FLUORESCENT ORANGE SIGN SHEETING EVALUATION BY A STATE DOT

Mark J. Poppe, P.E., PTOE (Corresponding Author)
Assistant Traffic Engineer
City of Glendale, Arizona
5850 W. Glendale Avenue
Glendale, AZ 85301
Ph: (623) 930-2940
Fax: (623) 915-1029
Email: m.poppe@worldnet.att.net

Craig A. Roberts, P.E.
Assistant Professor
Department of Civil and Environmental Engineering
Northern Arizona University
PO Box 15600
Flagstaff, AZ 86011-5600
Ph: (928) 523-0651
Fax: (928) 523-8951
Email: craig.roberts@nau.edu

Seth W. Chalmers, P.E.
Principal
Chalmers Engineering Service, Inc.
1451 North El Camino Drive
Tempe, Arizona 85281
Ph. (480) 949-0925
Fax: (480) 946-7755
Email: swchalmers@juno.com

ABSTRACT

The following paper describes the first phase of the Arizona Department of Transportation (ADOT) fluorescent orange sign sheeting evaluation. The general goal of the evaluation was to develop a reasonable list of approved fluorescent orange sign sheeting products while ensuring the approved products will function adequately for the intended purpose. The evaluation used a human factors type test with an expert panel of test subjects. The evaluation resulted in the conditional approval of three fluorescent orange prismatic sign sheeting materials for use on rigid/semi-rigid substrate materials and four prismatic roll-up sign materials. The approval of the materials is conditional, pending a follow-up evaluation after one year of outdoor exposure.

BACKGROUND

A recent Arizona Transportation Research Center study *(1)* indicates that 3% of all reported accidents occur in work zones. The study also indicates that work zone accidents on the Arizona state system are over-represented when compared to all accidents statewide. While this study did not recommend any specific counter measures related to the type of work zone signing, it is obvious that advance signing is the first line of defense for warning and directing traffic safely through the work zone. It is also widely held that conspicuity of the warning sign is an important factor in driver recognition of warning signs. Hummer and Scheffler *(2)* found fluorescent orange signs in work zones caused primarily positive changes in driver behavior while Jenssen and Brekke's work *(3)* lead them to recommend the use of fluorescent signing to improve work zones.

In an effort to improve work zone safety, the Deputy State Engineer for Operations, the Assistant State Engineer for Construction, and the Assistant State Engineer for Traffic developed a policy for the use of fluorescent sheeting for all orange work zone traffic control signing. This policy is outlined in ADOT Traffic Engineering Policies, Guidelines and Procedures Section 300 *(4)*. The policy states fluorescent orange sheeting will be the standard for all work zone signs beginning January 1, 2002.

Fluorescent Sign Sheeting

Color perception is chiefly a matter of subtraction or absorption. When we say 'this sign is orange" what we are actually saying is the molecular constitution of its surface is such as to absorb all visible light rays but those in the orange range of 590 – 640 nm *(5)*. The electromagnetic energy in this range is then reflected back to the eye. The perception or discrimination of color is related to the psychophysical response to the various wavelengths of visible light.

Fluorescence is a term used to describe a material that can absorb electromagnetic energy at shorter wavelengths and re-emit this energy at a longer wavelength. With this conversion or shift of electromagnetic energy to longer wavelength light, a fluorescent orange sign could theoretically have an apparent reflectance of over 100%. Fluorescent materials appear to have the greatest relative visibility under low light conditions such as at twilight or during overcast conditions *(6)*. It is suspected this may be due to the quality of light and the relative amount of shorter wavelength visible light under these conditions. Consequently, the quality of the light source can be a significant factor in the evaluation of fluorescence.

Fluorescent sign material holds great promise to increase conspicuity based on the results of several studies *(7, 8, 9, 10 and 11)*. These studies address the effectiveness of fluorescent signs as opposed to standard colored signs. However, they generally investigate a limited number of material types and focus on the relative merit of using fluorescent versus non-fluorescent materials. They do not evaluate the effectiveness of fluorescence at various levels or attempt to quantify a measurement for fluorescence. They do not evaluate the "minimum" effective levels of fluorescence or compare a large number of available fluorescent materials. One exception is a study by Burns and Johnson of the 3M / Traffic Control Materials Division *(6)* which does set forth a metric. However, their physical measurement for fluorescence does not use a full array of materials at various levels of fluorescence. Additionally, their work does not address the minimum levels of fluorescence required to provide an acceptable level of conspicuity or an appropriate luminance contrast ratio. Therefore the issue of acceptable levels of fluorescence is not addressed.

National standards for retroreflective sign sheeting are set forth in the American Society for Testing and Materials (ASTM) standard specification D4956-01. ASTM D4956-01 establishes minimum retroreflectivity values and color limits for various types of sheeting. The color, or chromatic requirements established in ASTM D4956-01 are based on measurements made using a simulated idealized daylight source designated CIE (Commission Internationale De L'Eclairage) standard illuminant D65.

The unique nature of fluorescent colorant makes the current ASTM D4956-01 methods for measuring color a poor test for the evaluation of fluorescent sign sheeting. Unfortunately, at this time, there is no single, generally accepted method for measuring fluorescence. The ASTM D4.38 and E12 committees are currently working to establish an industry-wide accepted measure and appropriate values for the selected measure. Until this work is completed, transportation agencies wishing to specify fluorescent materials must look to other methods for determining which products are appropriate for use on public roadways.

GOAL AND OBJECTIVES OF THE EVALUATION

Open and competitive bidding is a fundamental guiding principle of public procurement. As such, it is essential that public agencies develop specifications which conform to this principle unless there is compelling evidence to suggest that restrictions will better serve the public through greater safety or long term cost savings.

> *The general goal of this evaluation was to develop a reasonable list of approved fluorescent orange sign sheeting products while ensuring the approved products will function adequately for the intended purpose.*

Given the fact there is no consensus on the appropriate physical metric for the evaluation of fluorescent sign sheeting, it seemed appropriate that ADOT pursue strictly a human perception based evaluation of the available products. ADOT solicited sample materials from all known manufacturers of fluorescent orange signing materials. A sample of commonly used non-fluorescent (standard) orange sign sheeting materials were also obtained. A list of test materials used in the evaluation is presented in Table 1. The approval process was based on human perception and evaluation of the products. Therefore, the net result of the evaluation is a multiple brand name specification. It should be noted the evaluation is a snap shot of what the market had to offer, at the time of the evaluation, in the area of fluorescent orange sign sheeting materials. The ADOT goal was not to establish a physical metric for the evaluation of fluorescent sign sheeting materials. Nor was the goal to correlate the psychological response to some physical measure. Hopefully, the ASTM committees working on these issues will address these needs before there are a substantial number of new fluorescent orange signing materials on the market.

In order to achieve the stated goal three specific objectives were developed:

1. To identify a minimum of three approved manufacturers for fluorescent orange retroreflective prismatic materials for use on rigid/semi-rigid substrates.
2. To identify a minimum of two approved manufacturers for fluorescent orange retroreflective prismatic materials for use as a roll-up sign material.
3. To determine how fluorescent orange retroreflective SEG (super engineering grade) sign sheeting and other non-reflective sign sheeting materials compare to the prismatic materials under day time viewing conditions.

The information obtained as part of the third objective may be used to determine if lower cost products may be viable products under certain circumstances.

EVALUATION TEST PROCEDURE

The evaluation procedure was composed of two separate viewing techniques. The first viewing was based on the review and evaluation of each test sign individually, with no simultaneous comparison with other signs. The second viewing was based on a viewing of all signs at the same time for a relative ranking of all signs. This procedure will be completed a total of four times. The first phase of the study included two replications that were performed on December 6, 2000. One occurred under full day light conditions. The second under twilight conditions. The midday and twilight viewing procedures will be repeated under phase two of the study, after the signs have been exposed to outdoor weathering for approximately one year.

The signs were positioned in a 3x5 array as shown in Figure 1. As stated previously, fourteen types of materials were tested. One test material from the family of fluorescent orange prismatic for rigid substrate materials was selected at random as a duplicate material to fill the matrix. This also served as a check on the repeatability of the scoring. The sign positions within the matrix were randomly assigned. The sign positions were also reassigned after the midday viewing, in an attempt to mitigate the influence of neighbors in the array and viewing order. All test signs were 24 inch by 24 inch, W13-1 warning sign supplemental plates as shown in Figure 2. The "35" was 10 inch type E legend. The "mph" was 4 inch type E legend. The research team received input from various manufacturer representatives during the development of the evaluation procedure. A preliminary test procedure was first prepared. The research team then performed a pretest based on the preliminary procedure. The procedure was then refined as described below.

Repeated Measures Design

The following is a general overview of the experimental design and evaluation. A detailed discussion of the study is contained in an Arizona Transportation Research Center final report *(12)*. A repeated measures design was developed for the experiment (*13* and *14*). This design is used in experimental work where the experimental units are people, typically called *subjects* in classic literature. It takes its name because the subject is measured repeatedly. Because of differences in experience, training, or background, the differences in responses of different subjects to the same treatment may be very large in some experimental situations. This variability is controlled by using a design where all treatments are evaluated by each person. This analysis is equivalent to an analysis for a randomized complete block design, with the subjects considered to be the blocks. These evaluations assume that there is no interaction between the blocking variable (subjects) and the response variable (evaluation rankings and ratings).

The response variables in this experiment will be ranked and rated data. There are two ways to analyze ranked and rated data. One uses nonparametric analysis methods that don't rely on the assumption that the data is normally distributed. Nonparametric methods are not available for all situations, however. When nonparametric methods do not exist, the other way to is to simply use the standard methods of analysis of variance (ANOVA) on ranked data. This method is not theoretically pure but is supported by Conover (*15*), Montgomery (*13*) and Neter et al (*14*).

Questions were kept very short. The scale for each question was identical. A seven point scale was used, which is generally adequate for surveys using ordered responses (*16*). The middle point was assigned neutral position. The one-, three-, five-, and seven-points were anchored. The "adequate" anchor was assigned to the three-point.

One Sign at a Time Review

Each test sign was reviewed one sign at a time. The test subjects were asked to rate each sign on five categories:

1. Color
2. Large Legend Legibility
3. Small Legend Legibility
4. Conspicuity
5. Fluorescence

The instructions to the test subjects included a brief description of these categories, as follows.

Color evaluation is based on an assessment of how "orange" the sign appears based on the viewers notion of what type of orange should be used for work zone construction signs. Is the sign recognized as being primarily orange or primarily some other color?

Legibility evaluation is based on an assessment of how easily the sign may be read and how clean or crisp the legend appears. Each sign contains identical text with two different legend sizes on each sign. The viewer is asked to assess both sizes independently, recognizing that the larger legend should be more legible than the smaller legend.

Conspicuity evaluation is based on an assessment of how well the sign stands out from the background.

Fluorescence evaluation is based on an assessment of how much the color seems to radiate.

The test subjects were asked to rate each sign on a scale of one to seven with one being the highest or best score and seven being the lowest score. The instructions to the test subjects included a brief description of these ratings, as follows.

A *one (1)* rating means the sign is *very good*, clearly effective and suitable for use.
A three (3) rating means the sign is adequate, effective and suitable for use.
A five (5) rating means the sign is marginal, questionably effective and suitable for use. Generally speaking, a marginal sign would only be used if there were no adequate alternatives.
A seven (7) rating means the sign is poor and would not be considered suitable under these viewing conditions.

The subjects were asked to provide brief comments, when appropriate, in order to determine the reasoning for the rating.

All Signs at the Same Time Review

All signs were reviewed simultaneously. The test subjects were asked to rank the overall effectiveness of each sign. The ranking was based on one (1) being the best or highest ranking and fifteen (15) the lowest. The test subjects were instructed that the overall effectiveness of the sign should take into consideration all the factors previously used in the one-sign-at-a-time review: color, legibility, and visual impact (conspicuity & fluorescence).

It should be noted that there is a distinction between the physical phenomenon of fluorescence that takes place at the surface of a material and the unique sensory experience that occurs when viewing a fluorescent colored material. Scheiber (17) introduces the term “fluorence” to describe this physiological analogue of fluorescence. While the authors are not always clear with respect to this distinction, it is clear that what is actually being “measured” in this evaluation is the human experience rather than the physical phenomenon of fluorescence.

Selection of Panel

A useful panel could be drawn from either the motoring public or from experts in the field. Both approaches had appeal, but the expert panel was chosen. This approach allowed the evaluation of the technical characteristics of sign sheeting materials believed important in a work zone without having to either (a) educate panel members in how to evaluate these technical characteristics (e.g., *conspicuity*) in a fashion consistent with other signing materials or (b) devise “motoring public” evaluation methods that could then be mapped to these technical factors. The available testing budget also limited the resources, both time and money, that could be expended. For these reasons, the panel was drawn from the attendees of the December 2000 quarterly meeting of the ADOT Traffic Control Products Evaluation Committee (TCPEC). All panel members were current or past members of the TCPEC. Therefore, the panel members had a good working knowledge of roadway signing and sign sheeting materials. Additionally they could be reassembled for the longitudinal needs of the experiment with the expectation that there would be no dropout among the panel members. Clearly, the selection of this panel is not representative of the motoring public. Its composition is, however, well suited to the goal of establishing an approved products list.

DATA COLLECTION

The test panel was composed of ten experts. All subjects were males between 38 and 50 years of age. The mean age of the panel was 45.8 years. All test panel members reported normal color vision. The panel members were given a copy of the work plan well in advance of the test so they had an opportunity to familiarize themselves with the test procedure. The panel subjects were briefed as a group in a conference room prior to walking to the adjacent test rack area. While panel subjects were aware of which products were being tested, they were not told which sample numbers were assigned to the various materials. Panel members were also told there was one redundant material.

The test rack was located in the flat asphalt yard of an ADOT facility near the center of Phoenix. The panel made an evaluation at two different times on the initial test day, December 6, 2000. A large blank aluminum-colored sign was placed behind and above the test rack, which shielded the signs from backlighting glare. The background of the signs was cluttered. The subjects were placed in a line perpendicular to the face of the test rack. The axis of viewing was east-west, with the panel subjects facing west. The first evaluation period was between 1:00 and 1:30 p.m. during clear, sunny conditions. The sun was directly overhead on the east-west axis but slightly south of directly overhead on the north-south axis. The second evaluation period was between 4:45 and 5:15 p.m. during slightly overcast, twilight conditions. During the twilight evaluation, the sun was behind the test rack but low enough to be blocked by buildings in the background. A distance of 250 feet was selected for viewing based on results of the pretest. This resulted in a legibility index of one inch of text height per 25 feet of viewing distance for the large legend and one inch per 62.5 feet for the small legend.

The first technique viewed each sign individually. The signs were located within the test rack matrix based on a random selection. A single sign was uncovered (see Figure 3). In order to hold the composition of the light constant for all subjects, it was necessary to have all subjects evaluate each sign simultaneously, but independently. Each subject evaluated the sign on the five categories: color, large legend, small legend, conspicuity, and fluorescence. Once the evaluations were done, the sign was covered and another sign was uncovered and evaluated. Sixty seconds per sign was given to evaluate all the questions, but the total time was never needed.

The second technique viewed all the signs simultaneously and was done last (see Figure 4). The positions of the signs on the test rack remained the same as during the first viewing technique. Each subject was given a map of the sign layout and asked to rank the signs in order of best performance, with 1 being best and 15 being worst (no ties). They were asked to "*Please take into consideration: overall effectiveness (color, legibility and conspicuity).*"

The panel's second evaluation at twilight was done using exactly the same procedures as the midday evaluation. However, the location of the signs was changed within the test rack matrix, based on random selection.

Two follow-up demonstration tests were also performed, one for sign manufacturer representatives and a second for ADOT upper management. Both demonstrations were performed in January 2001. The demonstration test data were compiled and released to the demonstration test panelists along with the results from the official test panel. The primary purpose of the demonstration tests was to familiarize the manufacturers and top level decision-makers with the evaluation procedure prior to release of the results. While the demonstration test data were not included in the official data set used to establish the approved list, the results were used to verify the repeatability of the experiment.

After the data collection, the signs were left uncovered on the test rack, fully exposed to all the elements. They were oriented vertical south-facing. The longitudinal aspect of the experiment had initially placed the second set of midday and twilight evaluations at age 3 months. However just prior to reaching this time point, a thorough examination all the signs was unable to detect any noticeable deterioration. One exception was a sample material that was not rated for long-term outdoor exposure. Therefore, the next evaluations were delayed. Little noticeable deterioration was detected after six months except, again, in the short-term exposure rated material. Based on this, the next evaluation was scheduled at age 1 year and will be conducted at that time.

EVALUTION OF DATA

The raw data from each viewing were entered into spreadsheet workbooks. The One-Sign-at-a-Time review scores were reduced to mean scores for each rating category: color, large legend, small legend, conspicuity, and fluorescence. A composite rating was also calculated based on a mean taken across all five categories. The mean values for the One-Sign-at-a-Time ratings are shown in Table 2 and Table 3. The All-Signs-at-the-Same-Time review scores were reduced to a mean score for each sign. The mean scores for the midday and twilight reviews are shown in Table 4.

The following terms and abbreviations are used in the analysis description, tables, and figures.

Sign	=	factor name for the variable of different manufacturer's sign samples.
S1	=	one of the 15 levels of the *Sign* factor, i.e., one of the 15 sign samples *S1…S15*.
TOD	=	factor name for the variable of the times an evaluation was done on the same day.
noon	=	one of 2 factor levels of the *TOD* factor.
dusk	=	one of the 2 factor levels of the *TOD* factor.
Block	=	factor name given to subjects, which was used as a blocking variable.
R1	=	one of the 10 factor levels of the *Block* factor, i.e., one of the 10 subjects *R1…R10*.
Fluor	=	response variable for the characteristic "fluorescence."
Conspicuity	=	response variable for the characteristic "conspicuity."
Color	=	response variable for the characteristic "color."
Small.Legend	=	response variable for the characteristic "legibility" for the smaller legend on the sign sample.

Large.Legend = response variable for the characteristic "legibility" for the larger legend on the sign sample.
Composite = response variable for the mathematical mean of the measured response variables: *Fluor*, *Conspicuity*, *Color*, *Small.Legend*, and *Large.Legend*.
Ranking = response variable for the ranking of all signs against each other for "overall effectiveness."

Analysis of Variance (ANOVA)

Estimating Models for Fluor

An ANOVA model was estimated with *Fluor* as the response variable, *Sign* and *TOD* as the factors, and an interaction term of *Sign*:*TOD*. This design is a repeated measures on both factors and is analyzed as a two factorial design with replicates that are treated as blocks (*13* and *14*). The blocks are the subjects within the expert panel and each subject's evaluation is a replicate. It is a complete block design because each block (subject) contains all treatments. Such designs assume that treatments and blocks (subjects) do not interact. The results of this Model A are listed in Table 5 and indicate the *Sign* factor and the interaction term *Sign*:*TOD* are both significant at less than the 0.01 alpha level while the *TOD* factor is not significant, having a *p-value* of 0.23.

TOD not being significant simply indicates that the difference between the mean ratings of all signs at *noon* (3.0733) and at *dusk* (2.9733) cannot be detected statistically with only one degree of freedom. While this difference would be a crucial factor to explore in experiments aimed at the general effects of fluorescent sign materials, it is not of much interest in this experiment where the specific fluorescent qualities of individual signs is the subject of interest. For our purposes, the interaction between the *Sign* and *TOD* factors is much more interesting. It was expected that the mean rating of a *Sign* would change somewhat under the different *TOD* light conditions of *noon* and *dusk* due the nature of fluorescence materials reactions to different types of light. However, all things being equal, we would expect these shifts to be similar for all signs, which would not make the interaction term significant.

One reason the interaction term might be significant is due to the four signs that were made of non-fluorescent materials (*S3, S6, S8*, and *S9)*. A review of the interaction plot in Figure 5 is quite revealing. At first glance, one can readily see the signs are segregated into two groups. Even more important to our pass/fail criteria, these groups are completely on one side or the other of the 3.0 rating. Recall that the mean rating on fluorescence must be "satisfactory" or better, which translates into a <u>3.0 or less value</u>. Although not discernible on the plot, a review of the underlying data reveals that only four of the signs received a higher *Fluor* rating at *dusk* than at *noon*: *S4, S6, S8*, and *S9*. Since three of these counter-trend signs are non-fluorescent signs, it appears that all the fluorescent signs but one behaved in the same manner. It is worth noting that this behavior was expected since fluorescence is purported to have greatest relative visibility under low light conditions.

When the interaction is significant the main effects are essentially rendered meaningless, since any main effect will have to be qualified in reference to the second factor. Typically in the presence of a strong interaction, it is a logical strategy to test for *simple effects.* A test of a simple effect is essentially an analysis of variance evaluating all levels of one factor across only one level of the other factor. Recall, however that it appears that most of the interaction effect is caused by the presence in the test group of the 4 non-fluorescent signs. And in fact, if you remove these signs and estimate an ANOVA model (see Table 6), the interaction term becomes non-significant and both the *Sign* and the *TOD* effects become significant. The difference across levels of *Signs* is sufficiently large and across levels of *TOD* is sufficiently small in both models to render them effectively identical. Results from both models regarding the pass/fail criteria are identical. The full model with the non-fluorescent signs included for comparison purposes is more interesting, so that is what is reported here, not only for *Fluor* but also for the *Composite* and *Ranking* response variables.

Since the model effects are significant, multiple comparisons among the *Signs* can be made to determine the final pass/fail status of each sign individually. Rather than perform pair-wise comparisons, upper bound confidence intervals were constructed for each sign mean rating. Many methods can be used to construct the confidence intervals. The ones reported here used a simulation-based method and a family-wise error rate, so that the probability that all bounds hold is at least 1-alpha. This creates relatively small intervals but it is overkill as the data is such that the intervals constructed under the most conservative methods still yield identical results regarding the pass/fail boundary of a mean rating of 3.0. The results, shown in Figure 6, indicate that signs *S2, S3, S6, S8, S9*, and *S15* fail with mean *Fluor* ratings greater than 3.0.

Estimating Models for Composite

The same repeated measures on both factors ANOVA model was estimated for *Composite* as the response variable as was done for *Fluor*. The models are so similar in all respects that the discussion of the *Composite* model would be the same as already described for the *Fluor* model. Again multiple comparisons among *Signs* was made to determine the final pass/fail status of each sign individually as regards the *Composite* rating. The intervals were again constructed using the same method. The results, shown in Figure 7, indicate that signs *S2, S3, S6, S8, S9,* and *S15* fail with mean *Composite* ratings greater than 3.0. These are the same subset that failed the *Fluor* pass/fail evaluation.

Estimating Model for Ranking

The *Fluor* and *Composite* models were all that were needed to make the final pass/fail determination for each sample sign. However, a model with *Ranking* as the dependent variable was estimated for comparison purposes, and was designed to provide some validation to the process. This model is different from the two previous models in that it uses ranked values instead of rated values. This model has some peculiarities due to the use of rankings, but again evaluates quite similarly to the previous two models using rating response variables. The plot of interest is of the two-tailed confidence intervals around the *Ranking* means for each *Sign*. These were again done using a simulation-based method and a family-wise error rate, so that the probability that all bounds hold is at least 1-alpha. The results, shown in Figure 8, are gathered into groupings of "overlapping" confidence intervals. Sign rankings within one of these groups cannot be statistically distinguished from any other sign in the same group. Recall that a ranking of 1 is the best and 15 is the worst. The location of the signs that failed the *Fluor* and *Composite* ratings test, *S2, S3, S6, S8, S9,* and *S15*, appears to validate their failed evaluations.

Evaluation of Model Assumptions

Departures from ANOVA Repeated Measures Model Assumptions

A typical panel study uses a random sample of subjects drawn from the population of interest. This allows the effects of subjects to be viewed as random. In our case, experts were deemed to be the population of interest. The selection of the panel, however, was not taken randomly from all such experts. This departure from classic repeated measures design is justified based on resource limitations and the intended use of the results. Obviously the results of such a panel could not be generalized to the entire population of experts. However, the goal was to use experts intimately familiar with ADOT signing materials and the technical factors that were used in their selection. The panel drawn is believed to be representative of this group, although not drawn in a random fashion. Therefore this departure from underlying assumptions of a repeated measures design is not believed to unduly bias the experimental results for their intended purpose. Other parties, however, are cautioned that these results are most appropriate for ADOT transportation facilities for which the experts do have a strong experiential bias and may not be useful for other agencies facilities.

Ideally *order effect* interference is minimized by randomizing the order of treatments for <u>each</u> subject. Because of the desire to use natural lighting for the signs coupled with the need to hold lighting conditions constant for all subjects, it was too costly to randomize treatment order for each subject. Instead all subjects viewed the signs simultaneously, which means the results are subject to order effect interference. Having decided to accept this deficiency in the experimental design, this interference was, however, minimized to the degree possible over time. This was done by randomizing the presentation of treatments whenever the subjects viewed them at a different time.

Randomizing the order of treatments for each subject would also reduce *carryover effect*. For the reasons already stated, this was not done. Other steps could be taken, if deemed necessary, whereby the order of treatment presentations or even the number of times each treatment is preceded by any treatment is balanced. None of these were employed because it was deemed adequate to simply allow sufficient time between presentations of the treatments to reduce carryover effect.

Evaluation of the Appropriateness of the Estimated Models

The repeated measures models need to be checked for potential problems with the normality assumption, unequal error variance by treatment or block (subject), and block-treatment (subject-treatment) interaction. It is assumed in a repeated measures design that there is no subject treatment interaction, which is usually a reasonable assumption. This assumption was checked by a plot of responses by subjects, which exhibited strong parallelism thereby supporting the lack of subject-treatment interaction. Diagnostic plots of residuals were checked and generally support the assumption of equal variance across treatments and blocks. Finally a normal probability plot indicated a reasonably normal distribution of the error terms. These assumptions were upheld reasonably well for both the *Fluor* and *Composite* models. While the *Ranking* model was more problematic in theoretical underpinnings, diagnostic plots appear to support its assumptions and its “forced” differences in responses among *Sign* levels gave similar results to the other models.

As previously discussed for the *Fluor* model, the estimated models have a general lack of parallelism in the plot of responses by treatments, i.e., a statistically significant interaction term between the factors *Sign* and *TOD*. However, also as previously discussed, when a model was estimated which removed the non-fluorescent signs from the experiment, this problem was also removed giving much “cleaner” models. However, the models for *Fluor*, *Composite,* and *Ranking* without the non-fluorescent signs had identical results as those when all the signs were present. Therefore the models reported here contain all 15 sample signs, which allowed direct comparisons among fluorescent and non-fluorescent signs.

FINDINGS

Nine blind samples passed the pass/fail criteria. One of these was not retroreflective nor rated for long-term outdoor exposure. This material was included in the experiment for comparison purposes. One of the samples was a duplicate of another sample. Therefore ADOT approved seven sheeting materials (see Table 7), all of which were purported to be fluorescent orange materials by their respective manufacturers. This list served to fulfill the requirements of the first two objectives of the evaluation:

1. Three approved manufacturers for fluorescent orange retroreflective prismatic materials for use on rigid/semi-rigid substrate materials were identified.

2. Two approved manufacturers (four materials) for fluorescent orange retroreflective prismatic materials for use as a roll-up sign material were identified.

Six sign sheeting materials failed the pass/fail criteria. Of these, two were purported to be a fluorescent orange material by their respective manufacturers. This result serves to fulfill a portion of the third objective. The single fluorescent orange SEG material did not compare well with the approved materials, under the daytime viewing conditions of this evaluation. The single fluorescent orange prismatic material that failed is also the only material in this class with a metalized backing. This would seem to indicate that a metalized backing has a negative impact on the daytime color of fluorescent colorants. The single non-retroreflective material did compare well with the approved materials. This finding would seem to indicate that non-retroreflective fluorescent orange materials may be as effective as the retroreflective materials, under daytime viewing conditions. However, the color fastness of such materials may be a concern. The other four materials that failed were various types of non-fluorescent standard orange sign sheeting materials. This finding would be expected and tends to support the general evaluation procedure.

DISCUSSION

The results of the experiment fulfilled the charge to the TCPEC. While the use of a panel study is common in social and behavioral sciences including human factors analysis, it is avoided in materials science in lieu of physical measurements that are reproducible. However, at this time there is no single, generally accepted method within the sign sheeting manufacturing community for evaluating fluorescence. ADOT believed this material would provide benefits in work zone areas and adopted a policy requiring its use by January 1, 2002. It charged its Traffic Control Products Evaluation Committee to devise a solution that would (a) comply with public procurement policies, ideally though competitive bidding and (b) ensure the approved products will function adequately for the intended purpose. This was accomplished through the design and conduct of a repeated measures experiment.

ADOT implemented the results of the evaluation by adding those materials that were approved to the ADOT Approved Products List. Each material, however, was only given a “conditional approval.” The approval was conditioned on a product continuing to meet the pass/fail criteria at each time point in the longitudinal study. A second panel evaluation will be conducted at age 12 months. Based on the apparent rate of deterioration at that time, the next time point for evaluations will be established. The same panel of experts will be used to avoid introducing noise in the data that might confound future results. The continued availability of all these panel members is one of the primary reasons for using an expert panel closely associated with the TCPEC.

Areas for Further Study

The evaluation of materials by means of human factors testing as outlined in this paper is very useful in the absence of a physical metric for the evaluation of the materials. Such evaluations are most useful and long lasting when the population of available materials is static over a long period of time. This is not the case with fluorescent sign sheeting. ADOT has already identified three new fluorescent orange sign sheeting materials that should be evaluated in the near future for consideration as additions to the ADOT Approved Products List. There is also an increasing number of fluorescent yellow, yellow-green, and red sign sheeting materials. This strongly points to the need for additional research as proposed by Schieber *(17)* to develop an acceptable metric for the evaluation of fluorescent colors.

To be considered adequate, a sample had to receive a rating of 3.0 or less under a forced rating of all materials. This does not necessarily indicate the adequacy of the material in providing a message to the motoring public. Further research is needed on the human response to various levels of fluorescence under actual driving conditions to assess appropriate minimum levels of fluorescence.

The categories of conspicuity and fluorescence, while defined differently, have a large overlapping component. Therefore, one would expect a strong correlation between the ratings for these two categories. Both parametric and nonparametric tests confirm that a relationship exists between these two variables, with Pearson's product moment correlation coefficient of 0.88. Even though all the signs were nominally orange, there was a greater difference in the color of the various signs than initially expected by the research team. The differences are much more pronounced than what can be demonstrated using digital photograph (see Figure 4). So color may also contribute to the conspicuity rating (color and conspicuity have a Pearson's product moment correlation coefficient of 0.82). It is also difficult to clearly separate the color component from the fluorescent characteristics, so there is this further potential interaction between the ratings in the various categories (fluorescence and color have a Pearson's product moment correlation coefficient of 0.83). It is not possible to determine the extent to which color and fluorescence impact the ratings in other categories, based on the design of the experiment used in this research. This is a possible area for further study.

ACKNOWLEDGEMENTS

This work was funded by the Arizona Department of Transportation and directly supported by the Traffic Control Products Evaluation Committee, under the direction of Mike Manthey, State Traffic Engineer. Samples were provided by Avery Dennison, 3M Company, Nippon Carbide Industries, and Reflexite Corporation. The authors thank Ken Uding and Dennis Couzin (Avery Dennison), Sue Chrysler (3M), George Tamayori (Nippon Carbide), and Chris Gaudette (Reflexite) for providing technical information and input during the development of the evaluation procedure. ADOT Sign Factory personnel were critical to the research, particularly Tom Huey, Steve Slowik, and Rick Price. The research team would also like to thank Chris Williams and Raul Amavisca who assisted in the data collection.

REFERENCES

1. Walker, V., J. Upchurch, Effective Countermeasures to Reduce Accidents in Work Zones, Report FHWA-AZ99-467, Arizona Transportation Research Center, 1999.
2. Hummer, J. E. and C. R. Scheffler. Driver Performance Comparison of Fluorescent Orange to Standard Orange Work Zone Traffic Signs. In *Transportation Research Record 1657*, TRB, National Research Council, Washington, D.C., 1999, pp. 55-62.
3. Jenssen, G. D., and B. A. Brekke. Study Of The Visibility And Effectiveness Of Fluorescent Retroreflective Traffic Control Signing In Work Zone Applications. Presented at the Traffic Safety on Two Continents Conference at Lisbon, Spain, 1998. Sponsored by Swedish National Road and Transport Research Institute, Transportation Research Board, Forum of European Road Safety Research Institute, and Laboratorio Nacional de Engenharia Civil. Full report S-581 95 Linkoping available from Swedish National Road and Transport Research Institute.
4. ADOT Traffic Engineering Policies, Guidelines and Procedures, Section 300 -- Signs, p. 380-382, May 2001.
5. Itten, J. *The Elements of Color*, Van Norstrand Reinhold Company, New York, NY, 1970.
6. Burns, D. M. and N. L. Johnson. The Correlation of Measured Spectral Radiance of Fluorescent and Non-fluorescent Materials to Perceived Conspicuity under Natural Lighting. Paper presented at the AIC Interim Meeting on Colorimetry, Berlin, Sept. 3-6, 1995 also published in *Die Farbe*, Vol. 43, No. 4-6, 1997, pp. 185-203.
7. De Vos, A. P., J. W. A. M Alferdinck, and P.J. Bakker. Fluorescent Retroreflective Signing of Work Zones. Report TM-98-C044, TNO Human Factors Research Institute, Soesterberg, The Netherlands, August 1998.
8. Dutt, N., J. E. Hummer, and K. L. Clark. User Preference for Fluorescent Strong Ye llow-Green Pedestrian Crossing Signs. In *Transportation Research Record 1605*, TRB, National Research Council, Washington, D.C., 1997, pp. 17-40.
9. Dutt, N., J. E. Hummer, K. L. Clark, and S. Blakely. Controlled Environment Evaluation of Fluorescent Strong Yellow-Green Pedestrian Crossing Signs. In *Transportation Research Record 1553*, TRB, National Research Council, Washington, D.C., 1996, pp. 86-94.
10. Schnell, T., P. J. Ohme, K.F. Gulyuva, C. Donaubauer, E. Wiese, E. Derby, and D. Noelting. Driver Looking Behavior in School Zones with Fluorescent Yellow Green and Normal Yellow Signs. Paper presented at the 80th Annual Meeting of the Transportation Research Board, National Research Council, Washington D.C., 2001.
11. Schnell, T., K. Bentley, E. Hayes, and M. Rick. Legibility Distances of Fluorescent Traffic Signs and their Normal Color Counterparts. Paper presented at the 80th Annual Meeting of the Transportation Research Board, National Research Council, Washington D.C., 2001.
12. Poppe, M. J., C. A. Roberts, and S. W. Chalmers. *Fluorescent Orange Sign Sheeting Evaluation*. Report FHWA-AZ-01-116(1). Arizona Department of Transportation, Arizona Transportation Research Center, 2001.
13. Montgomery, D. C. *Design and Analysis of Experiments, Fourth Edition*. John Wiley & Sons, Inc., New York, 1997.
14. Neter, J., M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied Linear Statistical Models, Fourth Edition*. Irwin, Chicago, 1996.
15. Conover, W. J. *Practical Nonparametric Statistics, Third Edition*. John Wiley & Sons, Inc., New York, 1999.
16. Fink, A. *How to Ask Survey Questions*. Sage Publications, Thousand Oaks, CA, 1995.
17. Schieber, F. Modeling the Appearance of Fluorescent Colors, Research Paper Proposal Submitted to the Visual Performance Technical program Human Factors and Ergonomics Society 2001 Annual Meeting, Heimstra Human Factors Laboratories, Unversity of South Dakota, Vermillion, SD, 2001.

List of Tables and Figures

TABLES

Table 1 Test Sample Sign Sheeting Materials

Manufacturer	Brand	Type
3M	Diamond Grade 3924 F/G	Fluor. Prismatic Rigid ASTM Type VII
3M	High Intensity Grade 3864	Std. ASTM Type III
3M	Engineer Grade CW80	Std. ASTM Type I
3M	RS34 F	Fluor. Prismatic Roll-up ASTM Type VI
3M	Diamond Grade RS24	Fluor. Prismatic Roll-up ASTM Type VI
Avery Dennison	SL3005	Fluor. Non-retroreflective Vinyl
Avery Dennison	T-6504	Std. Prismatic ASTM Type III/IV
Avery Dennison	W-7514	Fluor. Prismatic Rigid Type VIII
Nippon Carbide	Crystal Grade 92547	Fluor. Prismatic Rigid
Nippon Carbide	Fluor. Super Egr. Grade 15947	Flour. ASTM Type II
Nippon Carbide	Super Engineer Grade 15037	Std. ASTM Type II
Reflexite	Marathon 3932068373	Fluor. Prismatic Roll-up ASTM Type VI
Reflexite	Super Bright 3932459374	Fluor. Prismatic Roll-up ASTM Type VI
Reflexite	Endurance 3961072373	Fluor. Prismatic Rigid Type Unspecified

Table 2 Midday One-Sign-at-a-Time Ratings

	Color		Large Legend		Small Legend		Conspicuity		Fluorescence		Composite	
Sample	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
S1	1.6	0.70	1.8	0.63	2.1	0.88	1.6	0.52	1.4	0.52	1.70	0.44
S2	4.1	1.20	3.2	1.14	3.3	0.95	3.8	0.79	4.7	1.06	3.82	0.89
S3	4.7	1.25	3.9	1.10	4.7	1.25	4.2	1.14	5.5	1.08	4.60	0.95
S4	1.8	0.63	2.2	0.63	2.5	0.71	1.8	0.63	1.8	0.42	2.02	0.48
S5	2.4	1.07	1.6	0.70	2.2	0.79	1.6	0.52	1.4	0.52	1.84	0.53
S6	3.4	0.97	3.2	0.79	3.7	0.48	3.5	0.53	4.4	0.84	3.64	0.53
S7	1.8	0.92	1.7	0.48	2.4	0.70	1.8	0.63	1.8	0.42	1.90	0.44
S8	5.1	1.20	2.8	1.23	3.7	0.95	4.1	1.10	5.4	0.84	4.22	0.85
S9	3.8	1.14	2.8	1.23	3.3	0.95	3.4	0.84	4.8	1.03	3.62	0.80
S10	1.8	0.63	1.9	0.74	2.2	0.79	1.8	0.63	1.4	0.52	1.82	0.48
S11	2.4	0.84	2.1	0.74	2.6	0.84	2.1	0.99	2.7	0.67	2.38	0.67
S12	2.0	0.94	2.6	0.97	2.9	1.10	2.2	0.79	2.3	0.67	2.40	0.73
S13	1.9	0.99	2.5	1.08	2.6	0.84	2.3	0.48	2.4	0.52	2.34	0.60
S14	1.3	0.48	1.8	0.79	2.5	0.85	1.7	0.48	1.5	0.53	1.76	0.43
S15	2.9	1.10	2.3	0.95	3.1	1.20	3.2	1.32	4.6	1.07	3.22	0.96

Table 3 Twilight One-Sign-at-a-Time Ratings

	Color		Large Legend		Small Legend		Conspicuity		Fluorescence		Composite	
Sample	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
S1	1.1	0.32	1.3	0.67	1.5	0.71	1.1	0.32	1.0	0.00	1.20	0.30
S2	3.5	1.08	2.7	0.67	3.2	1.14	3.2	1.03	4.0	1.15	3.32	0.91
S3	3.9	1.37	3.5	1.08	4.3	1.16	4.3	0.67	5.4	1.07	4.28	0.88
S4	1.9	0.74	2.0	0.67	2.3	0.48	2.1	0.57	2.3	0.48	2.12	0.48
S5	1.9	0.99	1.2	0.42	1.4	0.52	1.3	0.48	1.0	0.00	1.36	0.36
S6	4.2	1.32	3.3	1.06	4.1	1.10	4.1	0.74	5.4	1.26	4.22	0.89
S7	1.1	0.32	1.6	0.70	2.1	0.57	1.4	0.52	1.1	0.32	1.46	0.31
S8	5.1	0.88	3.5	1.27	4.1	1.37	4.6	0.84	6.1	0.88	4.68	0.83
S9	4.7	1.25	3.8	1.03	4.0	1.25	4.6	0.97	5.7	1.06	4.56	0.90
S10	1.3	0.48	1.5	0.53	1.9	0.74	1.2	0.42	1.1	0.32	1.40	0.35
S11	2.1	0.57	1.9	0.57	2.5	0.71	2.2	0.79	2.3	0.82	2.20	0.62
S12	1.4	0.52	1.6	0.70	2.5	0.97	1.7	0.67	1.4	0.52	1.72	0.56
S13	2.0	0.67	2.1	0.74	2.6	0.97	1.9	0.74	2.1	0.74	2.14	0.60
S14	1.5	0.53	1.5	0.53	1.9	0.74	1.4	0.52	1.3	0.48	1.52	0.40
S15	3.2	1.23	2.6	0.97	3.0	1.05	3.4	0.97	4.4	1.51	3.32	0.98

Table 4 All-Signs-at-the Same-Time Ranks

	Midday		Twilight	
Sample	Mean	S.D.	Mean	S.D.
S1	1.4	0.70	1.5	0.53
S2	12.3	1.57	12.5	1.84
S3	12.5	2.32	12.4	2.32
S4	6.0	1.25	6.3	0.82
S5	3.7	2.75	2.6	2.17
S6	11.6	1.96	12.7	1.49
S7	3.4	1.17	4.2	0.79
S8	13.9	0.99	14.4	0.84
S9	12.3	1.06	12.2	1.23
S10	5.0	1.41	4.5	0.85
S11	9.0	1.25	9.0	1.25
S12	6.2	1.14	6.3	1.06
S13	7.7	1.42	9.0	1.25
S14	3.0	1.05	2.7	0.82
S15	12.0	2.11	9.7	0.48

Table 5 ANOVA Model A: Fluorescence Full Model With Interaction Including Non-fluorescent Signs

```
Fluor ~ Sign + TOD +Sign:TOD + Error(Block)
data = test1and2

Error: Block
          Df  Sum of Sq  Mean Sq
Residuals  9  32.40333   3.60037

Error: Within
           Df  Sum of Sq   Mean Sq   F Value      Pr(F)
     Sign  14   876.8867  62.63476  118.3785  0.0000000
      TOD   1     0.7500   0.75000    1.4175  0.2348995
 Sign:TOD  14    24.7000   1.76429    3.3345  0.0000611
Residuals 261   138.0967   0.52911
```

Table 6 ANOVA Model B: Fluorescence Full Model With Interaction Excluding Non-fluorescent Signs

```
Fluor ~ Sign + TOD +Sign:TOD + Error(Block)
data = test1and2.less.stds

Error: Block
          Df  Sum of Sq   Mean Sq
Residuals  9  11.90909   1.323232

Error: Within
           Df  Sum of Sq   Mean Sq   F Value     Pr(F)
     Sign  10   285.0273  28.50273  64.21453  0.0000000
      TOD   1     7.2727   7.27273  16.38492  0.0000753
 Sign:TOD  10     6.6273   0.66273   1.49308  0.1445265
Residuals 189    83.8909   0.44387
```

Table 7 Approved Fluorescent Orange Prismatic Sign Sheeting

Fluorescent Orange Prismatic Rigid/Semi-rigid Substrate

Sample	Manufacturer	Brand
S7	3M	Diamond Grade 3924 F/G
S4 & S12	Avery Dennison	W-7514
S1	Nippon Carbide	Crystal Grade 92547

Fluorescent Orange Prismatic Roll-up

Sample	Manufacturer	Brand
S13	3M	RS34 F
S14	3M	Diamond Grade RS24
S10	Reflexite	Marathon 3932068373
S11	Reflexite	Super Bright 3932459374

FIGURES

Figure 1 Test Rack with Signs Covered

Figure 2 Typical Sample Sign Composition

Figure 3 One-Sign-at-a-Time Midday Rating Evaluation

Figure 4 All-Signs-at-the-Same-Time Ranking Evaluation

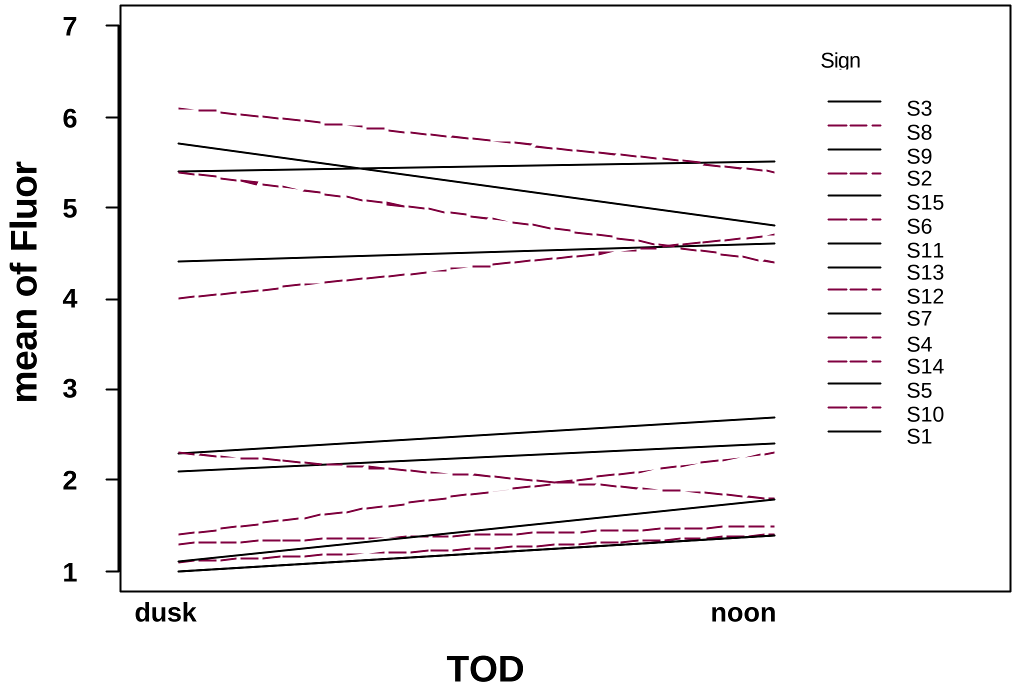

Figure 5 ANOVA Model A: Fluorescence Full Model Interaction Plot

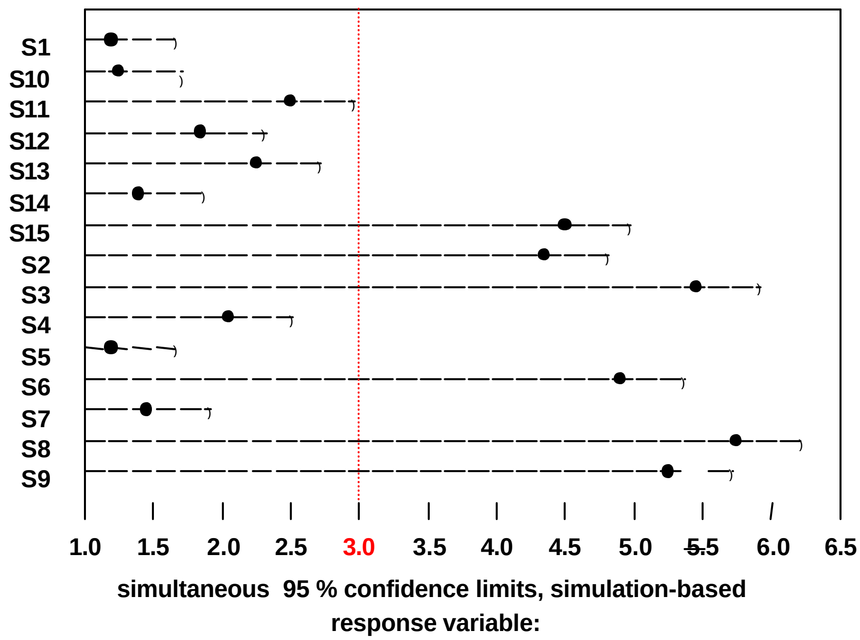

Figure 6 Comparison of 95% Confidence Intervals for *Fluor* Ratings

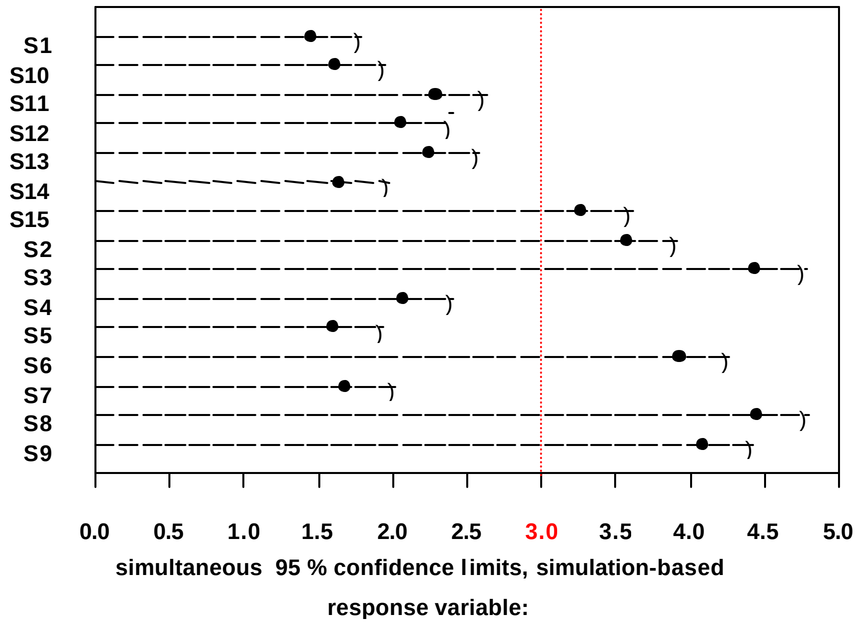

Figure 7 Comparison of 95% Confidence Intervals for *Composite* Ratings

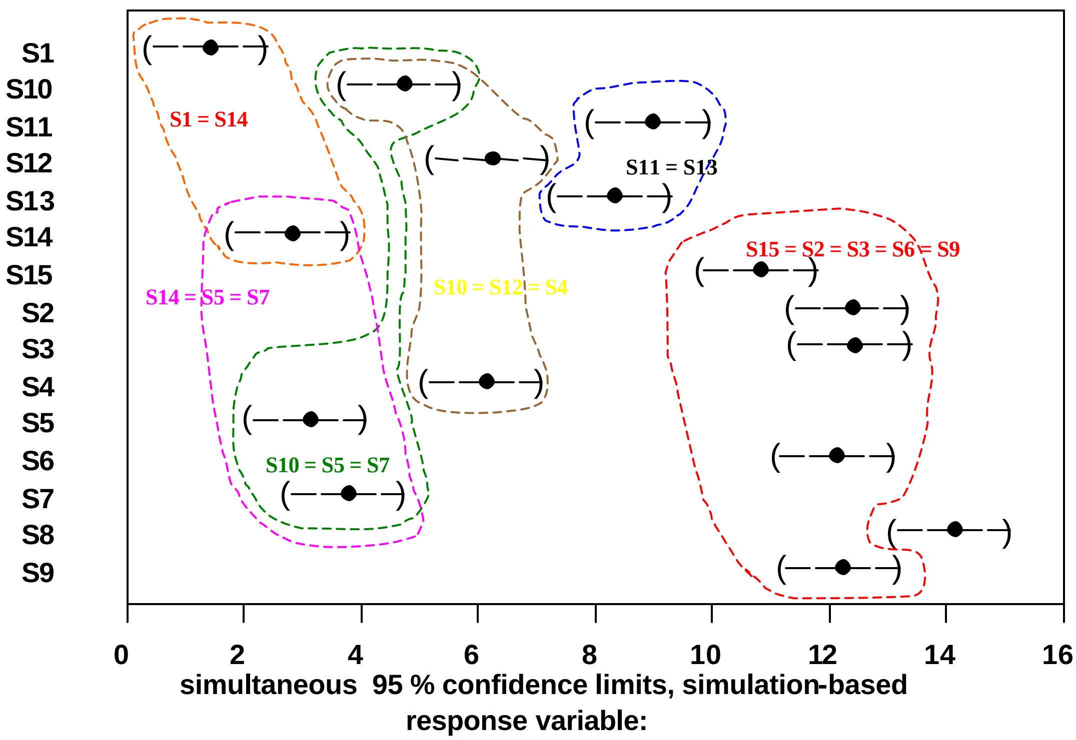

Figure 8 Comparison of 95% Confidence Intervals for Sign *Rankings*