To be presented at the 13th International Symposium on Aviation Psychology, Dayton, OH, 2005.

DEVELOPMENT AND VALIDATION OF A SURVEY TO ASSESS SAFETY CULTURE IN AIRLINE MAINTENANCE OPERATIONS[1]

Alyssa Mitchell Gibbons
Department of Psychology – University of Illinois at Urbana-Champaign
Champaign, Illinois
Terry L. von Thaden
Aviation Human Factors Division – University of Illinois at Urbana-Champaign
Savoy, Illinois
Douglas A. Wiegmann
Aviation Human Factors Division – University of Illinois at Urbana-Champaign
Savoy, Illinois

This paper describes the development and validation of a survey to assess safety culture in airline maintenance operations according to the five-factor model of safety culture proposed by Wiegmann et al. (2002). Maintenance technicians at two FAR Part 121 scheduled passenger airlines (N = 109 and 76) completed the original version of the survey. The results yielded useful diagnostic information about the safety culture of each airline, but factor analyses indicated that the five-factor model may not be adequate to describe the data. A more complex model is proposed and modifications to the survey are suggested.

Safety Culture

Aviation organizations are becoming highly interested in understanding safety culture and how it can be improved. Safety culture can be defined as "the enduring value and priority placed on worker and public safety by everyone in every group at every level of an organization" (Wiegmann, Zhang, von Thaden, Sharma, & Mitchell, 2002). Wiegmann et al. (2002) reviewed the safety culture literature across a number of industries and identified five critical indicators of an organization's safety culture:

Organizational Commitment (OC): the organization's commitment to safety, as expressed by upper management;
Managerial Involvement (MI): the active involvement of mid-level managers or supervisors in promoting safety;
Employee Empowerment (EE): the degree to which individual employees are empowered to make safety a priority;
Accountability System (AS): the system by which employees are held accountable for acting unsafely; and
Reporting System (RS): the quality and usability of the system for reporting and processing safety information.

While strength in one area can compensate to some extent for deficiency in another (e.g., strong employee empowerment may limit the negative impact of poor management involvement), strength in all areas is the hallmark of a culture that truly promotes safety to the fullest.

Commercial Aviation Safety Survey

The Commercial Aviation Safety Survey (CASS) was developed, based on the five-factor model described above, to assist airlines in diagnosing strengths and weaknesses within their safety cultures so that the weaknesses can be addressed. The flight operations version of the CASS was created first, with items based on existing safety culture inventories from a number of industries. Wiegmann, von Thaden, Mitchell, Sharma, & Zhang (2003) provide a detailed description of the development of the flight operations survey. The development of the maintenance survey is the focus of the present report. The maintenance version of the survey is designed to reflect the same structure as the flight operations version (that is, the same five indicators of safety culture), but to use terminology and describe behavior appropriate to the maintenance function.

Several steps were taken to develop the survey in such a way that it paralleled the structure of the flight operations survey but contained items directly relevant to maintenance professionals. The flight operations survey contained 89 items. Thirty-eight of these were judged as applicable for maintenance with minimal revisions. An additional 13 items were modified more extensively to reflect the intent of

[1] This research was supported by Federal Aviation Administration Award No. DTFA 01-G-015. Kathy Fazen was the technical monitor. The views expressed are those of the authors and do not necessarily reflect those of the Federal Aviation Administration. The authors thank William Geibel and Marcus Credé for their assistance in the survey development process.

each original flight operations item in a maintenance context. For example, the item "management expects pilots to push the weather" was replaced by the item "supervisors never pressure inspectors to sign-off on borderline work." Maintenance technicians are not at all likely to encounter the scenario described in the first item, but both items represent the same basic idea: a specific, common situation in which the responsible manager pressures a subordinate to behave in an unsafe manner.

At the time that the maintenance survey was being developed, one preliminary test of the flight operations survey had already been conducted, so general feedback from that test was incorporated into the revisions. Apparently confusing or ambiguous items were excluded, as were items that did not appear to have clear parallels in maintenance operations. Additional items were chosen and/or written to fill in the gaps left by the excluded items. The original safety culture inventories used to create the flight operations survey were consulted again, to see whether we had overlooked any items appropriate for maintenance. This search yielded six items. Twenty new items were written based on the extensive airline maintenance experience of one of the researchers, who pointed out situations and attitudes common in such an environment.

The final version of the maintenance survey contained 84 items. As in the flight operations survey, respondents were instructed to use a 7-point Likert-type response scale to indicate their agreement or disagreement with each item. A rating of 1 indicated that the respondent "strongly disagree(d)" with the item and a rating of 7 indicated that he or she "strongly agree(d)." The center point of the rating scale, 4, was labeled "neither agree nor disagree." Space was provided beside each item for respondents to write comments if they chose.

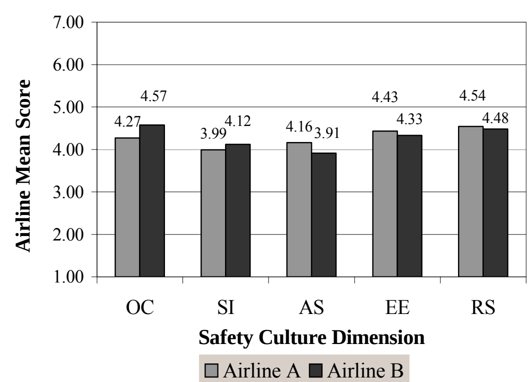
Initial Results from Two Airlines

Maintenance personnel from two FAR Part 121 passenger airlines completed the survey. Participants returned surveys directly to the researchers. They were assured that their responses would remain confidential and they were not asked to provide their names or other personally identifying information. No compensation was offered to participants or their organizations.

A total of 1148 surveys were distributed: 860 to employees of Airline A and 288 to Airline B. One hundred and nine of the Airline A surveys and 76 of the Airline B surveys were returned, for response rates of 13% and 26% respectively. At Airline A, most respondents (74%) described their primary job responsibility as "Aircraft Technician;" at Airline B, respondents were more evenly divided between technicians (40%) and supervisory positions (Line Manager, Lead Technician, Inspector, or Manager; 51% combined).

Dimension Scores. Scores for each airline were calculated for each of the five dimensions of safety culture as the mean of participants' responses to the items in each dimension scale. Items indicating a negative safety culture (e.g., "My airline is more concerned with making money than being safe.") were reverse coded. All five scales showed acceptable levels of reliability for both airlines (α = .74 – .94). Dimension scores for both airlines appear in Figure 1.

Figure 1. CASS scores for two FAR Part 121 airlines.



Both of these airlines appear to have "middle-of-the-road" safety cultures, with scores near the neutral point (4.0) in most areas. However, the pattern of the dimension scores suggests different areas of strength and weakness for each airline, implying that the actual safety cultures experienced by employees are quite different. Reporting systems are strong at both airlines; they are the strongest area at Airline A, while organizational commitment is Airline B's strongest dimension. Airline A needs improvement in supervisory involvement, while Airline B needs to improve its accountability system. Analyses of individual item responses and respondent comments further supported these overall impressions. All scales were negatively correlated with technicians' perceptions of risk at both airlines. The maintenance CASS appears to be a useful diagnostic tool. The items can be grouped together into reliable scales to provide a broad-level picture of the organization or analyzed individually to identify specific strengths and weaknesses, providing useful information to airline management seeking to improve safety culture.

Factor Analysis of the Maintenance CASS

Analytical Strategy

To validate the five-dimensional model of safety culture proposed above, we conducted confirmatory factor analyses (CFAs) using the Mx software package (Neale, 2002). We conducted an overall CFA for the five factor model and then tested single-factor models for each of the five dimension scales individually. In all analyses, model fit was tested by considering the overall chi-squared value for the model (X^2), the root mean squared error of approximation (RMSEA), the normed fit index (NFI), the Tucker-Lewis index (TLI) and the relative noncentrality index (RNI; see Neale, Boker, Xie, & Maes, 2004 for definitions and citations for all fit indices). Models are usually considered to fit well when the X^2 value is nonsignificant compared to the degrees of freedom, the RMSEA is below .10, and the NFI, TLI, and RNI are above .90 (McDonald & Ho, 2002). Given the relatively small sample used in this study, we considered a model to fit well when most of these criteria were met.

If a model did not fit well, we considered the matrix of residual discrepancies between the observed correlation matrix and that expected under the model. When an item showed large residual correlations (> .15) with other items, we considered whether the item might have been confusing or ambiguous, whether it was highly correlated with only one other item (introducing instability into the model), or whether it showed a pattern of large residuals with other items that might suggest the existence of another factor. In the first two cases, the item was discarded and the fit of the model without that item was assessed. In the third case, items were grouped logically into subfactors and the fit of the new multifactor model was assessed. Improvement in fit was assessed by means of chi-squared difference tests, which compare the fit of the original model with the fit of the revised model. A significant difference implies that the revised model fits significantly better than did the original model. If the large residuals within a scale could not be resolved through these methods, exploratory factor analyses (EFAs) were conducted using the COFA (see McDonald, 1999) software program to determine whether multiple factors were needed to describe the data. Exploratory factor analyses were subsequently retested with confirmatory factor analyses so that the fit obtained could be compared to that of the original models.

The data from Airline A were used in the initial analyses, because the small sample size from Airline B was not sufficient to allow a test of the full model. The data from Airline B were used as a cross-validation sample for the revised versions of the individual factor scales.

Overall Model Fit. The first model tested was the one hypothesized: a five-factor solution with each item loading on the dimension it was intended to measure. This model fit the data poorly: X^2_{2765} = 5660.27, $p <$ .01; RMSEA = .11; NFI = .30; TLI = .43; RNI = .45. Further, 10% of the residual correlations had absolute values greater than .15. One possible explanation for the poor fit of the five-factor model is that respondents did not discriminate between dimensions of safety culture when completing the survey, but rather based their responses on their overall perception of the safety culture as good or bad. If this were the case, a single-factor model in which all items simply reflect the overall positive or negative safety culture of the organization would fit well. The single-factor model also fit poorly, however: X^2_{2774} = 5711.40, p<.01; RMSEA = .11. In fact, the fit of the single-factor model was significantly poorer than that for the five-factor model, ΔX^2_9 = 51.14, $p <$ .01.As the data could not be described by either the five-factor model or a single general factor, the dimension scales were analyzed individually to identify specific sources of misfit.

Accountability System. The single-factor model for the accountability system scale showed acceptable fit: X^2_{27} = 32.93, p=.20; RMSEA = .05, NFI = .84, TLI = .95, RNI = .96. Examination of the residuals suggested that two pairs of items shared particularly high correlations, implying that the items in each pair may measure the same thing to such an extent as to be redundant. However, removing one item from each pair did not significantly improve the fit of the model, so the items were retained. Cross-validation with data from Airline B showed reasonably acceptable fit, X^2_{27} = 48.75, p=.01; RMSEA = .11; NFI = .87; TLI = .91; RNI = .94.

Reporting System. The single factor model showed acceptable fit for the reporting system scale, X^2_{35} = 42.63, p = .18; RMSEA = .05; NFI = .85; TLI = .96; RNI = .97. Low correlations between two pairs of items resulted in high residuals (> .15) for those pairs, but as all four items had high correlations with the other items in the scale and the overall fit of the model was good, they were retained. However, in the cross-validation sample from Airline B, the single-factor model did not fit as well, X^2_{35} = 70.13, p = .00; RMSEA = .12; NFI = .75; TLI = .81; RNI = .85. Removing the item “I am familiar with the system for formally reporting safety issues in my airline” improved the fit of the model in the Airline B sample, $X^2_{27} = 49.86$, $p = .01$; RMSEA = .11; $\Delta X^2_8 = 20.26$, $p<.01$; NFI = .80; TLI = .86; RNI = .89; and removing it from the Airline A data improved the fit slightly, but not significantly $X^2_{27} = 33.54$, $p = .18$; RMSEA = .05; $\Delta X^2_8 = 9.10$, $p = .33$. In light of that evidence, the item was retained.

Supervisory Involvement. The initial single factor model did not quite fit the supervisory involvement scale well, $X^2_{77} = 108.78$, $p = .01$; RMSEA = .07; NFI = .83; TLI = .93; RNI = .94. Examination of the residual matrix indicated that a large number of the discrepancies were related to two items. Respondent comments on one of the items indicated that the item was interpreted differently by different respondents, but the reason for the misfit of the other item was unclear. Removing both items, however, improved the fit of the model, so that the model containing twelve items fit acceptably, $X^2_{54} = 63.59$, $p = .17$; RMSEA = .04; $\Delta X^2_{11} = 24.11$, $p = .01$; NFI = .85; TLI = .95; RNI = .96. In the Airline B sample, the fit of the twelve-item model was similar, but not quite so good, $X^2_{54} = 84.73$, $p < .01$; RMSEA = .09; NFI = .80; TLI = .90; RNI = .92.

Employee Empowerment. The single factor model for the employee empowerment scale did not fit particularly well, $X^2_{54} = 87.42$, $p<.01$; RMSEA = .08, NFI = .73, TLI = .84, RNI = .87. An attempt to separate the items into two factors (technicians' authority to improve safety and their safety professionalism) based on large residuals and logical relationships among items yielded only slightly improved fit ($X^2_{53} = 76.26$, $p = .02$; RMSEA = .07) and a further division into three factors (authority, professionalism, and peer influence) did not fit better ($X^2_{51} = 75.50$, $p = .01$; RMSEA = .07). Consequently, an exploratory factor analysis (EFA) was conducted to investigate the structure of the scale. A three-factor model was tested first, because the three-factor model suggested above showed (though barely) the best fit of the three. The promax rotated solution identified three factors that were in many ways similar to the three factors suggested by conceptual grouping. The first factor appears to reflect supervisors' respect for technicians in safety matters (authority), the second describes technicians' personal pride in upholding safety standards (professionalism), and the third indicates a peer culture that supports safety (peer influence). In a CFA, this model showed acceptable fit, $X^2_{51} = 65.56$, $p = .08$; RMSEA = .05. One conceptual difficulty remained in that two items, "Everyone routinely performs the operational checks after the work is completed," and "Everyone routinely re-inspects each other's work or has someone inspect their work before return to service," were clearly similar in content, but loaded on different factors. However, the former item had near-equal loadings (.26 and .25, respectively) on both the professionalism and peer culture factors. Moving this item to the peer culture factor actually slightly improved the fit of the model, $X^2_{51} = 61.14$, $p = .16$; RMSEA = .05. In the interest of parsimony, a second exploratory analysis was conducted requesting only two factors. The factors identified by the promax rotation were identical to those suggested by the conceptual two-factor grouping. As that model had already been shown to fit poorly, the three-factor model for the employee empowerment dimension was retained.

The three-factor model appeared to fit the cross-validation data from Airline B well, $X^2_{51} = 45.54$, $p=.69$; RMSEA = .00. However, some of the fit indices were inappropriately high: NFI = .75; TLI = 1.06; RNI = 1.05. This suggests empirical underidentification, a condition that occurs when the observed correlations between variables in a sample are near zero. This is most likely a function of the small Airline B sample size, but it prevents us from being able to draw conclusions about the cross-sample validity of the three-factor employee empowerment model.

Organizational Commitment. The single-factor model did not fit the data well for the organizational commitment scale, $X^2_{434} = 732.23$, $p < .01$; RMSEA = .08; NFI = .55; TLI = .73; RNI = .75. Of the 465 residual correlations, ninety-nine were greater than .10, with 34 greater than .15. This suggests that a multi-factor model is necessary to describe the items in this scale – such pervasive residuals are not likely to be resolved by removing a few items. To identify a starting point for conceptually grouping these items, we looked to the parallel analysis that had previously been conducted for the flight operations survey. That investigation used an exploratory factor analysis to identify three factors: upper management attitude toward safety, use of preventive safety practices (such as safety training), and commitment of organizational resources to safety. The maintenance items were correspondingly grouped into similar factors and a three-factor model was tested. However, that model showed only small (but significant) improvement in fit over the single-factor model, $X^2_{431} = 704.48$, $p = .00$; RMSEA = .08.

A series of EFAs was then conducted using COFA. Two-, three-, four-, and five-factor solutions were tested, but the four-factor model showed the best fit in subsequent CFAs: $X^2_{399} = 550.69$, $p = .00$; RMSEA

= .06, with 21 residuals greater than .15. Fit indices for this model approached acceptable levels, NFI = .69, TLI = .88, RNI = .89. The first three factors in this model appeared to represent management attitude (e.g., "Unsafe behavior is not tolerated in my company"), allocation of resources (e.g., "Tool control, calibration, and equipment certification are closely monitored by my company"), and quality of safety training. .The fourth factor contained only three items, and no conceptual relationship between these was readily apparent, except perhaps that all were rather indirect assessments of safety culture. As this factor was not clearly interpretable and may have simply consisted of poor items, another confirmatory analysis was conducted excluding those three items (and thus the fourth factor). This new three-factor model (consisting of attitude, resource, and training factors) did not yield a significant improvement in fit over the four-factor model (X^2_{321} = 475.63, p = .00; RMSEA = .07; ΔX^2_{78} = 75.05, p=.57). However, in the revised model, it became apparent that many of the large residuals were associated with one item. Excluding this item from the new three-factor model resulted in a significant improvement in fit for that model (X^2_{296} = 421.30, p = .00; RMSEA = .07; ΔX^2_{25} = 54.33, p<.01), and the resulting model also fit significantly better than the four-factor model (ΔX^2_{103} = 129.38, p=.04). Fit indices for this model were similar to those for the four-factor model, NFI = .72, TLI = .88, RNI = .89. Eight large (> .15) residuals remained, but no item was connected with more than one of these, and no logical connections between pairs of items sharing large residuals were apparent. The revised three-factor model (attitude, resources, and training) was retained.

Again, data from Airline B were used to cross-validate the revised model. The three-factor model did not fit particularly well, X^2_{296} = 480.08, p = .00; RMSEA = .10, NFI = .61; TLI = .77; RNI = .79. This suggests that these factors should be used with caution in future research, as they may reflect idiosyncratic characteristics of Airline A rather than the general structure of organizational commitment across airlines.

Revised Model Overall Fit. When all revisions are taken into account, the new model contains a total of nine factors (the organizational commitment and employee empowerment scales were each divided into three factors). The original survey contained seventy-six items, but seven items were deleted in the revision process, so the revised model contained only sixty-nine. A confirmatory factor analysis was conducted to test the fit of the revised model. Again, however, the full model did not fit particularly well, X^2_{2246} = 4099.49, p < .01; RMSEA = .10; NFI = .38; TLI = .54; RNI = .56. While these values represent an improvement in fit compared to the original model, they still fall short of acceptable levels. Of the 2415 residuals, 231 (9.6%) had absolute values above .15.

To determine whether the observed improvement in fit was due to the regrouping of items or merely to the elimination of poor items, an additional confirmatory factor analysis was conducted using only the sixty-nine items in the revised model but grouping them into the original five factors. Again, this model fit poorly overall, X^2_{2267} = 4236.75, p < .01; RMSEA = .10. The fit of the revised (nine-factor) model was significantly better than that of this five-factor model, ΔX^2_{21} = 137.26, p < .01, but comparison of the other fit indices (NFI = .36; TLI = .52; RNI = .54.) suggests that the difference is slight.

Conclusions

While the results of the factor analysis generally supported the three of the five factors, the organizational commitment and employee empowerment factors remained problematic, and even the revised complete model did not show a good fit to the data. These findings are consistent, however, with the corresponding analysis of the flight operations survey. That survey also indicated a need to divide organizational commitment and employee empowerment into three subfactors each, and the subfactors identified in that analysis correspond conceptually in many ways to the subfactors identified here. The correspondence is not exact, but these findings do suggest two conclusions: (1) While the accountability system, reporting system, and management/supervisory involvement scales appear to represent well-defined, unitary constructs, the pilot/employee empowerment and organizational commitment scales represent more complex phenomena that require further consideration; and (2) within those two scales, several consistent themes emerge that provide insight into how those constructs might be better defined in future.

Specifically, the concept of employee (or pilot) empowerment seems to encompass several distinct elements: the authority granted to employees by the organization, the authority and personal responsibility assumed by employees, and the positive or negative impacts of the peer culture regarding safety. With respect to organizational commitment, respondents in both flight operations and maintenance appeared to distinguish between the "talk" (statements and policies) and the "walk" (actions and commitment of resources) of their organizations.

The analysis of the flight operations survey concluded with a conceptual revision of the scale, based on input from the factor analysis and from respondent comments. For example, the pilot empowerment subfactors were linked with the accountability system factor as aspects of an "Informal Safety System" second-order factor and the reporting system was similarly divided to indicate different parts of the reporting process. Given the strong conceptual similarity between the flight operations analysis and the results reported here, we considered whether a similar structure could be adopted for the maintenance survey. Again, we considered respondents' comments as well as the factor analyses to identify problematic items or areas of concern to technicians that might have been overlooked in the original survey. The revised model for maintenance appears in Table 1.

Table 1. Revised model of safety culture maintenance.

Construct	Factors	Subfactors
Overall Safety Culture	Organizational Commitment	Safety Values Safety Fundamentals Work Environment Safety Training
	Supervisors	Supervisory Involvement Maintaining Standards
	Informal Safety System	Accountability Technicians' Authority Professionalism
	Formal Safety System	Reporting System Response & Feedback Safety Personnel

The informal safety system and reporting system factors from the revised flight operations survey were retained for the revised maintenance version. For the organizational commitment factor, the safety values and safety fundamentals subfactors from the flight operations survey were kept, but subfactors for safety training and a safe work environment were added. The supervisory involvement factor was retitled "Supervisors" and contained only two subfactors: supervisory involvement and "maintaining standards". This latter subfactor referred to supervisors' consistent enforcement of high safety standards. This reorganization required the creation of several new, specific items to ensure that each subfactor had enough items to be stable in future analyses. Items were also excluded if they seemed less relevant than or redundant with other items in the same scale. We also revised item wording wherever it appeared that an item might have been ambiguous or confusing.

While this initial test of the maintenance version of the CASS did not provide solid support for the five-factor model of safety culture, it nevertheless provided useful information on which further revisions can be based. The five scales proved useful as a diagnostic tool for identifying strengths and weaknesses of two airlines' safety cultures. Detailed factor analyses indicated that the accountability system, reporting system, and supervisory involvement factors represented fairly unitary constructs, which the employee empowerment and organizational commitment factors were more complex. As this is consistent with the findings of the flight operations survey, it seems likely that this reflects true complexity in the construct rather than only measurement error. When combined with respondents' substantive comments on the items, the factor analyses yielded information that was useful in creating a revised model of maintenance safety culture parallel to that created for flight operations . This new model formed the basis for an extensive revision to the maintenance CASS that may be tested in future research.

References

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Laurence Erlbaum.

McDonald, R. P., & Ho, M-H. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64-82.

Neale, M. C. (2002). *Mx*. Richmond: Virginia Commonwealth University.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2004). Mx: Statistical Modeling. Richmond: Virginia Commonwealth University.

Wiegmann, D.A., von Thaden, T. L., Mitchell, A. A., Sharma, G., & Zhang, H., (2003). Development and initial validation of a safety culture survey for commercial aviation. Technical Report AHFD-03-3/FAA-03-1. Prepared for the Federal Aviation Administration, contract DTFA 01-G-015.

Wiegmann, D., Zhang, H., von Thaden, T., Sharma, G., & Mitchell, A. (2002). A synthesis of safety culture and safety climate research. Technical Report ARL-02-3/ FAA-02-2. Prepared for the Federal Aviation Administration, contract DTFA 01-G-015.