Why safety and Human Factors/Ergonomics standards are so difficult to establish

Paul Green
University of Michigan Transportation Research Institute (UMTRI)
Michigan, USA

Abstract

This paper summarises lessons from experience in developing international and national vehicle safety and human factors/ergonomics standards. Based on this experience, problems of (1) meeting organisation (failure to follow Robert's Rules of Order, lack of numbered documents, etc.), (2) committee structure (lack of member expertise, too many members), and (3) inadequate committee support (lack of secretarial support and consultants to prepare materials) are often more important than the technical contributions of members. This paper offers specific suggestions to overcome these problems. Some of these problems occur because safety and human factors/ergonomics expertise and contributions to design can be undervalued.

Further, when standards development organisations lack policies that establish the desired level of safety, technical committees may not be able to agree on the performance requirements and validation procedures for safety and usability standards. Key elements of such policies are the standard of care (the level of protection required, e.g., do no harm) and burden of proof (confidence in the outcome, e.g., safe beyond a reasonable doubt).

Introduction

This paper is based upon the author's experience in developing standards through the Society of Automotive Engineers Intelligent Transportation Systems Division (SAE ITS) and the International Standards Organization (ISO), and related work as a member of the U.S. National Academy of Sciences Committee on Motor Vehicle Rollover. The ISO efforts have been with Working Groups 5 (Symbols) and 8 (TICS (Traffic and Information Control Systems) on Board – MMI (Man Machine Interface)) of Technical Committee 22, Subcommittee 13 (ISO TC 22/SC 13, Ergonomics Applicable to Road Vehicles). The SAE work has been with the ITS Safety and Human Factors Committee and its Navigation Subcommittee. The symbols work began in 1975. The predominant recent SAE activity has been the development of two SAE Recommended Practices, J2364 and J2365. SAE J2364 ("The 15-Second Rule") describes the limits of driver use of navigation systems in motor vehicles (Green, 1999c; Society of Automotive Engineers, 1999). SAE J2365 (Green, 1999a;

In D. de Waard, K.A. Brookhuis, J. Moraal, and A. Toffetti (2002), *Human Factors in Transportation, Communication, Health, and the Workplace* (pp. xx-xx). Maastricht, the Netherlands: Shaker Publishing.

Society of Automotive Engineers, 2002) describes a method to calculate navigation system task times and estimate compliance with J2364.

This paper gives particular attention to working group administrative and procedural matters (see also Stewart, 1998, 2000) and the essential need for organisational safety policies that consider the standard of care and burden of proof. The absence of safety policies mires the development of safety standards.

Organisational lessons from ISO WG 5 and 8 and the National Academy of Sciences

Documents, support people, and meeting leaders

ISO working groups, subcommittees, and technical committees have well-structured meetings, with formality increasing at higher levels. Meetings follow agenda that are distributed beforehand and include previous meeting minutes. Minutes contain resolutions (items voted upon by the committee, usually action items) that keep the work program focused. (Example: "Resolution 84, Safety Assurance – Mr. Noy will consider the comments from the experts given at the December 98 meeting and develop a draft proposal by 30 April 99 for an HMI Safety Assurance Preliminary Work Item for discussion at the June 99 WG8 meeting.). A good action item has a person responsible for it, a specific task to complete, and a due date.

Further, every document distributed gets a unique number that is entered in a document log. Referencing documents by number focuses discussion and assures that everyone, even correspondence members or those absent, has all of the documentation, ("The delegates should refer to TC22/SC13/WG5, document N389.") Meeting discussions are most productive when documents are submitted in advance, so document numbers appear on copies, though document numbers can be assigned during a meeting.

Numbering requires the presence of a support person with a document log at all meetings (who should also take minutes). That person can serve as a link to the past and to other related national and international committees on which they serve. ("As you may recall in the Paris meeting three years ago, resolution N205 was passed. It said…We need to be consistent with that resolution.") Usually, the contributions of the support person are far more important than the technical contributions of any single committee member. Because human factors/ergonomics is sometimes undervalued, obtaining staff support may be more difficult than for traditional engineering disciplines.

In an ideal world, draft resolutions are distributed at the end of the meeting so assignments are acted upon immediately, for example on the plane ride home. This is particularly valuable in multi-day meetings typical of ISO where activities one day are based on decisions made the previous day.

> *Lesson 1:* All committee documents should be numbered.

Lesson 2: Numbering occurs only if there is a support person present with a document log.

Lesson 3: For consistency, the same support person should serve the committee and related committees for many years. So technical experts can remain focused on the discussion, the support person should be from the staff of the standards development organisation.

Some meetings generate so many documents that carrying hardcopies from previous meetings can be overwhelming for air travellers. Therefore, groups often rely on their imperfect memories or that of the support person for history. As a consequence, some issues that were thought to be resolved are discussed over and over again.

Lesson 4: Standards development organisations need to provide support personnel and members with an easy to use, portable means to access, manage, and update meeting records and access organisational policies in real time.

One solution could be a laptop-based, electronic database of documents that can function as a non-networked application. The database should (1) automatically generate document status summaries, (2) provide links to all versions of documents (3) identify the specific changes made in each revision (and why), (4) identify all relevant resolutions including those from other technical groups, and (5) provide links to organisational procedures and rules for ballots, document preparation, and other matters.

Meeting rules and management

To avoid meeting anarchy, meeting chairs must keep the discussion focused, especially when a motion is being considered. Control does not need to be as strict as in a legislative body such as the U.S. Congress ("I yield two minutes of my time to the representative from the great state of Michigan"). However, cordial reminders regarding time constraints must be made to keep the discussion from being monopolised and to avoid excessive reiteration. The rules for meetings are well defined in Robert's Rules of Order, the basic manual of parliamentary procedure (Robert, Evans, Honermann, Balch, and Robert, 2000), of which versions have existed since 1876. In some situations, the Sturgis Code of Procedures (Sturgis, 2000) or Dementer's Manual (Dementer, 1969) may be appropriate. Slavish adherence to the fine points of parliamentary procedure is generally not necessary and in fact, some leeway may be desired when groups are small (10 or less).

Lesson 5: Meetings should follow Robert's Rules of Order. Meetings should stay on topic, and when they stray off track, the chair needs to gently interrupt and redirect the discussion.

Lesson 6: To resolve matters of procedure, primarily in larger groups, a parliamentarian may be needed. Many committees do not know Robert's Rules, so someone with that knowledge may need to be provided by the standards organisation.

> *Lesson 7:* At times, delegates who are non-native English speakers, especially from Asian cultures, may fall silent during discussions. It is the duty of both the chair and the delegates to seek input from them if they have not spoken for some time.

In contrast to the SAE and ISO, committees of the National Academy of Sciences allow only members of the committee and Academy staff to sit at the meeting table. Visitors and others are farther away from the discussion, making it obvious whom is a member. Furthermore, non-member participation in discussion is limited. These steps assure that (1) committee members are heard and control the discussion, (2) the discussion is based upon technical input from the experts, and (3) only members vote.

> *Lesson 8: In meetings, there should be a clear physical separation between committee members and non-members to avoid merging of roles, and input should only come from members unless the committee members decide otherwise.*

What can happen when there are no data

Of the working groups observed, WG5 has been particularly challenged by a lack of data on which to base standards. At a working group meeting, a delegation identifies the need for a symbol. Having a graphics expert with computer tools to develop readily modified graphics can foster the initial discussion, as would tools to produce tones and masking sounds, warning messages, and other information sources for other contexts.

For symbols, the initial request usually leads to a resolution calling for candidate symbols from the national delegations for the next meeting, usually in six months. Some delegations return with neatly drawn candidates, though almost never any data on legibility or understandability due to the lack of time (and funding). Because there are some candidates, there is pressure to select a graphic image for standardization. What engineers envision to be appropriate is different from what the driving public envisions (Green, 1981). Engineers tend to think of equipment from an internal, mechanistic perspective. So engineers draw gears to represent a transmission and the cross section of a drum brake to represent a brake. The public thinks of what they can see, shift levers and brake pedals, and as a consequence, may not understand the symbols created by engineers. Symbol testing may involve holding pages of candidates at arms length and squinting at them or looking at out of focus images (from an overhead projector) from the back of a room, hardly the basis for an international standard.

> *Lesson 9:* Because of the time pressure to develop standards, sometimes standards are based on little or no data. (Note: This is not always so. In the case of J2364, the lack of data is used as an argument to challenge the development of a standard.)

Lesson 10: Methods that allow for instant human factors/ergonomics analysis in meetings are needed. For example, computer models to predict the legibility of an image based on a pixel map could be very useful in guiding decisions about symbols.

The saga of SAE Recommended Practice J2364

Use of consultants to assist standards efforts

Insights from the development of SAE J2364 (and J2365) have been numerous. In contrast to the past, these two recommended practices, and ITS standards in general, pertain to new technology and there are usually no preceding company, industry, or government standards that could serve as a basis for them.

The development of these documents formally began in December 1996, with a one-page request for a quote from the SAE Safety and Human Factors Committee. After a five-month delay due to complex federal contracting requirements, the University of Michigan received a small contract. There were 15 project deliverables, a number specified by program managers remote from the actual effort. This excessive number of deliverables proved to be a major burden and somewhat shifted the emphasis of the project from supporting the development of standards to documenting when the University should be paid.

Lesson 11: To maintain consultant focus on technical output useful to standards development, oversight of consultants needs to be assigned to the standards committee, not the funding organisation.

The U.S. Department of Transportation funded this and other projects to accelerate the development of telematics standards. Consultants provided literature reviews (e.g., Green, 1999b, d) and draft documents, so the committee volunteers could focus on evaluating synthesised materials, not on assembling the raw material. All of the other SAE ITS standards efforts have had similar positive experiences with deliverables from consultants.

Lesson 12: Hiring consultants to produce literature reviews and draft standards significantly enhances the quality of standards and accelerates development.

How long does it take to develop a standard?

The project schedule called for all of the work, including two draft standards, to be completed by SAE in less than two years, which was unrealistic. In fact, as of the writing of this paper (July 2002), only the SAE committee has approved a draft of SAE J2364 (but the Navigation Subcommittee is developed a new draft). Given the impasse with regard to its content, an acceptable revision is likely to require at least another 6 months of debate. SAE J2365, the second document, was just approved as an SAE Recommended Practice, about four years behind schedule. Had there been no consultant, there might not even be a draft practices for discussion.

> *Lesson 13:* Experience with ISO and other well managed efforts indicates that about two years are required for fundamental agreement on a new standards concept, and another three years are needed for development to be completed. Conversion of existing standards (e.g., an SAE standard to an ISO standard) takes less time.

Why has it taken so long to develop J2364 and J2365? There was no progress on J2365 because the decision was to focus on J2364 first, and reaching a consensus agreement on J2364 has been problematic. J2364 was initially conceived as a design standard. Paraphrasing the words of Gene Farber (former chair of the Safety and Human Factors Committee), "We need to say what we should not do. For example, tell me how many items on a menu is too many." This request occurred because the then best-known navigation system in the U.S., the Rockwell PathMaster, required drivers to utilise a time consuming process of scrolling through long lists of city and street names to enter addresses. It took about a year to convince the SAE Committee that J2364 should be a performance standard.

> *Lesson 14:* As standards are developed, their purpose, approach, and content may change.

How many people should be in a discussion group?

Progress has also been hampered by the size of the subcommittee developing J2364. The original proposal was for a Navigation Subcommittee of five people including the chair, the consultant, and three other human factors experts (two of whom would be current on standards efforts in Japan and Europe respectively). In contrast, the initial subcommittee had over a dozen members, some of who had no human factors background and were philosophically opposed to a standard. Agreeing on a common meeting time and place was not always easy because of the group size. The subcommittee now has about eight active members. Most recently, much of the editing of drafts has involved groups of three or four people, and the smaller groups have made significant progress.

> *Lesson 15:* The group that reaches fundamental agreement on the content of a standard should be small. Five or fewer is recommended if adequate representation of key parties can be obtained, though the literature on group dynamics might suggest other sizes.

Applying this lesson to ISO is difficult because of the need for several nations to be represented by one or more experts on working groups. A solution is for working groups to have smaller planning and editing task forces with the appropriate balance of divergent viewpoints.

Who is qualified to serve as an expert?

In the mid-1990s, there was a shift in the U.S. from government regulations developed through contention between the government, industry, and other parties to consensus standards developed by qualified, independent technical experts working

with standards development organisations as opposed to trade associations. If the qualifications or independence of experts are in doubt, then so too is the premise for standards development by nongovernmental organisations such as SAE and ISO.

There is a common misconception that being human makes one a human factors/ergonomics expert, so non-experts often engage in professional practice, including standards development. Non-experts are more likely to have mistaken opinions about the quality of data or the importance of particular findings, complicating standards development. Similar misconceptions about expertise are not as common for traditional engineering disciplines, medicine, or law.

How does one determine who is a human factors/ergonomics expert? Just as one cannot become an electrical engineer without formal training, the same is true for a human factors engineer/ergonomist. Since standards have legal force, an expert should be someone who could testify in court about their area of expertise. In an American court, the following evidence would be supportive of expertise. The answers from academic and industry experts will be different.

1. A degree from an accredited or well recognised program that may have human factors or ergonomics in its title
2. Courses on this topic. (Titles should include phrases such as “human factors,” “ergonomics,” “biomechanics,” “human engineering,” “human decision making,” “safety engineering,” and so forth. One or two courses do not make an expert.)
3. Being a full member (meaning documented experience) of the Human Factors and Ergonomics Society, the Ergonomics Society, or some similar organisation
4. Papers, articles, and presentations authored on this topic (and descriptions of them)
5. Detailed descriptions of the safety and usability evaluations conducted
6. Human factors/ergonomics courses taught

Lesson 16: The qualification of experts should be established, preferably using the criteria for expert witnesses in a court of law. Considered are the relevance of their degrees, courses taken, professional affiliations, and so forth.

Often, unqualified individuals can be identified by one simple test of minimal exposure. Name one human factors/ergonomics textbook you own and have read cover to cover.

Organisations that want to track standards efforts in a particular area but lack a technical expert should feel free to send observers to standards committee meetings. In most cases, there are mechanisms for the views of observers at such meetings to be communicated directly to the technical experts.

To qualify experts and eliminate those with tangential interest, the SAE Safety and Human Factors Committee now requires new members to establish their expertise via a resume reviewed by the committee chair before being permitted to join as a voting

member. There are no restrictions on visitors or mailing list recipients. SAE has also instituted a policy to delete individuals from committees if they do not participate in meetings or ballots. These changes have improved the SAE standards development process by pruning the Committee of those who do not participate, so that obtaining a quorum to act on Committee business and make progress is much less of a problem.

However, those without human factors expertise can serve a valuable role in framing human factors decisions. In making decisions, it is important to know what is technically feasible, practical to do, and likely to occur in the future. This information may not be the purview of human factors/ergonomics specialists.

Expertise balance is a problem for non-European delegations to ISO TC 22/SC 13 working groups. Most meetings are held in Europe, so travel costs (for experts from other parts of the world) are significant. This is a particular problem for experts from universities, consumer organisations, and others, and is reflected in their lack of participation in working group activities. This is also a challenge for all delegates from developing nations.

> *Lesson 17:* If ISO automotive committees are to be balanced, some long-term funding for travel (and time) of non-industry members (especially from academia) is needed.

Though not often discussed in public, there are some concerns regarding the independence of technical experts. The SAE Technical Standards Board Governance Policy is similar to that of most standards development organisations. Every SAE standards meeting begins with the reading of Section 4.1 of the SAE policy, "Members … function as individuals and not as agents or representatives of any organisation with which they may be associated." Nonetheless, a few individuals privately admit, "I voted against this because my boss said so." Standards development organisations should look for opportunities to reinforce the importance of independent experts to employers, for example via written reminders from standards development organisations to employers of experts. Written affirmation from a member's employer should be considered, though few employers may be willing to do so. Members who cannot abide by the policy should resign their committee membership.

> *Lesson 18:* Some way to assure the independence of those working on standards is needed.

What is the appropriate measure of safety and usability?

Within the SAE Navigation Subcommittee there are two distinct perspectives concerning (1) the appropriate measures of safety and usability, (2) the types of evidence that should be used to establish acceptance criteria, (3) how safe is safe enough, (4) the appropriate population for safety and usability tests, and (5) even the basic premise of design induced errors. In American product liability law, if a product

design induced driver errors, then the manufacturer has some responsibility for the consequences. The different perspectives make reaching a consensus a challenge.

Early disagreements concerned the selection of a performance measure, either task time or eyes-off-the-road time. Wierwille (1995, see also Green, 1999d) has shown that crash risk increases as eyes-off-the-road time increases. However, the selection of a performance measure was strongly influenced by the Gould and Lewis (1985) design principles that emphasise early testing and feedback to design. Glance data are very expensive and time consuming to collect and require a working system in a driving simulator or mock-up, something that is available only shortly before production. In contrast, static (parked) task time is well correlated with task time while driving, and eyes-off-the-road time (Green, 1999d; Nowakowski, Utsui, and Green, 2000) and task times can be obtained using a stopwatch and a simulated interface early in the design process.

In fact, an early concept for Recommended Practice J2364 was not to require testing at all, but to rely on calculations based on the task times in J2365. Most engineering decisions are based on calculations, not empirical testing. The task times in J2365 were derived from (1) the Keystroke-Level Model (Card, Moran, and Newell, 1983) popular in human-computer interaction, (2) MTM-1, the standard database for movement times used by industrial engineers, and (3) UMTRI testing of the Siemens Ali-Scout Navigation System. Using J2365, design modifications can be evaluated in minutes or hours, not the days or weeks required for empirical testing.

> *Lesson 19: There may be a need for more than one measure of safety and usability—some for initial evaluation and others for certification.*

> *How safe is safe enough (and how usable is easy to use)?*

> A major sticking point between the two perspectives has been the acceptable level of safety and usability. There are very limited data in the literature relating use of navigation systems to crashes, mainly because use of such systems is generally not coded on crash investigation forms. To establish such criteria, some have argued for postponing development of a standard until more safety data are available (but curiously, they have not argued for postponing production of navigation products). A lesson from the Firestone/Explorer rollover matter in the U.S. is that one must act before fatalities accumulate. Furthermore, the position that "do not do anything until enough people die" is philosophically inconsistent with a human factors/ergonomics perspective and is a practice that would be unacceptable for a medical device. It is equally unacceptable for transportation.

There is confusion between the criteria for acceptance in a scholarly journal and the criteria for real-world engineering (and standards). As Don Norman, a well-known usability expert, has said, "Good enough is good enough." Gene Farber was fond of an expression attributed to both von Clausewitz and Voltaire; "The perfect is the enemy of the good." For new technology, there is not the time or resources to have

all of the information needed—correlation data for a wide range of test conditions, test-retest reliability scores for different experimenters, representative baseline data for a wide range of devices, etc. Products need to be developed in a timely manner and for a reasonable cost, and standards are needed to guide product development.

Sometimes safety and usability decisions must be based on the judgments of human factors/ergonomics experts, judgments that need to be consistent with other engineering decisions. For example, to determine if a GPS signal sensor has adequate sensitivity, one or two sensors might be tested in a few representative worst-case situations; not just those of expected use. Using that information, an electrical engineer will make a production decision, even though their knowledge of sensor sensitivity and the assessment process is incomplete. The engineer will not delay production until the topic of GPS signal reception has been exhaustively researched. The same should be true when the criteria for safety and usability standards are established.

> *Lesson 20:* Some decisions about safety criteria will need to be made before significant numbers of fatalities have occurred and research is complete.

Legal arguments about safety consider two key concepts, the standard of care and the burden of proof. The standard of care refers to the level of protection required. Some examples are:

- Do no (additional) harm. (What physicians use, the Hippocratic oath.)
- Do minimal harm.
- It should be absolutely safe.
- It should be generally safe.
- It should be somewhat safe.
- Do more good than harm.
- When in doubt, protect the driver.
- Only provide protection for unavoidable risks.
- The design does not matter because it is ultimately driver judgment.

The burden of proof refers to the degree which something must be established as safe or unsafe. There are two aspects to this concept. First, is the product assumed to be safe or unsafe unless established otherwise? Second, how much proof is required? Some examples are:

- beyond any shadow of a doubt
- beyond a reasonable doubt
- the preponderance of evidence
- the balance of the evidence
- the most probable outcome
- there is a reasonable chance
- the possibility exists

Tests for standard of care and burden of proof vary with the legal context (a civil vs. criminal matter) and the product (motor vehicles, home appliances, medical devices, industrial machinery, toys, nuclear power, etc.). In part, these differences occur because the willingness to accept risk varies with the situation (Fischoff, Lichtenstein, Slovic, Derby, and Keeney, 1984).

In the context of technical standards, decisions about burden of proof, standard of care--policy decisions--are now made by the members of technical committees, subcommittees, and working groups, and vary between groups within an organisation. Governing boards, not others, should set organisational policy. Experience has shown that there are philosophical differences between those with and without human factors/ergonomics expertise in terms of acceptable levels of standard of care and burden of proof.

> *Lesson 21:* Governing boards (of ISO, SAE, corporations, government agencies) should establish unambiguous and uniform policies for standard of care and burden of proof to guide technical committees and product design, and to make most effective use of technical experts' time. This is probably the most important of the lessons learned.

As an example of how such policies might be applied, the Alliance of Automobile Manufacturers is writing a set of driver interface guidelines (www.umich.edu/~driving/guidelines). In a previous draft, Principle 2.1 stated the following:

"Systems with visual displays should be such that the driver can complete the desired task with sequential glances that are brief enough not to adversely affect driving."

Further, as an initial proposal, the document states the following:

While the vehicle is in motion, the total number of glances shall not exceed 10, with any single glance not to exceed 2 seconds, and a total time not to exceed 30 seconds.

Where do those numbers come from and do they make sense? Louis Tijerina (Ford) drafted a very well thought-out rationale for these values. Radio tuning is a task that affects driver eye glance behaviour, vehicle control, and critical event detection. (Question: Does that mean the radio-tuning task is on the borderline or over the borderline? See Stutts, Reinfurt, Staplin, and Rodgman, 2001.) The 85th percentile is commonly used in traffic engineering as a reasonable worst case. According to data from Dingus (1988), tuning a radio requires a mean of 6.91 glances, with a standard deviation of 2.39. Thus, the 85th percentile number of glances is 6.91 + 1.04 x 2.39 or 9.4 rounded to 10 glances according to Tijerina. (Should it be rounded to 9, the closest integer?)

Further, inspection of Rockwell's 1988 paper on glance durations from several on-road studies shows the 85th percentile glance duration to be 1.9 seconds, rounded to 2 seconds.

Multiplying the 2 numbers together (10 x 2) suggests an acceptable eyes-off-the-road time of 20 seconds, and if glances to the road are assumed to be 1 second, a total task time of 30 seconds. This criterion establishes when a task is probably unsafe, not when it is probably safe.

A different standard of care leads to substantially different acceptance limits. If the radio tuning task is the borderline, then one should use the mean number of glances from Dingus for tuning a radio (6.9 glances, rounded to 7) and the mean glance duration reported by Rockwell (1.44 seconds, rounded to 1.5). This suggests a maximum eyes-off-the-road time of 7 x 1.5 or 10.5 seconds, or slightly over half of the value suggested in the Alliance draft. Furthermore, if one were to quibble over the glance time to the road, Chiang, Brooks, and Weir (2001) suggest 0.5 seconds for a glance to the road while using a navigation system, suggesting an additional 3.5 seconds (7 x 0.5) would be needed to complete the task. Therefore, the total task time would be 10.5 + 3.5 or 14 seconds, a value consistent with J2364. Interestingly, the 14 seconds is just less than half of the Alliance proposal. Thus, even well-intentioned human factors engineers using the same data but different criteria for the standard of care can identify quite different criteria for acceptance, in this case differing by a factor of two, a difference of practical importance to engineers.

Policies for standard of care and burden of proof are particularly difficult to establish for international standards because of different national legal systems, legal heritage and laws of negligence, differences between traditional religious law and modern civil practice, and different philosophies of individual responsibility. For example, in the U.S., an aggrieved party would undertake a product liability action, whereas in Japan there would more effort to apologise for the problem and correct it quietly. In Germany, the philosophy is that actions are invariably the driver's responsibility. In France, there is an emphasis on individual liberty, so smoking is prevalent in public places even though illegal. There are international differences in helmet laws, compliance with speed regulations, the use of seat belts, and even the use of lane markings. (For example, some noted that in Italy, lane markings are viewed as a suggestion.)

Further problems with setting criteria are assumptions about the relationships between safety and performance, and usability and performance. For example, crash risk and motor vehicle fatalities increase as BAC (blood alcohol concentration) (Foley, 1998) and HIC (head impact criteria) increase. In most U.S. states, the legal maximum BAC value was changed from 0.10 to 0.08. Is a driver now unsafe at 0.081 and safe at 0.079? Are drivers in the U.S. less safe than in many countries in Europe, where the BAC limit is 0.05? Likewise, crash risk increases with eyes-off-the-road time, task time, and other measures, but there is no point at which performance dramatically changes from safe to unsafe. However, establishing an acceptable standard of care is much easier when the safety policy is clear.

What data should be considered in setting a standard?

During the development of SAE J2364, it has become apparent that those with a human factors background have been more likely to support use of a wider variety of evidence, not just crash statistics or use of radio tuning as a yardstick. Other data considered include how long drivers are willing to look away from the road, stress ratings from using systems, and findings from test tracks and simulators, not just on-the-road studies.

> *Lesson 22:* Human factors/ergonomics experts are more likely than non-experts to accept a broader range of evidence in setting criteria for safety and usability, an important advantage when evidence is limited as often is the case for new technology.

Who are the users?

Agreement on the user population is important when selecting whom should be tested to verify compliance with a standard. Some have argued for only testing the intended users for navigation products, for example only 40-year-old males who are technophiles. However, the human factors perspective is to test all those that might be expected to use a product, being certain to include reasonable worst cases. Thus, the author's 80-year-old mother who does not know how to use Windows or Mac operating system (a non-technophile) would be excluded from the population of intended users. But she does drive a late model Lexus with a navigation system, and specifically purchased that option because of its potential usefulness. From the perspective of consistency, the driver position of most motor vehicles is designed to accommodate the 5th to 95th percentile anthropometry of adults, regardless of the product market segment.

> *Lesson 23:* Without agreement on the user population, there will be no agreement on the test sample to validate compliance with a standard. The test sample should be a reasonable worst case of both the expected and intended users. For motor vehicles, this means that older drivers (over age 65) must be part of the test sample because they invariably drive every mass marketed motor vehicle.

The current status of J2364 and J2365

In 1999, SAE J2364 was approved by the Navigation Subcommittee and approved by the required two-thirds of the Safety and Human Factors Committee. The Committee chair replied to each objection and J2364 was forwarded to the ITS Division for approval. The Division rejected J2364, saying a consensus was not achieved. At the request of the SAE Technical Board, the ruling of the Division was appealed. As the hearing approached, the matter began to resemble the appeals in Florida of the U.S. Bush/Gore Presidential election happening at the time. Those opposed to the standard brought in a lawyer, so those in favor responded in kind. There were questions about how many voted, which ballots counted, etc. Unfortunately, the appeals panel ruled to return J2364 to committee (and the subcommittee) to build a greater consensus.

There has been discussion of a supplemental or alternative procedure that involves using visual occlusion to strengthen the consensus. Research is being conducted on this measure and there are many unresolved issues regarding test parameters and the merits of the procedure. Resolution of the research questions may occur this year. The development of an SAE safety policy would greatly assist in achieving consensus.

> *Lesson 24:* When standards are contested, the resolution can be a political process, not one of technical merit. Cynics have used the phrase "political circus."

Where there is disagreement about a proposed standard (as in J2364) or a standard has considerable detail, a rationale document helps to focus discussion by providing the supporting reasons for each element of a proposal. This was the case of the SAE adaptive cruise control standard and is evident in the elaboration document for the Alliance principles.

> *Lesson 23:* When standards are being discussed (and especially when they are balloted), an accompanying rationale document is needed. The document should include (1) the standard of care and burden of proof used as a basis for decisions, (2) a summary table with a bulleted list of supporting facts for each major decision (e.g., why the total task time limit is 15 seconds), and (3) linked explanatory text with references to the human factors literature for those facts.

Rationales are needed for the selection of the evaluation method, each dependent and independent measure, values for time parameters, the number and types of subjects, etc. Admittedly, some decisions will be based on human factors expert judgment. Although a rationale document takes time to prepare, having one should abbreviate committee debates and reduce the number of objections, thus shortening the time to develop a consensus standard. In hindsight, the consultant should have been funded to create a rationale document for J2364.

Final thoughts

Whereas some standards development organisations are beset by problems, others function very effectively. Some of these problems are exacerbated for safety and human factors/ergonomics standards because work on these topics is undervalued and expertise is not recognised as being unique.

Further, a major point of this paper is that the passage of human factors/ergonomic standards is hampered not just by a lack of data or technical matters, but often by flaws in the process by which standards are established. SAE working groups and ISO task forces can have members who lack adequate technical expertise or can have too many members. Having consultants to prepare literature reviews, drafts of standards, and rationale documents makes working groups and subcommittees much more effective, allowing volunteers to focus on evaluating alternative recommendations, not on creating them. Working groups and committees also need to have chairs keep discussions on track, use Robert's Rules of Order, and support staff from parent standards organisations to track documentation and assignments. Again, some of these problems, especially those related to staffing, are more common for safety and human factors/ergonomics activities because the contribution of these disciplines is sometimes viewed as less important than those of other engineering disciplines, and input from experts and non-experts is commingled.

However, the most significant lesson learned is that boards of standards development organisations such as SAE and ISO must set clear policies for the standard of care and burden of proof used as a basis for standards. Similarly, government organisations involved in rulemaking, and corporate boards of manufacturers and suppliers, must do the same. Without such guidance, no matter how well meaning and educated the committee members (or product engineers), those organisations become very expensive and non-productive debating clubs. The experience is frustrating to dedicated members, and the safety and usability standards (and ultimately products) developed, if developed at all, fall short of societal needs.

Although this document considers problems in detail, the intent is to emphasise lessons learned. There are many talented people with a sincere interest in making products and services safe and easy to use. People work on human factors/ergonomics problems because they want to make life better for all. This paper was written to assist those involved with human factors/ergonomics standards development reflect upon past activities so as to improve the quality of their work.

Acknowledgments

I would like to thank Jim Foley (Mitretek), Monica Milla (UMTRI), Christopher Nowakowski (PATH), Gary Rupp (Ford), Colleen Serafin (Johnson Controls), Omer Tsimhoni (UMTRI), and Jill Wilson (National Academy of Sciences) for their comments on drafts of this manuscript. The author would also like to acknowledge Walter Sicks (VDA), Karin Bolchezy and Penny Brown (SAE), and Susan Scott (JPL) for noteworthy support of ISO and SAE meetings, and Rudi Haller (BMW) and Francois Hartmann (Renault) for their adept management ISO TC 22/SC 13/WG 8 meetings.

References

Chiang, D. P., Brooks, A. M., and Weir, D. H. (2001). *An Experimental Study of Destination Entry with an Example Automobile Navigation System* (SAE paper 2001-01-0810), Warrendale, PA: Society of Automotive Engineers.

Dementer, G. (1969). *Dementer's Manual of Parliamentary Law and Procedure*, Boston, MA: Little, Brown.

Dingus, T.A. (1988). *Attentional Demand for an Automobile Moving-Map Navigation System* (unpublished Ph.D. dissertation), Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial Engineering and Operations Research.

Fischoff, B., Lichtenstein, S., Slovic, P., Derby, S.L., and Keeney, R.L. (1984). *Acceptable Risk*, London, U.K.: Cambridge University Press.

Foley, J. (1998). Personal communication.

Foley, J.P (2001). Issues in HMI Standard Development or The Rantings of a Mad Standards Developer, *ITS View*, September.

Gould, J.D. and Lewis, C. (1985). Designing for Usability: Key Principles and What Designers Think, *Communications of the ACM*, March, *28*, 300-311.

Green, P. (1981). Displays for Automotive Instrument Panels: Production and Rating of Symbols. *HSRI Research Review*, *12*(1), 1-12.

Green, P. (1999a). Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety, *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (CD-ROM), Santa Monica, CA: Human Factors and Ergonomics Society, available at http://www.umich.edu/~driving/publications.html.

Green, P. (1999b). *Navigation System Data Entry: Estimation of Task Times* (Technical Report UMTRI-99-17), Ann Arbor, MI, The University of Michigan Transportation Research Institute, available at http://www.umich.edu/~driving/publications.html.

Green, P. (1999c). The 15-Second Rule for Driver Information Systems, *ITS America Ninth Annual Meeting Conference Proceedings*, Washington, D.C.: Intelligent Transportation Society of America, CD-ROM, available at http://www.umich.edu/~driving/publications.html.

Green, P. (1999d). *Visual and Task Demands of Driver Information Systems* (Technical Report UMTRI-98-16), Ann Arbor, MI: The University of Michigan Transportation Research Institute, available at http://www.umich.edu/~driving/publications.html.

Nowakowski, C., Utsui, Y., and Green, P. (2000). *Navigation System Evaluation: The Effects of Driver Workload and Input Devices on Destination Entry Time and Driving Performance and Their Implications to the SAE Recommended Practice* (Technical Report UMTRI-2000-20), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Robert, S.C., Evans, W.J., Honermann, D.H., Balch, T.J. and Robert, H.M. (2000). *Robert's Rules of Order: Newly Revised* (10th edition), Cambridge, MA: Perseus, (http://www.robertsrules.com/default.html).

Rockwell, T.H. (1988). Spare Visual Capacity in Driving--Revisited: New Empirical Results of an Old Idea, in Gale, A.G., Freeman, M.H., Haslegrave, C.M., Smith, P., and Taylor, S.P. (eds.), *Vision in Vehicles II*, Amsterdam, Elsevier Science, 317-324.

Society of Automotive Engineers (2001). *SAE Technical Standards Board Governance Policy*, Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers (1999). *Navigation and Route Guidance Function Accessibility While Driving* (SAE Recommended Practice J2364, draft of December 15, 1999), Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers (2002). *Calculation of the Time to Complete In-Vehicle Navigation and Route Guidance Tasks* (SAE Recommended Practice J2365), Warrendale, PA: Society of Automotive Engineers.

Stewart, T. (1998). Ergonomics Standards – the Good, the Bad, and the Ugly, in Hanson, M.A., (ed.), *Contemporary Ergonomics 1999*, London, U.K.: Taylor and Francis, 3-7.

Stewart, T. (2000). Ergonomics User Interface Standards: Are They More Trouble than They Are Worth? *Ergonomics*, *43*, 1030-1044.

Stutts, J.C., Reinfurt, D.W., Staplin, L., and Rodgman, E.A. (2001). *The Role of Driver Distraction in Traffic Crashes* (technical report), Washington, D.C.: AAA Foundation for Traffic Safety.

Sturgis, A.F. (2000). *The Standard Code of Parliamentary Procedure* (4th ed.), New York: McGraw-Hill.

Wierwille, W.W. (1995). Development of an Initial Model Relating Driver In-Vehicle Visual Demands to Accident Rate, *Proceedings of the 3rd Annual Mid-Atlantic Human Factors Conference*, Blacksburg, VA: Virginia Polytechnic Institute and State University, 1-7.