

---

# Quality of Care

---

## Error Reduction and Performance Improvement in the Emergency Department through Formal Teamwork Training: Evaluation Results of the MedTeams Project

*John C. Morey, Robert Simon, Gregory D. Jay, Robert L. Wears, Mary Salisbury, Kimberly A. Dukes, and Scott D. Berns*

---

**Objective.** To evaluate the effectiveness of training and institutionalizing teamwork behaviors, drawn from aviation crew resource management (CRM) programs, on emergency department (ED) staff organized into caregiver teams.

**Study Setting.** Nine teaching and community hospital EDs.

**Study Design.** A prospective multicenter evaluation using a quasi-experimental, untreated control group design with one pretest and two posttests of the Emergency Team Coordination Course™ (ETCC). The experimental group, comprised of 684 physicians, nurses, and technicians, received the ETCC and implemented formal teamwork structures and processes. Assessments occurred prior to training, and at intervals of four and eight months after training. Three outcome constructs were evaluated: team behavior, ED performance, and attitudes and opinions. Trained observers rated ED staff team behaviors and made observations of clinical errors, a measure of ED performance. Staff and patients in the EDs completed surveys measuring attitudes and opinions.

**Data Collection.** Hospital EDs were the units of analysis for the seven outcome measures. Prior to aggregating data at the hospital level, scale properties of surveys and event-related observations were evaluated at the respondent or case level.

**Principal Findings.** A statistically significant improvement in quality of team behaviors was shown between the experimental and control groups following training ( $p = .012$ ). Subjective workload was not affected by the intervention ( $p = .668$ ). The clinical error rate significantly decreased from 30.9 percent to 4.4 percent in the experimental group ( $p = .039$ ). In the experimental group, the ED staffs' attitudes toward teamwork increased ( $p = .047$ ) and staff assessments of institutional support showed a significant increase ( $p = .040$ ).

**Conclusion.** Our findings point to the effectiveness of formal teamwork training for improving team behaviors, reducing errors, and improving staff attitudes among the ETCC-trained hospitals.

**Key Words.** Teams, teamwork, MedTeams, teamwork training, medical error

---

The MedTeams project is a translational research effort to apply crew resource management (CRM) behavioral principles developed in aviation to emergency medical care. Hospital emergency departments share many of the same characteristics with workplaces where CRM is effective, such as time-stress, dispersed and complex information, multiple players, and high-stakes outcomes. Preliminary observations in emergency departments (ED) established that the same CRM behaviors employed by highly effective aviation teams could be useful in the ED (Weiner, Kanki, and Helmreich 1993; Simon, Morey, and Locke 1997). A retrospective review of ED closed claims revealed that failure to engage in one or more of these teamwork behaviors was associated with an adverse event and indemnity payments. In 43 percent of the cases reviewed, teamwork behaviors would have prevented or mitigated the adverse event had they been applied (Risser, Rice et al. 1999). Similar analyses attribute about 80 percent of anesthesia mishaps to human error and 70 percent of commercial aviation accidents to crew errors (Gardner-Bonneau 1993; Taggart 1994).

Crew training has led to reductions in aviation mishaps beyond those produced by improvements in equipment and technology. The aviation community began introducing CRM training two decades ago and it is now required for all military and commercial U.S. aviation crews and air carriers operating internationally (Helmreich and Foushee 1993; Helmreich 1997). The basic principle of CRM is that crew communication and coordination behaviors are identifiable, teachable, and applicable to high-stakes environments. An additional principle is that those behaviors, although seen spontaneously, are not practiced reliably, regularly, or well unless specific training and reinforcement has established them. Specific CRM behaviors have been identified through experimentation and observation of high-reliability teams

---

This work was supported by U.S. Army Research Laboratory contract DAAL01-96-C-0091. The views, opinions, and findings are those of the authors and should not be construed as an official U.S. Department of Defense position, policy, or decision unless so designated by other official documentation.

Address correspondence to John C. Morey, Ph.D., Senior Research Psychologist, Crew Performance Group, Dynamics Research Corporation, 60 Frontage Road, Andover, MA, 01810. Robert Simon, Ed.D., is Chief Scientist, Systems Engineering Group, and Mary Salisbury, R.N., M.S., is Director, MedTeams Systems Group, both with Dynamics Research Corporation in Andover. Gregory D. Jay, M.D., Ph.D., is Assistant Professor, Department of Medicine, Section of Emergency Medicine, Brown University, Providence, RI. Scott D. Berns, M.D., M.P.H., is Assistant Professor of Pediatrics at Brown. Robert L. Wears, M.D., M.S., Professor, Department of Emergency Medicine, University of Florida Health Sciences Center, Jacksonville. Kimberly A. Dukes, Ph.D., is with DM-Stat, Inc., Medford, MA.

in demanding, time-stressed environments such as combat aviation and naval command and control centers (Leedom and Simon 1995; McIntyre and Salas 1995). Crew resource management training has been shown to be effective in these environments and is being extended into other domains (Helmreich and Foushee 1993; Salas et al. 1999).

Finally, an essential principle of CRM is that a team needs to be formally established for teamwork behaviors to be effective. In contrast to the notion of a team as any loosely coordinated group of caregivers and support staff, the formal teamwork structure of this study stipulates that a team be made up of between 3 to 10 members. A team is composed of physicians, nurses, and technicians who are organized for a shift. The number of designated teams for a shift depends on factors such as staffing levels and patient volume. From these larger teams, ad hoc teams are formed to respond to emergent events such as resuscitations. In this model, teamwork is sustained by a shared set of teamwork skills rather than permanent assignments that carry over from day to day.

Teamwork theory development has focused on input variables affecting team functioning such as task, work environment, and team member characteristics (Salas et al. 1992), what and how to train (Salas and Cannon-Bowers 2001), and outcome constructs of teamwork effectiveness (McIntyre and Salas 1995). Explanatory mechanisms of team processes exist in the form of constructs such as situational awareness and shared mental models, but relating team processes to work productivity outcomes has been limited by measurement difficulties with these constructs. Because this study examined the applicability of CRM to health care, a goal of this study was to generate a set of testable teamwork process–outcome propositions for future healthcare research. With respect to the training intervention, we sought to gain insight into the effectiveness of the training materials and methods, and features of the curriculum that needed revision. In addition, we sought to determine if the training intervention changed staff attitudes and behaviors and had an impact on patient care. These complementary perspectives are referred to as formative and summative evaluation in the education research literature (Bloom, Hastings, and Madaus 1971).

The first objective of this study was to adapt an aviation-oriented teamwork curriculum to the particular circumstances of EDs by developing and then implementing a training curriculum (Emergency Team Coordination Course [ETCC]) organized around five team dimensions (maintain team structure and climate, apply problem-solving strategies, communicate with the team, execute plans and manage workload, and improve team skills)

(Risser, Rice et al. 1999; Risser, Simon et al. 1999). The second objective was to evaluate the effectiveness of the intervention with measures developed to address three outcome constructs: Team Behaviors, Attitudes and Opinions, and ED Performance.

## METHODS

### *Study Design*

A prospective investigation using a quasi-experimental, untreated control group design with one pretest and two posttest measurements (Cook and Campbell 1979) was conducted from May 1998 to March 1999. Sixteen potential EDs were contacted by the authors for possible inclusion in the experiment. To participate, EDs had to agree to contract requirements to minimize changes to their physical facilities, staffing levels, and administrators for the yearlong study. Among the EDs contacted, nine agreed to participate and self-selected into the experimental or the control groups. Six EDs were in the experimental group and three EDs were in the control group. Two of the control group EDs needed to complete administrative actions that conflicted with the intervention, and the third enrolled late and required start-up time to participate. Data collection periods were 31 days in duration and occurred in May 1998 (Period 1–Pretest), October 1998 (Period 2–Posttest 1), and March 1999 (Period 3–Posttest 2). The intervention occurred between Period 1 and Period 2 for the experimental group. Control EDs delayed training until after Period 2, which precluded evaluation of the effect of the intervention. The measurements taken in the experimental and control groups at Period 1 and Period 2 were used to assess the effect of the intervention. The additional measurement in the experimental group at Period 3 addressed the effect of the intervention over time (program sustainment). Data collection was scheduled at four-month intervals to allow teamwork implementation activities to take effect.

The study participants were hospital staff, which included ED caregivers (physicians, nurses, and technicians) and admitting unit nurses, and patients receiving emergency care. Institutional Review Board approval or exemption and waiver of written consent were obtained at each of the participating hospitals. Staff and patients were informed of their right to decline to complete surveys. Hospital staff and patients were not identified on data collection instruments.

### *Intervention*

The teamwork training curriculum was developed by an expert panel, a working group of designated physician–nurse pairs from each of the participating EDs, several nationally known consultants and advisors, and behavioral scientists. The panel met on five occasions from 1996 to 1998 to design the ETCC, establish and test evaluation measures, and determine the study design.

The study intervention was the ETCC and the ED teamwork reorganization that followed. The ETCC curriculum is summarized in appendix Table A1 and is described elsewhere (Risser, Rice et al. 1999; Risser, Simon et al. 1999). The instructors were the physician–nurse pairs at the respective EDs who were participating in the expert panels. Emergency department staff at experimental hospitals completed the ETCC between Period 1 and Period 2. The training was conducted with mixed groups of physicians, nurses, and technicians. Following initial training, each ED created a team-based staffing pattern comprised of physician–nurse–technician teams.

The ETCC classroom instruction and workplace practicum involves 48 concrete teamwork behaviors. This behavioral orientation focuses on the processes of teamwork, the specific coordinating actions that caregivers must take with one another to work as an effective team. This is different from conventional health care team implementations that focus on caregiver roles, organizational structure, and care delivery functions (Lowe and Herranen 1982; Manion, Lorimer, and Leander 1996; van Weel 1994). The ETCC supports taskwork—the clinical tasks involved in emergency care delivery—with team behaviors within a reengineered organizational framework that facilitates patient care.

### *Outcome Measures and Scales*

From a suite of 14 survey instruments containing 17 measures hypothesized to quantify the effect of the intervention, 7 measures were selected a priori and reported here to represent 3 primary outcome constructs (Team Behavior, ED Performance, and Attitudes and Opinions). The 7 measures include at least 2 representative measures from each of the 3 constructs. Specifically, team dimension ratings and subjective workload measures represent the team behavior construct; observed errors and admission evaluation measures represent the ED performance construct; and staff attitudes toward teamwork, staff perceptions of support, and patient satisfaction measures represent the attitudes and opinion construct. The results from the Course Critique are

reported only for descriptive purposes, as they do not assess the effect of the intervention.

The first measure representing the team behavior construct was the qualitative assessment of team behaviors. This assessment was completed by the physician and nurse instructors (i.e., expert panel members) trained in using the Team Dimensions Rating Form that consisted of behaviorally anchored rating scales. The behaviorally anchored rating scales were validated in previous military aviation research using instructors within a unit to rate flight crews assigned to that unit (Leedom and Simon 1995). The physician and nurse rater training consisted of an intensive one-day multimedia workshop and open discussion to reach project-wide consensus and standards. Written scoring procedures and detailed, qualitative descriptions of teamwork behavior at the highest, midpoint, and lowest end of the scoring scale provided the criterion-referenced scoring criteria for the ratings. During data collection periods, the raters were instructed to conduct 50 team observations, 20 each of urgent and emergent cases selected at random (i.e., case-based observations) and 10 global observations of ED-wide teamwork across randomly selected shifts. Assessments were taken independently and were uniformly distributed over the 31 days of a data collection period. Each urgent and emergent case observation lasted 30 minutes and the global observation lasted one hour. For each team observed, the five teamwork dimensions (e.g., apply problem solving strategies) were rated separately. Each dimension was scored on a seven-point response scale (very poor to superior), and the ratings in the five dimensions were combined to a single teamwork behavior rating score.

The physician and nurse who performed the teamwork ratings in their respective EDs were observed for one day by an outside rater who traveled to each hospital to confirm rating calibration based on project standards and to resolve any rating discrepancies. Teamwork raters also completed two calibration exercises. Departing from their standard practice of individual observations of global and case-based teamwork, the physician and nurse raters jointly observed but separately rated a team. Two such joint observations were conducted during each period. The interrater reliability of the behaviorally anchored ratings for these teams was calculated using a method described by Winer (1971) and incorporating procedures from Clauser, Clyman, and Swanson (1999).

The second measure of the team behavior construct was the NASA Task Load Index (Hart and Staveland 1988), a measure of individual subjective workload experience. Subjective workload ratings were collected from each

team member caring for the 20 urgent and 20 emergent cases observed for teamwork ratings. Each ED chose types of presenting cases that occurred with predictable high frequency (e.g., heart attacks, abdominal pain, poisoning) that it used for teamwork ratings and subjective workload assessments during the data collection periods. The EDs controlled the type and number of cases sampled, but they did not control the number of staff members caring for the patients. At least two, and up to as many as six, ED caregivers were expected to provide workload ratings depending on the type and complexity of the case. Each of the six items comprising the subjective workload index (i.e., six psychological components of performing work such as mental demand and effort), were measured on a 21-point response scale (very low to very high). Responses on these six items were combined into a single subjective workload score.

Two measures, observed errors and admission evaluation, represented the ED performance construct. While conducting case-based and global single-rater observations for teamwork ratings, the physician or nurse also recorded any witnessed clinical errors (i.e., observed errors), defined as any clinical task that actually or potentially put a patient at risk. The physician or nurse observer completed a form that provided a narrative description of the observed error, a listing of the team behaviors that might have eliminated or reduced the error, and a description of the actions the observer took if he or she felt it necessary to intervene. A blind review of the observed error written reports were independently done by a physician and nurse to verify that reported events met the project definition of a clinical error. Interrater agreement on independent reviews of observed errors was measured using kappa statistics, which assess the percent agreement beyond that expected by chance. All discrepant cases were resolved by a third pair of reviewers and this dataset was used in the analysis. The Admission Evaluation Survey queried admitting unit nurses about preparation of patients from the ED for admission to their hospital unit. Surveys were completed for all inpatient admissions originating in the ED during the data collection periods. This newly developed, single item measured how well each patient was prepared for admission, scored on a 10-point response scale (poor to excellent).

Three measures represented the attitude and opinion construct. The ED Staff Attitude and Opinion Survey measured staff attitudes toward teamwork concepts (e.g., assigning roles and responsibilities in clinical situations) and ED staff perception of support from senior managers and peers to incorporate teamwork principles into clinical tasks. The measure of staff attitudes toward teamwork was based on 15 items, each measured on a seven-point response

scale (unlikely to likely). Responses to the 15 items were combined into a single measure. The staff perception of support measure addressed support relating to managers and peers and combined three items, each based on a seven-point response scale (unlikely to likely). All staff were asked to complete the Staff Attitude and Opinion Survey. The Patient Satisfaction Survey queried patients discharged from the ED to home about satisfaction with their ED visit. Patient Satisfaction Surveys were independently and randomly administered across shifts and days of the week. Each ED was asked to survey at least 160 patients using its existing patient satisfaction survey protocol during each data collection period. The 12 items (e.g., my caregivers knew what other caregivers had done for me; my caregivers took the time to explain things to me), each measured on a seven-point response scale (strongly disagree to strongly agree), were combined into a single patient satisfaction score.

The Staff Attitude and Opinion Survey and the Patient Satisfaction Survey were designed specifically for this study. The items in each survey were constructed so that the respondents (i.e., ED caregivers or patients) evaluated specific teamwork behaviors. Emergency department staff members rated whether they had positive attitudes toward these teamwork behaviors and they also judged perceived support by their peers or supervisors for performing these behaviors. Attitudes were assessed using techniques originally developed by Ajzen and Fishbein (1980) and applied to the evaluation of aviation CRM training (Morey, Grubb, and Simon 1997; Grubb, Morey, and Simon 1999). Patients rated whether these teamwork behaviors were evident during their care. For example, patients responded to the item: "My caregivers knew what other caregivers had done for me." This item, addressing the central concept of team communication taught in the course, reflected the expectation that effective communication should be perceived by patients as they experienced care provided by the team. The patient satisfaction survey also presented two items about overall satisfaction with care and the patient's willingness to recommend that particular ED to others.

### *Statistical Analyses*

Data were collected at each of the nine participating hospitals from physician and nurse observers, staff members, and patients. The unit of measurement is the respondent or case. The seven outcome measures were created and evaluated by combining data collected at all nine hospitals and over all data collection periods (i.e., Periods 1, 2, and 3). The assumption was that the



underlying psychometric properties of the measures did not vary across hospitals and periods. Using standard Likert scaling techniques (DeVellis 1991), multi-item scales were created for each of the measures, except for the observed error measure, which was dichotomous (i.e., whether or not at least one error occurred). For the six Likert-based outcome measures, psychometric properties were evaluated using Cronbach's alpha (internal consistency reliability) and factor analyses (construct validity). Once the items for a given scale were finalized, the items within a scale were aggregated and the computed score was transformed to a 0 to 100 scale (Ware 1993). For all measures, except subjective workload and observed errors, a score of 100 is the desirable response. The range of responses for subjective workload is 0 to 100, where 0 indicates no workload.

Once the seven case level outcomes were created, the data were summarized at the case or respondent level for each hospital and for each period considered separately. Prior to aggregating the error outcome measure at the ED level, the length of observation was taken into account (i.e., the numerator was whether or not at least one error had occurred and the denominator was the length of observation). The difference scores between Period 1 and Period 2 for each hospital and between Period 2 and Period 3 for the experimental hospitals only were also summarized using descriptive statistics (e.g., number of observations, means, standard deviations). In order to assess the effect of the intervention between Period 1 and Period 2 on each of the seven outcome measures, generalized estimating equations (GEE) were used to account for the correlation between the case-level data within each hospital as well as the repeated nature of the data (i.e., Period 1 and Period 2). In these analyses, the observed error outcome was modeled as binomial (and adjusted for length of observation) and the remaining six outcome variables were modeled as normal. We used the model-based standard errors (as opposed to the empirically based standard errors due to the small number of hospitals participating) and assumed an exchangeable correlation structure (i.e., similar pair-wise correlations between case-level data, which ensures asymptotically valid standard error estimates even when the correlations are not truly exchangeable). Zeger and Liang (1986) have shown that with a large number of clustering units or hospitals in these analyses, sound inferences can be made about specific effects. While this case-level analysis is considered more powerful, with relatively few hospitals we were concerned about the validity of these estimates.

Thus, respondent or case-level information was aggregated within each hospital to create hospital-level measures. A two independent samples t-test of

the difference scores was performed to test whether there was significantly more improvement among the experimental hospitals as compared to control hospitals between Period 1 and Period 2 for each of the seven outcome measures. In these analyses, the hospital ( $n = 9$ ) was the unit of analysis. Paired t-tests were used to test for significant improvement in mean scores between Period 1 and Period 2 within the experimental and control groups. In order to determine whether the effect of the intervention was sustained in the experimental group, paired t-tests were performed to test whether there was a significant difference in mean scores between Period 2 and Period 3 for each of the seven outcome measures.

Hospital level characteristics (e.g., number of ED staff) between the experimental ( $n = 6$ ) and control ( $n = 3$ ) group hospitals were compared using a chi-square analysis for nominal data and two independent sample t-tests for continuous data. All tests of significance were determined based on a two-sided  $\alpha = 0.05$ . Nonparametric analyses were not performed because the analysis of variance has been shown to be robust in the presence of nonnormality (Sullivan and D'Agostino 1996; Heeren and D'Agostino 1987). No adjustments were made for multiple testing. All analyses were performed using SAS statistical software (SAS Institute 1996).

## RESULTS

All clinical staff (684 physicians, nurses, and technicians) in the six experimental group EDs received training between Period 1 and Period 2 and all clinical staff (374 physicians, nurses, and technicians) in the three control group EDs did not. Hospital and ED characteristics are shown in Table 1. Emergency departments were approximately equally divided between military and civilian sites, and were predominantly teaching institutions. There were no significant differences between the control and experimental hospitals with respect to hospital type, annual ED patient visits, number of ED staff, and ED staff/visits ratio (per one thousand visits). Thus, even though hospitals were not randomized, the characteristics between the experimental and control groups were comparable. No hospitals reported changes to their physical facilities, staffing levels, or administrators during the experiment.

Demographic data on patients were obtained from the random sample of patients who completed the Patient Satisfaction Survey and descriptive statistics were summarized at the case or patient level. The mean age of ED

Table 1: Hospital and Emergency Department (ED) Characteristics

<i>Hospital</i>	<i>Group</i>	<i>Status</i>	<i>Hospital Type</i>	<i>Annual ED Patient Visits</i>	<i>Number of ED Staff</i>	<i>ED Staff/Visits Ratio (per 1,000 visits)</i>
A	Exp	Mil	Comm	22,850	50	2.2
B	Con	Civ	Comm	33,343	51	1.5
C	Exp	Civ	Teach	40,667	95	2.3
D	Exp	Mil	Teach	40,173	103	2.6
E	Exp	Civ	Teach	33,337	104	3.1
F	Con	Mil	Teach	47,437	108	2.2
G	Exp	Civ	Teach	86,229	199	2.3
H	Con	Civ	Teach	82,901	215	2.6
I	Exp	Mil	Teach	57,873	133	2.3

**Exp**=Experimental; **Con**=Control; **Mil**=Military; **Civ**=Civilian; **Comm**=Community; **Teach**=Teaching.

patients in the experimental hospitals was 38.9 and in the control hospitals was 41.9 in Period 1. Fifty-four versus 55 percent of the ED patients were female in the experimental and control hospitals, respectively, in Period 1. Emergency department patients were asked to rate their health over the past year (range 0–100, higher scores indicating better health) and the mean scores for Period 1 were 61.6 and 65.2 for the experimental and control hospitals, respectively. The control and experimental group patients who participated in the study were similar in both Period 1 and Period 2 (data not shown).

Table 2 provides a description of each of the seven outcome measures. Also shown are the descriptive statistics performed at the case level for all hospitals and periods combined. The number of respondents varied for each measure and for each hospital. Because five of the outcome measures were multi-item scales (except for observed error and admission evaluation), internal consistency reliability of items within an outcome measure was assessed utilizing Cronbach's alpha. High internal consistency reliabilities were observed for each outcome measure because Cronbach's alpha was well above the standard threshold of 0.80 (range: 0.81 to 0.97). Separate analyses (i.e., factor analyses) for items comprising each measure yielded one construct for each multi-item measure based upon multiple criteria (i.e., 100 percent variation explained, scree, and  $\lambda \geq 0$ ). Missing data was minimal, amounting to 8.1 percent or less for each of the outcome measures.

The interrater reliability of the team dimension ratings assessed during calibration observations was determined to be in the moderate range, from .61

Table 2: Study Measures' Descriptive Statistics: Respondent-Level Analysis

<i>Measure</i>	<i>Description</i>	<i>Number of Scale Items</i>	<i>Cronbach's Alpha</i>	<i>% Missing Responses</i>	<i>Mean or % (std)*</i>
<b>Team Behavior</b>					
Team dimension ratings	Higher score indicates more proficient teamwork as defined by the five team dimensions of the Behaviorally Anchored Rating Scales (BARS).	5	0.94	0.4%	45.5 (22.1)
Subjective workload	Lower score indicates lower level of perceived workload.	6	0.81	0.4%	33.7 (18.5)
<b>ED Performance</b>					
Observed errors	Higher score indicates more clinical errors defined as the ratio of the number of at least one witnessed error (global, urgent, emergent) to the length of observation time per case.	1	na	0.0%	11.9% (45.9%)
Admission evaluation	Higher score indicates better overall ED preparation of patients for admission into a hospital unit.	1	na	6.3%	80.2 (19.6)

<b>Attitude and Opinion</b>	Staff attitudes toward teamwork	Higher score indicates better attitude toward teamwork principles (e.g., assigning roles and responsibilities in clinical situations, communicating with staff both globally and for an individual patient, balancing caseload and cross-monitoring actions of coworkers).	15	0.95	2.3%	76.1 (17.4)
Staff perceptions of support	Higher score indicates perceived support of senior managers, ED administrative team, and peers with respect to performing teamwork behaviors.	3	0.85	1.7%	78.4 (21.2)	
Patient satisfaction	Higher score indicates more satisfaction with care received in the ED and willingness to recommend the ED to others.	12	0.97	8.1%	91.3 (16.6)	

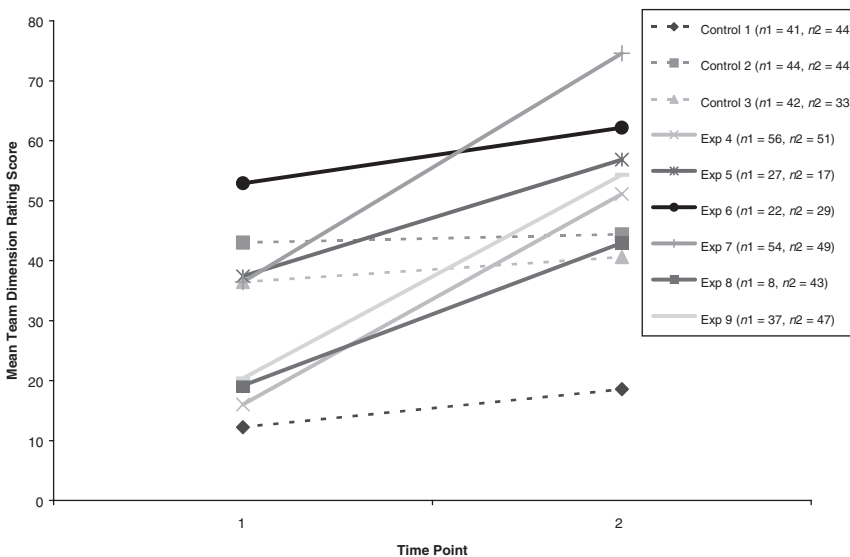
\* Measures are for all hospitals and periods combined. Range of values is 0–100.

to .81 across the five team dimensions. Blinded raters who evaluated the reporting of observed errors agreed with the original observers' judgments in 91.1 percent of the cases, resulting in a kappa statistic of 0.69 ( $p < 0.0001$ ).

Figure 1 provides descriptive information at the hospital level for one of the seven outcome measures, the Team Dimension Ratings for Period 1 and Period 2 (selected for example only). Each ED was asked to obtain 50 observations within each period for the Team Dimension Rating measure. Although the number of cases for each hospital varies from 8 to 56 for Period 1 and from 17 to 51 for Period 2, the standard errors are roughly equivalent among hospitals (0.91 to 2.38 for Period 1 and 1.09 to 3.44 for Period 2). Data not presented here (but posted as an appendix on <http://www.hsr.org>) show generally similar results for the other six outcome measures with roughly equivalent standard errors.

Generalized estimating equations (GEE) were used to test the effect of the intervention between the control and experimental hospitals using case-level data. Because the GEE results were similar to those of the two independent sample t-tests of differences scores between Period 1 and Period 2, we present the results of the simpler analysis. Thus, these results are based on the

Figure 1: Team Dimension Ratings: Hospital Level Mean Scores by Time Point



hospital as the unit of analysis ( $n = 9$ ). Specifically, the mean of each outcome measure represents a single data point. The two independent sample t-test approach does not account for the differences in standard errors within each hospital for a given outcome measure, but because the results from the generalized estimating equations and the two independent sample t-test of difference scores yielded equivalent results, it suggests that the standard errors are roughly equivalent for a given measure. Results of the analyses to investigate the effect of the intervention are shown in Tables 3 and 4.

Teamwork significantly improved in the experimental group between Period 1 and Period 2 as compared with the control group ( $p = 0.012$ , Table 3). The overall quality of teamwork improved in the experimental group as shown by the significant increase in the mean team dimension ratings from 30.4 in Period 1 to 57.0 in Period 2 ( $p = 0.002$ ). The control group did not show significant improvement with a mean of 30.6 in Period 1 and 34.5 in Period 2. As shown in Table 4, the effect of the intervention in the quality of teamwork measure was not different in the experimental group when the Period 2 mean of 57.0 was compared to the Period 3 mean of 58.3 ( $p = 0.710$ ). Caution should be used when interpreting the significance levels in Table 4 because the study was not designed to test equivalence of means in the experimental group from Period 2 to Period 3.

There was no significant difference in the mean subjective workload ratings between the experimental group and the control group for Period 1 and 2 ( $p = 0.668$ , Table 3). Workload was not significantly different from Period 2 to Period 3 in the experimental group ( $p = 0.081$ , Table 4).

There was no significant difference in the observed error rate from Period 1 to Period 2 between the experimental and control groups ( $p = 0.140$ ). The mean observed error rate was 30.9 in Period 1 and 4.4 in Period 2 for the experimental group and 16.8 in Period 1 and 12.1 in Period 2 for the control group. The observed clinical error rate was significantly reduced in the experimental group ( $p = 0.039$ ), but not in the control group ( $p = 0.591$ ), between Period 1 and Period 2. Content analysis of the error reports and the types of teamwork errors associated with the errors did not reveal reasons for the initial (Period 1) differences in the rate of errors between the experimental and control groups. The difference in observed error rates between Period 2 and Period 3 did not reach statistical significance in the experimental group (4.4 percent for Period 2 and 2.8 percent for Period 3,  $p = 0.720$ ). Note that one hospital was eliminated from this analysis because it did not provide pretest data, precluding the use of the intent to treat methodology. Examples of events reported as observed errors are shown in Table 5.

Table 3: Effect of The Teamwork Training Intervention

Measure	Group	n	Period 1		Period 2		Between Group Improvement (Exp vs. Con)		Within Group Improvement	
			Mean (std)*	Mean (std)*	Mean Difference <sup>§</sup>	p	Mean Difference	p		
<b>Team Behavior</b>										
Team dimension ratings										
	Exp	6	30.4 (14.3)	57.0 (10.7)	22.7	.012	26.6	.002		
	Con	3	30.6 (16.2)	34.5 (14.0)			4.0	.109		
Subjective workload										
	Exp	6	33.7 (10.6)	38.2 (8.2)	2.2	.668	4.4	.232		
	Con	3	29.7 (5.4)	32.0 (4.6)			2.2	.303		
<b>ED Performance</b>										
Observed errors										
	Exp	5	30.9% (22.5%)	4.4% (4.9%)	-21.8%	.140	-26.5%	.039		
	Con	3	16.8% (9.8%)	12.1% (11.4%)			-4.6%	.591		
Admission evaluation										
	Exp	6	80.5 (4.8)	83.4 (2.2)	2.7	.259	2.9	.067		
	Con	3	75.1 (3.9)	75.3 (7.2)			0.2	.935		



Attitude and Opinion								
Staff attitudes toward teamwork	Exp	6	75.0 (1.8)	78.5 (1.7)	4.6	.065	3.5	.047
	Con	3	77.6 (4.0)	76.5 (2.0)			-1.1	.477
Staff perceptions of support	Exp	6	76.2 (3.4)	83.1 (4.4)	4.2	.323	6.9	.040
	Con	3	78.6 (3.6)	81.3 (6.4)			2.8	.315
Patient satisfaction	Exp	6	90.7 (2.1)	92.6 (3.9)	2.8	.109	1.9	.243
	Con	3	90.0 (2.9)	89.1 (2.5)			-0.9	.066

\* Range of measures 0–100.

§ Between group improvement, two independent sample t-test of different scores.

Two-sided test. Bold values indicate  $p < 0.05$ .

Two-sided paired t-tests for group improvement in measure from Period 1 to Period 2. Bold values indicate  $p < 0.05$ .

Table 4: Effect of the Intervention over Time—Experimental Group Only

Measure	n	Period 2 vs. Period 3			Mean Difference	p*
		Period 1 Mean (std)	Period 2 Mean (std)	Period 3 Mean (std)		
<b>Team Behavior</b>						
Team dimension ratings	6	30.4 (14.3)	57.0 (10.7)	58.3 (14.5)	1.4	.710
Subjective workload	6	33.7 (10.6)	38.2 (8.2)	35.1 (10.1)	-3.1	.081
<b>ED Performance</b>						
Observed errors	5	30.9% (22.5%)	4.4% (4.9%)	2.8% (6.2%)	-1.6%	.720
Admission evaluation	6	80.5 (4.8)	83.4 (2.2)	85.7 (3.6)	2.4	.157
<b>Attitude and Opinion</b>						
Staff attitudes toward teamwork	6	75.0 (1.8)	78.5 (1.7)	73.2 (8.0)	-5.2	.200
Staff perceptions of support	6	76.2 (3.4)	83.1 (4.4)	76.6 (5.6)	-6.5	.131
Patient satisfaction	6	90.7 (2.1)	92.6 (3.9)	91.9 (3.4)	-0.7	.565

\*Two-sided value for paired t-test.

Table 5: Examples of Observed Errors

Observed Error Synopsis
<ul style="list-style-type: none"> <li>• An intern discusses a case with the attending physician, but fails to mention that the patient has a blood pressure of 149/106. None of the nurses or technicians in attendance mentions this to the attending, nor do they recheck the blood pressure.</li> <li>• The patient had been exposed to a toxic agent. The nurse with the patient did not inform the physician that the patient was designated a “contact precaution.” The physician did not check the chart, and examined the patient without gloves.</li> <li>• Nurse did not include ipratropium in a nebulizer treatment as ordered.</li> <li>• A trauma patient, brought in by ambulance, was receiving oxygen by mask. The mask was left in place but not hooked up to the oxygen flow meter in the room.</li> <li>• A patient with chest pain was placed in a bed by a technician who did an EKG and placed the strip in the patient’s chart. The patient was then unobserved for 25 minutes. The nurse and physician were unaware the patient was present in the bed.</li> </ul>

There was no significant improvement in the preparation of ED patients for admission to the hospital in the experimental group between Period 1 and Period 2 as compared to the control group ( $p = 0.259$ ). Further, the differences

in means between Period 2 and Period 3 did not reach statistical significance in the experimental group ( $p = 0.157$ ).

Staff attitudes toward teamwork did not significantly improve in the experimental group between Period 1 and Period 2 as compared to the control group ( $p = 0.065$ ), and increased significantly in the experimental group (75.0 in Period 1 to 78.5 in Period 2,  $p = 0.047$ ). No significant difference was detected in the experimental group's mean ratings from Period 2 to Period 3 ( $p = 0.200$ ).

The perception of ED staff members that management and peers support their applying teamwork principles improved in the experimental group ( $p = 0.040$ ) but not the control group ( $p = .315$ ). However, the test of difference scores between the control and intervention group for Period 1 and Period 2 did not reach statistical significance in the experimental group as compared to the control group ( $p = .323$ ). The level of support was not different in the experimental group from Period 2 to Period 3 ( $p = .131$ ).

No significant differences in patient satisfaction were obtained in the experimental group as compared to the control group ( $p = .109$ ) or for the experimental group for Periods 1 and 2 ( $p = .243$ ). Likewise, no significant differences in patient satisfaction ratings were detected in the experimental group from Period 2 to Period 3 ( $p = .565$ ).

The course critique completed at the close of the classroom training asked respondents to rate the overall value of the ETCC on a five-point scale, rescaled to 0–100 where 100 indicated the course was very useful. The mean rating for the course was 87.7 (standard deviation = 17.9;  $Q_1 = 75$ ,  $Q_2 = 100$ ) indicating that the course was rated as very useful among the 591 respondents in the experimental group.

## DISCUSSION

While the notion of teamwork in health care is familiar, the concept is vague and generally limited to promoting congenial working relationships among coworkers. Although rigorously trained in the individual execution of clinical tasks, physicians and nurses have little training to prepare them for the more tightly defined teamwork behaviors typical of aviation-based crew resource management (CRM). Yet, improved teamwork among caregivers has been identified as a fundamental principle of error reduction (Leape, Kabcenell, and Berwick 1998; Kohn, Corrigan, and Donaldson 1999). Some efforts have been made to foster team-oriented behaviors, most notably by Gaba and his

associates in promoting the implementation of CRM principles in anesthesia (Howard et al. 1992). However, teamwork is not a natural product of working together, as the disparities in teamwork attitudes among operating room staff and between intensive care physicians and nurses have shown (Helmreich and Schaefer 1994; Sexton, Thomas, and Helmreich 2000).

We sought to advance formal teamwork training by adapting behavioral features of CRM to the operational requirements of EDs and introducing organizational changes to further encourage teamwork behaviors. The results from this evaluation show that the MedTeams intervention (the ETCC and subsequent teamwork implementation) led to significant improvement in staff attitudes toward teamwork. More importantly, the quality of teamwork behaviors observed in the ED improved without the cost of increased caregiver workload, as assessed through caregiver ratings of their subjective workload.

Of particular importance was the finding that the number of observed clinical errors was significantly reduced in teamwork-trained EDs. Errors of a nonclinical nature, such as a failure to process an admission request within a prescribed time period, were not of interest in this context and were not reported. Witnessed clinical task errors that potentially or actually put a patient at risk were recorded. An example of a reported error before training was the situation of two nurses each administering the same dose of morphine IV after a verbal order for morphine IV was given during a burn resuscitation. The staff recognized the overdose when the patient's breathing slowed, at which point they intervened and the patient recovered. A verbal check-back to indicate acceptance of the verbal order, a teamwork behavior taught in the ETCC, may have avoided or "captured" this error.

Our findings indicate that the intervention was effective in each of the three domains (Team Behavior, ED Performance, Attitudes and Opinions). The positive impact of the intervention was in large part maintained over the eight months of posttraining observation. Since the results are positive and consistent across the outcome constructs, the failure to reach statistical significance for some measures is likely due to the relatively small number of hospitals in the study leading to decreased statistical power. Further, the number of comparisons performed (which increases the overall experimental  $\alpha$ ) is mitigated by the fact that the results consistently agreed with the expected improvements in performance.

The training was well received by physicians, nurses, and technicians alike, but implementation in the workplace requires concerted and sustained effort. Staff perceptions of support at first increased and then showed a

downward trend. Our experience is that the integration of effective teamwork skills into emergency care requires ongoing management efforts, which may not be immediately rewarded. Emergency departments are much less standardized in their physical layout and operations than are airliners, and successful team structures varied among the experimental group hospitals depending on staffing patterns and the physical flow of patients through their facilities.

A variety of qualitative findings developed from site visits and project summaries strengthen the quantitative findings and provide testable hypotheses for future research. The importance of leadership at the organizational and operational level became evident in our study. While our intervention focused on leadership within work teams, our study revealed that leadership functions in support of teamwork implementation needed to be performed at various levels of the organization. First, we underestimated the importance of upper level management in supporting a training initiative at the departmental level. A significant determinant of teamwork implementation success is the sustained commitment and active involvement of executive leaders. Likewise, leaders at the department level need to institute a reward system for teamwork successes, provide teamwork role models themselves, and appear in the workplace to observe and encourage staff to engage in teamwork behaviors. At the caregiver team level, each patient needs a designated or emergent leader to initiate and guide the care delivery process.

The vertical integration of leadership support for teamwork practices is supported by coaching and mentoring of teamwork behaviors in the department. Primary and associate instructors need to undertake the role of coaches in the workplace to help staff members identify opportunities to promote teamwork behaviors, critique teamwork performance, and reinforce teamwork processes on a team-by-team basis. While this was not realized early in the program development, coaching emerged as a principal mechanism for continuing the education process and enabling the change to team-based care.

The delivery of the training itself may be accomplished most effectively by teaching all teamwork behaviors during the class as we did in this study, but then “dosing” the introduction of behaviors into the workplace over time. Staff members may feel incapable of implementing all the teamwork behaviors at one time, so the phased introduction of subsets of behavior may facilitate the assimilation of the teamwork behaviors and modifications of work patterns. For example, verbal call-outs and check-backs are simple behaviors readily

incorporated into existing clinical protocols. However, engaging team members in the planning process is a more complex teamwork action that may entail developing new methods of information exchange. This teamwork action could be postponed until these methods are ready to be introduced.

Teamwork is promoted by visible changes in the work environment. Teams need to be physically identified by colored scrubs, armbands, or identification tags. This serves not only to assist staff in identifying their own team members, but also to benefit patients by knowing which team is responsible for their care. The physical layout of the department can also be enlisted to create teamwork system supports. Whiteboards with essential patient information emerged as a central information exchange medium, and created a focal point for periodic team situation updates and task prioritization discussions. Some departments reconfigured workspaces to eliminate barriers separating nursing and medical staff, thus promoting exchanges of information.

These physical changes, and the shift to a teamwork culture, became issues of staff resistance in the early phase of the teamwork implementation process. While many of the changes brought by a teamwork structure were seen by staff members as valuable, resistance to these changes nevertheless emerged. Examples of points of resistance were the wearing of team identifiers and the designation of physicians as the team leaders. While the creation of designated teams is a prescribed feature of teamwork, the exact means of team identification can be tailored by individual organizations. Likewise, while the leadership function is central to teamwork, a workable leadership solution at the caregiver level may take a variety of forms with both physicians and nurses assuming leadership roles depending on clinical, operational, and situational demands. In particular, our prescription of placing the physician in the leadership role was not always effective for the management of operational issues, since some physicians were not inclined to directly manage ED operations. Moreover, some physicians felt they did not have sufficient leadership training to manage both clinical cases and ED operations. We found that who performed specific leadership functions became less important than the requirement that clinical and operational management information be exchanged among physician and nurse leaders. Thus, it became apparent that a shift to a teamwork system needed to have both prescriptive and flexible features.

As is the case in aviation, which requires periodic CRM retraining and recertification, refresher training and ongoing efforts to incorporate teamwork into daily operations will be needed for these behaviors to become a

permanent part of the ED culture. Management initiatives that introduce teamwork considerations into team reviews, process improvements, morbidity and mortality conferences, and employee evaluations and promotions are essential to teamwork implementation.

This study has some limitations. The teamwork implementation entailed obvious cues such as color-coded teams, structured whiteboard rounds, and program-specific terminology. As a result, blinded ratings of teams in the experimental group were not possible. Therefore, the instructors were trained in criterion-referenced, behaviorally anchored rating techniques for rating teamwork. A trial of videotape review by observers blinded to group assignment proved unsatisfactory because important behaviors occurring outside camera or microphone range were missed. However, the teamwork raters were initially calibrated and subsequently completed joint observations for recalibration during the study to avoid drift. In addition, given the 91 percent agreement rate of observed errors that was significantly above chance, we feel that the lack of blinding was unlikely to introduce appreciable bias into the observed error results.

The quasi-experimental design introduced a limitation because of the possibility for alternative explanations for the obtained results (Cook and Campbell 1979). Significant challenges to validity appear to have been minimized in this study because the self-selection into experimental and control groups did not yield significant differences in hospital characteristics, hospitals reported no extraneous organizational changes that invalidated their participation agreements, teamwork ratings were shown to be reliable, and observers were forthcoming in reporting unflattering details of observed clinical errors.

The limitation with respect to statistical power posed by the small number of hospitals was anticipated and a future ETCC validation is planned with a larger cohort. However, small sample sizes are common for team research in operational settings with the number of teams evaluated typically in the range of 5 to 15 (e.g., Salas et al. 1999; Serfaty, Entin, and Johnston 1998), or when team data are aggregated into higher units of analysis as was the case in this study (McIntyre and Salas 1995).

In conclusion, teamwork training based on CRM was successful in increasing specific teamwork behaviors and indicated an effect of reducing clinical errors and enhancing staff attitudes toward teamwork. Although emergency departments are unique environments, it seems reasonable that other high-risk areas of care will benefit from similar training.

## APPENDIX

### ETCC<sup>TM</sup> CURRICULUM AND INTERVENTION

The delivery and implementation of the ETCC was divided into three phases that constitute the MedTeams program: site planning and preparation, ETCC training, and teamwork implementation. Site planning consisted of (a) a communication campaign to introduce staff and higher level management to the ED's impending change to a teamwork structure, (b) a schedule for ETCC training and subsequent teamwork roll-out, (c) determining the ED team structure and a team identification scheme such as colored scrubs, and (d) establishing a means for maintaining situational awareness by displaying patient information and other operational information for team use.

Led by a physician and nurse instructor pair, the ETCC teamwork training consisted of mixed classes of approximately 16 physician, nurses, technicians, and optionally, unit clerks, who completed eight hours of instruction organized into topic areas as shown in Table A1. The ETCC training day also provided for (a) behavioral modeling through videotaped segments of good and poor teamwork, (b) practical exercises to engage students in practicing components of teamwork, such as task prioritization and case review from a teamwork perspective, and (c) analysis and discussion of clinical vignettes conveying features of good and poor teamwork. This train-up phase lasted approximately two months depending of the size of the ED.

The ETCC curriculum is organized into seven components with an introduction, five main learning modules, and an integration unit. Each of the five main learning modules corresponds to one of the five Team Dimensions. The objective of the instruction is to train the process of how a team functions in terms of communication and coordination behaviors. Each of the 48 team behaviors is presented with respect to the situations that give rise to its use, the techniques for expressing the behavior, and the teamwork and operational outcomes expected from that behavior. The training emphasizes that the teamwork process does not unfold as a fixed series of steps, but rather as an adaptive, mutually supportive mix of responses to the demands of the situation.

Learning modules contain lecture and discussion of each of the behaviors included in a team dimension and are complemented by vignettes, descriptions of teamwork failures, and practical exercises that enhance participant understanding. All vignettes and teamwork failures presented are from actual accounts acquired through observation or open- and closed-case



Table A1: ETCC™ Curriculum Summary

<i>Course Component</i>	<i>Contents</i>
Introduction	The ETCC introduction justifies the need for teams and teamwork in the current health care environment. It begins with a discussion of four elements that distinguish work teams from work groups and continues with an overview of teamwork research findings. An explanation of the paradigm shift that occurs when traditional groups move to a team system is presented followed by a discussion of the attitude shifts necessary to build a strong teamwork system. The introduction concludes with a review of the five Team Dimensions, which are introduced as categories of observable behavior used to train and measure team performance.
Module 1. Maintain Team Structure and Climate	This module addresses ED composition and defines the Core Team in the ED context. This module describes the Designated Team Leader, Situational Leader, and Team Member roles. It highlights the characteristics of team climate and includes strategies for managing conflicts that arise within the team. It also provides an opportunity for course instructors to present the locally determined team structure defined for their ED. Examples of behaviors taught are (a) establish the leader, (b) hold each other accountable for team outcomes, and (c) resolve conflicts constructively.
Module 2. Apply Problem-Solving Strategies	This module addresses planning and decision making as central teamwork actions associated with performing clinical tasks. Forms of planning are presented, as are various decision-making models. Instruction includes a discussion of error and the role of the team in supporting the decision maker with information to reduce the occurrence of decision-making errors. Cross-monitoring, assertion, advocacy, and the Two-Challenge Rule are introduced as elements of the Teamwork Monitoring Cycle. A practical exercise is included that demonstrates the relationship between a shared mental model and team communication. Examples of behaviors taught are (a) engage team members in the planning process, (b) reports slips, lapses, and mistakes to the team, and (c) advocate and assert a position or corrective action.
Module 3. Communicate with the Team	This module addresses the quality of information exchange within the team. Situation awareness is presented as a central concept to explain the importance of effective communication. It addresses the practice of offering and requesting information that can improve situation awareness and impact clinical practice and patient outcomes. Standards of effective communication are provided as are a number of techniques for communicating vital information within the team. This module includes a number of practical exercises that raise awareness of the impact of language on shared mental models and reinforce the importance of validating information transfer. Examples of behaviors taught

Table A1: Continued

<i>Course Component</i>	<i>Contents</i>
Module 4. Execute Plans and Manage Workload	<p>are (a) provide situational awareness updates, (b) call out critical informational during emergent events, and (c) communicate decisions and actions to team members.</p> <p>This module addresses the impact of workload on work performance. It describes several strategies for managing workload including delegation, prioritization and secondary triage, resource management, and task assistance. A practical exercise is included that provides an opportunity for participants to practice resource reallocation in a rapidly changing environment. Examples of behaviors taught are (a) prioritize tasks for individual patients, (b) execute the team-established plan, and (c) request assistance for task overload.</p>
Module 5. Improve Team Skills	<p>This module addresses the issue of team skill improvement through coaching, situational learning, and team performance review. Performance goals, performance feedback, and performance improvement strategies are described as key elements of the team improvement process. The <i>Resuscitation Videotape Review</i> is an optional exercise for departments with video capabilities. Examples of behaviors taught are (a) engage in situational learning and teaching with the team, (b) conduct event reviews, and (c) participate in educational forums addressing teamwork.</p>
Integration Unit	<p>Customized by each participating hospital to address issues pertinent to the implementation of the teamwork system into that emergency department.</p>

reviews. Professional-quality video segments that demonstrate teamwork principles are presented in each module.

Once all staff completed the ETCC, the teamwork implementation phase was initiated on an established start date. This phase was characterized by (a) forming teams by shift and delivering care in a team structure, (b) each staff member completing a four-hour practicum in which teamwork behaviors were practiced and critiqued by an instructor, and (c) coaching and mentoring of teamwork behaviors by instructors to all staff during normal shifts. This phase lasted six months.

## REFERENCES

- Ajzen, I., and M. Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

- Bloom, B. S., J. T. Hastings, and G. F. Madaus. 1971. *Handbook of Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Clauser, B. E., S. G. Clyman, and D. B. Swanson. 1999. "Components of Rater Error in a Complex Performance Assessment." *Journal of Educational Measurement* 36 (1): 29–45.
- Cook, T. D., and D. T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- DeVellis, R. F. 1991. *Scale Development: Theory and Applications*. Newbury Park, CA: Sage.
- Gardner-Bonneau, D. J. 1993. "What Is Iatrogenics, and Why Don't Ergonomists Know? An Interview with Dr. Lowell Levin." *Ergonomics in Design* (July): 18–20.
- Grubb, G., J. C. Morey, and R. Simon. 1999. "Applications of the Theory of Reasoned Action Model of Attitude Assessment in the Air Force CRM Program." In *Proceedings of the Tenth International Symposium on Aviation Psychology*, edited by R. S. Jensen and L. A. Rakovan, pp. 298–301. Columbus, OH: Aviation Psychology Laboratory of the Ohio State University.
- Hart, S. G., and L. E. Staveland. 1988. "Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In *Human Mental Workload*, edited by P. S. Hancock and N. Meshkati, pp. 139–83. Amsterdam: North-Holland.
- Heeren, T., and R. D'Agostino. 1987. "Robustness of the Two Independent Samples T-Test When Applied to Ordinal Scaled Data." *Statistics in Medicine* 6 (1): 79–90.
- Helmreich, R. L. 1997. "Managing Human Error in Aviation." *Scientific American*, May, pp. 62–7.
- Helmreich, R. L., and H. C. Foushee. 1993. "Why Crew Resource Management? Empirical and Theoretical Bases of Human Factors Training in Aviation." In *Cockpit Resource Management*, edited by E. L. Weiner and B. G. Kanki, pp. 3–45. San Diego, CA: Academic Press.
- Helmreich, R. L., and H. -G. Schaefer. 1994. "Team Performance in the Operating Room." In *Human Error in Medicine*, edited by M. S. Bogner, pp. 225–53. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Howard, S. K., D. M. Gaba, K. J. Fish, G. Yang, and F. H. Sarnquist. 1992. "Anesthesia Crisis Resource Management: Teaching Anesthesiologists to Handle Critical Incidents." *Aviation, Space, and Environmental Medicine* 63 (9): 763–70.
- Kohn, L. T., J. M. Corrigan, and M. S. Donaldson, eds. 1999. *To Err is Human: Building a Safer Health Care System*. Washington, DC: National Academy Press.
- Leape, L. L., A. Kabacennell, and D. M. Berwick. 1998. *Reducing Adverse Drug Events and Medical Errors*. Boston: Institute for Healthcare Improvement.
- Leedom, D. K., and R. Simon. 1995. "Improving Team Coordination: A Case for Behavioral-Based Training." *Military Psychology* 7 (2): 109–22.
- Lowe, J. I., and M. Herranen. 1982. "Understanding Teamwork: Another Look at Concepts." *Social Work in Health Care* 7 (2): 1–11.
- Manion, J., W. Lorimer, and W. J. Leander. 1996. *Team-Based Health Care Organizations: Blueprint for Success*. Gaithersburg, MD: Aspen Publishers.

- McIntyre, R. M., and E. Salas. 1995. "Measuring and Managing for Team Performance: Emerging Principles from Complex Environments." In *Team Effectiveness and Decision-Making in Organizations*, edited by R. A. Guzzo, pp. 9–45. San Francisco: Jossey-Bass.
- Morey, J. C., G. Grubb, and R. Simon. 1997. "Towards a New Measurement Approach for Cockpit Resource Management Attitudes." In *Proceedings of the Ninth International Symposium on Aviation Psychology*, edited by R. S. Jensen and L. A. Rakovan, pp. 478–83. Columbus, OH: Aviation Psychology Laboratory of The Ohio State University.
- Risser, D. T., M. M. Rice, M. L. Salisbury, R. Simon, G. D. Jay, S. D. Berns, and the MedTeams Consortium. 1999. "The Potential for Improved Teamwork to Reduce Medical Errors in the Emergency Department." *Annals of Emergency Medicine* 34 (3): 373–83.
- Risser, D. T., R. Simon, M. M. Rice, and M. L. Salisbury. 1999. "A Structured Teamwork System to Reduce Clinical Errors." In *Error Reduction in Health Care*, edited by P. L. Spath, pp. 235–78. San Francisco: Jossey-Bass.
- Salas, E. and J. A. Cannon-Bowers. 2001. "The Science of Training: A Decade of Progress." In *Annual Review of Psychology*, edited by S. T. Fiske, D. L. Schacter, and C. Zahn-Waxler. Palo Alto, CA: Annual Reviews.
- Salas, E., T. L. Dickinson, S. A. Converse, and S. I. Tannenbaum. 1992. "Toward an Understanding of Team Performance and Training." In *Teams, Their Training and Performance*, edited by R. W. Swezey, pp. 3–29. Norwood, NJ: Ablex.
- Salas, E., J. E. Fowlkes, R. J. Stout, D. M. Milanovich, and C. Prince. 1999. "Does CRM Training Improve Teamwork Skills in the Cockpit?: Two Evaluation Studies." *Human Factors* 41 (2): 326–43.
- SAS Institute. 1996. *SAS statistical software* (release 6.12). Cary, NC: SAS Institute.
- Serfaty, D., E. E. Entin, and J. H. Johnston. 1998. "Team Coordination Training." In *Making Decisions under Stress: Implications for Individual and Team Training*, edited by J. A. Cannon-Bowers and E. Salas, pp. 221–45. Washington, DC: American Psychological Association.
- Sexton, J. B., E. J. Thomas, and R. L. Helmreich. 2000. "Error, Stress, and Teamwork in Medicine and Aviation: Cross Sectional Surveys." *British Medical Journal* 320: 745–9.
- Simon, R., J. C. Morey, and A. Locke. 1997. *Full Scale Development of the Emergency Team Coordination Course and Evaluation Measures*. Andover, MA: Dynamics Research Corporation.
- Sullivan, L. M., and R. B. D'Agostino. 1996. "Robustness and Power of Analysis of Covariance Applied to Data Distorted from Normality by Floor Effects: Homogeneous Regression Slopes." *Statistics in Medicine* 15 (5): 477–96.
- Taggart, W. R. 1994. "Crew Resource Management: Achieving Enhanced Flight Operations." In *Aviation Psychology in Practice*, edited by N. Johnston, N. McDonald, and R. Fuller, pp. 309–39. Brookfield, VT: Ashgate.
- van Weel, C. 1994. "Teamwork." *Lancet* 334: 1276–9.
- Ware, J. E. 1993. *SF-36 Health Survey: Manual & Interpretation Guide*. Boston: The Health Institute, New England Medical Center.

- Weiner, E. L., B. G. Kanki, and R. L. Helmreich, eds. 1993. *Cockpit Resource Management*. San Diego, CA: Academic Press.
- Winer, B. J. 1971. *Statistical Principles in Experimental Design*, 2d ed., pp. 283–9. New York: McGraw-Hill.
- Zeger, S. L., and K. Y. Liang. 1986. “Longitudinal Data Analysis for Discrete and Continuous Outcomes.” *Biometrics* 42 (1): 121–30.