THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, *13*(2), 153–171

Using Schematic Aids to Improve Recall in Incident Reporting: The Critical Event Reporting Tool (CERT)

Douglas A. Wiegmann
Institute of Aviation
University of Illinois at Urbana–Champaign

Terry L. von Thaden
Graduate School of Library and Information Science
University of Illinois at Urbana–Champaign

Incident reports are intended to facilitate problem identification and aid in allaying accidents, thus improving safety. Unfortunately, most incident reports are generally unstructured, providing little or no guidance to the reporter. Therefore, most reports contain information only about what happened, as opposed to why an incident happened, making identification of intervention and prevention strategies extremely difficult. This study seeks to help remedy this problem with incident reports by developing and testing a method for improving pilots' reporting of incidents. This method, coined the critical event reporting tool (CERT), is described, and research supporting its potential as an instrument for improving the quality of incident reports is provided.

As noted by Heinrich (1959), incidents, or deviations from safe operations, are precursors to accidents, consisting in orders of magnitude significantly higher than accidents. Incidents then indicate the presence of problems in systems that, if left unresolved, have the potential to result in an accident (Heinrich). Incident reporting, therefore, has long been utilized in the aviation realm as a learning tool for proactively analyzing and treating unsafe conditions and actions before they become accidents (Connell, 1999; Fitts & Jones, 1947). One widely known incident-reporting system in aviation is the National Aeronautics and Space Administration's (NASA)

Requests for reprints should be sent to Douglas A. Wiegmann, Aviation Research Lab, University of Illinois, #1 Airport Road, Savoy, IL 61874. E-mail: dwiegman@uiuc.edu

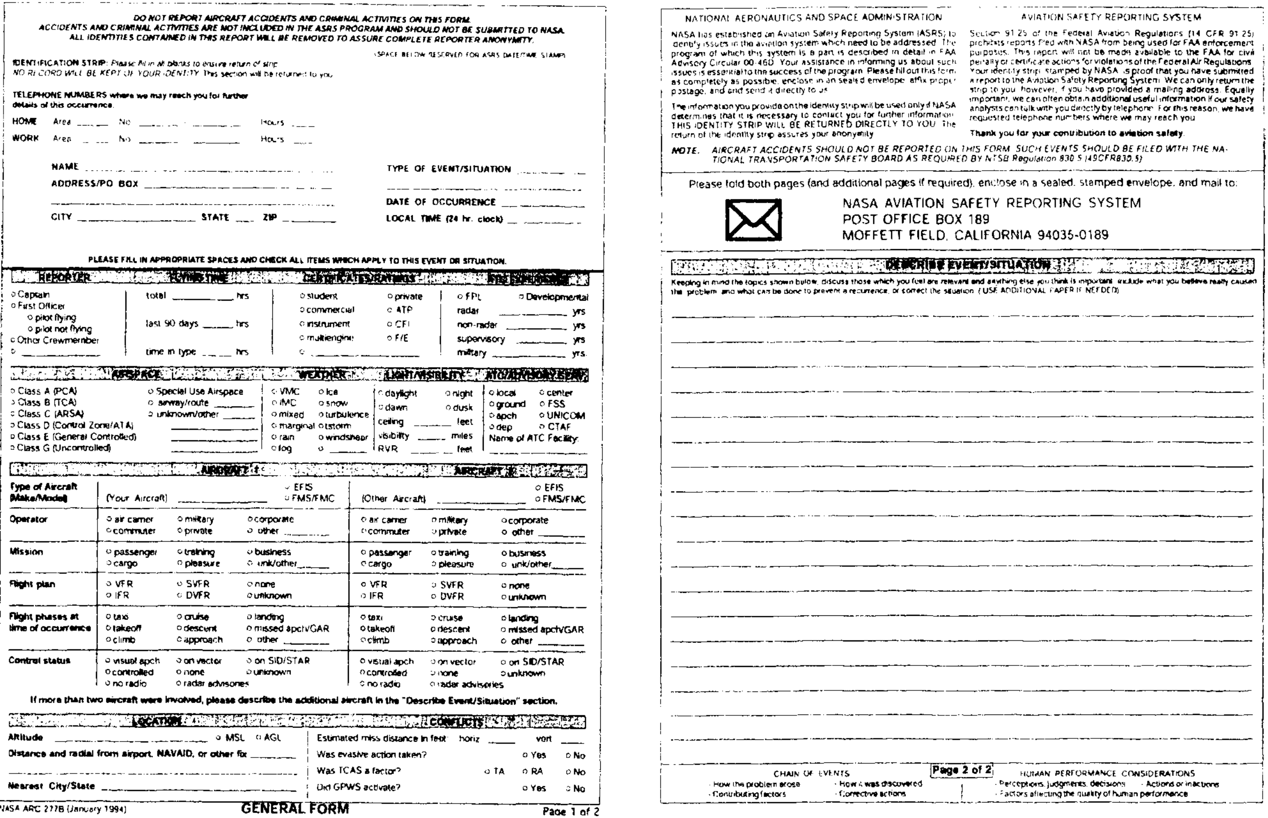
Aviation Safety Reporting System (ASRS) (Chappell, 1997). Connell (1999) noted that such anonymous incident reporting systems are crucial to accident prevention because incident reports often provide information about safety hazards that is frequently lost or not obtained from accident investigations due to crew member fatality.

Unfortunately, however, most incident reporting systems do not always gather rich enough event data to fully explain why unsafe incidents occur. Indeed, most aviation incident-reporting forms, including the one used in the ASRS, collect a plethora of factual information surrounding an incident, such as the type of aircraft, time of day, and weather conditions. The narrative portion of most of these incident-reporting forms is often a free format essay that provides little or no guidance to the reporter on how to describe the critical events of the incident (see Figure 1 for an example of the ASRS form). As a result, most reports contain only information on *what* happened, as opposed to *why* an incident happened. In addition, these reports seldom contain information concerning the circumstances or nature of the incident that prevented it from becoming an accident. Consequently, the identification of intervention strategies based on incident report information is often onerous.

FACTORS AFFECTING INFORMATION RECALL

As chronicled by Ericsson and Simon (1980), the process of recalling an event is limited by the capacity of short-term memory (STM), where only the most recently attended-to information is directly reachable. A portion of the contents of STM, however, is often instantiated in long-term memory (LTM) before it is lost from STM. It is this portion that can, at a later date, be retrieved from LTM. What is recalled and how well it is remembered depends on a variety of factors, including the period between when the information is acquired and when it is recalled. In general, the longer the delay between knowledge encoding and retrieval, the poorer recall tends to be (Loftus, Greene, & Doyle, 1989). According to Tversky and Kahneman (1973), who investigated the transitional process from event to retrospective description, events that are readily recalled are judged to be representative and frequent, yet lead to large errors in estimation. Furthermore, people often fall prey to the fundamental attribution error, in which they attribute the causes of other peoples' behavior to internal characteristics but attribute their own behavior to uncontrollable circumstances in the environment (Ross & Nisbett, 1991).

The type of information recalled may also be a function of one's expertise or experience with a given task. For example, Hall, Gott, and Pokomy (1995) noted that experts are able to describe intricate aspects of their problem-solving procedures, yet sometimes provide limited insight into the principles employed, failing



DO NOT REPORT AIRCRAFT ACCIDENTS AND CRIMINAL ACTIVITIES ON THIS FORM.
ACCIDENTS AND CRIMINAL ACTIVITIES ARE NOT INCLUDED IN THE ASRS PROGRAM AND SHOULD NOT BE SUBMITTED TO NASA.
ALL IDENTITIES CONTAINED IN THIS REPORT WILL BE REMOVED TO ASSURE COMPLETE REPORTER ANONYMITY.

(SPACE BELOW RESERVED FOR ASRS DATE/TIME STAMP)

IDENTIFICATION STRIP: Please fill in all blanks to ensure return of strip.
NO RECORD WILL BE KEPT OF YOUR IDENTITY. This section will be returned to you.

TELEPHONE NUMBERS where we may reach you for further details of this occurrence:

HOME Area ____ No ________ Hours ____
WORK Area ____ No ________ Hours ____

NAME ________ TYPE OF EVENT/SITUATION ________
ADDRESS/PO BOX ________
DATE OF OCCURRENCE ________
CITY ________ STATE ____ ZIP ____ LOCAL TIME (24 hr. clock) ____

PLEASE FILL IN APPROPRIATE SPACES AND CHECK ALL ITEMS WHICH APPLY TO THIS EVENT OR SITUATION.

REPORTER	FLYING TIME	CERTIFICATES/RATINGS		ATC EXPERIENCE	
○ Captain	total ____ hrs	○ student	○ private	○ FPL	○ Developmental
○ First Officer		○ commercial	○ ATP	radar ____ yrs	
○ pilot flying	last 90 days ____ hrs	○ instrument	○ CFI	non-radar ____ yrs	
○ pilot not flying		○ multiengine	○ F/E	supervisory ____ yrs	
○ Other Crewmember	time in type ____ hrs	○ ____		military ____ yrs.	
○ ____					

AIRSPACE		WEATHER		LIGHT/VISIBILITY		ATC/ADVISORY SERVICE	
○ Class A (PCA)	○ Special Use Airspace	○ VMC	○ ice	○ daylight	○ night	○ local	○ center
○ Class B (TCA)	○ airway/route ____	○ IMC	○ snow	○ dawn	○ dusk	○ ground	○ FSS
○ Class C (ARSA)	○ unknown/other ____	○ mixed	○ turbulence	ceiling ____ feet		○ apch	○ UNICOM
○ Class D (Control Zone/ATA)		○ marginal	○ tstorm	visibility ____ miles		○ dep	○ CTAF
○ Class E (General Controlled)		○ rain	○ windshear	RVR ____ feet		Name of ATC Facility:	
○ Class G (Uncontrolled)		○ fog	○ ____				

	AIRCRAFT 1			AIRCRAFT 2		
Type of Aircraft (Make/Model)	(Your Aircraft) ____		○ EFIS ○ FMS/FMC	(Other Aircraft) ____		○ EFIS ○ FMS/FMC
Operator	○ air carrier ○ commuter	○ military ○ private	○ corporate ○ other ____	○ air carrier ○ commuter	○ military ○ private	○ corporate ○ other ____
Mission	○ passenger ○ cargo	○ training ○ pleasure	○ business ○ unk/other ____	○ passenger ○ cargo	○ training ○ pleasure	○ business ○ unk/other ____
Flight plan	○ VFR ○ IFR	○ SVFR ○ DVFR	○ none ○ unknown	○ VFR ○ IFR	○ SVFR ○ DVFR	○ none ○ unknown
Flight phases at time of occurrence	○ taxi ○ takeoff ○ climb	○ cruise ○ descent ○ approach	○ landing ○ missed apch/GAR ○ other ____	○ taxi ○ takeoff ○ climb	○ cruise ○ descent ○ approach	○ landing ○ missed apch/GAR ○ other ____
Control status	○ visual apch ○ controlled ○ no radio	○ on vector ○ none ○ radar advisories	○ on SID/STAR ○ unknown	○ visual apch ○ controlled ○ no radio	○ on vector ○ none ○ radar advisories	○ on SID/STAR ○ unknown

If more than two aircraft were involved, please describe the additional aircraft in the "Describe Event/Situation" section.

LOCATION	CONFLICTS
Altitude ____ ○ MSL ○ AGL	Estimated miss distance in feet: horiz ____ vert ____
Distance and radial from airport, NAVAID, or other fix ____	Was evasive action taken? ○ Yes ○ No
	Was TCAS a factor? ○ TA ○ RA ○ No
Nearest City/State ____	Did GPWS activate? ○ Yes ○ No

NASA ARC 277B (January 1994) GENERAL FORM Page 1 of 2

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA has established an Aviation Safety Reporting System (ASRS) to identify issues in the aviation system which need to be addressed. The program of which this system is a part is described in detail in FAA Advisory Circular 00-46D. Your assistance in informing us about such issues is essential to the success of the program. Please fill out this form as completely as possible, enclose in an sealed envelope, affix proper postage, and send it directly to us.

The information you provide on the identity strip will be used only if NASA determines that it is necessary to contact you for further information. THIS IDENTITY STRIP WILL BE RETURNED DIRECTLY TO YOU. The return of the identity strip assures your anonymity.

AVIATION SAFETY REPORTING SYSTEM

Section 91.25 of the Federal Aviation Regulations (14 CFR 91.25) prohibits reports filed with NASA from being used for FAA enforcement purposes. This report will not be made available to the FAA for civil penalty or certificate actions for violations of the Federal Air Regulations. Your identity strip, stamped by NASA, is proof that you have submitted a report to the Aviation Safety Reporting System. We can only return the strip to you, however, if you have provided a mailing address. Equally important, we can often obtain additional useful information if our safety analysts can talk with you directly by telephone. For this reason, we have requested telephone numbers where we may reach you.

Thank you for your contribution to aviation safety.

NOTE: AIRCRAFT ACCIDENTS SHOULD NOT BE REPORTED ON THIS FORM. SUCH EVENTS SHOULD BE FILED WITH THE NATIONAL TRANSPORTATION SAFETY BOARD AS REQUIRED BY NTSB Regulation 830.5 (49CFR830.5).

Please fold both pages (and additional pages if required), enclose in a sealed, stamped envelope, and mail to:

NASA AVIATION SAFETY REPORTING SYSTEM
POST OFFICE BOX 189
MOFFETT FIELD, CALIFORNIA 94035-0189

DESCRIBE EVENT/SITUATION

Keeping in mind the topics shown below, discuss those which you feel are relevant and anything else you think is important. Include what you believe really caused the problem, and what can be done to prevent a recurrence, or correct the situation. (USE ADDITIONAL PAPER IF NEEDED)

CHAIN OF EVENTS
- How the problem arose
- Contributing factors
- How it was discovered
- Corrective actions

Page 2 of 2

HUMAN PERFORMANCE CONSIDERATIONS
- Perceptions, judgments, decisions
- Actions or inactions
- Factors affecting the quality of human performance

FIGURE 1 Example of NASA's ASRS reporting form.

to establish the relationship between their domain knowledge and the strategies used to solve the problems. Presumably, experts possess elaborate cognitive schemas and scripts, or mental models of the domain, which improve problem solving yet make their knowledge more implicit than explicit. Novices, on the other hand, tend to report what is immediately (superficially) available to their awareness. When pressed for an explanation, novices generate random reasoning. Because of their lack of understanding of the domain, novices are unable to think in terms of configuring effective plans for efficient analysis.

KNOWLEDGE ELICITATION TECHNIQUES

The process of how to best elicit knowledge is multifaceted and must be strategically organized, so that it can capture the intricate knowledge structures of experts and still elicit thorough procedural explanations from novices. Indeed, there are strategies to improve knowledge elicitation. For example, Diaper (1989) outlines a systematic elicitation process in which the elicitor devises a system that ensures all the relevant material is identified and gathered, regardless of expertise. One such method is careful goal decomposition, in which the solving of a problem is broken down into subgoals or subtasks. Large problems are broken down into smaller ones until they can be discussed in enough detail to show the interrelation of the higher- and lower-level corollaries of the problem.

Cognitive task analysis (CTA) is another method that has been developed to elicit an operator's strategies and decisions by analyzing the various steps involved in an operator's performance of a task. One such CTA technique is the critical decision method (CDM) (Klein, 1993), in which an interviewer uses cognitive probes to understand the processes underlying decisions made during nonroutine events. Table 1 provides a list of some of the factors that are addressed during this CTA process. Using cognitive probes provides the means to focus on key decisions, cues, and options used during a critical event. Probing for the factors allows the interviewer to uncover a variety of aspects of the operator's decision processes, including why certain choices (as opposed to other choices) were made at key points in the event, what aided the decision process, and what might have been done had the scenario been different. Depending on the domain, different probes may be used, rather than the full complement. CDM also delineates the differences between novice and expert performance of a task by noting the cues and inferences experts might make when performing a task and comparing these strategies to those of novices. Organizing experts' representations of how they relate to a complex system allows interventions to be developed that facilitate performance and safety.

A variety of other knowledge elicitation techniques also exist, including cognitive interviewing (Geiselman & Fisher, 1989; Loftus et al., 1989; Memon & Bull, 1991),

TABLE 1
A List of the Cognitive Probing Factors Addressed During CTA[a]

Probe Type	*Probe Content*
Cues	What were you seeing and hearing?
Knowledge	What information did you use in making this decision, how was it obtained?
Goals	What were your specific goals at the time?
Situation assessment	If you had to describe the situation to someone else at this point, how would you summarize it?
Options	What other courses of action were considered, or were available to you?
Basis of choice	How was this option selected/other options rejected?
Experience	What specific training or experience was necessary or helpful in making this decision?
Aiding	If the decision was not the best, what training, knowledge, or information could have helped?
Hypotheticals	If a key feature of the situation had been different, what difference would it have made in your decision?

Note. CTA = cognitive task analysis.
[a]From Klein (1993).

critical incident reporting techniques (Flanagan, 1954; Kirwan & Ainsworth, 1992), and even hypnosis (Hiland & Dzieszkowski, 1984). However, all of these require one-on-one interviews conducted by "knowledge extraction" experts and often require a considerable amount of time to employ. Therefore, none are suitable for incident reporting within aviation, given that incidents are not generally investigated and interviewing all aircrew involved in incidents would be too costly and logistically not feasible given the large number of incidents that occur in the industry. For example, NASA's ASRS Department receives an average of over 33,000 reports per year (NASA, 2000). Consequently, most incident information in the aviation domain is generated using paper-and-pencil forms that pilots complete on their own. Therefore, what is needed is a knowledge elicitation tool that captures the expertise of elicitors yet can be utilized by individual pilots who are naïve to psychological and human factors methodologies.

Schematic Maps

One possible tool for improving the information recalled in aviation incident reports is a schematic map. Schematic maps use nodes (boxes) and links to spatially represent experts' mental models of a domain or process (Brooks & Dansereau, 1983; Dansereau, 1978). Schematic maps depict, through visual representation, the system's main components and interrelationships, as well as the causes and consequences that various inputs have on system performance. Users of these maps fill in the boxes with the details of a particular event to facilitate recall and understand the factors that influenced their behavior.

Schematic maps have been shown to help people analyze their own personal knowledge and experiences in a variety of contexts and to aid in the production of more informative reports, as compared to reports from those who do not use mapping (N. G. Wiegmann, 1992). However, the majority of this work has been in the context of behavioral counseling, using schematic maps to help people identify and recall the factors that influence and motivate their behavior (Dees, 1991). The application of schematic maps to incident reporting has yet to be examined.

Potentially, schematic organizers could be employed as a knowledge elicitation tool for pilots to use prior to writing an incident narrative. Indeed, schematic maps eliminate many of the drawbacks of traditional knowledge elicitation techniques. Because schematic maps represent experts' models of the domain, and provide a form that pilots can complete by themselves, schematic maps essentially eliminate the need for a one-on-one interview by an expert. Furthermore, research in other domains suggests that schematic maps can be understood and utilized with minimal training and effort (D. A. Wiegmann, Dansereau, Skaggs, & Gordon, 1992). As such, schematic maps have the potential not afforded by traditional knowledge elicitation methodologies for widespread application within the aviation industry.

Critical Event Reporting Tool

The critical event reporting tool (CERT) has been developed to improve the type of information reported within the narrative portion of aviation incident reports. CERT draws upon previous work on CTA and the CDM (Klein, Calderwood, & MacGregor, 1989; Militello & Hutton, 1998), as well as upon research on schematic maps (D. A. Wiegmann, Dansereau, et al., 1992; N. G. Wiegmann, 1992). Specifically, CERT was designed to serve as a knowledge elicitation tool that pilots can use for recalling event-related information prior to writing an essay. In particular, CERT was designed to encourage pilots to think of why an incident happened and the factors that affected their actions during an incident, in addition to describing what events occurred during an incident.

Figure 2 provides a depiction of CERT. The center of the form contains the four main components of an action sequence, as generally defined by CDM. The first box represents the process of situation assessment, which has been shown to be a key factor in expert performance in time-critical, high-risk environments (Klein, 1993). The second box in the sequence represents decision planning, or the course of actions considered for remedying the unsafe situation. The third box refers to actions or activities that were performed to execute the plan, and the forth box highlights the possible good and bad outcomes that may have resulted from these actions. Along the top of the form are boxes, which represent external performance-shaping factors

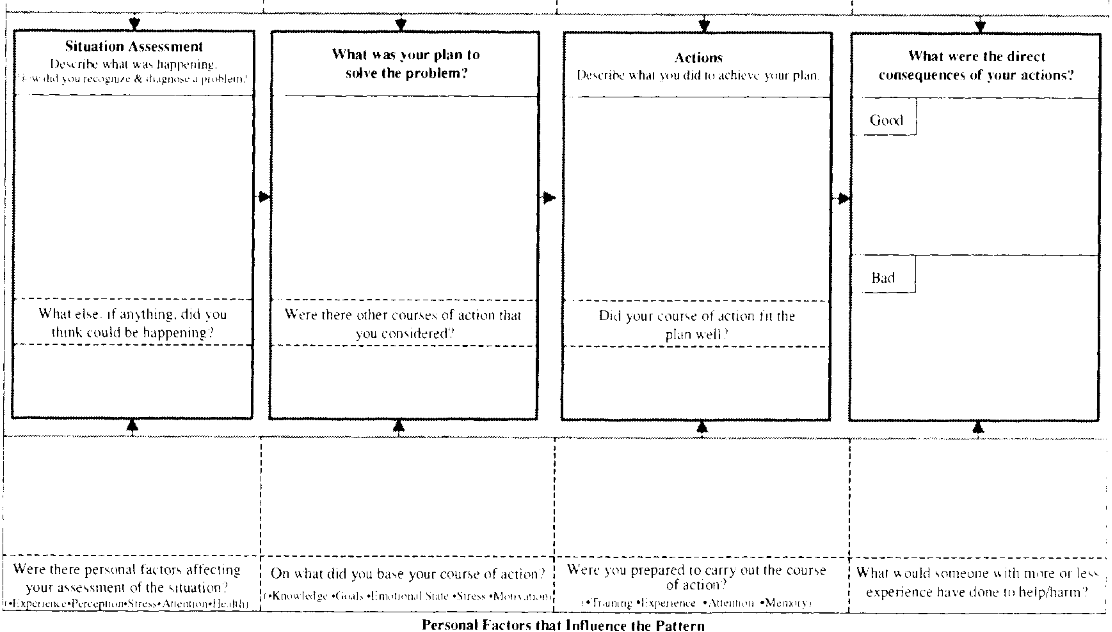


FIGURE 2 Example of the critical event reporting tool (CERT). Original printed on 11 in × 17 in paper.

that may have influenced the operator's success during each of these stages of action (Miller & Swain, 1987). These include such factors as workload, distraction, time pressure, equipment design, and task difficulty. The outside bottom portion of the form represents internal performance-shaping factors or operator preconditions that may have influenced performance at each stage (Shappell & Wiegmann, 2001). These include such factors as experience, stress, memory, and expertise.

Potentially, CERT provides a structure that prompts or cues the recall of important event information. In addition, the spatial layout of the form may help highlight the interrelationships among factors and the consequences of operator actions in response to the critical event. Empty boxes within the organizer may also provide feedback to the users with respect to gaps in their recollection or description of the incident. Finally, the generic-structure CERT may prove beneficial in its application across a variety of incidents or domains. Nonetheless, a framework such as CERT may also have potential drawbacks for users. For example, this type of schematic organizer may be too inflexible to capture all of the information that users need to report. In addition, the static nature of the CERT form may stagnate the reporting of dynamic, iterative events that often occur in the aviation domain. Finally, the complexity of the form or its "busyness" may also prove too complex or difficult for novice users to employ effectively.

PURPOSE OF THIS STUDY

The purpose of this study was to empirically evaluate the potential strengths and weaknesses of CERT as a knowledge elicitation tool for event reporting within the context of aviation. Pilots were exposed to in-flight emergencies during a simulated cross-country flight; they then completed an ASRS narrative report of the incident. Half of the pilots completed CERT prior to writing the report; the other half of the pilots read only an example incident report. Pilots' impressions of CERT were assessed, as well as the impact that CERT had on essay content and quality.

METHOD

Participants

We tested a group of general aviation pilots (n = 34) from the University of Illinois' Institute of Aviation. Participants were, on average, 20 years of age (SD = 3.5) and had an average of 121 total flight hours (minimum = 50, maximum = 371). The pilots were paid $6/hr for their time, with a maximum of 2 hr spent on the experiment.

Task and Procedure

Participants began the experiment by completing a consent form and a preexperimental questionnaire that asked them to provide basic background information, including information about their previous flight experience. Upon completion of these documents, participants performed a simulated visual flight rules, cross-country flight. The simulation was developed using Elite software (Version 6) for the personal computer and was presented using a 550-MHz PC and a 20-in. color monitor. The system was also equipped with a yoke, rudders, throttle and trim controls, sound, and a tracking ball that could be used to access other instruments and controls. The simulator was configured to approximate a Cessna 172.

The estimated duration of the flight was approximately 45 min. However, the flight simulation was preprogrammed to produce mechanical failures in the aircraft at specific points in the flight. At 15 min into the flight, each pilot was given a preprogrammed alternator failure. Then at 20 min, each received a drop in engine revolutions to 50% of full power. Given the pilot had not successfully diagnosed the problem or diverted to an alternate airport, the airplane engine was preprogrammed to fail completely at 25 min into the simulation. All failures were properly represented on the instrument and annunciator panel and through simulated sound when applicable. At each instance of programmed airplane problems, one or more airports were well within landing distance of the plotted course. The airplane was controllable and able to be flown with minimal problems despite the alternator failure and RPM drop. However, once the complete engine failure occurred, there was no chance for restart, and an immediate emergency landing was required.

After completion of the flight simulation, participants were brought into a debriefing room, where they were asked to complete a report about the event using the essay portion of the ASRS form. However, prior to completing this report, participants were assigned to either the CERT-group (n = 17) or the control group (n = 17), with an equal number of pilots from the different university flight courses being assigned to each group to control for flight experience. Participants in the CERT group received a brief description and blank copy of CERT, which they were allowed to study for 5 min. They were then provided an example of a completed CERT that was mapped to a fictitious automobile accident (5 min). Next, they read an example essay/incident report that corresponded to the fictitious automobile accident (5 min). During this same time period, pilots in the control group completed a short distractor task (10 min) that involved reading about the ASRS system. As with the CERT group, they were given the example essay from the same fictitious automobile accident (5 min). Pilots in both groups then began the incident reporting task, describing the events that they had experienced during the flight simulation. Pilots in the CERT group were given 15-min to complete the form prior to writing the essay, whereas pilots in the control group wrote only the essay. Hence, pilots in both groups began recalling the simulation events after a 15-min delay. Pilots in both groups were also given as much time as they needed to complete the essay portion of the report.

Following completion of the essay, pilots in both groups were given a set of eleven 7-point Likert-scale questions and asked to rate their essays in terms of content, format, and how well their essay described the incident. CERT pilots were also administered an extra 7-point Likert-scale questionnaire with eleven questions and were asked to rate CERT in terms of its format and effectiveness and as an aid in their recall of the event. They were also asked to provide written comments about their likes or dislikes of the CERT form. Upon completion of the experiment, participants were thanked, compensated for their time, and then dismissed.

Essay Coding

Pilots' essays were evaluated by a certified flight instructor (CFI) who was blind to group assignment. These evaluations by the CFI were done using the same eleven 7-point Likert-scale questions that participants used to evaluate their own essays. The content of the pilots' essays was also analyzed by categorizing statements into one of three broad categories, including *what* happened (i.e., descriptive statements about events), *why* something happened (i.e., analytical statements about the causes of events), and *context* statements (i.e., preamble and postscript statements). These categorizations were made by raters who held advanced human factors degrees as well as private pilot certificates and who were naïve to group membership.

RESULTS

Subjective Evaluation of CERT

All subjective evaluations of CERT were positive, with no responses to the rating questions falling below a rating of 4 (i.e., neutral) on the 7-point Likert scale. Average ratings for each item are presented in Table 2. Note that negatively worded questions have been reverse-coded so that higher scores reflect more positive ratings for each item. Asterisks next to items indicate that the average rating is significantly higher than neutral (a score of 4). Overall, pilots felt that the CERT form aided them in their recall of the incident ($M = 4.65$, $SD = 1.22$), $t(16) = 2.18$, $p < .05$. They also felt that the form was flexible enough to adapt to their specific requirements ($M = 4.9$, $SD = 1.41$), $t(16) = 2.58$, $p < .05$. Pilots found the form helpful in highlighting where important information may have been left out when recalling the event ($M = 5.12$, $SD = .86$), $t(16) = 5.37$, $p < .01$, and generally felt it had adequate size and space for writing in the boxes ($M = 5.4$, $SD = 1.94$),

TABLE 2
Mean Ratings of Subjective Evaluations of CERT

Abridged Question	*Mean Rating*
Was setup of form constraining?	4.41
Did form aid in recall process?	4.65*
Spatial layout help determine relationships?	4.35
Form flexible enough to adapt?	4.88*
Language too technical?	6.00*
Form confusing at first glance?	4.23
Fields generic enough?	5.59*
Did form highlight gaps in info?	5.12*
Adequate space to write?	5.41*
Did you struggle to fill in all boxes?	4.47
Did form prepare you to write essay?	4.41

Note. CERT = critical event reporting tool.
*Indicates that mean rating is significantly higher than neutral ($p < .05$). Note that negatively worded items have been reverse coded. Higher scores indicate more positive ratings on each item.

$t(16) = 3.0, p < .01$. They did not find the terminology in the form to be too technical to understand ($M = 6.0$, $SD = 1.22$), $t(16) = 6.73$, $p < .01$, nor were the terms judged to be too generic ($M = 5.6$, $SD = 1.3$), $t(16) = 5.13, p < .01$.

When asked to qualitatively comment on the positive and negative aspects of the form, pilots were generally more positive, with 71% of the participants reporting positive aspects about the form (29% having no opinion), in contrast to 53% responding to negative aspects of the form (47% having no negative commentary). Table 3 delineates a summary of the positive and negative statements of the respondents. In general, many of the positive comments concerning the form related to its aid in recalling the events, explaining the causes of the events, and organizing one's thoughts for writing the essay. The negative comments generally focused on issues concerning the constraints the form placed on the description of the event and its inflexibility to the dynamic nature of the flight environment.

Subjective Evaluation of Essays

Self-evaluations. Pilots in both the CERT and control group rated the content and quality of their essays using a set of eleven 7-point Likert-scale questions (again items were coded so that higher scores reflected more positive ratings). These items were analyzed using a stepwise, forward-entry logistic regression procedure in which the treatment group (CERT vs. control) served as the criterion variable and responses on the questionnaire items served as the predictor variables. The results of this analysis revealed a significant logistic regression function, $\chi^2(2, N = 34) = 11.83$, $p < .01$. In particular, two questionnaire items

TABLE 3
Summary of Positive and Negative Comments about CERT

Positive	
	"Big Boxes."
	"How it flowed."
	"It helps when organizing your thoughts."
	"It is nice to have the outline of points you want to present in the report."
	"It reminded me of some key points that I would have left out without it."
	"It made sense and would probably help jog memory when used long after an incident."
	"I like the parenthesis where it would say (knowledge, emergency procedures) and such."
	"It was helpful in preparing me for the essay."
	"Some of the things it listed were helpful and reminded me of the incident."
	"The general boxes above where you could scratch notes, and then the bigger box in the middle where you could elaborate on the above points."
	"The smaller words (suggestions) that helped explain what factors to think about."
	"Very detailed."
Negative	
	"Boxes seemed repetitive."
	"Boxes were too small."
	"I felt like I was saying the same thing over and over."
	"It felt like all the questions were the same so I don't know if I was answering them correctly."
	"I was told what to put in the boxes, should have just been offered ideas."
	"It didn't cope well with a single event that had multiple problems."
	"Many boxes were repetitive. Confusing when I first saw the form."
	"The form can be cut down in half."
	"Too constraining for multiple emergencies. Would have been better to have one sheet per emergency."
	"My orientation would be to fill in the boxes from the top and work down. The problem is that the bottom should be filled out before the middle, I think."
	"Some of the arrows are confusing."

Note. CERT = critical event reporting tool.

combined to significantly differentiate between groups. Item 1 pertained to pilots' impressions of how well they felt their essays described *what* happened during the incident ($\beta = -1.2002, p < .05$), whereas the Item 2 pertained to their impressions of how well they felt their essays described *why* the events happened ($\beta = .7422$, $p < .05$). As seen in Figure 3, pilots in the control group (essay only) felt that they were better at relaying to the reader a sense of *what* happened during the incident ($M = 5.94, SD = .83$), as opposed to those who used CERT ($M = 5.06, SD = 1.43$). In contrast, pilots who used CERT felt they were better at relaying *why* they performed the way they did ($M = 5.53, SD = 1.23$), as opposed to those who wrote the essay only ($M = 4.65, SD = 1.87$).

Expert evaluation. The CFI ratings of participants' essays were also analyzed using a stepwise, forward-entry logistic regression procedure in which the treatment group (CERT vs. control) served as the criterion variable, and responses

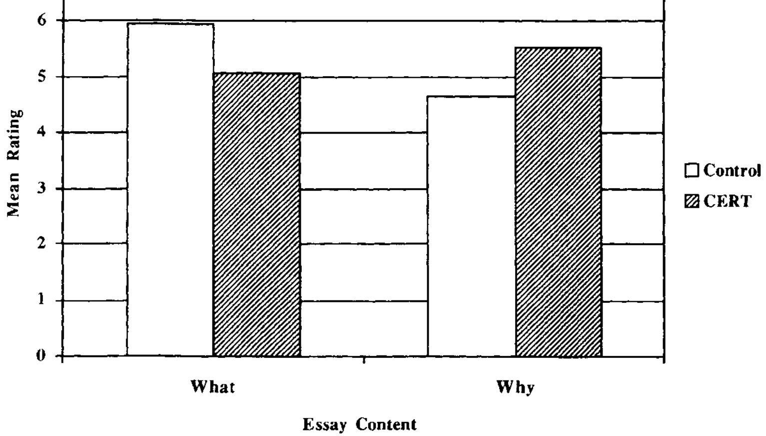


FIGURE 3 Pilots' subjective evaluations of the essay performance.

on the questionnaire items served as the predictor variables. The results of this analysis revealed a significant logistic regression function, $\chi^2(1, N = 34) = 7.38$, $p < .01$. However, only one item relating to "ease of reading" significantly differentiated between the groups ($\beta = -.5507$, $p < .01$). Surprisingly, the expert rater judged the control group's essays as easier to read ($M = 5.6$, $SD = 1.3$) than the CERT group's essays ($M = 3.9$, $SD = 2.1$).

Objective Evaluations of Essays

Quantitative analysis. The amount of time to complete the essays for pilots in the CERT group ($M = 15.82$ min, $SD = 4.17$) did not differ significantly from completion times of pilots in the control group ($M = 15.41$ min, $SD = 6.12$). There was no significant difference in the number of words used by the CERT group ($M = 276$, $SD = 88.14$) and the number of words per essay in the control group ($M = 284$, $SD = 98.74$). There was also no significant difference between the groups on indexes of reading difficulty or sophistication level, with control essays scoring 67.09 ($SD = 8.7$) on the Flesch reading-ease score and CERT essays scoring 68.05 ($SD = 6.9$). Likewise, control essays scored 7.9 ($SD = 1.7$) and CERT essays scored 7.6 ($SD = 1.5$) on the Flesch–Kincaid Grade level score.

Qualitative analysis. Figure 4 presents the percentage of statements within pilots' essays that fell within each of the three content categories (what, why, and context). As can be seen from the figure, the majority of statements contained in the essays were descriptive, with essays of pilots in the control group ($M = 69\%$,

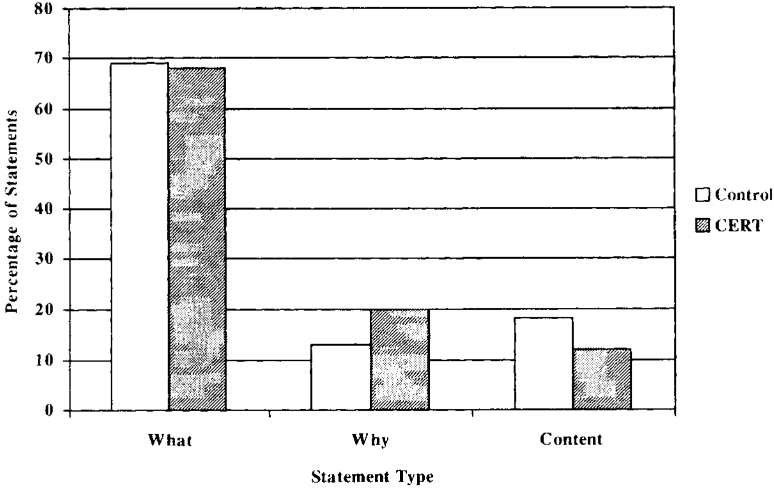


FIGURE 4 Comparison of essay content between the CERT-group and the control-group.

SD = 8.98) having a slightly higher percentage of descriptive statements than those written by the CERT group (M = 68%, SD = 6.42). Essays of participants in the control group (M = 18%, SD = 3.88) also tended to have a higher percentage of context statements than the essays generated by pilots in the CERT group (M = 13%, SD = 7.41). However, essays by pilots in the CERT group tended to have a higher percentage of analytical statements (M = 19%, SD = 4.67) than the essays by pilots in the control group (M = 13%, SD = 2.66).

These differences between groups were again analyzed using a stepwise, forward-entry logistic regression procedure in which treatment group (CERT vs. control) was the criterion variable and scores for each statement type (what, why, and context) were the predictor variables. The results of this analysis revealed a significant logistic regression function, $\chi^2(1, N = 34) = 4.674$, $p < .05$, with scores on only the *why* statements significantly discriminating between groups ($\beta = 8.27$, $p < .05$). No significant difference between groups was obtained for descriptive and context statements.

DISCUSSION

Evaluation of CERT

Pilots' evaluations of the value and utility of CERT were, overall, very positive. CERT was judged to aid the recall of the incident and was viewed as being helpful in highlighting where important information may have been overlooked during the recollection of the event. In addition, CERT was not seen as too confusing, nor did the pilots who participated in this study view it as being too generic to fit the event they had experienced. Pilots also stated that CERT helped them organize their thoughts and helped them determine the relationship between events or "how things flowed." They also stated that CERT reminded them of key factors to consider when analyzing the event.

However, pilots noted some problems with CERT. In general, these problems related to inflexibility or constraints that CERT placed on the recall of the events. Apparently, the static nature of the CERT form was not completely compatible with the dynamic nature of the events related to the incident. Consequently, revisions may need to be made to the form that allow more flexibility or make the reporting process more fluid. This could possibly be accomplished by providing pilots with multiple copies of CERT, which pilots could use to analyze every iteration of the critical event sequence. A better way, however, may be to automate the form so that pilots can complete it using a computer. A computerized version of the form might allow for greater flexibility and adaptability of the form to dynamic situations, ultimately enhancing user performance and satisfaction.

Evaluation of Essay Reports

Information content. Pilots' subjective evaluations of their essays revealed that those who used CERT felt as though they had included more analytical information in their essays than pilots who did not use CERT prior to writing the essay. In contrast, the pilots who did not use CERT reported that they provided more descriptive information about the incident in their essays than did pilots who used CERT. Objective content analysis confirmed that these subjective evaluations by the pilots were generally correct. Essays written by pilots who used CERT did contain, on average, more analytical statements about why things had occurred during the incident than did essays written by pilots who did not use CERT. Essays of pilots who did not use CERT had a slightly larger percentage of descriptive statements than those who did use CERT.

These findings support the results of previous research showing that schematic maps facilitate users' insights into the causes and consequences of their behavior (Dees, 1991). However, this study is the first to demonstrate that schematic maps can be effective in improving the recall of analytical information in incident reports. It should be noted that these benefits of CERT occurred after only a brief training period. Furthermore, pilots were only allowed to use the tool for a maximum of 15 min prior to writing their essays. This procedure was implemented in order to address issues related to the real-world application of the tool, which would generally be constrained by pilots' willingness to spend time preparing to write the report. These findings also suggest that experts' models or CTA techniques can be represented in a way that novices can use to improve their recall of factors that contribute to their actions and the outcome of incidents. Consequently, subsequent reports are more likely to contain analytical information that might ultimately facilitate the identification of effective interventions.

Information quantity. Pilots who used CERT did not recall more information about the event than did pilots who did not use CERT. Essays written by both groups of pilots contained a relatively equal number of words and sentences. One possible reason for this apparent lack of impact of CERT on the amount of information recalled is that the time between experiencing the critical event and completing the report was rather short (only 15 min). Research on human memory indicates that decrements in memory increase with recall delay, and that memory aids are more effective in improving recall after longer rather than shorter delays (Bransford & Stein, 1984). Another possible reason is that participants in both groups were instructed on how to write a report by giving them an example essay about a fictitious car accident that was generally very elaborate. They also used the actual narrative portion of the ASRS form to write their essays. This form contained some instructions about the types of human factors issues to consider when describing the event. Consequently, the relatively short recall delay, combined with an elaborate example and retrieval cues in the ASRS form, may have served to improve the recall of pilots above what they might normally recollect after longer delays or when left to their own devices. This increase in recall could have produced a "ceiling effect," affording little room for CERT to enhance the amount of information recalled.

Readability. An objective index of readability revealed no difference between essays written by pilots who used CERT and those written by pilots who did not. Surprisingly, however, subjective ratings of the essays by a CFI indicated the essays written by the CERT group were more difficult to read than were those written by the control group. Perhaps since the essays written by pilots in the CERT group generally contained less contextual information, as well as slightly less descriptive information, these essays were unusual or different from normal types of reports or essays. Therefore, they were seen as more difficult to read. Indeed, no instructions were given to the pilots who used CERT as to how to transition or translate the information from the CERT form into an essay. Possibly, users of CERT could benefit by having some sort of strategy for organizing and relaying the information generated from the form into a narrative, making the essay more readable.

Other Applications of CERT

The generalizability of CERT to the reporting of critical incidents other than the type used in this study needs to be explored. Given its theoretical basis and generic structure, CERT should be applicable to other types of incidents, as well as incidents outside the cockpit, such as in air traffic control or maintenance. Its application to incidents in other types of dynamic, high-risk systems, such as nuclear power and medicine, might also be explored. However, the internal and external performance-shaping factors contained in CERT were derived for the general aviation pilot. Therefore, specific examples used in the form may need to be modified to fit a particular domain. Additional performance-shaping factors such as supervisory and organizational factors known to impact operator behavior (Reason, 1990; Shappell & Wiegmann, 2001) may also need to be included in CERT when transitioning to incidents involving commercial aircrew or other operators in an organizational context.

Although CERT was designed as a tool for aiding report writers in recalling and writing incident reports, it may have other applications for eliciting knowledge. For example, CERT might prove useful as an interviewing tool that safety professionals could use to follow up on incident reports filed by aircrew. In other words, rather than having aircrew complete the form prior to writing the report, CERT could be used by interviewers as a tool for extracting additional details about selected incidents. CERT may also prove useful as an aid for conducting cognitive task analyses. Indeed, because CERT is based partially on the CDM (Klein, 1993), it may serve to provide structure or consistency when performing such analyses. In either case, whether used as an interviewing tool for incidents or as a CTA technique, CERT potentially reduces the need for the person performing these tasks to be an expert in human factors. Since CERT schematically represents an expert model of the action process, perhaps it can be used for these purposes without the normally required expertise. Clearly, however, research is needed to test the effectiveness of CERT or other schematic maps in these contexts.

CONCLUSION

The critical event reporting tool (CERT) appears to have potential as an aid in critical incident reporting. CERT incorporates previous work on CTA and on CDM (Klein et al., 1989; Militello & Hutton, 1998), as well as research on schematic maps (D. A. Wiegmann, Dansereau, et al., 1992; N. G. Wiegmann, 1992). Results of this study indicate that CERT encourages pilots to think of *why* an incident happened and the factors that affected their actions during an incident in addition to describing *what* events occurred during an incident. Additional research is needed, however, to explore the effectiveness of CERT in other contexts and with other incident types. Enhancements to the form are also needed, including improvements to its flexibility. These efforts could eventually lead to the collection of better real-world incident data and ultimately to more effective intervention and accident prevention programs.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the Federal Aviation Administration (DTFA 99-G-006). The views expressed in this article are those of the authors and do not necessarily reflect those of the FAA. We acknowledge the invaluable contribution of Donald Talleur in setting up the simulation used in this study. An abbreviated version of this article was previously published in the *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (Santa Monica, CA: Human Factors & Ergonomics Society, 2001).

REFERENCES

Bransford, J. D., & Stein, B. S. (1984). *The ideal problem solver: A guide for improving thinking, learning and creativity.* New York: W. H. Freeman and Co.

Brooks, L. W., & Dansereau, D. F. (1983). Effects of structural schema training and text organization on expository prose processing. *Journal of Educational Psychology, 75*, 811–820.

Chappell, S. (1997). Using voluntary incident reports for human factors evaluations. In N. Johnston, N. McDonald, & R. Fuller (Eds.), *Aviation psychology in practice* (pp. 149–169). VT: Ashgate.

Connell, L. (1999). Incident reporting: The NASA Aviation Safety Reporting System. *GSE Today*, 66–68.

Dansereau, D. F. (1978). The development of learning strategy curriculum. In H. F. O'Neil, Jr. (Ed.), *Learning strategies* (pp. 1–29). New York: Academic Press.

Dees, S. M. (1991, August). *Using knowledge maps to develop understanding of personal behavior and behavior related to drug use: An overview.* Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

Diaper, D. (1989). *Knowledge elicitation: Principles, techniques and application.* New York: Wiley.

Ericsson, K., & Simon, H. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251.

Fitts, P. M., & Jones, R. E. (1947). Analysis of 270 "pilot error" experiences in reading and interpreting aircraft instruments. In *Air material command.* Dayton, OH: Wright Patterson AFB.

Flanagan, J. (1954). The critical incident technique. *Psychological Bulletin, 54*(4), 327–358.

Geiselman, R. E., & Fisher, R. P. (1989). The cognitive interview technique for victims and witnesses of crime. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 191–215). New York: Springer-Verlag.

Hall, E., Gott, S., & Pokomy, R. (1995). *A procedural guide to cognitive task analysis: The PARI methodology* (Armstrong Laboratory Technical Report AL/HR-TR-1995-0108). San Antonio, TX: Brooks AFB.

Heinrich, H. (1959). *Industrial accident prevention: A scientific approach* (4th ed.). New York: McGraw-Hill.

Hiland, D. N., & Dzieszkowski, P. A. (1984). Hypnosis in the investigation of aviation accidents. *Aviation, Space and Environmental Medicine, December*, 1136–1142.

Kirwan, B., & Ainsworth, L. (1992). *A guide to task analysis.* Washington, DC: Taylor & Francis.

Klein, G. (1993). *Naturalistic decision making: Implications for design.* Dayton, OH: CSERIAC SOAR 93–1.

Klein, G., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Systems, Man & Cybernetics.* SMC-19, 462–472.

Loftus, E. F., Greene, E. L., & Doyle, J. M. (1989). The psychology of eyewitness testimony. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence.* New York: Springer-Verlag.

Memon, A., & Bull, R. (1991). The cognitive interview: Its origins, empirical support, evaluation and practical implications. *Journal of Community & Applied Social Psychology, 1*, 291–307.

Miller, D. P., & Swain, A. D. (1987). Human error and human reliability. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 219–250). New York: Wiley.

Militello, L., & Hutton, R. (1998). Applied cognitive task analysis: A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(2), 17–23.

NASA (2000). *Aviation safety reporting system program overview*. Retrieved from http://asrs.arc.nasa.gov/overview_nf.htm.

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

Ross, L., & Nisbett, R. (1991). *The person and the situation: Perspectives of social psychology*. New York: McGraw-Hill.

Shappell, S. A., & Wiegmann, D. A. (2001). Applying reason: The human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety, 1*(1), 59–86.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232.

Wiegmann, D. A., Dansereau, D. F., Skaggs, L. P., & Gordon, N. B. (1992, April). *Comparing techniques for processing schematic maps*. Paper presented at the annual meeting of the Southwestern Psychological Association, Austin, TX.

Wiegmann, N. G. (1992). *Schematic organizers and cooperative review: Impact on recall and metacognition*. Unpublished master's thesis: Texas Christian University.

Manuscript first received November 2001