Reconsidering the Role of Design Standards in Developing Effective Safety Labeling: Monolithic Recipes or Collections of Separable Features?

Michael J. Kalsher, William G. Obenauer, & Christopher F. Weiss

Michael J. Kalsher
Corresponding Author
Rensselaer Polytechnic Institute
Department of Cognitive Science
School of Humanities, Arts, and Social Sciences
110 8th Street, Troy, NY, 12180
301-F Carnegie Bldg
kalshm@rpi.edu

The proper citation for this article is:

Kalsher MJ, Obenauer WG, Weiss CF. Reconsidering the Role of Design Standards in Developing Effective Safety Labeling: Monolithic Recipes or Collections of Separable Features? Human Factors. 2019;61(6):920-952. doi:10.1177/0018720818820413

The published journal article can be accessed at:

https://doi.org/10.1177/0018720818820413

Reconsidering the Role of Design Standards in Developing Effective Safety Labeling: Monolithic Recipes or Collections of Separable Features?

Michael J. Kalsher

William G. Obenauer

Christopher F. Weiss

Rensselaer Polytechnic Institute, Troy, NY, USA

Precis: This research investigated whether safety labeling design guidelines, such as the American National Standards Institute (ANSI) Z535 series, contribute to better warnings. Results showed that choice of experimental design and analytical methods can dramatically influence a study's results and conclusions drawn.

Running head: SAFETY LABEL FEATURES

Manuscript Type: Extended Multi-Phase Study

Word Count: 11204

Correspondence concerning this article should be addressed to Michael J. Kalsher; Department of Cognitive Science; School of Humanities, Arts, and Social Sciences; 301-F Carnegie Bldg.; Rensselaer Polytechnic Institute; 110 8th St, Troy, NY 12180.

Contact: kalshm@rpi.edu

Abstract

Objective. This research investigated whether safety labeling design guidelines, such as the American National Standards Institute (ANSI) Z535 series, contribute to better warnings.

Background. Studies investigating the impact of safety label formatting on warning effectiveness have produced mixed findings. Additionally, research has failed to find a consistent relationship between measures of predicted and actual compliance. One commonality is that all of these studies have investigated the ANSI Z535 guidelines as a binary variable rather than as an integrative system of separable features.

Method. We measured predicted compliance using both a within-subjects and a between-subjects design, but actual compliance using only a between-subjects design. Data were analyzed using both ANOVA and linear / probit regressions to test the relationships between warning features recommended in the ANSI Z535 guidelines and measures of behavioral compliance.

Results. Predicted compliance assessed via a within-subjects design differed greatly from predicted compliance assessed via a between-subjects design. Levels of predicted and actual compliance were most similar when both measures were assessed using a between-subjects design. Consistent with previous research, location had a strong relationship with actual compliance, but surprisingly, presence of an ANSI-style orange warning header had a negative relationship with compliance.

Conclusion. The choice of experimental design and analytical methods can dramatically influence a study's results and conclusions drawn. This research identified several aspects of experimental design that should be considered in future research on warning effectiveness.

Application. Testing features recommended in the ANSI Z535 guidelines under varying conditions can contribute to the development of more effective warnings.

Keywords: warning systems, warnings, warning compliance, experimental design, statistics and data analysis

Introduction

Lawsuits arising from product defect liabilities can often result in multi-million dollar jury awards and settlements (Levy Konigsberg LLP, 2012) and can occur in response to defects in a variety of products and equipment ranging from kitchen appliances (City News Service, 2015) to automobiles (Sandler, 2015), exercise equipment (Watts, 2015), and even furniture (Nadolny, 2015). Lawsuits stemming from product defects can extend to situations in which there is an allegation of failure to warn users about a product's residual hazards (Barrett & Kott, 1994). To increase product safety and counter the potential for product liability lawsuits that often follow product-related deaths and injuries, voluntary safety guidelines, such as the ANSI Z535 safety labeling standards, have been developed to standardize the design and placement of product warnings with the related goal of communicating safety information more effectively (ANSI, 2011).

Researchers have long contemplated the importance of developing scientifically justifiable guidelines for how warnings should be designed and how their effectiveness should be evaluated (Laughery & Wogalter, 2006, 2014; Lehto, 1992; Wogalter, Conzola, & Smith-Jackson, 2002). Studies investigating the effectiveness of warning guidelines, such as the guidelines set forth in ANSI Z535, have approached the question in different ways. Some research in this area has focused on evaluating isolated warning components (e.g., the use of color, font size, choice of signal word, pictorials, placement), or on combinations of warning components, specified in the standard. Most of these studies have relied on perceptual measures to gauge effectiveness, including noticing, reading, memory, perceived hazardousness, preference, and behavioral intentions (e.g. Kalsher, Wogalter, Brewster, & Spunar, 1995; Loring & Wiklund, 1988; Matthews, Andronaco, & Adams, 2014). Only a relatively few studies have assessed actual compliance behavior (Braun & Silver, 1995; Dingus, Wreggit, & Hathaway, 1993; Hunn & Dingus, 1992; Kalsher, Gallo, Williams, & Wogalter, 2000; Wogalter, Kalsher, & Racicot, 1993). A growing number of studies have used this focused approach to identify warning features recommended by ANSI Z535 that alone, and in combination, are consistently effective across different cultural groups (e.g. Borade, Bansod, & Gandhewar, 2008; Lesch, Rau, & Choi, 2016; Lesch, Rau, Zhao, & Liu, 2009; Smith-Jackson & Essuman-Johnson, 2002; Yu, Chan, & Salvendy, 2004). Other researchers have extended this approach to investigate the potential interactions between characteristics of the warning, the person interacting with the warning, and the product use context (i.e. Rogers, Lamson, & Rousseau, 2000).

Another approach has focused on investigating warning configurations described as “similar to”, “in accordance with”, “ANSI formatted”, or “varying in terms of the extent of conformance to ANSI Z535 (e.g. Arghami, Kian, & Mohammadfam, 2009; Cheatham & Wogalter, 1999; Dingus et al., 1993; Wogalter et al., 1999; Wogalter, Kalsher, Frederick, Magurno, & Brewster, 1998). Overall, research investigating the impact of safety labeling configuration on measures of warning effectiveness has produced somewhat conflicting findings. Some previous studies have reported that warning labels that are consistent with the ANSI Z535 guidelines are more effective than alternative warnings in terms of both predicted compliance (Laughery et al., 2002) and actual behavioral compliance (Smith & Wogalter, 2010). Other studies, however, have reported that the effectiveness of warning labels is contingent upon perceived risk (Heckman, Harley, Scher, & Young, 2010) and that ANSI-style warning labels can be less effective than both older OSHA-style warning labels (Kim & Wogalter, 2009) and more generic non-ANSI style warnings (Frantz, Young, Rhoades, & Wisniewski, 2005; Young, Frantz, Rhoades, & Darnell, 2002).

The primary aim of the present research is to investigate the basis for these previous discrepant findings by considering differences in the experimental designs employed in these previous studies and other potentially relevant study characteristics. We do so in the context of a series of three experiments. These experiments constitute both direct and conceptual replications of experimental designs utilized by Shaver et al. (2006); Frantz and Rhoades (1993); and Frantz, Rhoades, Young, and Schiller (2000). The use of replication in our experimental design provides the following benefits. First, replication allows us to contribute to an ongoing discussion in the warnings literature that utilizes a consistent design template to examine how aspects of warning design and placement influence behavioral compliance. Additionally, replication of prior experimental methodology provides greater clarity as to how experimental design influences the consistency/inconsistency of our findings with those of other research within the subset of studies that have utilized this design.

The first issue that we consider in our replication studies relates to the following question: What constitutes a warning that comports with ANSI guidelines? A review article by Young, Frantz, Rhoades, and Darnell (2002) described a study of racquetball players (Hathaway & Dingus, 1992) in which one of the experimental conditions was described as including an "ANSI-style sign". Closer inspection of the warning, however, reveals that it did not conform to the ANSI-recommended use of a safety orange background with the signal word "WARNING", a concise message in the message panel, or space between multiple messages in the message panel (ANSI, 1991). Frantz, Young, Rhoades, and Wisniewski (2005) compared compliance with "less ANSI" warnings (Frantz & Rhoades, 1993) to compliance with "more ANSI" warnings" (Frantz et al., 2000), but the "more ANSI" warnings in their study did not comport to the ANSI Z535 recommendation to avoid prepositional phrases and use symbols and pictorials whenever practical (ANSI, 1998).

Similarly, Huntley-Fenner, Harley, Trachtman, and Young (2007) investigated recall and compliance with warnings that varied with respect to the degree of conformance to ANSI Z535.6. Three formats of a product manual were created, termed “non-ANSI”, “partial-ANSI” and “full-ANSI”. The “partial-ANSI” and “full-ANSI” variants were identical to the “non-ANSI” format (an unaltered portion of the manufacturer’s user guide) except for the presence of additional ANSI-recommended features. For example, in the “partial-ANSI variant, signal words were accompanied by the signal-alert icon and placed inside a bolded text box atop the ANSI-recommended colored background (DANGER presented in white caps on a safety red background; WARNING presented in black caps on a safety orange background). The results indicated that compliance was not significantly related to the presence, absence or degree of ANSI formatting. It is noteworthy that the “non-ANSI” variant actually incorporated a number of ANSI-recommended features, including the use of bolded text, numbering, and (all-cap, bolded) signal words. Finally, Kim and Wogalter (2009) tested an “ANSI format” warning that did not include the recommended feature of bulleted text (ANSI, 2007). To address this issue, we incorporate a variety of warning labels into our research and test the effectiveness of different, separable features of the ANSI Z535 guidelines as opposed to treating the ANSI safety labeling guidelines as a binary variable.

Another issue that may contribute to the discrepant findings in compliance studies relates to sample size. The generic warning condition in Frantz, Young, Rhoades, and Wisniewski's (2005) study, for example, had a relatively small number of participants (n=8) in it. In fact, because of this small sample size, the authors of the original study to use this data elected to use a p-value of 0.10 to denote statistical significance (Frantz & Rhoades, 1993). Shaver et al. (2006) had a similar issue with sample size as this study used fewer than 15 participants per condition. Even when large numbers of participants have been reported in warning compliance studies, the number of participants can be misleading. For example, Young, Frantz, Rhoades, and Darnell's (2002) literature review described Shaver and Braun's (2000) study as having 4,620 participants, but Shaver and Braun grouped their data into 140 blocks for analysis. Given that their study incorporated four different warning configurations and warning color was manipulated in some of those conditions, it is likely that despite having 4,620 participants, after grouping data and removing outliers from analysis, they actually had a sample size of approximately 20 observations per condition. We address the issue of sample size in the context of a compliance study that incorporates 131 subjects and utilizes a logistic regression to test the relationships between different, separable warning features and compliance with warnings, thereby allowing for greater variation in exposure to these features. For example, in Study 2 of this paper, 63 participants were exposed to warning labels with orange warning headers, while 68 participants were exposed to warning labels without orange warning headers.

Although previous research has shown higher rates of warning compliance for students (Frantz & Rhoades, 1993) than for non-students (Frantz et al., 2000), comparisons of predicted compliance versus actual compliance have tested predicted compliance using only students and actual compliance using only non-students (Shaver et al., 2006). We address this, and related, problems by partly replicating Shaver et al.’s (2006) study in Study 1, but we incorporate both students and non-students into our test of actual compliance in Study 2. Additionally, we incorporate both students and non-students into our final test of predicted compliance in Study 3.

Our results provide interesting insight into the predicted effectiveness of various components of the ANSI safety standards. These results, however, were not replicated in our test of actual compliance. In fact, the warning label feature that was predicted to have the strongest positive effect on compliance in Study 1 was shown to have a negative relationship with actual compliance in Study 2. This observation motivated Study 3 in which we diverted from the research trend of measuring predicted compliance using a within-subjects design (Frantz et al., 2005; Shaver et al., 2006) by recording measures of predicted compliance using a between-subjects design that mimicked the experimental design of our observed (actual) compliance study. The results indicated that the experimental design used when assessing predicted compliance may play an important role in how closely these results can be replicated in studies of actual behavioral compliance.

Study 1

Our first study was a partial replication of Shaver et al.'s (2006) study. The results of the study provided us with a baseline value for predicted compliance. In addition to testing three warning label formats used in the original study, we included five new formats. This resulted in a diverse set of warning label formats that allowed us to evaluate the relative contributions of individual warning features to measures of effectiveness as compared to considering label design from a binary perspective (i.e., as either ANSI-style or non-standard).

Methods

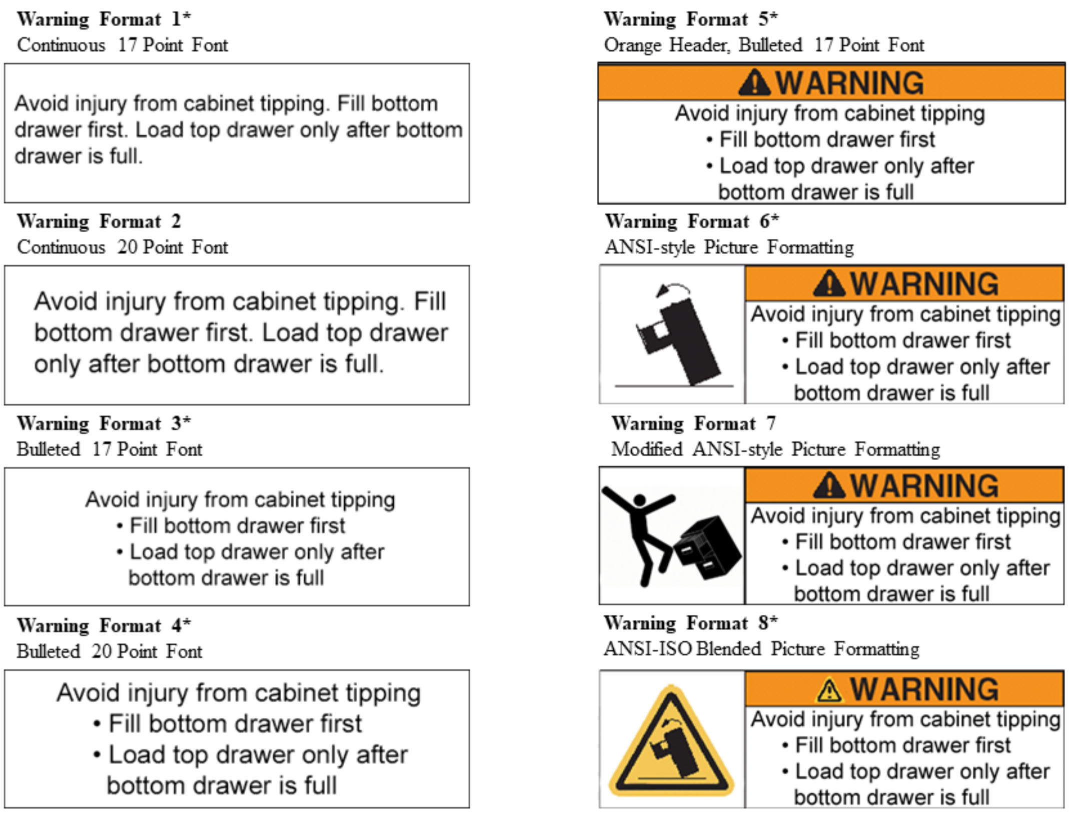
Participants. A total of 533 students (*freshman = 58.35%, sophomore = 17.45%, junior = 12.20%, senior = 11.63%, graduate = 0.38%*) from 43 different primary majors at a small, private college in the northeastern United States participated in this study for course credit. The sample was composed of 409 males and 124 females. Participants' ages ranged from 17 to 26 years (*M = 19.4, SD = 1.33*). Approximately half (*56.66%*) of the participants reported owning a file cabinet, but only 0.94 percent of participants reported having been previously injured using a file cabinet. Comparative t-tests showed no differences in mean values of our dependent variables of predicted noticing and predicted compliance based upon gender or file cabinet ownership ($ps > 0.10$), justifying the pooling of observations for subsequent analyses. Variance comparisons showed no differences in predicted compliance between participants based upon file cabinet ownership or gender ($ps > 0.10$). While we found no differences in variances for predicted noticing based upon ownership [$F(1847, 2415) = 0.98, p > 0.10$], the variance for predicted noticing did differ significantly by gender [$F(3271, 991) = 0.88, p < 0.05$]. For the purpose of consistency with prior research, data for males and females was still pooled. The implications of this decision are addressed in our robustness checks.

Procedure. We will provide a summary of the procedure as it has been described in detail by Kalsher, Obenauer, & Weiss (2016). Participants were given a six-page packet that included a set of instructions, images of warning labels, an image of a file cabinet, and questions pertaining to both the participants' experience with file cabinets and their demographic characteristics. The final page of the packet asked participants to estimate how many people out of 100 would notice and comply with each warning label.

Each packet included images of the same eight warning labels for a within-subjects design (*see Figure 1*). Three of these images were replicated from Shaver et al. (2006), while the other five images were constructed based upon recommendations in the ANSI (2011) guidelines. The file cabinet image was also manipulated as 268 packets included an image of a two-drawer file cabinet and 265 packets included an image of a four-drawer file cabinet. There was no significant difference in means for predicted noticing ($p = 0.10$) or predicted compliance ($p >$

0.10) between the two-drawer and four-drawer cabinet conditions, so these observations were pooled for subsequent analyses. This research complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at the university. Informed consent was obtained from each participant.

Figure 1: Warning Label Formats Included in Studies



*included in Studies 2 & 3

Results

Predicted Noticing. Consistent with previous research that has investigated warning effectiveness using measures of precursors to behavioral compliance through comparisons of means (Frantz et al., 2005; Laughery et al., 2002; Shaver et al., 2006), our initial analyses compared mean values for predicted noticing / compliance for each label design using a within-subjects ANOVA. Mauchly's test was significant, indicating that the assumption of sphericity had been violated, thus a Greenhouse-Geiser adjustment was applied to the degrees of freedom. The results show that the predicted rate at which people would notice a warning label on a file cabinet was significantly affected by the format of the label [$F(2.36, 1255.84) = 986.98, p < 0.001, R^2 = 0.81$]. As shown in *Table 1*, predicted noticing was greatest for the label with ANSI-ISO blended picture formatting ($M = 77.52, SD = 20.68$) and lowest for the label with continuous 17-point font ($M = 37.44, SD = 23.30$).

TABLE 1: Comparison of Predicted and Actual Compliance

Study	Prior Research		Current Study	
Outcome Variable	Predicted Noticing	Predicted Compliance	Predicted Noticing	Predicted Compliance
Continuous 17 point font			37.4%	30.9%
Continuous 20 point font			41.9% ***	34.4% ***
Bulleted 17 point font	62.0% [f3]	52.0% [f3]	41.6%	34.8%
Bulleted 20 point font	41.4% [s]	24.8% [s]	46.7% ***	39.3% ***
Orange header, bulleted 17 point font	80.4% [f3]	72.1% [f3]	66.3% ***	52.9% ***
ANSI-style picture formatting	73.8% [s]	54.5% [s]	75.5% ***	62.8% ***
Modified ANSI-style picture formatting			75.6%	64.0%
ANSI-ISO blended picture formatting	91.0% [s]	71.8% [s]	77.5% [t]	64.9%
Effect of Label[z]			$F(2.36, 1255.84) = 986.98$ ***	$F(2.66, 1412.78) = 887.91$ ***
Omega-squared (label)			*0.35*	*0.26*
R-squared (model)			*0.81*	*0.85*

Significance values for current study are for post-hoc tests comparing each label format to the format listed directly above using a Bonferroni correction.
$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$ *for two-tail test*

[z] *Degrees of freedom adjusted using Greenhouse-Geisser correction*

f3 Data from Frantz et al. (2005) study. Note that "Bulleted 17 point font" was numbered and capitalized and text in "Orange header" was continuous

s Data from Shaver et al. (2006) study. Methodology and design matches the present study.

Bonferroni post hoc tests revealed a significant increase in predicted noticing ($\Delta = 4.50$, $p < 0.001$) from the label with continuous 17-point font to the label with continuous 20-point font ($M = 41.95$, $SD = 23.57$). There was no significant difference ($p > 0.10$) in predicted noticing between the label with 20-point continuous font and the label with 17-point bulleted font ($M = 41.62$, $SD = 23.87$). There was, however, a significant increase in predicted noticing ($\Delta = 5.07$, $p < 0.001$) from the label with 17-point bulleted font to the label with 20-point bulleted font ($M = 46.69$, $SD = 23.79$). There was a large increase in predicted noticing ($\Delta = 19.57$, $p < 0.001$) from the label with 20-point bulleted font to the label with 17-point bulleted font and an orange warning header ($M = 66.26$, $SD = 22.20$). There was also a significant increase in predicted noticing ($\Delta = 9.23$, $p < 0.001$) from the label with an orange warning header to the label with ANSI-style picture formatting ($M = 75.49$, $SD = 20.57$). The difference in predicted noticing between the labels with ANSI-style picture formatting and a modified ANSI-style picture ($M = 75.57$, $SD = 20.67$) was not significant ($p < 0.10$). The increase in predicted noticing from the label with a modified ANSI-style picture to the ANSI-ISO label ($M = 77.52$, $SD = 20.68$) was marginally significant ($\Delta = 1.95$, $p < 0.10$).

Predicted Compliance. As with our dependent variable of predicted noticing, Mauchly's test was significant so a Greenhouse-Geiser adjustment was applied to the degrees of freedom in our analysis. Similarly, results show that the predicted rate at which people would comply with a warning label on a file cabinet was significantly affected by the format of the label [$F(2.66, 1412.78) = 887.91$, $p < 0.001$, $R^2 = 0.85$]. Predicted compliance was greatest for the labels with ANSI-ISO blended picture formatting ($M = 64.90$, $SD = 23.63$) and modified ANSI-style picture formatting ($M = 63.96$, $SD = 23.48$). Bonferroni post hoc tests showed no significant difference

($p > 0.10$) in predicted compliance across these two conditions. Again, predicted compliance was lowest for the label with continuous 17-point font ($M = 30.88$, $SD = 22.65$).

Similar to our post hoc analyses for predicted noticing, Bonferroni post hoc tests showed a significant change in predicted compliance ($\Delta = 3.54, p < 0.001$) between the label with 17-point continuous font and the label with 20-point continuous font ($M = 34.42$, $SD = 23.28$). The difference in predicted compliance between the 20-point continuous font label and the 17-point bulleted font label ($M = 34.84$, $SD = 23.12$) was not significant ($p > 0.10$). The change in predicted compliance from the 17-point bulleted font label to the 20-point bulleted font label ($M = 39.26$, $SD = 23.49$) was significant ($\Delta = 4.24, p < 0.001$) as was the change in predicted compliance ($\Delta = 13.62, p < 0.001$) from the 20-point bulleted font label to the label with 17-point bulleted font and an orange warning header ($M = 52.88$, $SD = 23.35$). Finally, the increase in predicted compliance from the label with an orange warning header to a label with an ANSI-style picture formatting ($M = 62.85$, $SD = 23.60$) was significant ($\Delta = 9.97, p < 0.001$).

Warning Features. While prior research has described warning labels as being either consistent ("ANSI-style") or inconsistent with ANSI guidelines (Frantz et al., 2005; Shaver et al., 2006; Smith & Wogalter, 2010), the ANSI (2011) guidelines are actually comprised of a collection of separable features. In order to analyze the relationships between these features and predicted noticing / compliance, we created binary variables for font size (0 = 17 point, 1 = 20 point), paragraph format (0 = continuous font, 1 = bulleted font), warning header (0 = orange header not present, 1 = orange header present), presence of a pictogram (0 = not present, 1 = present), and the presence of an alternative (modified image or ANSI-ISO blend) pictogram style (0 = not present, 1 = present). Because subjects were exposed to all eight conditions, we then incorporated these variables into a linear fixed-effects regression using robust standard errors.

This was done using the xtreg function in Stata14. The fixed-effects model controls for both observable and unobservable individual differences amongst subjects, thus accounting for within-person correlations in responses (*see Appendix I*) and more effectively isolating the predicted effects of the variables of interest (Angrist & Pischke, 2009). Additionally, unlike the ANOVA, the fixed-effects regression has no underlying assumptions of normality or homogeneity of variance (Stock & Watson, 2015).

Our model using predicted noticing as the dependent variable was significant [$F(6, 532) = 292.97$, $p < 0.001$, $R^2 = 0.35$]. As shown in *Table 2*, the coefficients on binary variables for font size ($\beta = 4.79$, $SE = 0.28$, $p < 0.001$) and paragraph format ($\beta = 4.46$, $SE = 0.54$, $p < 0.001$) were both positive and significant, indicating that increasing font size from 17 point to 20 point or introducing bulleting on a warning label would result in an increase in predicted noticing. For the dependent variable of predicted noticing, the feature of an orange warning header had the largest coefficient ($\beta = 24.50$, $SE = 0.91$, $p < 0.001$). The inclusion of an image also had a positive and significant coefficient ($\beta = 9.23$, $SE = 0.49$, $p < 0.001$). In terms of alternative pictograms, the use of a modified image had no significant relationship with predicted noticing ($p > 0.10$) beyond that of simply incorporating an image, but the additional variance explained by the feature of an ANSI-ISO blended image was significant ($\beta = 2.03$, $SE = 0.50$, $p < 0.001$).

Our model using predicted compliance as the dependent variable was also significant [$F(6, 532) = 275.14$, $p < 0.001$, $R^2 = 0.26$]. The positive coefficients on increased font size ($\beta = 3.98$, $SE = 0.25$, $p < 0.001$) and the use of bulleted text, ($\beta = 4.40$, $SE = 0.49$, $p < 0.001$) were similar to those in the predicted noticing model. The coefficient for the orange warning header was still positive and significant ($\beta = 17.82$, $SE = 0.71$, $p < 0.001$), though it was smaller than in the predicted noticing model. The coefficient on the relationship between an image ($\beta = 9.97$, *SE = 0.48, p < 0.001*) and predicted compliance was similar to that in the noticing model. Both the use of a modified image ($\beta = 1.12$, $SE = 0.52$, $p < 0.05$) and the use of an ANSI-ISO blended image ($\beta = 2.05$, $SE = 0.51$, $p < 0.001$) were predicted to have a positive effect on compliance beyond that of simply including an image.

TABLE 2: Linear Fixed-Effects Regressions (Controlling for Individual Fixed Effects) Examining the Relationships Between Warning Features and Predicted Noticing / Predicted Compliance

Warning Feature	Predicted Noticing		Predicted Compliance	
Constant	37.30 ***	(0.53)	30.66 ***	(0.47)
20 point font	4.79 ***	(0.28)	3.98 ***	(0.25)
Bulleted font	4.46 ***	(0.54)	4.40 ***	(0.49)
Orange header	24.50 ***	(0.91)	17.82 ***	(0.71)
Image	9.23 ***	(0.49)	9.97 ***	(0.48)
Modified image	0.08	(0.51)	1.12 *	(0.52)
ANSI-ISO blend	2.03 ***	(0.50)	2.05 ***	(0.51)
R-Squared	0.35		0.26	
Model fit	$F(6, 532) = 292.97$ ***		$F(6, 532) = 275.14$ ***	

****p<.001, **p<.01, *p<.05, †p<.10 for two-tail test with robust standard errors adjusted for clustering*

Robustness Tests. Although the size of our sample should alleviate concerns regarding the assumption of normality in linear models such as ANOVA (Field, 2012; Lumley, Diehr, Emerson, & Chen, 2002), for robustness, we tested this assumption using Shipiro-Wilk normality tests. ANOVA's assumption of normality states that the error terms should have a normal distribution (Glass, Peckham, & Sanders, 1972). The results of our Shipiro-Wilk tests indicated that this assumption may have been violated for both the dependent variable of predicted noticing ($W=0.993$, $p<0.001$) and predicted compliance ($W=0.996$, $p<0.001$). These high values for W taken in conjunction with the large sample size, however, indicate that these tests may have been significant due to the high power of the tests detecting small departures from normality. Histograms of the residuals support this argument, but non-parametric replications of our analyses were conducted for robustness.

Friedman's ANOVA for predicted noticing was significant $[X^2(7) = 2806.83, p<0.001]$. Post-hoc comparisons using Wilcoxon Signed Ranks Tests with Bonferroni corrections yielded results consisted with those reported in *Table 1*. For predicted compliance, Friedman's ANOVA was significant once again $[X^2(7) = 2800.92, p<0.001]$. Post-hoc analyses were also consistent with those reported in *Table 1*.

Additionally, as discussed above, an F-test indicated that the variance for predicted noticing differed significantly for male and female participants. To address this, we conducted subsample analyses for predicted noticing by gender. In the male subsample, Mauchly's test was significant ($p<0.001$) so a Greenhouse-Geiser correction was utilized. An ANOVA revealed that the predicted effect of warning label format on noticing was significant for male participants $[F\ (2.47, 1009.24) = 752.24, p<0.001]$. The results of post-hoc comparisons using Bonferroni corrections were consistent with the reported results with one exception. While the difference in predicted noticing between the modified ANSI-style picture formatting and the ANSI-ISO blended picture formatting was marginally significant for the full sample ($p<0.10$), this difference was not significant within the male subsample ($p>0.10$).

Mauchly's test was only marginally significant ($p<0.10$) in the female subsample, but a Greenhouse-Geiser correction was utilized for robustness. Consistent with the male subsample, the ANOVA revealed that the predicted effect of warning label format on noticing was significant for female participants $[F\ (2.06, 253.67) = 242.37, p<0.001]$. With two minor exceptions, the post-hoc comparisons were consistent with the overall results. First, the difference in mean predicted compliance for the continuous 17-point font and the continuous 20-point font was only marginally significant ($p<0.10$) in the female subsample after a Bonferroni correction. Additionally, similar to the male subsample, the difference in predicted noticing between the modified ANSI-style picture formatting and the ANSI-ISO blended picture formatting failed to reach statistical significance ($p>0.10$).

Discussion

Unlike previous research that has treated conformation to ANSI (2011) as a binary variable (i.e. ANSI-style or non-standard) this study examined the differential predicted outcomes associated with the separable features recommended by the ANSI guidelines. Our results indicated that as different components to the ANSI guidelines are added to a warning label, rates of both predicted noticing and predicted compliance increase. Across various model specifications and analytical techniques, we found robust evidence that participants associated bulleted text, increased font size, an orange warning label, and warning images with higher levels of noticing and compliance with warning labels.

However, unlike Shaver et al. (2006), who reported “estimates of noticing and compliance appeared to increase in a linear and monotonic fashion as the number and/or salience of the included formatting elements increased,” (p. 2198) we found that increases in estimates of noticing and compliance were dependent upon the presence of specific features. For example, as shown in *Table 2,* the predicted impact on compliance for an orange header was more than four times that of increasing font size or adding bulleting to the warning’s text.

Beginning to isolate the effects of the separable features of warning labels is critical to the development of optimal warnings as space constraints often require warning designers to consider tradeoffs of different features (i.e. there may only be space for larger font *or* a warning image on a label). Additionally, the positive predicted effects of bulleted text, increased font size, an orange header, and use of graphics in warnings provide some confirmation that people perceive the components of ANSI Z535 warning label guidelines as appropriate measures for improving safety compliance. Warnings research would benefit from future studies that treat warning label components as separable features.

One possible reason that we did not find the linear increase in estimated compliance described above may be that our study included twice as many warning labels as previous studies, allowing us to more clearly isolate estimated predicted effects. Additionally, presenting participants with eight different warning configurations may have caused them to more carefully consider the differences in warning design than they would when evaluating a smaller number of configurations. These results should be interpreted cautiously, however, as our experimental design was not a full factorial. For example, there was not a warning label that included an image and also had continuous text.

Study 2

Extant research has failed to replicate the results of predicted warnings compliance studies in studies of actual compliance (Frantz et al., 2005; Shaver et al., 2006; Young et al., 2002). These studies have created opportunities for future research. First, as noted previously, some of the research in this area has tended to conceptualize warning labeling in a binary fashion (i.e., ANSI-style or non-standard). Instead, we suggest that the correct approach is to view the ANSI Z535 guidelines as an integrative system of separable components that can be tested using well-established measures of effectiveness both separately and in combination. Second, we observed that some of the previous findings were derived from relatively small sample sizes. Frantz et al. (2005), for example, relied on only eight subjects exposed to the “less ANSI” condition and Shaver et al. (2006) relied on sample sizes of 12-14 subjects per condition. While the use of small sample sizes does not necessarily invalidate the results of these previous studies, it does suggest a need for additional testing for purposes of verification. Study 2 addresses both of these opportunities.

Methods

Participants. A total of 142 students and non-students (staff members) at the same small, private college described in Study 1 participated in this study. Some students received course credit for their participation. One participant's data was omitted because the person revealed prior knowledge of the experiment and 10 other participants' data was omitted because these individuals did not follow the researchers' directions. The final sample consisted of 112 students and 19 staff members. The students in the final sample consisted of 57 freshmen, 24 sophomores, 19 juniors, and 9 graduate students, with three students not reporting their year of study. Participants' ages ranged from 16 to 58 years ($M = 23.05, SD = 9.73$). The sample was composed of 68 males and 63 females each of whom were randomly assigned to one of six experimental conditions. Eighty-nine percent of subjects claimed English as their primary language. Distribution of subjects by gender, staff member status, student year, and native language were all consistent across conditions, thus data were pooled for analyses. A demographic breakdown of the sample by condition is presented in Appendix II. Comparisons of means and variances of our dependent variables by gender, native language, and student status showed no significant differences with only one exception. The mean value of one of our compliance measures differed significantly for males and females. These comparisons are shown in Appendix III. Concerns regarding the pooling of data are discussed further in our robustness checks.

Procedure. In this study, we replicated the experimental design used by Frantz and Rhoades (1993); Frantz, Rhoades, Young, and Schiller (2000); and in the second study of Shaver et al. (2006). Participants were told that the purpose of the experiment was to analyze how people arrange their work space. The true purpose was to investigate the impact of different warning configurations on compliance and precursors to compliance. The informed consent form that was completed by all subjects prior to their participation included the following description intended to underscore the mild deception:

> *"As a participant in this study, you will be asked to set up a fictitious office according to your preferences…the point of the study is to assess how you go about "designing" your office…"*

After completing the informed consent form, participants were taken to the experiment room. For consistency with previous studies (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006), a 1.8 m x 2.4 m section of the floor of the experiment room was taped off and designated as the work space. Within that workspace, there was a table on casters, a two-drawer file cabinet in its original packaging (the bottom of the box was cut off for ease of removal), a chair, a computer, a telephone, and a box of office supplies. The office supplies included a telephone box and a stack of files with a mass of 3.2 kg. Though other studies have included files with a mass of 4.8 – 6.8 kg (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006), the mass of files was reduced in this study to ensure that the file cabinet was not at risk of tipping over as such risk could result in compliance that was not influenced by the warning label.

Participants were randomly assigned to one of six warning label design conditions (*see Figure 1*). For approximately half (*53.44%*) of the observations, a 3.8 cm x 12.7 cm warning label with a 1 cm flap warning label was affixed to the front of the file cabinet drawers and wrapped around the side of the file cabinet, consistent with Shaver et al. (2006) and Frantz et al. (2000). For the rest of the observations, the label was placed on a cardboard bridge across the inside of the top drawer. This placement was similar to that of Frantz and Rhoades' (1993) Condition 4, though we did not include a second label in this condition.

On the far side of the experiment room, there was a laptop computer with a built-in camera. This laptop computer streamed video into a room where two experimenters observed the study. The screen of the laptop was white with text that read, “When you are done, please wave to the camera.” After being led to the experiment room, participants were read the same set of instructions used by Shaver et al. (2006). They were also told that the location of windows, doors, outlets, phone jacks, etc. did not matter. The final section of the directions instructed them to wave to the camera when they were done setting up their office.

After leaving the experiment room, the experimenter joined a second experimenter in an observation room. Watching a live stream video of the experiment room, the experimenters made notes regarding the amount of time the participant appeared to spend reading the label, the order in which the participant placed files in the file cabinet, and other relevant observations (i.e. participant did not use the file cabinet, participant discarded phone book, etc.). Once the participant waved to the camera, the first experimenter brought her or him to a separate room for debriefing and to complete a questionnaire. The second experimenter recorded the weights of files in each drawer and reset the experiment room according to a written set of experimenter instructions that was developed for purposes of procedural consistency.

Results

In order to compare our findings to those of previous research (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006), our initial analysis compared noticing the label, reading the label, and compliance with the label through a comparison of means. Summary statistics for this study as well as the aforementioned studies are shown in *Table 3*. Because of the binary nature of our dependent variables, this analysis violates the ANOVA's assumption of a linear dependent variable (Field, 2012; Gaito, 1980). The results of parametric tests on models using binary dependent variables, however, are frequently consistent with those of non-parametric models (Chatla & Shmueli, 2016). Additionally, as shown in *Appendix IV*, our homogeneity of variance tests provide inconsistent results regarding whether or not our analysis violates ANOVA's assumption of homogeneity of variance. Concerns regarding the effect of violations of assumptions on our findings are discussed in our robustness checks below. *Table 4* reports the results of our one-way between-subjects ANOVA for all observations, allowing for direct comparison with the results of previous research (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006). Bonferroni post hoc tests are reported in *Table 5*.

Noticing. During the debriefing session, participants were asked whether or not they noticed any labels or markings on the file cabinet. The first model shows that mean values for self-reported noticing rates across label format conditions for participants in the "outside" location condition differed significantly [$F(5, 64) = 2.73$, $p < 0.05$, *omega-squared* $= 0.11$]. The rate of self-reporting noticing was lowest for labels with an orange warning header and no image ($M = 45.45$, $SD = 52.22$). All warnings without an orange header had self-reported noticing rates above 75 percent. However, as shown in *Table 5*, because the ANSI-ISO blended label had the highest rate of self-reported noticing ($M = 100.00$, $SD = 0.00$), Bonferroni post hoc tests indicated that the presence of an orange warning header had no significant impact

TABLE 3: Summary of Reported Noticing, Reported Reading, and Observed Compliance by Label Format and Location

		Prior Research				Current Study				
Warning Format	Warning Location	Noticed Label	Read Some	"Lax" Compliance	"Strict" Compliance	Noticed Label	Read Some	Read Most	"Lax" Compliance	"Strict" Compliance
Continuous 17 point font	Outside of file cabinet					76.9%	46.2%	30.8%	61.5%	46.2%
Bulleted 17 point font	Outside of file cabinet	93.3% [f1]	66.7% [f1]	53.3% [f1]	40.0% [f1]	75.0%	58.3%	25.0%	50.0%	33.3%
Bulleted 20 point font	Outside of file cabinet	92.9% [s]	78.6% [s]	7.1% [s]		92.3%	84.6%	84.6%	69.2%	61.5%
Orange header	Outside of file cabinet	97.6% [f2]	35.7% [f2]	16.7% [f2]	3.6% [f2]	45.5%	18.2%	9.1%	27.3%	9.1%
ANSI-style pic	Outside of file cabinet	83.3% [s]	33.3% [s]	16.7% [s]		60.0%	10.0%	0.0%	10.0%	10.0%
ANSI-ISO blended	Outside of file cabinet	100.0% [s]	35.7% [s]	7.1% [s]		100.0%	27.3%	0.0%	18.2%	0.0%
Continuous 17 point font	Inside file cabinet drawer					60.0%	60.0%	60.0%	90.0%	70.0%
Bulleted 17 point font	Inside file cabinet drawer	93.3% [f1]	66.7% [f1]	73.3% [f1]	53.3% [f1]	100.0%	90.0%	90.0%	100.0%	100.0%
Bulleted 20 point font	Inside file cabinet drawer					90.0%	80.0%	70.0%	90.0%	70.0%
Orange header	Inside file cabinet drawer					72.7%	54.5%	27.3%	54.5%	54.5%
ANSI-style pic	Inside file cabinet drawer					50.0%	40.0%	20.0%	60.0%	40.0%
ANSI-ISO blended	Inside file cabinet drawer					70.0%	60.0%	30.0%	70.0%	30.0%
Continuous 17 point font	All Locations (Overall)					69.6%	52.2%	43.5%	73.9%	56.5%
Bulleted 17 point font	All Locations (Overall)	93.3% [f1]	66.7% [f1]	63.3% [f1]	46.7% [f1]	86.4%	72.7%	54.5%	72.7%	63.6%
Bulleted 20 point font	All Locations (Overall)					91.3%	82.6%	78.3%	78.3%	65.2%
Orange header	All Locations (Overall)					59.1%	36.4%	18.2%	40.9%	31.8%
ANSI-style pic	All Locations (Overall)					55.0%	25.0%	10.0%	35.0%	25.0%
ANSI-ISO blended	All Locations (Overall)					85.7%	42.9%	14.3%	42.9%	14.3%
All Formats (Overall)	Outside of file cabinet					75.7%	42.9%	27.1%	41.4%	28.6%
All Formats (Overall)	Inside file cabinet drawer					73.8%	63.9%	49.2%	77.0%	60.7%

f1 Data from Frantz and Rhoades (1993) study. Note that for "outside" location, warning was present in two locations. Also, font size of warning was not published

f2 Data from Frantz et al. (2000) study. Note that text was not bulleted.

s Data from Shaver et al. (2006) study. Methodology and design matches the present study.

TABLE 4: ANOVA Model Summaries for Reported Noticing, Reported Reading, and Observed Compliance by Label Format and Location

Warning Location	Dependent Variable	Source	Partial SS	df	MSE	F	omega-squared
Outside	Noticed Label	Label	2.26	(5, 64)	0.17	2.73 *	0.11
Outside	Read Some or More	Label	4.58	(5, 64)	0.20	4.67 **	0.21
Outside	Read Most or More	Label	6.22	(5, 64)	0.12	10.45 ***	0.41
Outside	"Lax" Compliance	Label	3.42	(5, 64)	0.21	3.23 *	0.14
Outside	"Strict" Compliance	Label	3.50	(5, 64)	0.17	4.16 **	0.19
All	Noticed Label	Label	2.59	(5, 119)	0.17	2.98 *	0.07
		Location	0.00	(1, 119)		0.03	0.00
		Label#Loc	1.43	(5, 119)		1.65	0.03
All	Read Some or More	Label	5.28	(5, 119)	0.21	5.03 ***	0.14
		Location	1.77	(1, 119)		8.41 **	0.06
		Label#Loc	0.69	(5, 119)		0.66	0.00
All	Read Most or More	Label	7.99	(5, 119)	0.16	10.06 ***	0.27
		Location	1.97	(1, 119)		12.39 ***	0.09
		Label#Loc	1.82	(5, 119)		2.29 *	0.05
All	"Lax" Compliance	Label	4.49	(5, 119)	0.19	4.73 ***	0.13
		Location	4.70	(1, 119)		24.74 ***	0.17
		Label#Loc	0.53	(5, 119)		0.56	0.00
All	"Strict" Compliance	Label	5.45	(5, 119)	0.19	5.84 ***	0.16
		Location	3.77	(1, 119)		20.17 ***	0.14
		Label#Loc	1.10	(5, 119)		1.18	0.01

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$

TABLE 5: Post Hoc Comparisons of Mean Values (Study 2)

Comparison #	Warning Label	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Comparisons for "Outside Location" ANOVAs Only						
1	All Labels With No Header	81.58%	63.16%	47.37%	60.53%	47.37%
	All Labels With Header	68.75%	18.75%	3.13%	18.75%	6.25%
	Difference	12.83%	44.41% ***	44.24% ***	41.78% ***	41.12% ***
2	17pt Bulleted With No Header	75.00%	58.33%	25.00%	50.00%	33.33%
	17pt Bulleted With Header	45.45%	18.18%	9.09%	27.27%	9.09%
	Difference	29.55%	40.15%	15.91%	22.73%	24.24%
3	17pt Bulleted With No Header	75.00%	58.33%	25.00%	50.00%	33.33%
	20pt Bulleted With No Header	92.31%	84.62%	84.62%	69.23%	61.54%
	Difference	-17.31%	-26.28%	-59.62% ***	-19.23%	-28.21%
Comparisons for complete sample ANOVA						
1	All Labels With No Header	82.35%	69.12%	58.82%	75.00%	61.76%
	All Labels With Header	66.67%	34.92%	14.29%	39.68%	23.81%
	Difference	15.69% t	34.20% ***	44.54% ***	35.32% ***	37.96% ***
2	17pt Bulleted With No Header	86.36%	72.73%	54.55%	72.73%	63.64%
	17pt Bulleted With Header	59.09%	36.36%	18.18%	40.91%	31.82%
	Difference	27.27% t	36.36% *	36.36% **	31.82% *	31.82% *
3	17pt Bulleted With No Header	86.36%	72.73%	54.55%	72.73%	63.64%
	20pt Bulleted With No Header	91.30%	82.61%	78.26%	78.26%	65.22%
	Difference	-4.94%	-9.88%	-23.72%	-5.53%	-1.58%

(*ps>0.10*) on self-reported noticing. Results were similar in the full sample [$F(5, 119) = 2.98$, $p < 0.05$, *omega-squared* $= 0.07$], though post-hoc analyses revealed a marginally significant difference in predicted noticing based upon the presence of an orange warning header (*ps<0.10*).

Reading. Participants were also asked how much (none, some, most, all) of the label they read. Consistent with Shaver et al. (2006), our first model used "read some or more" as the dependent variable. In the "outside" location subsample, this model showed differences in mean values for self-reported reading across label conditions [$F(5, 64) = 4.67$, $p < 0.01$, *omega-squared* $= 0.21$]. The rate of self-reporting reading was highest for the 20-point bulleted text label ($M = 84.62$, $SD = 37.55$) and lowest for labels with an orange warning header and ANSI-style pictogram ($M = 10.10$, $SD = 31.62$). Bonferroni post hoc tests revealed that the three warnings without orange headers had significantly higher rates of self-reported reading than the three warnings with orange headers ($p<0.001$). Warning label format was also shown to have a significant effect on self-reported reading in the full sample [$F(5, 119) = 5.03$, $p < 0.01$, *omega-squared* $= 0.14$]. In addition to finding that collectively warnings without an orange header performed better than those with an orange header ($p<0.001$), Bonferroni post hoc tests in the full sample indicated that the 17-point bulleted font label without an orange header performed significantly better than the 17-point bulleted font label with an orange header ($p<0.05$).

In order to examine the impact of variable definition on findings in warnings research, we also constructed a reading model using "read most or more" as the dependent variable. The effect of label on reading was significant for both the "outside" location subsample [$F(5, 64) = 10.45$, $p < 0.001$, *omega-squared* $= 0.41$] and the full sample [$F(5, 119) = 10.06$, $p < 0.001$, *omega-squared* $= 0.27$]. This tightening of the requirement for self-reported reading resulted in decreased reading rates across conditions, with the exception of the 20-point font warning label, where the rate of self-reported reading was the same as the prior model. The results of Bonferroni post hoc tests were consistent with the prior model with one exception. The revised reading criteria resulted in a significant difference in rates of self-reported reading between the 17-point bulleted font ($M = 25.00, SD = 45.23$) and 20-point bulleted font conditions for participants in the "outside" location condition ($p<0.001$).

Compliance. Because this study was a conceptual replication of the work performed by Frantz and Rhoades (1993), Frantz et al. (2000), and Shaver et al. (2006), we employed the same measures of "lax" and "strict" compliance that had been used in prior studies through the use of two separate dependent variables. Consistent with prior work (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006), participants were considered to be in compliance with the warning label in a "lax" sense if they loaded the bottom drawer first or the weight of files in the bottom drawer was greater than or equal to the weight of files in the top drawer (1 = loaded bottom drawer first or loaded majority of files into bottom drawer, 0 = loaded top drawer first and loaded majority of files into top drawer). They were considered compliance in a "strict" sense if all files were placed in the bottom drawer (1 = all files placed in bottom drawer, 0 = some or all files placed in top drawer).

As shown in *Table 4*, in our ANOVA using "lax" compliance as the dependent variable, the effect of warning label format was significant in both the "outside" location subsample [$F(5, 64) = 3.23, p < 0.05$, *omega-squared* $= 0.14$] and the full sample [$F(5, 119) = 4.73, p < 0.001$, *omega-squared* $= 0.13$]. Compliance rates were highest for the three warning label formats that did not have orange headers. Bonferroni post hoc tests revealed that the rates of compliance were significantly higher for the three warnings without orange headers than they were for the three warnings with orange headers in both the subsample and full sample ($ps<0.001$).

Additionally, in the full sample, the 17-point bulleted font label without an orange header performed better than the 17-point bulleted font label with a header ($p<0.05$). Results using “strict” compliance as the dependent variable were similar.

Location. *Table 4* shows that the effect of warning location was significant for each model ($ps<0.01$) except for predicted noticing ($p>0.10$). Higher rates of self-reported reading and actual compliance were observed in the condition in which the warning was placed on a cardboard bridge inside of the file cabinet, therefore interfering with the task at hand. For the self-reported reading dependent variables, the effect sizes on location as reported in terms of omega-squared (*0.06 and 0.09*) were considerably smaller than the effect sizes on label format (*0.14 and 0.27*). For models with compliance dependent variables, however, the effect sizes on location (*0.17 and 0.14*) were comparable to those on label format (*0.13 and 0.16*) indicating that warning location plays as important a role in compliance as warning label format does. Location did not interact with format to influence compliance outcome variables ($ps>0.10$).

Warning Features. In order to evaluate the effects of the separable features of the ANSI Z535 warning guidelines, we conducted a series of probit regressions testing the relationships between our five dependent variables and the following warning features: bulleted font, font size, an orange warning header, the presence of an image, an ANSI-ISO blended image, and location. Because the between-subjects design of this experiment did not allow us to control for latent variables using a fixed-effects model, we also controlled for subjects’ gender, age, and primary language. The marginal effects, which reflect the change in probability of a binary outcome variable based upon a change in the value of an explanatory variable, are shown in *Table 6*. The model using self-reported noticing as the dependent variable approached significance [$X^2(9) = 15.41, p < 0.10, pseudo\ R^2 = 0.11$]. Models using read some or more [$X^2(9) = 35.26, p < 0.001,$

pseudo R^2 *= 0.22*], read most or more [$X^2(9) = 46.74, p < 0.001$, *pseudo* $R^2 = 0.35$], “lax” compliance [$X^2(9) = 38.20, p < 0.001$, *pseudo* $R^2 = 0.25$], and “strict” compliance [$X^2(9) = 48.50, p < 0.001$, *pseudo* $R^2 = 0.29$] as dependent variables were all significant.

TABLE 6: Marginal Effects (Probit) for Reported Noticing, Reported Reading, and Observed Compliance by Label Features and Location (Study 2)

Outcome Variable	Noticed Label		Read Some or More		Read Most or More		"Lax" Compliance		"Strict" Compliance	
Warning Feature										
Bulleted font	21.35%	(0.16)	27.56% ^{t}	(0.15)	14.56%	(0.12)	-2.22%	(0.17)	11.22%	(0.14)
Size 20 point font	4.65%	(0.14)	5.35%	(0.17)	19.44%	(0.17)	7.53%	(0.16)	-3.04%	(0.16)
Orange header	-29.83% *	(0.13)	-49.81% ***	(0.13)	-53.00% ***	(0.11)	-44.08% **	(0.15)	-48.92% ***	(0.12)
Image present	-1.92%	(0.12)	-9.45%	(0.17)	-3.13%	(0.18)	-1.87%	(0.17)	1.54%	(0.17)
ANSI-ISO blended	21.95% **	(0.07)	22.92%	(0.15)	8.13%	(0.20)	8.43%	(0.16)	-20.23%	(0.15)
Location	-4.81%	(0.08)	18.27% ^{t}	(0.10)	25.58% *	(0.10)	46.55% ***	(0.09)	43.65% ***	(0.10)
Control Variables										
Gender	-1.87%	(0.08)	8.16%	(0.11)	20.63% *	(0.10)	9.43%	(0.10)	27.06% **	(0.10)
Age	-0.48%	(0.00)	-1.50% **	(0.01)	-1.47% *	(0.01)	-0.01%	(0.01)	-0.09%	(0.01)
Primary Language	-3.30%	(0.11)	-14.86%	(0.15)	-14.15%	(0.17)	13.60%	(0.18)	-5.58%	(0.18)
Pseudo R-Squared	0.11		0.22		0.35		0.25		0.29	
Model fit	*X(9) = 15.41* ^{t}		*X(9) = 35.26* ***		*X(9) = 46.74* ***		*X(9) = 38.20* ***		*X(9) = 48.50* ***	

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$ for two-tail test, robust standard errors in parentheses

The results of our probit model show that while location of the warning label did not significantly impact self-reported noticing ($p > 0.10$), the marginal effect on reading some or more of the label approached significance ($\beta = 0.18, SE = 0.10, p < 0.10$), and location had a significant marginal effect on reading most or more ($\beta = 0.26, SE = 0.10, p < 0.05$), “lax” compliance ($\beta = 0.47, SE = 0.09, p < 0.001$), and “strict” compliance ($\beta = 0.44, SE = 0.10, p < 0.001$). The marginal effect of an orange header on the warning label was negative and significant for noticing ($\beta = -0.30, SE = 0.13, p < 0.05$), reading some or more ($\beta = -0.50, SE =$

0.13, p < 0.001), reading most or more (*β = -0.53, SE = 0.11, p < 0.001*), “lax” compliance (*β = -0.44, SE = 0.15, p < 0.01*), and “strict” compliance (*β = -0.49, SE = 0.12, p < 0.001*). The marginal effect of using an ANSI-ISO blended pictogram was significant only for noticing (*β = 0.22, SE = 0.07, p < 0.01*) and the marginal effect of bulleted font on reading some or more approached significance (*β = 0.28, SE = 0.15, p < 0.10*). The marginal effects of all other warning features were non-significant (*ps > 0.10*).

Self-Reported Measures. The fact that location had such a strong marginal effect on compliance while having no relationship with noticing raises questions about how to interpret the meaning of information obtained from self-reported measures of noticing and reading. Similar questions have been raised in the past as Shaver et al., (2006) had only five subjects report that they did not notice a warning label, but when asked why they did not comply with the warning, seven subjects claimed that they did not notice it. To gain insight into how we should interpret self-reported measures of noticing and reading, we examined the relationships between self-reported noticing, self-reported reading, and explanations for not complying with the warning. Several interesting findings emerged and are described below.

When asked to explain why he did not read the entire warning, one subject who initially reported not noticing or reading the warning label claimed that he did read the entire warning. Another subject who initially reported not noticing the warning label claimed to have read some of it. Three subjects who initially reported noticing the warning label later explained that they did not read the entire warning because they did not notice it. Twelve subjects who reported both noticing and reading some of the warning said that they did not read the entire warning label because they thought it was an unimportant part of the packaging. Another subject who initially reported reading none of the label later claimed to have read enough to get the information he needed before he stopped reading. Finally, one subject claimed to have read only the "bold part" of the label, but this subject was not in an experimental condition that included a warning label with any bold font. The conflicted information that arose during this examination suggests that results of tests for self-reported measures should be interpreted cautiously.

Robustness Tests. As described above, although justified by prior research, the use of ANOVA with a binary dependent variable does violate one of the core assumptions associated with the use of this statistical technique. Although a comparison of means was necessary for compatibility with prior research, in order to ensure that violating the assumptions of ANOVA did not influence our results, our analyses were replicated in a non-parametric model. Kruskal-Wallis tests for all five of our dependent variables in both the "outside" location subsample and the full sample were significant ($ps<0.05$). These results are shown in *Appendix X*. Post-hoc Mann-Whitney (*Appendix XI*) tests with Bonferroni corrections were similar to the results reported in *Table 5*. Specifically, in both the "outside" subsample and the full sample, there were significant differences in both measures of reading and both measures of compliance for the three warning labels without orange headers and the three warning labels with orange headers ($ps<0.01$). Additionally, there were significant differences in compliance rates ("lax" and "strict") for participants in the "outside" location and participants in the "inside" location.

In order to address concerns regarding the pooling of participants of different genders, student status, or native languages, we conducted a series of replications of our primary ANOVA model and post-hoc comparisons. ANOVA results from subsamples of students only (*Appendix V*) and native English speakers only (*Appendix VI*) were consistent with our primary analyses. Post-hoc analyses for these ANOVAs (*Appendices VIII and IX*) were consistent with the results presented in *Table 5*. To address concerns regarding the pooling of male and female participants, we replicated *Table 4* introducing a covariate for gender and interacting that covariate with label and location (*Appendix VII*). Consistent with our primary analyses, the effect of label remained significant in all models (*ps*<0.05) and the effect of location was significant for both measures of reading and both measures of compliance (*ps*<0.01). None of the interaction terms that included gender were significant (*ps*>0.10).

Discussion

The results of this study indicate that when warning location interferes with the task at hand, compliance with the warning's instructions increase. Additionally, contrary to our expectations, we found that warning label formats with a standardized orange header had lower levels of self-reported reading and observed compliance than warnings without an orange header. These results were present in an ANOVA model, probit regression, and a non-parametric model as well as a variety of subsample analyses. The consistency of our findings across model specifications and analytical techniques provide support for the robustness of our results.

The insights gained from our finding of a negative relationship between the presence of an orange warning header on a warning label and compliance with the warning's instructions present an important contribution to the warnings literature while also offering a pathway for future research. This finding may only be generalizable to conditions in which the perceived risk of a task is minimal as the mean estimation of likelihood that a future participant in the study would suffer an accident was 0.10 with a standard deviation of 0.12 on a scale of 0.00 to 1.00 with no significant difference across conditions. Based upon this finding, we propose that when the perceived risk of the task at hand is low, as in this instance, participants may view safety-related information as unimportant. In such instances, standardized formatting that draws attention to the fact that a label contains safety-related information may actually reduce the likelihood that an individual will read it, and ultimately, comply with it. Future research should investigate how standardized warnings perform compared to unstandardized warnings as perceptions of risk associated with a task vary.

The contribution of this study is strengthened by the fact that our replication of previous experimental designs allows for direct comparison to prior research. Observed compliance with warning labels with 17-point bulleted text, orange headers, and images in the "outside" location in this study were similar to observed compliance in studies using file cabinets with external warning labels (Frantz & Rhoades, 1993; Frantz et al., 2000; Shaver et al., 2006). This study's observed compliance (strict) for a warning label with 17-point bulleted font and no warning header located inside the top drawer of the file cabinet (100%) was considerably higher than both the observed lax (73%) and strict (53%) levels of compliance reported by Frantz and Rhoades (1993) for that same warning placement and configuration. Additionally, as shown in *Table 3,* compliance with the 20 point bulleted font label of 69.23% that was observed in the present study was considerably higher than the compliance rate of 7.1% observed by Shaver et al. (2006) for that same warning placement and configuration.

Consistent with previous studies (Frantz et al., 2005; Shaver et al., 2006), variation in observed compliance with warning labels in this study did not mirror that of variation in predicted compliance. Unlike Shaver et al. (2006), however, we found that compliance was significantly related to warning label features (We do not provide a detailed comparison of our results to those of Frantz et al. (2005) because their "less ANSI" format was simultaneously placed in two separate locations on the file cabinet whereas their "more ANSI" format was only used in one location). As described above, we found that warnings with recognizable standardized formatting, such as an orange warning header, yielded significantly lower levels of compliance than those without the standardized orange header.

Additionally, consistent with Frantz and Rhoades (1993), we found that both self-reported reading and observed compliance were both significantly impacted by warning location. This finding is interesting for the following reason: ANSI (2011) guidelines call for warning labels to “be readily visible to the intended viewer,” (p. 7) but warning studies often focus exclusively on features of the actual warning label, giving little attention to warning location. This is particularly relevant to this context because despite early findings that warning labels on a bridge inside the top drawer of a file cabinet are more effective than external warnings (Frantz & Rhoades, 1993), subsequent warning research using file cabinets has utilized external warning locations (Frantz et al., 2000; Shaver et al., 2006), thus failing to test different designs in their optimal locations. The findings of this study demonstrate that warning location should serve as a critical component in future warnings research.

Study 3

In our final study, we sought to investigate the apparent misalignment between predicted outcomes and actual behavior reported in Study 1 and Study 2. While this apparent misalignment could serve as the basis for arguing that measures of predicted compliance cannot be used to accurately predict behavior (Frantz et al., 2005; Shaver et al., 2006), in this study, we explore an alternative explanation. An initial goal of the ANSI Z535 series was to standardize the appearance of safety labels to help people differentiate safety related and non-safety related information (Young, Frantz, Rhoades, & Hall, 2006). Consequently, when participants are presented with multiple warning label variants to evaluate, as is commonly done using a within-subjects design, their predicted compliance ratings may be biased toward warnings that appear most like standardized safety labeling. Such a bias would not exist in studies examining actual compliance as these studies generally use a between-subjects design, thus contributing to differences between predicted and actual compliance (Frantz et al., 2005; Shaver et al., 2006). In order to examine this possibility, we replicated Study 1 using a between-subjects design.

Methods

Participants. Participants were recruited using Amazon's Mechanical Turk (mTurk). A total of 508 participants from the United States with ages ranging from 19 to 80 ($M = 37.27, SD = 12.15$) completed the study. The sample consisted of 324 females and 184 males with participants reporting their race/ethnicity as White (73.57%), Black/African-American (8.28%), Asian (5.92%), Hispanic (4.73%), other (1.19%), or two or more races (6.33%). Participants reported a variety of employment conditions with the most common responses being employed full-time only (52.66%), self-employed only (11.44%), employed part-time only (9.66%), unemployed (5.33%), and not working or looking for work (4.73%). Participants' education was reported as graduate degree (16.40%), bachelor's degree (35.56%), some college (35.77%), high school (11.07%), and less than high school (0.20%). There were no significant differences in participants' demographic characteristics across conditions.

Procedure. The procedure used in this study followed that of Study 1 with four notable exceptions. First, because participants were recruited through mTurk, all materials and instructions were presented electronically. Second, the file cabinet was only presented as a two-drawer unit (there was no four-drawer condition). Next, and most importantly, this study followed a between-subjects experimental design, so each participant was only exposed to one warning label format. Finally, because our interest in this study was to compare predicted compliance of warning labels with the actual compliance observed in Study 2, this study only included the six different warning label formats used in Study 2. Once participants accepted the assignment, called a HIT in mTurk, they were directed to click a hyperlink that randomly assigned them to one of six conditions. The number of participants completing the study in each condition ranged from 83 to 86.

As a final note, precautions were taken to identify participants who may not have completed the task with high quality. Observations were flagged if a participant incorrectly entered a confirmation code or completed the task in under 13 seconds. In total, 34 participants were flagged as potentially providing low-quality work. We ran our analysis both with the full sample and excluding these participants and the results were consistent. Results below are reported using the full sample.

Results

Predicted Noticing. The mean value for the predicted number of people out of 100 who would notice the warning label was 58.71 ($SD = 26.88$). An ANOVA testing for differences in these values across conditions was non-significant [$F(5, 502) = 1.03, p > 0.40$], demonstrating that there was no difference in predicted noticing of the warning label across conditions when using a between-subjects experimental design.

Predicted Compliance. The mean value for the predicted number of people out of 100 who would comply with the warning label was 52.55 ($SD = 24.46$). An ANOVA testing for differences in these values across conditions was non-significant [$F(5, 502) = 0.36, p > 0.80$], once again showing no evidence of differences in predicted compliance with the warning label across conditions when using a between-subjects experimental design.

Warning Features. Consistent with our first two studies, we examined the relationships between the separable features of the ANSI Z535.4 guidelines and our dependent variables using linear regressions. Our first model regressed the same explanatory and control variables as used in Study 2 on predicted noticing. This model was non-significant [$F(7, 470) = 1.26, p > 0.20$]. We then re-ran the regression using predicted compliance as the dependent variable. This model was also non-significant [$F(5, 470) = 0.50, p > 0.80$]. For robustness, several alternative models with different control variables were run and results were consistent, providing strong evidence that predicted noticing and compliance were not influenced by warning label features when tested using a between-subjects design.

Discussion

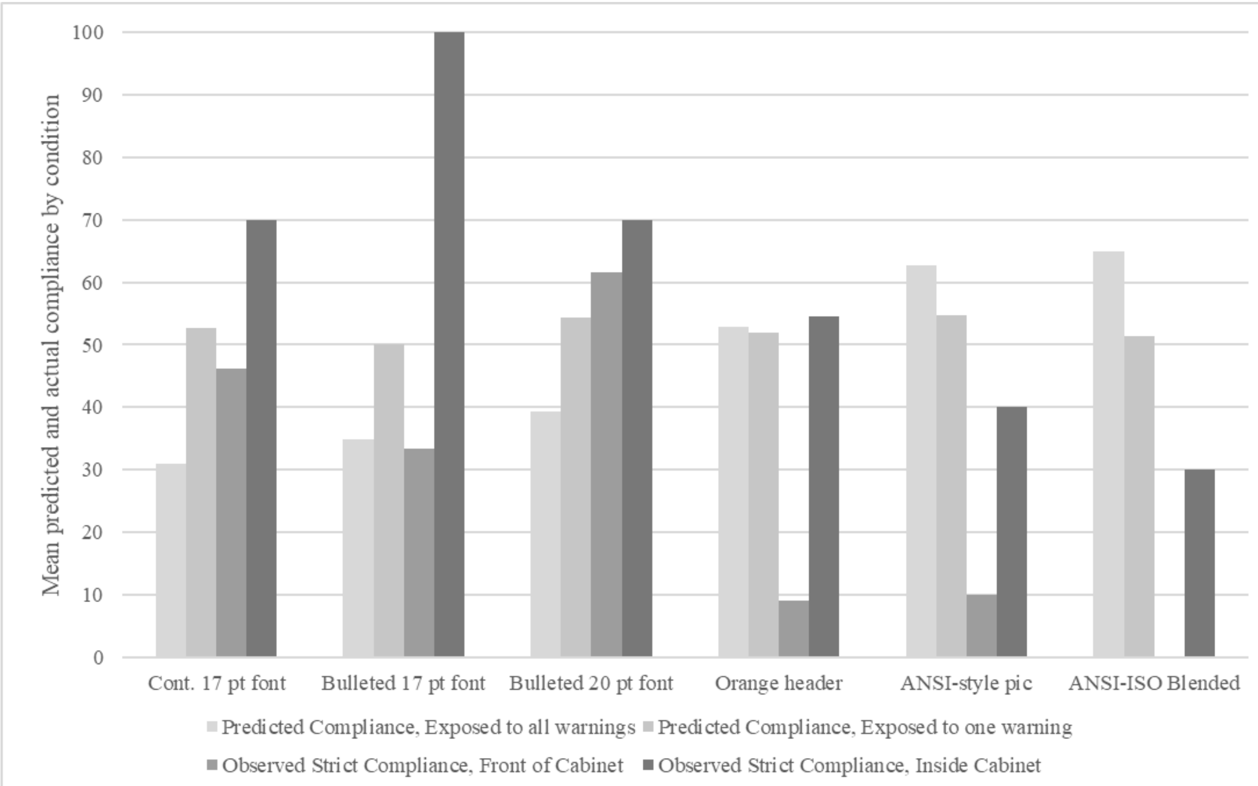
In Study 3, we sought to address the incongruent findings of Study 1 and Study 2. Similar to our findings in these studies, previous research has shown that the impact of different warning label formats on predicted compliance is not consistent with their impact on actual compliance (Frantz et al., 2005). We explored the possibility that differences in findings from Study 1 to Study 2 (as well as other studies that compare predicted compliance with warning labels to actual compliance) were influenced by experimental design as Study 1 utilized a within-subjects design, but Study 2 utilized a between-subjects design. Although a within-subjects design minimizes bias resulting from the differences between subjects exposed to each condition, it also creates threats to internal validity in that participants may identify the true purpose of the experiment, thus altering their behavior (Singleton Jr. & Straits, 2005). This potential threat is particularly relevant to the context of warnings studies due to the standardization of warning labels. Our results showed a very different pattern in predicted compliance in this study than was evident in Study 1. In fact, the results of Study 3 showed no significant difference in predicted noticing or compliance with a warning across different warning label designs, suggesting that reported perceptions of warning effectiveness are strongly influenced by experimental design.

General Discussion

While each of our studies provides interesting insight into the effects of warning label features on levels of both predicted and actual compliance, the most important contribution of this research stems from the integration of the three studies. One of the most interesting observations across studies was that while Study 1 showed that the orange warning header was the warning label feature with the highest predicted positive impact on compliance, this feature actually had a negative relationship with observed compliance in Study 2. As discussed previously, the low perceived risk of injury in Study 2 may have caused standardized formatting to have a negative effect on reading and complying with a warning. It is possible, however, that when measuring estimates of predicted compliance, people may fail to consider how low perceptions of risk will impact the search for safety-related information, and consequently, they may less accurately predict differential outcomes for a variety of warning label formats. Future research could investigate this by looking at the relationship between actual and predicted compliance in both high-risk and low-risk contexts.

Although past research has compared predicted and actual compliance, by measuring predicted compliance through two different experimental designs and comparing the results to recorded measures of compliance for various warning formats in two different locations, as in this research, we are able to provide unique insight into the relationship between predicted and actual compliance. *Figure 2* shows rates of predicted compliance for both experimental conditions and rates of observed strict compliance for both warning label locations by warning label format.

Figure 2: Comparison of Predicted and Actual Compliance



Neither method for measuring predicted compliance accurately reflected observed compliance, but predicted compliance as measured using a between-subjects design was generally closer to measures of observed compliance than was predicted compliance measured through a within-subjects design. In addition to visually reviewing graphs of results, this statement was tested by calculating the differences between predicted compliance in both experimental designs and observed compliance for each location across warning label formats. The standard deviations of these differences indicate that the variation in differences between predicted compliance and observed compliance across warning conditions was lower when using predicted compliance as measured in the between-subjects study. These results were consistent for both “lax” and “strict” measures of observed compliance. This appears to suggest that the relationship between predicted compliance and actual behavioral compliance is strongly influenced by how predicted compliance is measured.

Building on this, our Study 3 finding that predicted compliance did not vary based upon warning label format may lead one to the conclusion that warning format is not perceived as playing an important role in compliance and that the relationship between predicted compliance in a between-subjects design and actual compliance was simply driven by the fact that only one warning feature influenced actual compliance rates. Such a conclusion, however, would be drawn in haste. Participants in the between-subjects predicted compliance study were directed to look at a specific warning label, and therefore they were unable to replicate the behavior of laboratory subjects in the actual compliance study who may have dismissed standardized warnings when they were not searching for safety related information. Further, our findings came from a setting where perceptions of risk were low, and people were habituated to the processes being observed. Both predicted effectiveness and actual effectiveness of warning label features could differ in an alternative context. In addition to exploring how measures of predicted compliance can be collected to further improve upon their accuracy, as discussed above, future research should also investigate how experimental context influences the relationship between predicted compliance and observed behavioral compliance with warnings.

It should also be noted that while measures of predicted compliance frequently overestimated behavioral compliance when a warning label was positioned on the exterior of the file cabinet, measures of predicted compliance actually tended to underestimate “strict” compliance with warning labels when warnings were positioned inside of the top drawer of the file cabinet. While past studies measuring estimates of predicted compliance have communicated the location of the warning label (Frantz et al., 2005; Shaver et al., 2006), the locations have not been varied. Both this study and past research (Frantz & Rhoades, 1993) have shown evidence that warning location influences behavioral compliance, but this research stream provides little insight as to how the communicated location of a warning label influences estimates of predicted compliance. Consequently, differences between estimates of predicted compliance and observations of behavioral compliance may be influenced by how individuals incorporate information about warning location into their estimates, as opposed to being strictly driven by perceptions of the effectiveness of different warning label formats. This research stream would benefit from future research that tests the effects of communicated warning location on predicted estimates of compliance.

Building upon the importance of warning label location within the context of our research, Figure 2 provides some evidence that differences in compliance rates across different label formats may vary based upon label location. For example, for participants in the “inside” location condition, compliance rates for 17-point bulleted text label condition were higher than compliance rates for the 17-point continuous text or 20-point bulleted text conditions. Conversely, for participants in the “outside” location condition, compliance rates for 17-point bulleted text label condition were lower than compliance rates for the 17-point continuous text or 20-point bulleted text conditions. This observation may suggest that the optimal warning label design is contingent upon the intended placement of the warning. Our data analysis does not support this argument, however, as *Table 4* shows that the interaction between warning location and label format was not statistically significant for our behavioral compliance variables. Even so, consideration of the significant interaction of location and format for the dependent variable of self-reported reading along the pattern observed in *Figure 2* suggests that future research should explore this possibility further.

One limitation of this research comes from the fact that our samples in Studies 1 and 2 comprised predominantly college students and thus our results may have limited generalizability. This concern could be eased by the fact that in Study 2 we found no significant difference in the mean compliance rates of students and non-students. Despite this, because the non-students in our sample came from the same private college, generalizability could be limited. When considered within the context of prior research, however, with a few exceptions as described above, *Tables 1 and 3* indicate that our findings are largely consistent with the results of previous studies. This observation is relevant to generalizability of results as Shaver et al. (2006) and Frantz et al. (2000) drew their samples from the general population while Frantz and Rhoades (1993) included both students *and* custodians in their sample. Consequently, the results of our experiment using predominantly student participants to not appear to differ meaningfully from studies that have utilized members of the general population. Even so, if the generalizability of our findings was considered to be limited to student populations, that would still be highly relevant to our knowledge of warning effectiveness as it is likely that individuals with limited experience using office equipment have less knowledge regarding safety issues pertaining such equipment and thus may be at greater risk of accidents resulting from misuse due to lack of information.

Finally, this study offers a very important contribution by investigating the effects of a variety of components to the ANSI safety system. By testing the effectiveness of features rather than simply comparing ANSI-style warnings to more generic warnings, we have opened the door to a stream of research that can make meaningful contributions to the refinement of this system and the development of more effective warnings. These results should also place entities responsible for product safety (e.g., product and equipment manufacturers and distributors) on notice that the best way to ensure warning effectiveness is to carry out systematic testing on these materials with a representative sample of the target audience in advance of their implementation.

Conclusion

This study compared measures of predicted compliance with warning labels to observations of actual compliance. Our findings indicate that the experimental design for predicted compliance studies plays an important role in the relationship between predicted compliance and actual behavior. Further, the important role of warning location in Study 2 indicates that warnings should be tested in a variety of locations and under a variety of conditions in order to accurately assess their effectiveness. Finally, by examining the effects of individual warning features, we have shown how researchers can treat ANSI safety standards as an integrative system in order to support meaningful contributions to safety research.

KEY POINTS

- Warnings guidelines should be treated as collections of separable features rather than monolithic recipes.
- Experimental design and analytic methods can strongly influence the outcomes associated with warnings studies.
- The best way to ensure warning effectiveness is to carry out systematic testing on warning materials with a representative sample of the target audience.

REFERENCES

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

ANSI. (1991). American National Standard for Product Safety Signs and Labels.

ANSI. (1998). American National Standard for Product Safety Signs and Labels.

ANSI. (2007). American National Standard for Product Safety Signs and Labels.

ANSI. (2011). *ANSI Z535.4-2011 American National Standard for Product Safety Signs and Labels*.

Arghami, S., Kian, M. P., & Mohammadfam, I. (2009). Effects of safety signs on the modification of unsafe behaviours. *ZUMS Journal*, *17*(68), 93–98.

Barrett, M. J., & Kott, D. R. (1994). Warnings: Duty to retrofit / Subsequent duty to warn. In J. L. McGoldrick & F. T. Smith (Eds.), *New Jersey Product Liability Law* (pp. 97–125). Newark: New Jersey Law Journal Books.

Borade, A. B., Bansod, S. V., & Gandhewar, V. R. (2008). Hazard perception based on safety words and colors: An Indian perspective. *International Journal of Occupational Safety and Ergonomics*, *14*(4), 407–416. http://doi.org/10.1080/10803548.2008.11076777

Braun, C. C., & Silver, N. C. (1995). Interaction of signal word and colour on warning labels: differences in perceived hazard and behavioural compliance. *Ergonomics*, *38*(11), 2207–2220.

Chatla, S., & Shmueli, G. (2016). Linear probability models (LPM) and big data : The good, the bad, and the ugly. *Indian School of Business Research Paper Series*.

Cheatham, D. B., & Wogalter, M. S. (1999). Connoted hazard and perceived conspicuity of warning configurations. *Proceedings of the Human Factors and Ergonomics Society*, *43*(16), 883–887.

City News Service. (2015, October 2). Canyon country couple sue Kohl's, pressure cooker maker after explosion. *Los Angeles Daily News*.

Dingus, T. A., Wreggit, S. S., & Hathaway, J. A. (1993). Warning variables affecting personal protective equipment use. *Safety Science*, *16*(5–6), 655–673. http://doi.org/10.1016/0925-7535(93)90029-D

Field, A. (2012). *Discovering statistics using IBM SPSS Statistics*. (M. Carmichael, Ed.) (Fourth). Washington, DC: Sage Publications.

Frantz, J. P., & Rhoades, T. P. (1993). A task-analytic approach to the temporal and spatial placement of product warnings. *Human Factors*, *35*(4), 719–730.

Frantz, J. P., Rhoades, T. P., Young, S. L., & Schiller, J. A. (2000). Assessing the effects of adding messages to warning labels. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting*, *44*, 818–821. http://doi.org/10.1177/1541931200044028113

Frantz, J. P., Young, S. L., Rhoades, T. P., & Wisniewski, E. C. (2005). Predicted versus actual response to warning signs and labels: Examining the role of ANSI Z535 features. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, *49*, 1785–1789.

Gaito, J. (1980). Measurement scales and statistics: Resurgence of an old misconception. *Psychological Bulletin*, *87*(3), 564–567. http://doi.org/10.1037/0033-2909.87.3.564

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288. http://doi.org/10.3102/00346543042003237

Hathaway, J. A., & Dingus, T. A. (1992). The effects of compliance cost and specific consequence information on the use of safety equipment. *Accident Analysis and Prevention*, *24*(6), 577–584.

Heckman, G. M., Harley, E. M., Scher, I., & Young, D. E. (2010). Helmet use in sledding: Do users comply with manufacturer warnings? *Proceedings of the Human Factors and Ergonomics Society 54h Annual Meeting*, *54*, 733–737.

Hunn, B. P., & Dingus, T. A. (1992). Interactivity, information, and compliance cost in a consumer product warning scenario. *Accident Analysis and Prevention*, *24*(5), 497–505. http://doi.org/10.1016/0001-4575(92)90058-Q

Huntley-Fenner, G., Harley, E., Trachtman, D., & Young, D. (2007). ANSI Z535.6 and Conspicuity: A Test of the New State of the Art Format for Instructions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *51*(17), 1029–1033. http://doi.org/10.1177/154193120705101707

Kalsher, M. J., Gallo, J. B., Williams, K. J., & Wogalter, M. S. (2000). High levels of behavioral compliance in a realistic product assembly task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(28), 822–825. http://doi.org/10.1177/154193120004402814

Kalsher, M. J., Obenauer, W. G., & Weiss, C. F. (2016). Evaluating ANSI Z535-formatted warning labels as an integrative system. In *Proceedings of the Human Factors and Ergonomics Society 2016 Annual Meeting* (pp. 1642–1646). Washington, DC.

Kalsher, M. J., Wogalter, M. S., Brewster, B. M., & Spunar, M. E. (1995). Hazard level perceptions of current and proposed warning sign and label panels. *Proceedings of the Human Factors and Ergonomics Society*, *39*(5).

Kim, S., & Wogalter, M. S. (2009). Habituation, dishabituation, and recovery effects in visual warnings. *Human Factors and Ergonomics Society 53rd Annual Meeting Proceedings*, *53*, 1612–1616. http://doi.org/10.1518/107118109X12524444080675

Laughery, K. R., Paige, D. L., Laughery, B. R., Wogalter, M. S., Kalsher, M. J., & Leonard, S. D. (2002). Guidelines for warnings design: Do they matter? *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, *46*, 1708–1712. http://doi.org/10.1177/154193120204601904

Laughery, K. R., & Wogalter, M. S. (2006). Designing effective warnings. *Reviews of Human Factors and Ergonomics*, *2*(1), 241–271. http://doi.org/10.1177/1557234X0600200109

Laughery, K. R., & Wogalter, M. S. (2014). A three-stage model summarizes product warning and environmental sign research. *Safety Science*, *61*, 3–10. http://doi.org/10.1016/j.ssci.2011.02.012

Lehto, M. R. (1992). Designing warning signs and warning labels: Part II - Scientific basis for initial guidelines. *International Journal of Lndustrial Ergonomics*, *10*, 115–138. http://doi.org/10.1016/S1572-347X(00)80021-X

Lesch, M. F., Rau, P.-L. P., & Choi, Y. (2016). Effects of culture (China vs. US) and task on perceived hazard: Evidence from product ratings, label ratings, and product to label matching. *Applied Ergonomics*, *52*, 43–53. http://doi.org/10.1016/j.apergo.2015.06.010

Lesch, M. F., Rau, P.-L. P., Zhao, Z., & Liu, C. (2009). A cross-cultural comparison of perceived hazard in response to warning components and configurations: US vs. China. *Applied Ergonomics*, *40*(5), 953–961. http://doi.org/10.1016/j.apergo.2009.02.004

Levy Konigsberg LLP. (2012). LK wins historic youth smoking verdict in NY against RJ Reynolds on behalf of family of lung cancer victim. Retrieved from http://www.levylaw.com/lung-cancer-lawyers.php

Loring, B. A., & Wiklund, M. E. (1988). Improving swimming pool warning signs. *Proceedings of the Human Factors and Ergonomics Society*, *32*(15), 910–914.

Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, *23*(1), 151–169. http://doi.org/10.1146/annurev.publheath.23.100901.140546

Matthews, B., Andronaco, R., & Adams, A. (2014). Warning signs at beaches: Do they work? *Safety Science*, *62*, 312–318. http://doi.org/10.1016/j.ssci.2013.09.003

Nadolny, T. (2015, May 28). IKEA sued over dresser tip-over that killed Chesco toddler. *Philly.Com*.

Rogers, W. A., Lamson, N., & Rousseau, G. K. (2000). Warning research: An integrative perspective. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *42*(1), 102–139. http://doi.org/10.1518/001872000779656624

Sandler, L. (2015, April 15). GM can't be sued for making faulty switches. *Bloomberg Business*.

Shaver, E. F., & Braun, C. C. (2000). Effects of warning symbol explicitness and warning color on behavioral compliance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(27), 290–293. http://doi.org/10.1177/154193120004402702

Shaver, E. F., Young, S. L., Frantz, J. P., Rhoades, T. P., Hall, S. M., & Shah, R. J. (2006). Comparison of ANSI and ISO standard formats on people's response to product warnings. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, *50*, 2197–2201.

Singleton Jr., R. A., & Straits, B. C. (2005). *Approaches to social research* (Fourth). New York: Oxford University Press.

Smith-Jackson, T. L., & Essuman-Johnson, A. (2002). Cultural ergonomics in Ghana, West Africa: A descriptive survey of industry and trade workers' interpretations of safety symbols. *International Journal of Occupational Safety and Ergonomics*, *8*(1), 37–50. http://doi.org/10.1080/10803548.2002.11076513

Smith, J. J., & Wogalter, M. S. (2010). Behavioral compliance to in-manual and on-product warnings. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, *54*, 1846–1850. http://doi.org/10.1177/154193121005402104

Stock, J. H., & Watson, M. W. (2015). *Introduction to econometrics* (Third edit). New York: Pearson.

Watts, E. (2015, October 7). Sen. Harry Reid suing exercise band maker over eye injury. *Fox 5 Vegas*. Las Vegas.

Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics*, *33*(3), 219–230. http://doi.org/10.1016/S0003-6870(02)00009-1

Wogalter, M. S., Kalsher, M. J., Frederick, L. J., Magurno, A. B., & Brewster, B. M. (1998). Hazard level perceptions of warning component and configurations. *International Journal of Cognitive Ergonomics*.

Wogalter, M. S., Kalsher, M. J., & Racicot, B. M. (1993). Behavioral compliance with warnings: effects of voice, context, and location. *Safety Science*, *16*(5–6), 637–654. http://doi.org/10.1016/0925-7535(93)90028-C

Wogalter, M. S., Paine, C. S., Mills, B. J., Carolina, N., Carolina, N., & Smith-jackson, T. L. (1999). Application of Cognitive Principles to the Design of Direct-to-Consumer Advertising of Prescription Medications. *Proceedings of the Human Factors and Ergonomics Society*, *43*(6), 515–519.

Young, S. L., Frantz, J. P., Rhoades, T. P., & Darnell, K. R. (2002). Does compliance with ANSI Z535 increase compliance with warnings ? *Professional Safety*, (September), 18–23.

Young, S. L., Frantz, J. P., Rhoades, T. P., & Hall, S. M. (2006). Development and objectives of the ANSI Z535 series of standards for safety signs and colors: A historical perspective. In M. S. Wogalter (Ed.), *Handbook of warnings* (pp. 445–454). Mahwah, NJ: Lawrence Erlbaum.

Yu, R., Chan, A. H. ., & Salvendy, G. (2004). Chinese perceptions of implied hazard for signal words and surround shapes. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *14*(1), 69–80.

Appendix I: Summary Statistics and Correlation Table (Study 1)

	Warning Style	Mean Values	Correlations 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	Continuous 17 point font, noticed	37.44 *(23.30)*	1.00															
2	Continuous 17 point font, followed	30.88 *(22.65)*	0.56	1.00														
3	Continuous 20 point font, noticed	41.95 *(23.57)*	0.95	0.57	1.00													
4	Continuous 20 point font, followed	34.42 *(23.28)*	0.56	0.96	0.61	1.00												
5	Bulleted 17 point font, noticed	41.62 *(23.87)*	0.86	0.50	0.83	0.50	1.00											
6	Bulleted 17 point font, followed	34.84 *(23.12)*	0.50	0.87	0.52	0.86	0.55	1.00										
7	Bulleted 20 point font, noticed	46.69 *(23.79)*	0.82	0.50	0.82	0.53	0.93	0.56	1.00									
8	Bulleted 20 point font, followed	39.26 *(23.49)*	0.48	0.85	0.52	0.86	0.52	0.94	0.57	1.00								
9	Orange header, bulleted 17 point font, noticed	66.26 *(22.20)*	0.63	0.41	0.67	0.44	0.58	0.37	0.63	0.37	1.00							
10	Orange header, bulleted 17 point font, followed	52.88 *(23.35)*	0.44	0.77	0.48	0.79	0.41	0.75	0.46	0.77	0.62	1.00						
11	ANSI-style picture formatting, noticed	75.49 *(20.57)*	0.51	0.35	0.55	0.37	0.49	0.33	0.53	0.34	0.86	0.56	1.00					
12	ANSI-style picture formatting, followed	62.85 *(23.61)*	0.34	0.66	0.39	0.68	0.32	0.66	0.36	0.69	0.55	0.89	0.64	1				
13	Mod ANSI-style picture formatting, noticed	75.57 *(20.67)*	0.47	0.29	0.51	0.32	0.52	0.33	0.59	0.34	0.80	0.52	0.84	0.54	1			
14	Mod ANSI-style picture formatting, followed	63.96 *(23.48)*	0.31	0.60	0.35	0.62	0.32	0.67	0.39	0.70	0.53	0.83	0.57	0.87	0.62	1		
15	ANSI-ISO blended pic formatting, noticed	77.52 *(20.68)*	0.45	0.29	0.49	0.31	0.47	0.31	0.51	0.32	0.79	0.53	0.85	0.57	0.86	0.56	1	
16	ANSI-ISO blended pic formatting, followed	64.90 *(23.63)*	0.31	0.57	0.35	0.60	0.29	0.63	0.34	0.67	0.54	0.82	0.58	0.87	0.55	0.88	0.59	1

Standard deviation of the mean in parentheses, all reported correlations significant at p<0.001

Appendix II: Subjects per Condition (Study 2)

		Warning Style					
Warning Location		Continuous 17 point font	Bulleted 17 point font	Bulleted 20 point font	Orange header	ANSI-style pic	ANSI-ISO blended
Outside of Cabinet	Female	8	5	6	6	3	5
	Non-native English	2	0	1	1	0	1
	Non-student	5	4	2	4	2	2
	Overall Sample	13	12	13	11	10	11
Inside of Cabinet	Female	4	5	6	6	4	5
	Non-native English	3	0	1	2	3	0
	Non-student	0	0	0	0	0	0
	Overall Sample	10	10	10	11	10	10
All Locations (Overall)	Female	12	10	12	12	7	10
	Non-native English	5	0	2	3	3	1
	Non-student	5	4	2	4	2	2
	Overall Sample	23	22	23	22	20	21

Appendix III: Differences in mean values and standard deviation by demographic characteristics (Study 2)

	Dependent Variable				
	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Male Mean Value	0.76	0.51	0.31	0.53	0.32
Female Mean Value	0.73	0.54	0.44	0.63	0.56
Mean Comparison	T(129) = 0.45	T(129) = -0.28	T(129) = -1.61	T(129) = -1.22	T(129) = -2.73**
Male Std Dev	0.43	0.50	0.47	0.50	0.47
Female Std Dev	0.45	0.50	0.50	0.49	0.50
Variance Comparison	F(67,62) = 0.91	F(67,62) = 1.00	F(67,62) = 0.86	F(67,62) = 1.07	F(67,62) = 0.89
Non-Native English Speaker Mean Value	0.71	0.64	0.50	0.57	0.57
Native English Speaker Std Dev	0.75	0.51	0.36	0.58	0.42
Mean Comparison	T(129) = -0.31	T(129) = 0.92	T(129) = 1.03	T(129) = -0.07	T(129) = 1.09
Non-Native English Speaker Mean Value	0.47	0.50	0.52	0.51	0.51
Native English Speaker Std Dev	0.43	0.50	0.48	0.50	0.50
Variance Comparison	F(13, 116) = 1.17	F(13, 116) = 0.98	F(13, 116) = 1.16	F(13, 116) = 1.07	F(13, 116) = 1.07
Non-Student Mean Value$^{\#}$	0.63	0.32	0.16	0.47	0.32
Student Mean Value$^{\#}$	0.80	0.47	0.31	0.39	0.27
Mean Comparison	T(68) = -1.50	T(68) = -1.16	T(68) = -1.30	T(68) = 0.61	T(68) = 0.34
Non-Student Std Dev$^{\#}$	0.50	0.48	0.37	0.51	0.48
Student Std Dev$^{\#}$	0.40	0.50	0.47	0.49	0.45
Variance Comparison	F(18, 50) = 1.53	F(18, 50) = 0.90	F(18, 50) = 0.64	F(18, 50) = 1.08	F(18, 50) = 1.12

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$ for two-tail test

$^{\#}$Values reported for "label outside file cabinet" condition only as all "non-students" were exposed to this condition

Appendix IV: Means, Standard Deviations, and Homogeneity of Variance Tests (Study 2)

Warning Location	Dependent Variable	N	Mean	Std Dev	Bartlett's Test (Chi-Sq)	Levene's Test-Mean (F)	Levene's Test-Median (F)
Outside	Noticed Label	70	0.76	0.43	4.85	13.79 ***	2.09 ^{t}
Outside	Read Some or More	70	0.43	0.50	3.55	4.65 **	1.31
Outside	Read Most or More	70	0.27	0.45	2.53	10.22 ***	1.60
Outside	"Lax" Compliance	70	0.41	0.50	2.87	4.45 **	1.38
Outside	"Strict" Compliance	70	0.29	0.46	5.18	19.04 ***	2.48 *
All	Noticed Label	131	0.75	0.44	7.54	12.94 ***	2.08 *
All	Read Some or More	131	0.53	0.50	6.86	4.95 ***	1.04
All	Read Most or More	131	0.37	0.49	5.33	7.34 ***	1.11
All	"Lax" Compliance	131	0.58	0.50	7.31	8.77 ***	1.69 ^{t}
All	"Strict" Compliance	131	0.44	0.50	5.87	16.89 ***	1.67 ^{t}

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{t}p<.10$

Appendix V: Replication of ANOVA Model Summaries for Reported Noticing, Reported Reading, and Observed Compliance by Label Format for Students Only (Study 2)

Warning Location	Dependent Variable	Source	Partial SS	df	MSE	F	omega-squared
Outside	Noticed Label	Label	1.79	(5, 45)	0.14	2.58 *	0.14
Outside	Read Some or More	Label	3.71	(5, 45)	0.20	3.71 **	0.21
Outside	Read Most or More	Label	4.74	(5, 45)	0.14	6.83 ***	0.37
Outside	"Lax" Compliance	Label	3.13	(5, 45)	0.20	3.13 *	0.18
Outside	"Strict" Compliance	Label	3.20	(5, 45)	0.15	4.13 **	0.24
All	Noticed Label	Label	2.41	(5, 100)	0.16	2.96 *	0.09
		Location	0.06	(1, 100)		0.38	0.00
		Label#Loc	1.21	(5, 100)		1.48	0.02
All	Read Some or More	Label	4.89	(5, 100)	0.21	4.57 ***	0.15
		Location	1.00	(1, 100)		4.66 *	0.04
		Label#Loc	0.49	(5, 100)		0.46	0.00
All	Read Most or More	Label	7.26	(5, 100)	0.18	8.28 ***	0.26
		Location	1.21	(1, 100)		6.93 **	0.06
		Label#Loc	1.07	(5, 100)		1.22	0.01
All	"Lax" Compliance	Label	3.97	(5, 100)	0.18	4.39 **	0.14
		Location	4.19	(1, 100)		23.23 ***	0.18
		Label#Loc	0.90	(5, 100)		1.00	0.00
All	"Strict" Compliance	Label	4.96	(5, 100)	0.18	5.40 ***	0.17
		Location	3.31	(1, 100)		17.99 ***	0.14
		Label#Loc	1.28	(5, 100)		1.39	0.02

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$

Appendix VI: Replication of ANOVA Model Summaries for Reported Noticing, Reported Reading, and Observed Compliance by Label Format for Native English Speakers Only (Study 2)

Warning Location	Dependent Variable	Source	Partial SS	df	MSE	F	omega-squared
Outside	Noticed Label	Label	1.81	(5, 59)	0.17	2.09 ^{t}	0.08
Outside	Read Some or More	Label	4.37	(5, 59)	0.19	4.52 **	0.22
Outside	Read Most or More	Label	5.56	(5, 59)	0.12	9.37 ***	0.40
Outside	"Lax" Compliance	Label	2.97	(5, 59)	0.22	2.74 *	0.12
Outside	"Strict" Compliance	Label	2.90	(5, 59)	0.17	3.39 **	0.16
All	Noticed Label	Label	2.34	(5, 105)	0.17	2.70 *	0.07
		Location	0.01	(1, 105)		0.07	0.00
		Label#Loc	1.38	(5, 105)		1.59	0.03
All	Read Some or More	Label	5.18	(5, 105)	0.21	5.03 ***	0.15
		Location	1.38	(1, 105)		6.68 *	0.05
		Label#Loc	0.74	(5, 105)		0.72	0.00
All	Read Most or More	Label	7.20	(5, 105)	0.15	9.37 ***	0.28
		Location	1.43	(1, 105)		9.34 **	0.07
		Label#Loc	1.90	(5, 105)		2.48 *	0.06
All	"Lax" Compliance	Label	4.14	(5, 105)	0.19	4.41 **	0.13
		Location	4.11	(1, 105)		21.91 ***	0.16
		Label#Loc	0.38	(5, 105)		0.40	0.00
All	"Strict" Compliance	Label	5.32	(5, 105)	0.18	5.84 ***	0.18
		Location	2.91	(1, 105)		15.98 ***	0.12
		Label#Loc	1.07	(5, 105)		1.18	0.01

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{\dagger}p<.10$

Appendix VII: Replication of ANOVA Model Summaries for Reported Noticing, Reported Reading, and Observed Compliance by Label Format and Location with Gender Controls and Interactions (Study 2)

Warning Location	Dependent Variable	Source	Partial SS	df	MSE	F	omega-squared
Outside	Noticed Label	Label	2.23	(5, 58)	0.18	2.51 *	0.11
		Gender	0.00	(1, 58)		0.00	0.00
		Label#Gen	0.31	(5, 58)		0.35	0.00
Outside	Read Some or More	Label	4.53	(5, 58)	0.20	4.46 **	0.22
		Gender	0.01	(1, 58)		0.04	0.00
		Label#Gen	0.76	(5, 58)		0.75	0.00
Outside	Read Most or More	Label	6.28	(5, 58)	0.11	10.96 ***	0.44
		Gender	0.02	(1, 58)		0.16	0.00
		Label#Gen	0.94	(5, 58)		1.64	0.05
Outside	"Lax" Compliance	Label	2.93	(5, 58)	0.22	2.64 *	0.12
		Gender	0.16	(1, 58)		0.72	0.00
		Label#Gen	0.55	(5, 58)		0.49	0.00
Outside	"Strict" Compliance	Label	3.54	(5, 58)	0.16	4.48 **	0.22
		Gender	0.59	(1, 58)		3.73 t	0.04
		Label#Gen	1.03	(5, 58)		1.30	0.02
All	Noticed Label	Label	3.04	(5, 112)	0.17	3.48 **	0.10
		Location	0.00	(1, 112)		0.00	0.00
		Gender	0.13	(1, 112)		0.77	0.00
		Label#Gen	0.96	(5, 112)		1.09	0.00
		Loc#Gen	0.08	(1, 112)		0.46	0.00
		Label#Loc	1.40	(5, 112)		1.60	0.03

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{t}p<.10$

Appendix VII (continued)

Warning Location	Dependent Variable	Source	Partial SS	df	MSE	F	omega-squared
All	Read Some or More	Label	5.77	(5, 112)	0.21	5.47^{***}	0.16
		Location	1.87	(1, 112)		8.86^{**}	0.07
		Gender	0.01	(1, 112)		0.04	0.00
		Label#Gen	1.29	(5, 112)		1.22	0.01
		Loc#Gen	0.02	(1, 112)		0.09	0.00
		Label#Loc	0.77	(5, 112)		0.73	0.00
All	Read Most or More	Label	7.96	(5, 112)	0.15	10.49^{***}	0.29
		Location	2.07	(1, 112)		13.64^{***}	0.10
		Gender	0.37	(1, 112)		2.47	0.01
		Label#Gen	1.28	(5, 112)		1.69	0.03
		Loc#Gen	0.23	(1, 112)		1.54	0.00
		Label#Loc	2.13	(5, 112)		2.80^{*}	0.07
All	"Lax" Compliance	Label	4.63	(5, 112)	0.20	4.70^{***}	0.14
		Location	4.65	(1, 112)		23.59^{***}	0.17
		Gender	0.15	(1, 112)		0.75	0.00
		Label#Gen	0.33	(5, 112)		0.34	0.00
		Loc#Gen	0.01	(1, 112)		0.04	0.00
		Label#Loc	0.51	(5, 112)		0.52	0.00
All	"Strict" Compliance	Label	5.41	(5, 112)	0.18	5.98^{***}	0.18
		Location	3.58	(1, 112)		19.76^{***}	0.14
		Gender	1.37	(1, 112)		7.58^{**}	0.06
		Label#Gen	0.47	(5, 112)		0.52	0.00
		Loc#Gen	0.03	(1, 112)		0.15	0.00
		Label#Loc	1.09	(5, 112)		1.20	0.01

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{t}p<.10$

Appendix VIII: Replication of Post Hoc Comparisons of Mean Values for Students Only

Comparison #	Warning Label	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Comparisons for "Outside Location" ANOVAs Only						
1	All Labels With No Header	88.89%	70.37%	55.56%	59.26%	48.15%
	All Labels With Header	70.83%	20.83%	4.17%	16.67%	4.17%
	Difference	18.06%	49.54% **	51.39% ***	42.59% **	43.98% **
2	17pt Bulleted With No Header	87.50%	75.00%	37.50%	50.00%	25.00%
	17pt Bulleted With Header	42.86%	28.57%	14.29%	42.86%	14.29%
	Difference	44.64% t	46.43%	23.21%	7.14%	10.71%
3	17pt Bulleted With No Header	87.50%	75.00%	37.50%	50.00%	25.00%
	20pt Bulleted With No Header	90.91%	81.82%	81.82%	63.64%	54.55%
	Difference	-3.41%	-6.82%	-44.32% *	-13.64%	-29.55%
Comparisons for complete sample ANOVA						
1	All Labels With No Header	85.96%	73.68%	64.91%	77.19%	64.91%
	All Labels With Header	67.27%	38.18%	16.36%	41.82%	25.45%
	Difference	18.69% *	35.50% ***	48.55% ***	35.37% ***	39.46% ***
2	17pt Bulleted With No Header	94.44%	83.33%	66.67%	77.78%	66.67%
	17pt Bulleted With Header	61.11%	44.44%	22.22%	50.00%	38.89%
	Difference	33.33% *	38.89% *	44.44% **	27.78%	27.78%
3	17pt Bulleted With No Header	94.44%	83.33%	66.67%	77.78%	66.67%
	20pt Bulleted With No Header	90.48%	80.95%	76.19%	76.19%	61.90%
	Difference	3.97%	2.38%	-9.52%	1.59%	4.76%

Appendix IX: Replication of Post Hoc Comparisons of Mean Values for Native English Speakers Only

Comparison #	Warning Label	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Comparisons for "Outside Location" ANOVAs Only						
1	All Labels With No Header	80.00%	62.86%	45.71%	60.00%	45.71%
	All Labels With Header	70.00%	16.67%	3.33%	20.00%	6.67%
	Difference	10.00%	46.19% ***	42.38% ***	40.00% **	39.05% **
2	17pt Bulleted With No Header	75.00%	58.33%	25.00%	50.00%	33.33%
	17pt Bulleted With Header	50.00%	20.00%	10.00%	30.00%	10.00%
	Difference	25.00%	38.33%	15.00%	20.00%	23.33%
3	17pt Bulleted With No Header	75.00%	58.33%	25.00%	50.00%	33.33%
	20pt Bulleted With No Header	91.67%	83.33%	83.33%	66.67%	58.33%
	Difference	-16.67%	-25.00%	-58.33% ***	-16.67%	-25.00%
Comparisons for complete sample ANOVA						
1	All Labels With No Header	81.97%	68.85%	57.38%	75.41%	60.66%
	All Labels With Header	67.86%	32.14%	12.50%	39.29%	21.43%
	Difference	14.11%	36.71% ***	44.88% ***	36.12% ***	39.23% ***
2	17pt Bulleted With No Header	86.36%	72.73%	54.55%	72.73%	63.64%
	17pt Bulleted With Header	63.16%	36.84%	15.79%	42.11%	31.58%
	Difference	23.21%	35.89% *	38.76% **	30.62% t	32.06% *
3	17pt Bulleted With No Header	86.36%	72.73%	54.55%	72.73%	63.64%
	20pt Bulleted With No Header	90.48%	80.95%	76.19%	80.95%	66.67%
	Difference	-4.11%	-8.23%	-21.65%	-8.23%	-3.03%

Appendix X: Kruskal-Wallis Model Summaries for Reported Noticing, Reported Reading, and Observed Compliance by Label Format and Location

Warning Location	Dependent Variable	N	df	Chi-Sq
Outside	Noticed Label	70	5	12.13 *
Outside	Read Some or More	70	5	18.45 **
Outside	Read Most or More	70	5	31.02 ***
Outside	"Lax" Compliance	70	5	13.90 *
Outside	"Strict" Compliance	70	5	16.92 **
All	Noticed Label	131	11	21.05 *
All	Read Some or More	131	11	30.54 **
All	Read Most or More	131	11	49.88 ***
All	"Lax" Compliance	131	11	37.96 ***
All	"Strict" Compliance	131	11	40.33 ***

***$p<.001$, **$p<.01$, *$p<.05$, †$p<.10$

Appendix XI: Post Hoc Comparisons of Mean Ranks

Comparison #	Warning Label	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Comparisons for "Outside Location" Only						
1	All Labels With No Header	*37.55*	*42.61*	*42.58*	*42.18*	*42.08*
	All Labels With Header	*33.06*	*27.06*	*27.09*	*27.56*	*27.69*
	z	*1.24*	*3.71* ***	*4.12* ***	*3.51* **	*3.77* ***
	Mann-Whitney U	*530.0*	*338.0*	*339.0*	*354.0*	*358.0*
	r	*0.15*	*0.44*	*0.49*	*0.42*	*0.45*
2	17pt Bulleted With No Header	*13.63*	*14.21*	*12.88*	*13.25*	*13.33*
	17pt Bulleted With Header	*10.23*	*9.59*	*11.05*	*10.64*	*10.55*
	z	*1.42*	*1.93*	*0.98*	*1.09*	*1.38*
	Mann-Whitney U	*46.5*	*39.5*	*55.5*	*51.0*	*50.0*
	r	*0.30*	*0.40*	*0.21*	*0.23*	*0.29*
3	17pt Bulleted With No Header	*11.88*	*11.29*	*9.13*	*11.75*	*11.17*
	20pt Bulleted With No Header	*14.04*	*14.58*	*16.58*	*14.15*	*14.69*
	z	*-1.16*	*-1.43*	*-2.94* **	*-0.96*	*-1.38*
	Mann-Whitney U	*64.5*	*57.5*	*31.5*	*63.0*	*56.0*
	r	*0.23*	*0.29*	*0.59*	*0.19*	*0.28*

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{t}p<.10$ *with Bonferroni corrections*

Italicized values reppresent "mean ranks"

"All Labels With No Header" = Warning Formats 1, 3, and 4 as shown in Figure 1

"All Labels With Header" = Warning Formats 5, 6, and 8 as shown in Figure 1

"17pt Bulleted With No Header" = Warning Format 3 as shown in Figure 1

"17pt Bulleted With Header" = Warning Format 5 as shown in Figure 1

"20pt Bulleted With No Header" = Warning Format 4 as shown in Figure 1

Appendix XI (continued)

Comparison #	Warning Label	Noticed Label	Read Some or More	Read Most or More	"Lax" Compliance	"Strict" Compliance
Comparisons for complete sample						
1	All Labels With No Header	*70.94*	*76.77*	*80.03*	*77.13*	*77.96*
	All Labels With Header	*60.67*	*54.37*	*50.86*	*53.99*	*53.10*
	z	*2.06*	*3.90* ***	*5.24* ***	*4.08* ***	*4.36* ***
	Mann-Whitney U	*1806.0*	*1409.5*	*1188.0*	*1385.5*	*1329.0*
	r	*0.18*	*0.34*	*0.46*	*0.36*	*0.38*
2	17pt Bulleted With No Header	*25.50*	*26.50*	*26.50*	*26.00*	*26.00*
	17pt Bulleted With Header	*19.50*	*18.50*	*18.50*	*19.00*	*19.00*
	z	*2.01*	*2.39* ^{t}	*2.48* ^{t}	*2.11*	*2.09*
	Mann-Whitney U	*176.0*	*154.0*	*154.0*	*165.0*	*165.0*
	r	*0.30*	*0.36*	*0.37*	*0.32*	*0.31*
3	17pt Bulleted With No Header	*22.43*	*21.86*	*20.27*	*22.36*	*22.82*
	20pt Bulleted With No Header	*23.54*	*24.09*	*25.61*	*23.61*	*23.17*
	z	*-0.52*	*-0.79*	*-1.67*	*-0.43*	*-0.11*
	Mann-Whitney U	*240.5*	*228.0*	*193.0*	*239.0*	*249.0*
	r	*0.08*	*0.12*	*0.25*	*0.06*	*0.02*
4	All Labels Outside Location	*66.59*	*59.57*	*59.28*	*55.14*	*56.21*
	All Labels Inside Location	*65.32*	*73.38*	*73.71*	*78.47*	*77.23*
	z	*0.26*	*-2.40* ^{t}	*-2.59* *	*-4.11* ***	*-3.68* ***
	Mann-Whitney U	*2093.5*	*1685.0*	*1664.5*	*1374.5*	*1450.0*
	r	*0.02*	*0.21*	*0.23*	*0.36*	*0.32*

$^{***}p<.001$, $^{**}p<.01$, $^{*}p<.05$, $^{t}p<.10$ with Bonferroni corrections

Italicized values reppresent "mean ranks"

"All Labels With No Header" = Warning Formats 1, 3, and 4 as shown in Figure 1

"All Labels With Header" = Warning Formats 5, 6, and 8 as shown in Figure 1

"17pt Bulleted With No Header" = Warning Format 3 as shown in Figure 1

"17pt Bulleted With Header" = Warning Format 5 as shown in Figure 1

"20pt Bulleted With No Header" = Warning Format 4 as shown in Figure 1

Biographies

Michael J. Kalsher, Rensselaer Polytechnic Institute, PhD in Industrial/Organizational Psychology, 1988, Virginia Tech

William G. Obenauer, Rensselaer Polytechnic Institute, MBA, 2014, Rensselaer Polytechnic Institute

Christopher F. Weiss, Rensselaer Polytechnic Institute, JD, 1999, Saint Louis University School of Law