Doing Better Driving Research: Suggestions from a Reviewer

Paul Green
University of Michigan Transportation Research Institute (UMTRI)
and Department of Industrial and Operations Engineering
Ann Arbor, Michigan 48109 USA
pagreen@umich.edu, +1 734 763 3795, www.umich.edu/~driving

ABSTRACT:

This paper describes 11 guidelines to improve publications describing human factors studies of driving. These guidelines include: writing as a native speaker; allocating most pages to new material; allocating half of the abstract text to results; stating the research issues as who, what, when, where, why, and how questions; providing numerical predictions for outcomes; providing images of what subjects saw; providing interaction and dual dependent measure plots; balancing the results between practical and statistical significance; and formatting the references as required. Most importantly, the author suggests that publications be required to (1) use the definitions for measures and statistics in SAE Recommended Practice J2944 so that studies can be compared and (2) list relevant standards and guidelines as keywords to help translate the research reported into practice.

1. INTRODUCTION

Recently, the author has shifted his publication focus from reporting individual experiments to overarching publications to improve the field of human factors, and in particular, driving research. Publications include (1) a history of automotive human factors [1], which should be useful to those new to driving research and graduate students, (2) a writing guide for undergraduate students [2], (3) recommendations on how to quickly develop course materials [3], written for new faculty but useful to many others, (4) an SAE Recommended Practice defining driving performance measures [4], and (5) a summary of driving performance measures, presented at the most recent Automotive User Interface conference [5].

In that spirit, one overarching question is how publications on human factors and driving can be improved. Several publications provide statistics on why manuscripts are rejected [6 - 8] for journals concerning the social sciences, education, and medicine, but not engineering. The reasons why vary with the field.

This publication provides 11 guidelines to improve human factors (engineering) publications concerning driving based on the author's experience as a reviewer of journal articles (e.g., from *Human Factors*, *Applied Ergonomics*), conference papers, and student manuscripts. Selected papers from the 2010 Humanist conference (the most recent offering online) provide examples of following (and not following) the guidelines.

2. THE GUIDELINES

2.1 Guideline 1: Make sure the manuscript is written as if the author was a native speaker.

It almost goes without saying that a manuscript should be easy to read and understand, and not have grammatical or spelling errors, even if the manuscript is not written in the manuscript author's native language. Approximately 5% of the manuscripts recently reviewed are so badly written they are returned without any comments, other than suggesting the submitter seek the help of a technical editor. This is a troublesome situation as Microsoft Word, when set to English as the default language, flags repeated words, incorrect use of plurals, sentence fragments, and violations of the that-which rule. For additional guidance, the classic source is Strunk and White [9], though Plainlanguage.gov and related web sites are helpful. Further, even the best authors benefit from a review of their manuscripts by an experienced technical editor.

2.2   Guideline 2: Allocate the most pages to what you have done.

Most journals and conferences limit manuscripts to five to eight pages. When planning a manuscript, create a page budget (e.g., Table 1) to avoid wasting time writing and then cutting excessively lengthy sections. Emphasize what was learned and what is new, not the literature. If the introduction of a manuscript exceeds 25% of the page count, rejection is highly likely. The literature is not the story (unless the manuscript is a literature review).

Table 1: Suggested Page Allocation for a Five-Page Manuscript

Topic	Pages
title, author, affiliations, abstract	1/3 – ½
introduction	1
method	1 (or less)
results	1-1/2 (or more)
conclusions and discussion	½
references	½

Poor page allocation is primarily a problem for manuscripts, not publications such as those from the 2010 Humanist Conference. In fact, there are several thorough but brief literature reviews (e.g., [10]), whose brevity facilitates the desired page allocation.

2.3   Guideline 3: List relevant standards, guidelines, policies, and procedures as keywords.

Human factors research publications often do not make the connection between research and practice, which is a major problem. This connection is important to practitioners, who, for example, are estimated to be more than 80% of the members of the Human Factors and Ergonomics Society. One way to make the connection for practitioners is to list relevant standards and guidelines as keywords. In fact, the author believes that a standards keyword line should be required for journals and conferences such as this one.

Here is how it will work. If someone was reporting research on forward collision warnings, he or she should list SAE J2400 [11] and ISO Standard 15623 [12] in the standards-related keyword entry. Even better would be if the manuscript authors had read those standards before they conducted their research, and their conclusions provided the new text for those standards based on the research conducted. This would benefit (1) researchers (whose research gets into practice), (2) standards writers (who, when searching Google for their standard by its number, would find new research to include), and (3) practitioners (who will learn of applicable standards). Approximately half of human factors research is basic (but very worthy) research, so for those publications, there may be no connection. Nonetheless, the author feels very strongly that this listing should be required because it connects research to practice.

Implementing such a requirement will take some effort. The scope statements of most standards are inadequate, and after purchasing many standards whose titles seemed relevant, one finds that only a few are actually relevant. More complete and readily accessed information on standards, guidelines, and other requirements documents is needed. For instance, the author has published lists of standards related to driver interfaces with summaries of them in several places [13 - 15]. There is also a webinar on the Human Factors and Ergonomics Society web site, a benefit to members, providing an overview of human factors standards [16]. Furthermore, for the U.S., other potentially relevant documents include the NHTSA visual-Manual Distraction Guidelines, state laws (concerning cell phone use driver licensing), handbooks produced by the states and the National Safety Council on safe driving, guidelines for road design in the American

Association of State Highway and Transportation Officials (AASHTO) green book [17], guidance for sign, signal, and marking design in the Manual of Uniform Traffic Control Devices (MUTCD) handbook [18], and rules for hours of service of truck drivers (to avoid fatigue) from the Federal Motor Carrier Safety Administration. Certainly, there are similar rules, laws, and guidelines elsewhere. Papers at the 2010 Humanist conference did cite standards, primarily ISO standards, but some papers could have done more (e.g., [19]).

2.4 Guideline 4: At least half of the words in the abstract should concern the results and include numeric data.

As expressed by Baue [20], "writing a good abstract is not abstract writing." A good abstract summarizes what was done, with about half of the abstract being results, so the reader can decide if they should read the entire paper. To comply with word limitations for abstracts, do not repeat the title in the first sentence, or describe the measures and conditions. Emphasize numeric data related to the procedure (e.g., the number of subjects) and driver performance (e.g. "The mean response time during the day was 1100 ms and at night was 50% greater"). If there is no substance in the abstract, then the publication probably lacks substance as well. Many abstracts in the 2010 Humanist Conference did not meet this content guideline.

2.5 Guideline 5: Present the research issues as 3-6 questions using the words who, what, when, where, why, and how.

A good introduction should say what the problem is, why it is important, and for publications, identify the relevant literature and what one should expect to find. Humanist Conference publications often refer to goals to identify the issues to be addressed, and are better than most publications in describing them. A more direct approach is to state the issues as who, what, when, where, why, and how questions. "What is the relative frequency of a driver being labeled inattentive versus attentive? … What is the relative … crash risk of eyes off the forward roadway?" [21].

Furthermore, as a corollary, verify that every question is also addressed in the results and in the conclusions. Subheadings for each question can help assure such.

2.6 Guideline 6: Provide numeric predictions of experiment outcomes.

Too often, authors just describe tests that were done with some discussion of how theories could explain the outcomes, explanations that appear tacked on in hindsight. Text such as, "Using so and so's method, the mean task time with a unique tone as feedback should be 100 ms less than when no tone is provided" is desired. Methods such as GOMS [22] and SAE J2365 [23, 24] can be used to estimate task times, and there are many other methods to estimate driver performance. The lack of predictions makes research on human factors and driving look weak. The lack of predictions is not acceptable for other fields of science and engineering, and should not be acceptable here, as well. Admittedly, there are often times when so little is known about a topic that predictions cannot be made. Nonetheless, such predictions are uncommon in Humanist Conference publications.

2.7 Guideline 7: Show pictures or drawings of what subjects saw.

A picture is worth 1000 words. If a visual interface was examined, show example screens and label them (e.g., character sizes, lighting levels) to save text.

If subjects drove on a particular road, show example road scenes. Most Americans, even driving researchers, are unlikely to know what a particular motorway near London is like to drive. Similarly, few Europeans, Japanese, Koreans, Chinese, or others will have any sense of what Interstate 94 west of Ann Arbor is like. Most Americans would be clueless as well.

The screens subjects saw are already in computer format, and road scenes are available from video recordings and Google street view. If not, one can use a cell phone camera to record the desired images.

2.8   Guideline 8: Define dependent measures using SAE J2944 [4] for driving performance measures and SAE J2396 [25] and ISO 15008 [26] for glance related measures.

For the last seven years, the author, with considerable help from Gary Rupp, has been writing an SAE Recommended Practice (J2944) to define driving performance measures and statistics [4]. This effort was stimulated by an extensive literature review [27], which showed that many measures had 10 or more names and were only defined 10-15% of the time. The current version of J2944 is more than 170 pages long with 300 references. For each measure and statistic there is a definition, guidance on the use of the measure or statistics based on the literature, in some cases sample data from naturalistic driving, and a specification for how the terms should be cited. When creating definitions, the approach was to include all likely alternatives (options), so as to encourage referencing J2944. Thus, as shown in Table 2, many terms have more than two alternative definitions; with a recommendation being provided where there was a consensus.

Table 2. Selected Terms Having Operational Definitions in SAE J2944 [4]

Term (options)	Comments
response time	15 terms … until accelerator moved, … until brake lights on, … until maximum jerk while accelerating, etc. with 2 or 3 options for each one
longitudinal measures	distance and time for each, see Figure 1; also range
time to collision (2)	plus minimum time to…, adjusted time to…, time exposed time to…, time integrated time to…, inverse time to…
required deceleration	
coherence	plus gain, phase angle, time delay
steering…	reaction time, movement time, response time, reversals (2 options)
lateral position (3)	
lane departure (11)	plus number of.., duration, magnitude, time integrated magnitude
time to line crossing (3)	plus minimum …, inverse …
lane change (5)	plus number of …, duration of…, severity, urgency
roadway departure (3)	plus number, duration of..,. magnitude of… , pavement departure, time integrated magnitude.

As an example of why J2944 is needed, highway engineers have studied highway capacity for almost a century, defining headway as the time from when one vehicle passes until the next one passes. Human factors researchers studying crash avoidance are interested in the space between vehicles, which they mistakenly and routinely call headway, not gap, its correct name. If the lead vehicle is a tractor-trailer, then the difference between the two measures, the error, is the length of that vehicle, approximately 55 feet (16.7 m).

However, the current situation is even more complicated (Figure 1). Some driving simulators that provide "headway" as an output measure, use it as the distance from the center of gravity of one vehicle to another (which makes perfect sense for vehicle dynamics calculations). Other simulators use the spatial/geometric center as the reference. If the users of these simulators are asked what they are measuring, they will say "headway," but refer to it as the distance between vehicles (gap), when in fact it is a third measure.
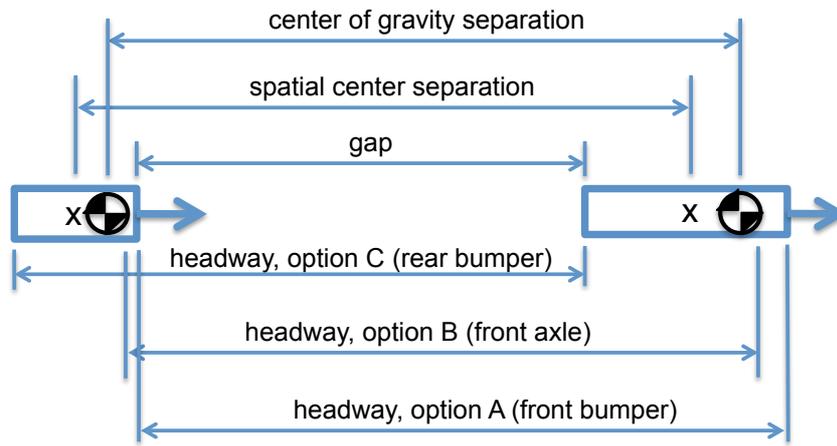
Figure 1: Vehicle Longitudinal Measures from SAE J2944 [4]

2.9 Guideline 9: Use interaction plots for independent measures and dual dependent measures instead of single variable plots.

A good publication provides a complete representation of the results. Often, one will see one figure showing the effect of age (just two bars, young and old), a second figure showing the effect of gender (again, just two bars), and a third figure with the interaction. Interaction plots, even when some factors are not significant, give a more complete representation of the results and require less space.

Similarly, in many driving studies, multiple dependent measures are collected, but they are examined individually, providing a nonintegrated picture. The reader is much better informed by figures that combine likely pairs of dependent measures, such as standard deviation gap and standard deviation of lane position. Figure 2 shows an example of steering angle and throttle angle from the SAVE-IT project. The two ellipses represent p=0.90 and p=0.95 if steering angle and throttle angle were from a joint normal distribution.
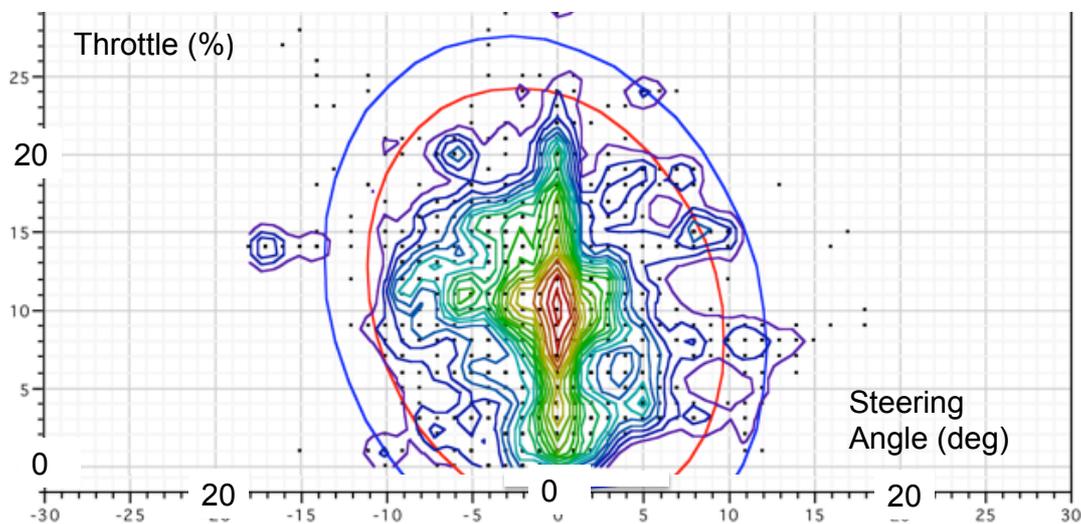

Figure 2: Bivariate Measures: Steering Angle vs. Throttle Angle for Freeways [28]

2.10 Guideline 10: Focus the analysis on regression analysis and means, not on ANOVA and statistical details of significance.

In psychology, a field that is a foundation for human factors, ANOVA is traditionally used to describe performance differences. However, human factors research is often conducted to predict performance, so regression analysis, a related method, is more appropriate. Furthermore, when regression equations are reported, the percentage of variance accounted for is also provided, which is useful information. For example, Whitehurst [29] describes a study of the factors affecting reading gauges (Table 3). In real designs, there are always compromises, but this table indicates that scale number progression should not be compromised as it accounts for the most variability in performance.

Table 3: Percent of Variance Accounted for by Various Design Factors from Whitehurst [29]

Factor	% Variance	Factor	% Variance
Numerical progression	38	Pointer design	0
Interpolation	9	Scale number location	0
Scale unit length	4	2-way interactions	5
Scale orientation	1	Subjects	17
Marker width	0	Residual	25
Clutter	0	Total	100

After statistical significance has been established, a publication should focus on mean differences or effect sizes. Do the results pass the "so-what test?" "There was a statistically significant difference between men and women (men were 3% less) for the rainy – nighttime condition, when subjects were tired." … "So what?"

There are many times when exhaustive reporting of statistical significance, listing the degrees of freedom, F value, significant level, eta value, and other statistics for one measure after another, including those that are not statistically significant, gets in the way of understanding what was found. Place those details in a table or in an appendix, or omit them. Those that are more statistically savvy than the author may disagree.

2.11 Guideline 11: Make sure the references are complete and follow the required format.

Problems in complying with this guideline occur when the manuscript is not written in the native language of the manuscript author. Manuscript authors follow the appropriate basic style, either the Harvard (author, year) style favored by psychological, educational, and many human factors publications, or the Vancouver (number reference style) favored by IEEE [30]. However, manuscript authors appear to ignore the specific format required (American Psychological Association, Modern Language Association, American Medical Association, University of Chicago, etc., http://www2.liu.edu/cwis/cwp/library/workshop/citation.htm). As with poor grammar, reviewers do not like spending time on such low-level corrections -- formatting references. Particularly troublesome is where references are formatted inconsistently, (e.g., some titles are in quotes and some are not, some journal names and book titles are underlined and some are italicized, author names are provided in multiple ways). Bibliographic software (e.g., Endnote) can easily resolve these problems.

3. CONCLUSIONS

This paper lists 11 guidelines to improve the ease of use and usefulness, and the overall quality of human factors publications on driving. Some of these guidelines are quite straightforward, such as writing as a native speaker, allocating page content to new material, checking reference formats, and providing pictures of stimuli and roads. Five guidelines, two of which are linked, are particularly important.

First, many papers, especially by novice authors (e.g., students) fail to clearly identify what the question or questions to be addressed are (for example, using the words who, what, when, where, why, and how). Further, even experienced authors often do not provide quantitative predictions of the expected outcomes.

Second, the focus on analysis needs to shift towards balancing the statistical and practical significance by emphasizing regression analysis and by reporting mean differences. What really matters, and by how much?

Third, much more needs to be done to translate research into practice. One mechanism is to add an entry --- relevant standards, guidelines, policies, and procedures -- below the existing keyword entry at the beginning of most papers.

Fourth, the author also strongly advocates that papers reporting driving studies must comply with SAE J2944, or provide definitions of equivalent detail.

Following these guidelines will improve the quality of research concerning human factors and driving. Most publications provide sufficient descriptions of the test method and proper tests of statistical significance. However, many publications fall short in terms of replicability and applicability, and as a consequence, are viewed by researchers in other fields of science and engineering, as not as good as they should be. These guidelines provide the direction needed to improve the quality of research on the human factors of driving.

4. REFERENCES

[1] Akamatsu, M., Green, P.A. and Bengler, K. (2013), 'Automotive technology and human factors research: Past, present, and future,' International Journal of Vehicular Technology, 2013, Article ID 749089.

[2] Green, P., 'Superman Solves a Puzzle,' handout for Industrial and Operations Engineering 436 and 437, University of Michigan, Department of Industrial and Operations Engineering, 2013.

[3] Green, P., 'Estimating compliance with the 15-second rule for driver-interface usability and safety,' Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1999, pp. 987-991.

[4] SAE draft Recommended Practice J2944, 'Operational definitions of driving performance measures and statistics,' 2014.

[5] Green, P., 'Using standards to improve the replicability and applicability of driver interface research,' Proceedings of Automotive UI 2012, 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH, October, 2012, pp. 15-22.

[6] Bordage, G., 'Reasons reviewers reject and accept manuscripts: the strengths and weaknesses in medical education reports,' Academic Medicine, 2001, 76, (9), pp. 889-896.

[7] Onwuegbuzie, A. J., & Daniel, L. G., 'Editorial: Evidence-based guidelines for publishing articles,' Research in the Schools, 2005,12, (2), pp. 1-11.

[8] Ehara, S., & Takahashi, K., 'Reasons for rejection of manuscripts submitted to AJR by international authors,' American Journal of Roentgenology, 2007, 188, (2), pp. W113-W116.

[9] Strunk, W., Jr., and White, E.G., 'The elements of style,' Allyn and Bacon, 1979.

[10] Engstrom, J., 'Scenario criticality determines the effect of working memory load on brake response time,' European Conference on Human Centred Design for Intelligent Systems, Berlin, Germany, April, 2010, pp. 25-35.

[11] SAE Information Report J2400, 'Human factors in forward collision warning systems: Operating characteristics and user interface requirements,' 2005.

[12] ISO Standard 15623, 'Transport information and control systems -- Forward vehicle collision warning systems -- Performance requirements and test procedures,' 2013.

[13] Green, P., 'Human factors and ergonomics in motor vehicle transportation,' in Salvendy, G. (ed.), 'Handbook of Human Factors and Ergonomics,' (Wiley, 2012, 4th edn.), pp. 1596-1614.

[14] Green, P., 'Motor vehicle driver interfaces," in Jacko, J. (Ed.), 'Handbook of Human-Computer Interaction,' (CRC Press, 2012, 3rd edn.), pp. 749-770.

[15] Jeong, H. and Green, P., 'SAE and ISO standards for warnings and other driver interfaces elements: A summary,' technical report, University of Michigan Transportation Research Institute, 2013.

[16] Green, P., 'Human factors/ergonomic standards,' webinar, March 20, 2012, http://www.hfes.org/web/webinars/marchgreen.html.

[17] American Association of State Highway and Transportation Officials (AASHTO), 'A policy on the geometric design of highways and streets' (AASHTO, 6th edn.), 2011.

[18] U.S. Department of Transportation, 'Manual of uniform traffic control devices,' 2009.

[19] Strand, N., Nilssson, J., Karlsson, I.C.M., and Nilsson, L, 'Exploring end-user experiences: Self-perceived notions on use of adaptive cruise control systems,' European Conference on Human Centred Design for Intelligent Systems, Berlin, Germany, April, 2010, pp. 487-496.

[20] Baue, A.E., 'Writing a good abstract is not abstract writing,' Archives of Surgery, 114, (1), 11.

[21] Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., and Ramsey, D.J., 'The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data,' technical report DOT HS 810 594, U.S. Department of Transportation, National Highway Traffic Safety Administration, 2006.

[22] Card, S. K., Moran, T. P., & Newell, A. (eds.), 'The psychology of human-computer interaction,' CRC Press, 1986.

[23] Nowakowski, C. and Green, P., 'Prediction of menu selection times parked and while driving using the SAE J2365 method,' technical report 2000-49, University of Michigan Transportation Research Institute, 2001.

[24] Kang, T-P., Lin, B.T-W., Green, P., Pettinato, S., and Best, A., 'Usability of a Hyundai-Kia generation 4 prototype navigation radio: Evidence from an occlusion experiment and SAE J2365 and Pettitt's method,' technical report UMTRI- 2013-11, University of Michigan Transportation Research Institute, 2013.

[25] SAE Recommended Practice J2396, 'Definitions and measures related to driver visual behavior using video techniques,' 2005.

[26] ISO Committee Draft Standard 15008, 'Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications and test procedures for in-vehicle visual presentation,' 2014.

[27] Savino, M.R., Standardized names and definitions for driving performance measures, masters thesis, Tufts University, 2009.

[28] Eoh, H., Green, P.A., Schweitzer, J. and Hegedus, E., 'Driving performance analysis of the ACAS FOT data and recommendations for a driving workload manager,' technical report UMTRI-2006-18, University of Michigan Transportation Research Institute, 2006.

[29] Whitehurst, H.O., 'Screening Designs Used to Estimate the Relative Effects of Display Factors on Dial Reading,' Human Factors, 1982, 24, (3), pp. 301-310.

[30] Sirkumar, B.N., 'Reference style and common errors with referencing,' in Srikumar, B.N, "Biomedical research: from ideation to publication,' (Wolters Kluwer, Lippincott Williams & Wilkins, 2010), pp. 479-488.