management and open communication are an important part of any MRM program. Variation in instructional quality will effect the degree to which that awareness is enhanced and the eagerness to apply it is kindled. The newly validated MRM/TOQ and the automated Evaluation Results Calculator (ERC) can provide timely and accurate measurements and control points to test and improve instructional quality.
3. Enlarge MRM education to include skill training. The MRM training in written turnover included hands-on exercises in writing technique and practical communication. This training focus was shown to have some influence on intentions to write turnovers and reports of having done so. Our data also suggest that targeted performance training, however well delivered, will not make much difference in management support and guidance in that performance is not forthcoming.
4. Find ways for management to provide coordinated, unequivocal, and unambiguous support. This recommendation has been a repeated theme in the reports from this program for many years. As long ago as 194 we noticed the positive effect on MRM programs of the personal guidance and constant attention by the Executive Maintenance VP (Taylor & Robertson, 1994). Once that senior executive turned his attention to other matters and stopped urging his subordinate managers to actively support MRM, the results began to fade and then reverse (Taylor & Christensen, 1998; p. 127).
   Several years later the negative consequences of management not supporting a program was documented in another company. AMTs, at first enthusiastic about MRM became frustrated in the months to follow and expressed antagonism to the program when surveyed and interviewed about it (Taylor, 1998; Taylor & Christensen, 1998, pp. 160-161).
   Despite this evidence the airline company sponsoring the training in written turnover (described in section I above) did not heed the advice and repeated

warnings to actively and visibly support their grogram's aims and intentions. Instead , top management seemed satisfied to continue the training when and as other priorities did not interfere. No top management guidance or constraint on middle management to vocally and visibly support the MRM program was ever reported.

To succeed well and for the long term, all management must lead and guide MRM efforts.

References:

ATA (U.S. Air Transport Association) (2001). *Spec 113: Maintenance Human Factors Program Guidelines.* Retrieved December 3, 2001, from http://www.airlines.org/public/publications/display1.asp?nid=938.

Brown, J.R. (1991). The retrograde motion of planets and children: Interpreting percentile rank. *Psychology in the Schools, 28, 345-353.*

Downie, N.M. & Heath, R.W. (1974). Basic Statistical Methods, 4th ed., Harper & Row, New York.

Choi, S. (1995). *The Effects of A Team Training Program and Inferences for Computer Software Development.* Ph.D. Dissertation, Education Department. Los Angeles: University of Southern California.

Gregorich, S.E.; Helmreich, R.L. & Wilhelm, J.A., (1990). The structure of cockpit management attitudes. *Journal of Applied Psychology, 75*, 682-690.

Hansen, J.S. & Oster, C.V., Jr. (eds.) (1997) *Taking Flight.* Washington, DC: National Academy Press.

Helmreich, R.L.; Foushee, H.C.; Benson, R. & Russini, R., (1986). Cockpit management attitudes: exploring the attitude-performance linkage. *Aviation, Space and Environmental Medicine,* 57, 1198-1200.

Helmreich, R.L. & Merritt, A.C. (1998). *Culture at Work in Aviation and Medicine.* Aldershot, Hants:Ashgate Publishing.

Hutchinson, III C.R. (1997) "Aviation speedometers, metrics on the hangar floor." *Ground Effects,* Jan-Feb, 1-5.

Jian, J., Bisantz, A.M. & Drury, C.G. (1998). Towards an empirically determined scale of trust in computerized systems: distinguishing concepts and types of trust. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, Santa Monica: HFES, 501-505.

Kirkpatrick, D. (1983)." Four steps to measuring training effectiveness." *Personnel Administration, 28*, (11), 19-25.

Kramer, R.M. & Tyler, T.R. (1996). "Whither Trust?" In Kramer, R.M. & Tyler, T.R (eds.), *Trust in Organizations.* Thousand Oaks:Sage Publications.

Marske, C.E. & Taylor, J.C. (1997) "The Socio-Cultural Transformation of Transportation Safety." In *Proceedings of the Symposium on Corporate Culture and Transportation Safety.* Washington,DC: National Transportation Safety Board.

Mishra , A.K. (1996). "Organizational Responses to Crisis: The Centrality of Trust." In Kramer, R.M. & Tyler, T.R (eds.), *Trust in Organizations.* Thousand Oaks:Sage Publications.

Norusis, M.J., (1990). *SPSS Base System User's Guide.* Chicago: SPSS.

NTSB (National Transportation Safety Board), (1997) *Aircraft Accident Report: In-flight fire and impact with terrain, Valujet Airlines Flight 592, Douglas DC-9-32, N904VJ, Everglades, near Miami FL, May 11, 1996.* NTSB/AAR-97/06. Washington, DC.

Patel, S., Drury, C.G. and Lofgren, J. (1994). Design of workcards for aircraft inspection. *Applied Ergonomics*, 25 (5), 283-293.

Patankar, M & Taylor, J. (2000). *Targeted MRM Programs: Setting ROI Goals and Measuring the Results.* SAE Technical Paper 2000-01-2127. SAE Advances in Aviation Safety Conference & Exposition, Daytona Beach, FL.

Perrow, C. (1999) *Normal Accidents, Revised Ed.* Princeton University Press: Princeton, NJ.

Seltiz, C.; Wrightsman, L.S. & Cook, S.W. (1976). *Research Methods in Social Relations 3rd edition.* New York: Holt, Rinehart and Winston.

Sherman, P.J. (1992). "New Directions of CRM Training." *Proceedings of the Human Factors Society 36 Annual Meeting,* p.896.

Stapleton, C.D. (1997). *Basic concepts in exploratory factor analysis (EFA) as a tool to evaluate score validity: A right-brained approach.* Retrieved December 3, 2001, from http://ericae.net/ft/tamu/efa.htm.

Taggart, W., (1990). "Introducing CRM into maintenance training." *Proceedings of the Third International Symposium on Human Factors in Aircraft Maintenance and Inspection.* Washington, D.C.: Federal Aviation Administration, 93-110.

Taylor, J.C. (1994) "Using Focus Groups to Reduce Errors in Aviation Maintenance"(Original title: Maintenance Resource Management [MRM] in Commercial Aviation: Reducing Errors in Aircraft Maintenance Documentation, Technical Report -- 10/31/94) Los Angeles: Institute of Safety & Systems Management, University of Southern California (available at "www.hfskyway.com/document.htm").

Taylor, J.C. (1998) *Evaluating the effects of maintenance resource management (MRM) interventions in airline safety* (Annual Report FAA Grant #96-G-003). Santa Clara University (available at "www.hfskyway.com/document.htm")

Taylor, J.C., (2000a). "The Evolution And Effectiveness Of Maintenance Resource Management (MRM)." *International Journal of Industrial Ergonomics, 26* (2), 201-215.

Taylor, J.C., (2000b). "Reliability And Validity of the 'Maintenance Resource Management, Technical Operations Questionnaire' (MRM/TOQ)." *International Journal of Industrial Ergonomics, 26* (2), 217-230.

Taylor, J.C. (2000c) Evaluating The Effects Of Maintenance Resource Management (MRM) In Air Safety. Report of Research Conducted under NASA-Ames Cooperative Agreement No. NCC2-1025 (SCU Project # NAR003). Santa Clara University.

Taylor, J.C. and Christensen, T.D. (1998). *Airline Maintenance Resource Management: Improving Communication.* Society of Automotive Engineers: Warrendale, PA.

Taylor, J.C. & Patankar, M.S. (1999). "Cultural Factors Contributing To The Success Of Macro Human Factors In Aviation Maintenance." *Proceedings of The Tenth International Symposium on Aviation Psychology.* The Ohio State University.

Taylor, J.C. & Patankar, M.S. (2001). "Four Generations of Maintenance Resource Management Programs in the United States: An Analysis of the Past, Present, and Future" *The Journal of Air Transportation World Wide,* Vol *6* (2), 3-32.

Taylor, J.C. & Robertson, M.M. (1994). "Successful Communication for Maintenance." *The CRM Advocate.* October, pp. 4-7.

Taylor, J.C. & Robertson, M.M. (1995). *The Effects of Crew Resource Management (CRM) Training in Airline Maintenance: Results Following Three Years Experience.* NASA Contractor Report 196696, Washington, D.C.:National Aeronautics and Space Administration.

Taylor, J.C. & Thomas, R.L. (2001a). "Written Communication Practices as Impacted by a Maintenance Resource Management Training Intervention" Project Report, MRM Research Program, Engineering School, Santa Clara University.

Taylor, J.C. & Thomas, R.L. (2001b). "Toward measuring safety culture in aviation maintenance: The structure of trust and professionalism." Project Report, MRM Research Program, Engineering School, Santa Clara University.

Warr, P.; Allen, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72 (3), 351-375.

Wrightsman, L.S. (1974). *Assumptions about human nature: A social-psychological analysis.* Monterey, CA: Brooks/Cole.

Appendix A: Calculator Scales and Survey Questions that Comprise Each Scale

Supervisor Trust and Safety

My supervisor can be trusted

My suggestions about safety would be acted upon if I expressed them to my lead or supervisor

My supervisor protects confidential or sensitive information

I know the proper channels to route safety questions

Mechanics' ideas are carried up the line

Value Communication and Trust in Coworkers

Having the trust and confidence of my coworkers is important

A debriefing and critique of procedures and decisions after a significant task is completed is an important part of developing and maintaining effective crew coordination

Employees should make the effort to foster open, honest and sincere communication

Start of shift crew meetings are important for safety and for effective crew management

My coworkers value consistency between words and actions

Assertiveness

Maintenance personnel should avoid disagreeing with one another

It is important to avoid negative comments about the procedures and techniques of other team members

Effects of My Stress

Even when fatigued, I perform effectively during critical phases of work

A truly professional team member can leave personal problems behind when working

Personal problems can adversely affect my performance

Appendix B: Developmental MRM/TOQ
(item numbers are the same as those in Table 4)

<<Maintenance Resource Management/Technical Operations Questionnaire (MRM/TOQ)

Maintenance management is interested in your comments regarding human factors and safety within the department. The success of this survey depends on your contribution, so it is important to answer as honestly and fairly as you can. All answers are confidential. There are no right or wrong answers. This survey is part of a NASA-sponsored study regarding maintenance safety throughout the USA. Additional comments are welcome throughout the survey. ***Completed surveys will be sent directly to Santa Clara University for analysis.>>***

I. BACKGROUND INFORMATION: Today's Date: ___/___/___

1. Job Title: ____________ ________________
2. Years in Maintenance at this company: _____
3. City or Station:______
4. Present Shift: ________
5. Gender Male Female
6. Year of birth:_________
7. Past Experience or Training: (# of years: fill in below)
   Military: _____ Trade School: _____ College: ____ Other Aviation: ____
   (Specify other company if "Other Aviation":__________________)
8. Non-Contract Contract
9. Where do you work? Line Hangar QC Planning Shop Stores Engineering Appearance Other

II. TECHNICAL OPERATIONS ATTITUDE MEASUREMENT:

1 Strongly Disagree	2 Slightly Disagree	3 Neutral	4 Slightly Agree	5 Strongly Agree

Using the scale above, please circle the number that best describes your opinion.

1 2 3 4 5 (1) My supervisor can be trusted.

1 2 3 4 5 (3) My suggestions about safety would be acted on if I expressed them to my lead or supervisor.

1 2 3 4 5 (4) My supervisor protects confidential or sensitive information

1 2 3 4 5 (6) Mechanics' ideas are carried up the line.

1 2 3 4 5 (7) I know the proper channels to route questions regarding safety practices.

1 2 3 4 5 (8) Having the trust and confidence of my coworkers is important.

1 2 3 4 5 (9) A debriefing and critique of procedures and decisions after a significant task is completed is an important part of developing and maintaining effective crew coordination

1 2 3 4 5 (11) Start of shift crew meetings are important for safety and for effective crew management.

1 2 3 4 5 (13) Employees should make the effort to open, honest, and sincere communic

1 2 3 4 5 (16) Personal problems can adversely aff performance.

1 2 3 4 5 (18) Maintenance personnel should avoid disagreeing with one another.

1 2 3 4 5 (20) Even when fatigued, I perform effec during critical phases of work.

1 2 3 4 5 (22) A truly professional team member c personal problems behind when wor

1 2 3 4 5 (23) . It is important to avoid negative co about the procedures and techniques team members.

1 2 3 4 5 (24) My coworkers value consistency bet words and actions.

Appendix C

Maintenance Resource Management/Technical Operations Questionnaire (Pre-training)

Your maintenance organization is interested in your comments regarding human factors and safety within the department. The success of this survey depends on your contribution, so it is important to answer as honestly and fairly as you can. All answers are confidential. There are no right or wrong answers. This survey is part of a FAA and NASA-sponsored study regarding maintenance safety throughout the USA. Additional comments are welcome throughout the survey.

I. BACKGROUND INFORMATION: Today's Date: ___/___/___

1. Job Title: ______________________________
2. Years in Maintenance at this company: _____
3. City or Station:______
4. Present Shift: _______
5. Gender   Male   Female
6. Year of birth:_________
7. Past Experience or Training: (# of years: fill in below)
   Military: _____ Trade School: _____ College: ____ Other Aviation: ____
   (Specify other company if "Other Aviation":_________________)
8. Non-Contract   Contract
9. Where do you work? Line   Hangar   QC   Planning   Shop   Stores   Engineering   Appearance   Other

II. TECHNICAL OPERATIONS ATTITUDE MEASUREMENT:

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

Using the scale above, please circle the number that best describes your opinion.

Scale	No.	Statement
1 2 3 4 5	1.	Maintenance personnel should avoid disagreeing with one another.
1 2 3 4 5	2.	Even when fatigued, I perform effectively during critical phases of work.
1 2 3 4 5	3.	My suggestions about safety would be acted on if I expressed them to my lead or supervisor.
1 2 3 4 5	4.	My supervisor protects confidential or sensitive information
1 2 3 4 5	5.	It is important to avoid negative comments about the procedures and techniques of other team members.
1 2 3 4 5	6.	Mechanics' ideas are carried up the line.
1 2 3 4 5	7.	I know the proper channels to route questions regarding safety practices.
1 2 3 4 5	8.	Having the trust and confidence of my coworkers is important.
1 2 3 4 5	9.	A truly professional team member can leave personal problems behind when working.
1 2 3 4 5	10.	We should always provide both verbal turnover to the oncoming shift.
1 2 3 4 5	11.	Employees should make the effort to f honest, and sincere communication.
1 2 3 4 5	12.	My supervisor can be trusted.
1 2 3 4 5	13.	My work impacts passenger satisfacti
1 2 3 4 5	14.	A debriefing and critique of procedure decisions after a significant task is co an important part of developing and m effective crew coordination
1 2 3 4 5	15.	Personal problems can adversely affec performance.
1 2 3 4 5	16.	My coworkers value consistency betw and actions.
1 2 3 4 5	17.	Start of shift crew meetings are impor safety and for effective crew manage
1 2 3 4 5		

THANK YOU FOR YOUR PARTICIPATION IN THIS SURVEY.

Appendix D
Maintenance Resource Management/Technical Operations Questionnaire (Post-training)

Your maintenance organization is interested in your comments regarding human factors and safety within the department. The success of this survey depends on your contribution, so it is important to answer as honestly and fairly as you can. All answers are confidential. There are no right or wrong answers. This survey is part of a FAA and NASA-sponsored study regarding maintenance safety throughout the USA. Additional comments are welcome throughout the survey.

I. BACKGROUND INFORMATION: Today's Date: ___/___/___

1. Job Title: ______________ ________________
2. Years in Maintenance at this company: _____
3. City or Station:______
4. Present Shift: _______
5. Gender Male Female
6. Year of birth:_________
7. Past Experience or Training: (# of years: fill in below)
   Military: _____ Trade School: _____ College: ____ Other Aviation: ____
   (Specify other company if "Other Aviation":_________________)
8. Non-Contract Contract
9. Where do you work? Line Hangar QC Planning Shop Stores Engineering Appearance Other

II. TECHNICAL OPERATIONS ATTITUDE MEASUREMENT:

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

Using the scale above, please circle the number that best describes your opinion.

1 2 3 4 5 1. Maintenance personnel should avoid disagreeing with one another.

1 2 3 4 5 2. Even when fatigued, I perform effectively during critical phases of work.

1 2 3 4 5 3. My suggestions about safety would be acted on if I expressed them to my lead or supervisor.

1 2 3 4 5 4. My supervisor protects confidential or sensitive information

1 2 3 4 5 5. It is important to avoid negative comments about the procedures and techniques of other team members.

1 2 3 4 5 6. Mechanics' ideas are carried up the line.

1 2 3 4 5 7. I know the proper channels to route questions regarding safety practices.

1 2 3 4 5 8. Having the trust and confidence of my coworkers is important.

1 2 3 4 5 9. A truly professional team member can leave personal problems behind when working.

1 2 3 4 5 10. We should always provide both verbal turnover to the oncoming shift.

1 2 3 4 5 11. Employees should make the effort to f honest, and sincere communication.

1 2 3 4 5 12. My supervisor can be trusted.

1 2 3 4 5 13. My work impacts passenger satisfacti

1 2 3 4 5 14. A debriefing and critique of procedure decisions after a significant task is co an important part of developing and m effective crew coordination

1 2 3 4 5 15. Personal problems can adversely affec performance.

1 2 3 4 5 16. My coworkers value consistency betw and actions.

1 2 3 4 5 17. Start of shift crew meetings are impor safety and for effective crew manage

Please go on to the other side-

1	2	3	4	5
Strongly Disagree	Slightly Disagree	Neutral	Slightly Agree	Strongly Agree

III. Human Factors Training QUESTIONS:

Using the scale above, please circle the number that best describes your opinion about each item.

1 2 3 4 5 **1.** This training has the potential to increase aviation safety and crew effectiveness.

1 2 3 4 5 **2.** This training will be useful for others

3. Is the training going to change your behavior on the job? (circle one from the list below)

No Change A Slight Change A Moderate Change A Large Chang

4. How will you use the information from the Human Factors training on your job?

5. What aspects of the Human Factors training were particularly good?

6. What do you think could be done to improve the training?

THANK YOU FOR YOUR PARTICIPATION IN THIS SURVEY.