Published in final edited form as:
J Safety Res. 2016 December ; 59: 97–104. doi:10.1016/j.jsr.2016.10.002.

Experimental analysis of using examples and non-examples in safety training☆

Matthew A. Taylor[a,b,*], **Oliver Wirth**[a], **Marc Olvina**[b], and **Alicia M. Alvero**[b]

[a]Health Effects Laboratory Division, National Institute for Occupational Safety and Health, USA

[b]Department of Psychology, Queens College and the Graduate Center, City University of New York, USA

Abstract

Introduction—The effects of training content consisting of examples and/or non-examples was studied on the acquisition of safety-related skills.

Method—Participants ($N = 160$) were randomly assigned to first receive computer-based training on office ergonomics that included either no examples of safe or at-risk postures, safe examples only, at-risk examples only, or both safe and at-risk examples. Participants then attempted to classify as safe or at-risk various postures depicted in short video clips and demonstrate with their own posture the range of safe postures.

Results—Groups that were trained with both safe and at-risk examples showed greater classification accuracy and less error in their demonstration of safe postures. Training with only safe or at-risk examples resulted in a moderate amount of error and a consistent underestimation of risk.

Conclusion—Training content consisting of both examples and non-examples improved acquisition of safety-related skills.

Practical applications—The strategic selection of training content may improve identification of risks and safe work practices.

Keywords

Education; Observation; Performance; Error; Ergonomics

1. Introduction

Training is an important component of occupational safety and health programs (OSHA, 1998; Burke et al., 2006; Robson et al., 2010; Burke et al., 2011). The primary purpose of training is to provide workers with the knowledge and skills necessary to avoid illness, injury, or death. Because of its importance, there is a continuous need for safety researchers

☆The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

*Corresponding author at: NIOSH, 1095 Willowdale Road (MS 4050), Morgantown, WV 26505, USA. taylormatthewa@yahoo.com (M.A. Taylor).

to evaluate training content and methods to improve its efficiency and effectiveness (Robson et al., 2010; Arthur, Bennett, Edens, & Bell, 2003; Cohen & Colligan, 1998).

1.1. Training with examples and non-examples

Experts in psychology, education, and instructional design have recommended incorporating examples into training to facilitate concept learning and skill acquisition (e.g., Clark, 1971; Brethower, 2000; Markle & Tiemann, 1970; Merrill, Tennyson, & Posey, 1992; Foshay, 2010). In concept learning, *examples* refer to objects, events, or instances that have one or more defining characteristics or qualities of a concept (Merrill et al., 1992). Examples are said to be members of a *concept class*. For example, cakes, cookies, pies, and candy are members of the class of *dessert foods*. Examples are usually necessary for concept learning to occur, but they are not always sufficient. Mastery of some concepts may require the use of non-examples. Non-examples are objects, events, or instances that do not have the defining characteristics or qualities of the concept and, therefore, do not belong to the concept class. Wheat bread, hot dogs, broccoli, and crackers are non-examples of the class *dessert food*. Research has shown that mastery of a concept is greatest when training includes both examples and non-examples (e.g., Derenne, 2006; Durkin & Rittle-Johnson, 2012; Grobe & Renkl, 2007; Stark, Kopp, & Fischer, 2011; Wisniewski, Church, & Mercado, 2009).

The importance of training with examples and non-examples seems to extend equally well to safety concepts; however, the explicit use of examples and non-examples in safety training is rarely discussed—if at all—in the safety literature. Consider the problem of teaching a contractor's apprentice safe and hazardous electrical conditions. To best illustrate the distinction between safe and hazardous conditions, the apprentice may be shown several safe conditions (i.e., examples) and several hazardous conditions (i.e., non-examples). Safe examples might include the presence of extension cords with insulated wire and a grounding conductor, wiring enclosed in panels and machinery, use of ground fault circuit interrupters, and use of electric tools in dry conditions. The hazardous instances or non-examples might include extension cords that are frayed, cut, or without a grounding conductor, damaged machinery with exposed wiring, use of an overloaded outlet, and use of electric tools in damp conditions. It seems intuitive that the apprentice shown only one type of example may not learn to recognize all possible safe and hazardous electrical conditions, and yet the safety training literature is mostly devoid of the topic of examples and non-examples. Furthermore, we can find no authoritative recommendations in the safety literature concerning the use of safe and at-risk examples, despite a common concern among safety experts that providing both examples and non-examples of safe conditions or practices may create confusion about what is safe and what is not safe.

Using both examples and non-examples may be important in safety training not only to increase accuracy in learning concepts but also to minimize bias. Research in the psychology of learning has shown that training with examples only can result in overgeneralization of the concept (e.g., Wisniewski et al., 2009). For example, a study that investigated learning in a driving simulator showed that training with safe driving examples only, when compared with both safe and at-risk examples, resulted in greater speeds and other risky maneuvers at a traffic signal (Ivancic & Hesketh, 2000). Indeed, the biased training in that study may have contributed to an overgeneralization of safe driving conditions and underestimation of risks, but the effects of biased training is not well understood. More research is needed to evaluate the effects of training with safe and at-risk examples to better understand the conditions under which biased training leads to an overestimation or underestimation of hazards and risks.

1.2. Verbal skills versus performance skills

Safety training is used to improve different types of safety-related skills associated with hazard and risk identification and safe work practices. Many of the skills trained are *verbal* (i.e., classification, recognition, discrimination, comprehension, detection, and identification) in the sense that they help workers report differences between safe and at-risk work conditions. As an example, a worker who is trained to inspect scaffolding for sound wooden planks is expected to visually evaluate the planks and accurately report whether they are safe (e.g., straight, consistent, and complete with clean surface, etc.) or hazardous (e.g., splits or warps greater than 1/4 inch, gouges, mold, etc.). Because workplace safety and health depend on verbal skills, it is imperative that the effects of examples and/or non-examples be considered in the development of safety training as they have the potential to either help or impede worker's learning of hazards and risks.

It is also important to determine how training with examples and/or non-examples affects safety-related *performance*, which can be defined as kinesthetic or physical repertoires (Wan, 2014; Tiemann & Markle, 1990). For example, courses on driver safety often use pictures and videos to teach people how to respond during a loss of vehicle control. In response to hydroplaning on a straight road, drivers are taught to keep the wheels straight and to let off of the accelerator or gently apply the brakes. Safe driving programs, like many other classroom and computer-based training programs often incorporate examples of safe practices with the assumption that the ability to recognize correct or incorrect responses will result in the ability to perform the appropriate safe responses. This transfer of learning from verbal skills to performance skills is an example of *vertical transfer* (Blume, Ford, Baldwin, & Huang, 2010).

The transfer of learning among skills seems to be an important consideration for safety training programs, especially those in which safety-related verbal skills are directly targeted and are assumed to also result in acquisition of associated performance skills. The necessary or boundary conditions under which this type of transfer of learning may occur has not been systematically studied in safety research. This void highlights the need for basic research to elucidate the extent to which training with examples and/or non-examples affects acquisition of safety skills. The results of such research could lead to more effective and efficient safety training programs.

1.3. Purpose and hypotheses

The main objective of this study was to evaluate the use of examples (safe leg angles) and non-examples (at-risk leg angles) in a computerized training module on postural ergonomics. For the purposes of this experiment, the content in the training module was simplified to focus only on safe and at-risk knee angles when seated at a computer workstation. Four different training modules were tested experimentally. The modules consisted of either (a) no safe or at-risk examples, (b) only safe examples, (c) only at-risk examples, (d) or both safe and at-risk examples. These training conditions were assessed on participants' acquisition of a posture-related verbal skill (i.e., classifying postures as safe or at-risk) and a performance skill (i.e., demonstrating safe postures). It was hypothesized that training with only safe or only at-risk examples will result in more accurate classification than training with no examples, but training with both safe and at-risk examples will result in the most accurate and least biased classification. Similarly, it was hypothesized that training with no examples will result in more error in demonstrations of safe postures than training with either safe or at-risk examples alone, but that training with both safe and at-risk examples will produce the least amount of error. Finally, we explored the transfer of learning by examining the effects of training with both safe and at-risk examples on the correspondence between classification and demonstration skills.

2. Method

2.1. Participants and settings

Participants (n = 160) were recruited from undergraduate psychology courses at Queens College. Each individual participated in one 40-min to 70-min session that took place in a private room equipped with a computer workstation. The study was approved by the college's institutional review board, and all participants signed a consent form.

2.2. Experimental design

A randomized group design was used to test the effect of the different training conditions. Participants were randomly assigned in a balanced manner to one of four groups that received training with either: (a) no safe or at-risk examples of knee angles (No Ex); (b) only safe examples (S Ex); (c) only at-risk examples (A Ex); or (d) both safe and at-risk examples (S&A Ex).

2.3. Procedure

Participants completed computer-based training and assessment as outlined in Table 1. The S Ex, A Ex, and S&A Ex groups were presented the training and assessment materials with an automated PowerPoint slideshow. The slideshow consisted of pictures and videos recycled from previous studies (e.g., Taylor & Alvero, 2012; Taylor, Skourides, & Alvero, 2012). The sets of pictures and videos depicted a person seated at a computer workstation with their lower leg in one of several neutral, flexion, and extension positions. Leg angles 77° to 100° were classified a priori as *safe*. Angles 40° to 76° (flexion or backward position) and 101° to 165° (extension or forward position) were classified as *at-risk* (ranges of safe and at-risk leg angles were adapted from materials provided by the U.S. Department of Labor, 2011). The pictures and videos of leg angles were measured in angular degrees using a digital protractor (Iconico Screen Protractor; v. 4; New York, NY). The No Ex group received no training and participated in the assessment phase only.

2.3.1. Training phase—Training began with an *information* component that displayed operational definitions of safe or at-risk leg angles to supplement the use of examples in the subsequent training components (cf. Klausmeier & Feldman, 1975). Safe leg angles were defined as “lower legs that are nearly perpendicular to the floor” and at-risk leg angles were defined as “lower legs that are angled substantially forward or backward.” The S Ex group was presented with the safe definition only, the A Ex group was presented with only the at-risk definition, and the S&A Ex group was presented with both definitions.

During the *presentation* component of training, participants viewed a set of pictures and a set of 5-s videos that were equal in number and that showed examples of various leg angles. The S Ex group viewed 30 examples of safe leg angles from 15 pictures and 15 videos, the A Ex group viewed 30 examples of at-risk leg angles, and the S&A Ex group viewed 30 examples of safe leg angles and 30 examples of at-risk leg angles.

During the *classification* component of training, participants completed a matching-to-sample procedure that consisted of a series of trials that showed various safe or at-risk leg positions. Two sets of trials were classified—one with 15 pictures and one with 15 5-s videos for a total of 30 trials. The S Ex group viewed 30 trials with only safe leg angles, the A Ex group viewed 30 trials with only at-risk leg angles, and the S&A Ex group viewed 30 trials of safe leg angles and 30 trials of at-risk leg angles. Participants were instructed to classify the leg angle seen on each trial as either safe or at-risk and then record their answer on a check sheet. To maximize the participants learning and attention to the training content, both sets of trials were repeated to provide self-administered feedback on accuracy (order of trials: classify set of pictures, feedback on pictures, classify set of 5-s videos, and then feedback on 5-s videos). Participants were instructed to compare the correct answers displayed on the screen to their answers on the check sheet and use a marker to highlight the correct trials.

2.3.2. Assessment phase—At the start of this phase, instructions preparing the participants for the assessment were presented. In the *classification* assessment, verbal skills were assessed with a randomized sequence of 30 30-s video trials, comprised of 15 trials with examples of safe leg angles and 15 trials with examples of at-risk leg angles. Consistent with the whole interval recording method (Wirth, Slaven, & Taylor, 2014), participants were instructed to classify each video as safe only if leg angles were safe for the entire video; otherwise, the videos were to be classified as at-risk. Participants recorded their classification of each video on a check sheet.

In the *demonstration* assessment, participants were instructed to physically demonstrate their estimates of the lower and upper limits of the safe range—that is, to move their right lower leg backwards to a flexion position as close as possible to 77° and then forwards to an extension position as close as possible to 100°. Participants were also instructed to hold each position for a few seconds, which was long enough to capture the position with a digital photo. This assessment consisted of two trials per participant—one trial for each position conducted in a randomly determined sequence. Each trial began with the participant’s lower right leg positioned perpendicular or approximately 90° to the floor.

2.4. Data reduction and analysis

2.4.1. Classifica tion assessment—A primary measure of the classification skill was the number of trials (30-s videos) correctly classified as safe and at-risk (examples and non-examples). Signal detection analysis was also conducted to yield indices of classification accuracy and bias, which are easy to interpret because they are standardized and largely independent of each other. First, participants' responses were categorized as *hits*, *false alarms*, *correct rejections*, or *misses*. A hit was a safe leg angle correctly classified as safe, and a correct rejection was an at-risk leg angle correctly classified as at risk. A false alarm was an at-risk leg angle incorrectly classified as safe, and a miss was a safe leg angle incorrectly classified as at-risk. The statistics d' and c were then calculated for indices of classification accuracy and bias, respectively (see Macmillan & Creelman, 2005 for instruction on calculating d' and c). These statistics were calculated for each participant and then averaged within each group.

A value of near zero for d' indicates a low level of classification accuracy, whereas a value near 4.65 (the ceiling) indicates a high level of accuracy (Macmillan & Creelman, 2005). In most applications, d' ranges between 0.5 and 2.0 (Wickens & Hollands, 2000). A value of zero for c indicates no classification bias, a positive value indicates a bias towards an at-risk classification (i.e., overestimation of risk), and a negative value indicates a bias towards a safe classification (i.e., underestimation of risk). The limits of c are ±2.33, and a greater deviation from zero indicates a greater magnitude of bias (Macmillan & Creelman, 2005). To evaluate both magnitude and direction of bias, real (positive and negative) c values were analyzed. Absolute (non-negative) c values were analyzed to evaluate the magnitude of bias, independent of direction.

2.4.2. Demonstration assessment—The primary measure in the demonstration assessment was participants' estimates of the limits of the range of safe leg angles—77° and 100°. The estimate of safe leg angles were captured in photos and measured in angular degrees using a digital protractor. Leg angles closer to the trained lower and upper limits (i.e., 77° and 100°) indicate more accurate estimates. Leg angles beyond either the lower or upper limit of the safe range indicate an underestimation of risk because participants overestimated the range of safe leg angles (i.e., some at-risk leg angles were identified as safe). In contrast, leg angles within the lower and upper limit of the safe range indicate an overestimation of risk because the participants underestimated the range of safe leg angles (i.e., some safe leg angles were identified as at-risk).

Reliability of the leg-angle measurements was assessed by comparing measurements from the primary experimenter to those of a second, independent observer. Among all 320 measurements, 308 (96.25%) met the reliability criterion set at less than a two-degree difference. Measurements that did not meet the criterion were reevaluated and resolved by consensus.

During debriefing, a few participants, for which English was a second language, reported misinterpreting the oral instructions in the demonstration assessment. As a result, some of these participants moved their leg to extreme flexion and extension positions that were well outside the range of safe leg angles. Taking this into account, a modified boxplot method

(Wilcox, 2003) was used to identify outliers. Data from 13 participants (8.13% or approximately 3–5 participants in each training group) were removed from the analysis of the demonstration assessment in an effort to reduce errors that may distort or mislead the outcomes (Wilcox, 2001).

2.4.3. Data analysis—The resulting data sets ($n = 160$ from classification assessment and $n = 147$ from demonstration assessment) were analyzed with SPSS (v. 21; Armonk, NY). Significant main effects and interactions are reported in the text, and the figures show the simple effects. The simple effects were evaluated using one of three post-hoc tests: Tukey's HSD, Tamhane's T2, or multiple *t*-tests where appropriate. Each test used a .05 alpha level.

Transfer of learning between verbal and performance skills is revealed by comparing outcomes in the classification and demonstration assessments. The outcomes were assessed qualitatively across and within each group, except No EX, by comparing the magnitude of error (i.e., mean difference in degrees from safe limit) in the demonstration assessment with (a) the magnitude of classification accuracy (mean d'), (b) the magnitude of classification bias (mean absolute c), and (c) the direction of classification bias (mean real c). Evidence of transfer of learning between skills for a group would include, but not limited to, any of the following three examples of correspondence: (a) high d' and low demonstration error, (b) low d′ and high demonstration error, and (c) either overestimations or underestimations of risk across both classification and demonstration performances. Each of these examples would indicate a correspondence by revealing a similar and related outcome in both the classification assessment and demonstration assessment.

3. Results

3.1. Classification assessment

Effects of the different training conditions on the mean number video trials classified correctly are plotted in Fig. 1. Data were analyzed using a 4 (training: No Ex, S Ex, A Ex, and S&A Ex) × 2 (class of leg angles: safe and at risk) mixed factorial ANOVA with training as a between-groups factor and class of leg angles as a within-subjects factor. The analysis revealed a significant main effect of training [$F(3,156) = 22.21$, $p < .001$] and significant interaction [$F(3,156) = 99.01$, $p < .001$] between training and class of leg angles. An analysis of the simple effects showed that among all groups the S Ex group correctly classified significantly more safe trials, whereas the A Ex group correctly classified significantly more at-risk trials. Another analysis of simple effects revealed that in each training group, the mean number of trials correctly classified was significantly greater in one class. For example, the No Ex and S Ex groups correctly classified more safe trials, whereas the A Ex and S&A Ex groups correctly classified more at-risk trials.

Fig. 2 (top panel) shows the results of the signal detection analysis with mean d' (i.e., classification accuracy) plotted by group. Single-sample *t*-tests revealed that mean d' was significantly greater than zero in all training groups. A one-way independent samples ANOVA revealed a significant difference in mean d' among the groups [$F(3,156) = 17.281$, $p < .001$]. A test of the simple effects showed that mean d' was significantly greater in the S&A Ex group than in all other groups.

Fig. 2 (middle panel) shows the magnitude and direction of bias with mean real *c* plotted by training group. A single-sample *t*-test revealed that mean real *c* was significantly different from zero in all training groups, and a one-way independent samples ANOVA revealed significant differences in mean real *c* across the groups [$F(3,156) = 96.46$, $p < .001$]. Analyses of simple effects revealed that mean real *c* in each group was significantly different from that in all other groups. Real *c* was closest to zero in the S&A Ex group. Real *c* was negative in the No Ex and S Ex groups indicating a bias towards safe classifications or an underestimation of risk; it was positive in the A Ex and S&A Ex groups indicating a bias towards at-risk classifications or an overestimation of risk.

The magnitude of bias, independent of direction, is plotted with absolute *c* in the bottom panel of Fig. 2. A single-sample *t*-test revealed that mean absolute *c* was significantly different from zero in all groups, and a one-way independent samples ANOVA revealed that absolute *c* was significantly different across the groups [$F(3,156) = 16.76$, $p < .001$]. Analysis of simple effects revealed that the mean absolute *c* in the No Ex group was significantly less than in the A Ex group; however, mean absolute *c* in the S&A Ex group was less than in all groups.

3.2. Demonstration assessment

Participants' demonstrated estimates of the upper and lower limits of the range of safe leg angles are shown in Fig. 3. First, single-sample *t*-tests revealed that the mean estimates were significantly different from the trained limits in all training groups except the S&A Ex group. Thus, all groups except the S&A Ex group consistently demonstrated leg angles beyond the safe limits. Data were also analyzed using a 4 (training: No Ex, S Ex, A Ex, and S&A Ex) × 2 (safe range limit: upper and lower) mixed factorial ANOVA with training as a between-groups factor and limits of the safe range as a within-subjects factor. The analysis revealed a significant main effect of training [$F(3,143) = 37.02$, $p < .001$] and a significant interaction between training and range limit [$F(3,143) = 10.37$, $p < .001$]. Analysis of simple effects revealed that the mean estimates for the upper limit in S&A Ex group were significantly less than the estimates of the other training groups. Estimates of the lower limit in the S&A Ex group were significantly greater than the estimates of the A Ex group. In the No Ex group, mean estimates of the upper and lower limits, respectively, were significantly greater than and less than the limits of the other training groups. Overall, this result shows that the No Ex group underestimated risk significantly more than all groups, yet all groups underestimated risk more than the S&A Ex group.

3.3. Correspondences between classification and demonstration skills

Qualitative comparisons of the outcomes in the classification and demonstration assessments revealed correspondences directly related to the manipulation of examples and non-examples in the training content. Correspondences were revealed by comparing demonstration error with the magnitude of classification accuracy (i.e., mean d'), magnitude of classification bias (i.e., mean absolute *c*) and the direction of bias as an under- or overestimation of risk (i.e., real *c*) across the groups. For example, the S&A Ex group showed the greatest mean d' and estimated leg angles that were closest to the limits of the safe range indicating high accuracy in both the classification assessment and demonstration assessment. The S&A Ex group also showed the least magnitude of bias in the classification assessment and the least error in the demonstration assessment. Finally, for the S&A Ex group, the direction of bias was somewhat inconsistent with a significant mean overestimation of risk in the classification assessment and a non-significant mean under- and overestimation of risk in the demonstration assessment.

The S Ex and A Ex groups each showed a moderate magnitude of accuracy and bias in the classification assessment and moderate error in the demonstration assessment. Furthermore, direction of bias in the S Ex group was consistent with underestimations of risk in both classification and demonstration assessments; however, this effect was not consistent in the A Ex group, which showed overestimations of risk in the classification assessment and underestimations of risk in the demonstration assessment.

4. Discussion

4.1. Implications

As hypothesized, the current study showed that participants who were trained with both safe and at-risk examples showed the greatest accuracy and least under- or overestimations of risk in both the classification assessment and demonstration assessment. These findings also support the hypothesis that the effects of training, at least when described qualitatively, result in a correspondence between skills, where the learning of safe postures transferred from the classification skill to the demonstration skill. Correspondence was further supported in the groups trained with only safe examples or only at-risk examples. These groups showed moderate accuracy and moderate under- or overestimation of risk in both the classification and demonstration assessments.

The effects of training with safe and at-risk examples are consistent with previous research on examples and/or non-examples (e.g., Derenne, 2006; Durkin & Rittle-Johnson, 2012; Grobe & Renkl, 2007; Stark et al., 2011; Wisniewski et al., 2009) and extend the study of exemplars in training content to a safety-related context. Also, the findings from the present study are largely consistent with a meta-analysis of research that assessed correspondences between performance and procedural knowledge (Taylor et al., 2005). Although this meta-analysis only targeted research evaluating the efficacy of Behavior Modeling Training (Baldwin, 1992), which relies on hands-on-training of performance skills, its general conclusions are consistent with the present finding that verbal and performance skills may be similarly affected by training content.

Training programs commonly depend on correspondence between skills—where one skill is directly trained and is expected to result in a transfer of knowledge and learning to a different skill (Doo, 2006). Although a correspondence between skills is often a presumed outcome with many training programs, especially with computer or textbook-based training modules, our findings show that correspondence among different behavioral repertoires is not inevitable. The present study showed that training with only *at-risk* examples resulted in an overestimation of risk in the classification assessment and an underestimation of risk in the demonstration assessment. These findings reveal that there are conditions where correspondence between skills could be moderate to low and consequently would hinder acquisition of the desired secondary skill (Jacobs & Thompson, 2000). In contrast, training with only *safe* examples showed a consistent underestimation of risk across both the classification and demonstration assessments. Although a correspondence between skills is often desirable, trainers should avoid using only safe examples, as it may result in an underestimation of risk and subsequently could lead to a greater rate of incidents. In occupations with a potential for serious adverse events, it may be desirable to train workers to overestimate risk, as workers performing a job with extra caution may manifest as a decrease in the rate of safety-related incidents (Kontos, 2004). The present study showed that training content with both safe and at-risk examples resulted in the most consistent overestimation of risk across both the classification and demonstration assessments. Furthermore, although some research has documented the occurrence of workers' underestimation of risk, (e.g., Mullen, 2004; Choudhry & Fang, 2008; Zimolong, 1985), no study quantified the degree to which workers underestimated risk. The present study showed that participants most severely underestimated risk when trained with only safe examples.

Assuming that our findings generalize to the broader field of occupational safety, then training content that includes both safe and at-risk examples would help workers more accurately identify and avoid tasks and events that could lead to illness, injury or death. The findings also corroborate previous assertions that trainers should not assume that training and mastery of one skill would result in or indicate mastery of a different skill (Cohen & Colligan, 1998; Doo, 2005). Instead, trainers should directly train and measure the desired skills to ensure that they have been adequately acquired (Arthur et al., 2003; Jones, Ollendick, & Shinske, 1989), especially when skills are associated with preventing higher levels of risk.

4.2. Limitations and future research

The current study used a simplified and somewhat contrived training scenario to experimentally study the effects of examples and non-examples on the acquisition of skills. The extent to which our findings extend to other more complex or real-world safety scenarios requires further study. Nevertheless, our findings are consistent with previous studies on the use of examples and non-examples in other domains, and thus it is reasonable to assume that the findings generalize to other contexts.

The ergonomic task—classifying and demonstrating safe knee angles—is of minimal complexity (each a single-step task and one dimensional) when compared with other tasks that consist of a chain of steps, are multi-dimensional, and require greater learner memory and processing (e.g., donning a respirator, driving a vehicle, operating a circular saw, tungsten inert gas welding). Research has shown that task complexity can either inhibit or facilitate acquisition of skills depending on the conditions of learning (Wulf & Shea, 2002). In the current study, it is possible that performance was relatively high due to the minimal complexity of the task.

Task difficulty is another factor of interest and applies to the similarity between safe and at-risk examples. Research in stimulus discrimination learning found that distinguishing members of different stimulus classes, which are very similar on one or more defining physical dimensions, can be improved by selecting training examples and non-examples that also are very similar. In other words, training fine discriminations of subtle stimulus differences between examples and non-examples can increase the precision of discrimination (e.g., Baron, 1973; Carnine, 1980; Derenne, 2006). In the present study, training with greater similarity between safe and at-risk examples (e.g., 2-degree difference vs. 10-degree difference between safe and at-risk knee angles) might have resulted in greater classification accuracy and reduced underestimations of risk. Also, it is likely that more misclassifications occurred with leg angles that were near the limit between the safe and at-risk classes; however, the classification assessment was not designed to identify the frequency of misclassification errors across the range of leg angles. Future research should use a classification assessment with greater resolution, so that the pattern of misclassification across the range of angles can be more thoroughly evaluated.

Research has shown that an elevated level of perceived risk associated with a task can facilitate skill acquisition (Burke et al., 2011). It seems reasonable to speculate that risk perception affects learning through attentional or motivational mechanisms. In the present study, level of attention and motivation to learn the subtle discriminations between safe and at-risk knee angles were likely limited because musculoskeletal disorders can be perceived as low risk and the task of identifying ergonomic postures posed little or no risk to the participants. To confirm the validity of these speculations, applied research examining the effects of training content on attention and motivation should be conducted with a range of jobs that vary in actual risk and with actual workers learning safety-related skills that are applicable and relevant to their own work.

Finally, future research should evaluate how content-related factors, such as the number and ratio of safe and at-risk examples, affects acquisition of safety-related skills. For example, there is evidence from basic research in psychology that the number of examples relative to non-examples can systematically bias individuals' identification of stimuli (e.g., Thomas & Vogt, 1983; Thomas, Windell, Williams, & White, 1985). Thus, manipulating the ratio of at-risk and safe training content in favor of at-risk examples (e.g., 2:1 or 3:1) might still yield accurate, but somewhat biased perceptions towards an overestimation or heightened sensitivity of hazards and risks. Indeed, errors are often unavoidable; however, safety training that limits errors to false positives (i.e., identifying risks when there are none) would be acceptable and perhaps desirable, especially in hazardous environments.

5. Practical applications

Safety professionals and others involved in developing content for safety training programs are in need of further guidance on selecting and arranging the training content. The present study provides insights into how training with or without safe and at-risk examples can impact the acquisition of safety-related skills. Although the findings show that using both examples and non-examples of safety improves learning outcomes and reduces bias towards over- or underestimations of risk, it might be appropriate to manipulate the number of at-risk examples relative to the number of safe examples to encourage heightened sensitivity to hazards and risks. Further applied research is needed to test these principles across a range of task and situation specific variables. Empirically driven designs of safety training materials will help maximize learning and, ultimately, reduce safety-related illness, injury, and fatalities.

Acknowledgments

The authors would like to thank Andreas Skourides, Angelica Cruells, Abu Mohammad, Emily Linder, and Jessica Parisi for their help with data collection. Also, the authors thank Bruce Brown and Robert Lanson for their input during the development of the study.

References

Arthur W Jr, Bennett W Jr, Edens PS, Bell ST. Effectiveness of training in organizations: a meta-analysis of design and evaluation features. Journal of Applied Psychology. 2003; 88:234–245. [PubMed: 12731707]

Baldwin TT. Effects of alternative modeling strategies on outcomes of interpersonal-skills training. Journal of Applied Psychology. 1992; 77:147–154. [PubMed: 1592758]

Baron A. Postdiscrimination gradients of human subjects on a tone continuum. Journal of Experimental Psychology. 1973; 101:337–342. [PubMed: 4753858]

Blume BD, Ford JK, Baldwin TT, Huang JL. Transfer of training: A meta-analytic review. Journal of Management. 2010; 36:1065–1105.

Brethower, DM. Organizational behavior management and instructional systems. In: Austin, J.; Carr, JE., editors. Handbook of applied behavior analysis. Reno, NV: Context Press; 2000. p. 423-448.

Burke MJ, Salvador RO, Smith-Crowe K, Chan-Serafin S, Smith A, Sonesh S. The dread factor: How hazards and safety training influence learning and performance. Journal of Applied Psychology. 2011; 96:46–70. [PubMed: 21244129]

Burke MJ, Sarpy SA, Smith-Crowe K, Chan-Serafin S, Salvador RO, Islam G. Relative effectiveness of worker safety and health training methods. American Journal of Public Health. 2006; 96:315–324. [PubMed: 16380566]

Carnine D. Relationships between stimulus variation and the formation of misconceptions. Journal of Educational Research. 1980; 74(2):106–110.

Choudhry RM, Fang D. Why operatives engage in unsafe work behavior: Investigating factors on construction sites. Safety Science. 2008; 46:566–584.

Clark DC. Teaching concepts in the classroom: A set of teaching prescriptions derived from experimental research. Journal of Educational Psychology Monograph. 1971; 62(3):253–278.

Cohen, A.; Colligan, MJ. Publications dissemination. 4676 Columbia Parkway, Cincinnati, Ohio 45226–1998: 1998. Assessing occupational safety and health training: A literature review. NIOSH. DHHS (NIOSH) publication no. 98–145. June

Derenne A. Effects of S+ and S− separation on gradient shifts in humans. (The Journal of) General Psychology. 2006; 133(2):163–173. [PubMed: 16705909]

Doo MY. The effects of presentation format for behavior modeling of interpersonal skills in online instruction. Journal of Educational Multimedia and Hypermedia. 2005; 14(3):213–235.

Doo MY. A problem in online interpersonal skills training: Do learners practice skills? Open Learning. 2006; 21(3):263–272.

Durkin K, Rittle-Johnson B. The effectiveness of using incorrect examples to support learning about decimal magnitude. Learning and Instruction. 2012; 22:206–214.

Foshay, WR. Some principles underlying the cognitive approach to instructional design. In: Silber, KH.; Foshay, WR., editors. Handbook of improving performance in the workplace: Instructional design and training delivery. San Francisco, CA: Pfeiffer; 2010. p. 3-22.

Grobe CS, Renkl A. Finding and fixing errors in worked examples: Can this foster learning outcomes? Learning and Instruction. 2007; 17:612–634.

Ivancic K, Hesketh B. Learning from errors in a driving simulation: Effects on driving skill and self-confidence. Ergonomics. 2000; 43(12):1966–1984. [PubMed: 11191780]

Jacobs BJ, Thompson CK. Cross-modal generalization effects of training non-canonical sentence comprehension and production in agrammatic aphasia. Journal of Speech, Language, and Hearing Research. 2000; 43:5–20.

Jones RT, Ollendick TH, Shinske FK. The role of behavioral versus cognitive variables in skill acquisition. Behavior Therapy. 1989; 20:293–302.

Klausmeier HJ, Feldman KV. Effects of a definition and a varying number of examples and nonexamples on concept attainment. Journal of Educational Psychology. 1975; 67:174–178.

Kontos AP. Perceived risk, risk taking, estimation of ability and injury among adolescent sport participants. J Pediatr Psychol. 2004; 29:447–455. [PubMed: 15277587]

Macmillan, NA.; Creelman, CD. Detection theory: A user's guide. 2. Hillsdale, NJ: Erlbaum; 2005.

Markle SM, Tiemann PW. Problems of conceptual learning. British Journal of Educational Technology. 1970; 1(1):52–62.

Merrill, MD.; Tennyson, RD.; Posey, LO. Teaching concepts. 2. Englewood. Cliffs, NJ: Educational Technology Publications; 1992.

Mullen J. Investigating factors that influence individual safety behavior at work. Journal of Safety Research. 2004; 35:275–285. [PubMed: 15288561]

Occupational Safety and Health Administration. Training requirements in OSHA standards and training guidelines. Washington D.C: Occupational Safety and Health Administration, U.S. Department of Labor; 1998. (OSHA Rept. 2254)

Robson, L.; Stephenson, C.; Schulte, P.; Amick, B.; Chan, S.; Bielecky, A.; … Grubb, P. A systematic review of the effectiveness of training & education for the protection of workers. Toronto: Institute for Work & Health; Cincinnati, OH: National Institute for Occupational Safety and Health; 2010. DHHS (NIOSH) publication no. 2010-127

Stark R, Kopp V, Fischer MR. Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education. Learning and Instruction. 2011; 21:22–33.

Taylor MA, Alvero AM. The effects of safety discrimination. Training and frequent safety observations on safety-related behavior. Journal of Organizational Behavior Management. 2012; 32:169–193.

Taylor PJ, Russ-Eft DF, Chan DWL. A meta-analytic review of behavior modeling training. Journal of Applied Psychology. 2005; 90:692–709. [PubMed: 16060787]

Taylor MA, Skourides A, Alvero AM. Observer error when measuring safety-related behavior: Momentary time sampling versus whole-interval recording. Journal of Organizational Behavior Management. 2012; 32:307–319. http://dx.doi.org/10.1080/01,608,061.2012.729389.

Thomas DR, Vogt MB. Stimulus generalization in humans viewed from a signal detection perspective. New Zealand Journal of Psychology. 1983; 12:1–6.

Thomas DR, Windell BT, Williams JL, White KG. Stimulus presentation frequency in brightness discrimination and generalization: A test of adaption-level and signal-detection interpretations. Perception & Psychophysics. 1985; 37:243–248. [PubMed: 4022755]

Tiemann, PW.; Markle, SM. Analyzing instructional content: A guide to instruction and evaluation. 4. Champaign, IL: Stipes Publishing Company; 1990.

U.S. Department of Labor, Occupational Safety & Health Administration (OSHA). Computer workstations. 2011i. Retrieved from http://www.osha.gov/SLTC/etools/computerworkstations/

Wan, M. Incidental trainer: A reference guide for training design, development and delivery. 1. Boca Raton, FL: Taylor & Francis Group; 2014.

Wickens, CD.; Hollands, JG. Engineering psychology and human performance. 3. Upper Saddle River, NJ: Prentice Hall; 2000.

Wilcox, RR. Fundamentals of modern statistical methods: Substantially improving power and accuracy. New York: Springer-Verlag; 2001.

Wilcox, RR. Applying contemporary statistical techniques. San Diego, CA: Academic Press; 2003.

Wirth O, Slaven J, Taylor MA. Interval sampling and measurement error: A computer simulation. Journal of Applied Behavior Analysis. 2014; 47(1):83–100. [PubMed: 24127380]

Wisniewski MG, Church BA, Mercado E. Learning-related shifts in generalization gradients for complex sounds. Learning & Behavior. 2009; 37(4):325–335. [PubMed: 19815929]

Wulf G, Shea CH. Principles derived from the study of simple skills do not generalize to complex skill learning. Psychonomic Bulletin & Review. 2002; 9:185–211. [PubMed: 12120783]

Zimolong, B. Hazard perception and risk estimation in accident causation. In: Eberts, RE.; Eberts, CG., editors. Trends in ergonomics/human factors II. Amsterdam, The Netherlands: North-Holland; 1985. p. 463-470.

Biographies

Matthew A Taylor recently became Patient Safety Fellow in the Center for Medical Product End-user Testing at the Veterans Affairs Pittsburgh Healthcare System. Prior to this position, he was a Service Fellow in the Health Effects Laboratory Division of the National Institute for Occupational Safety and Health. He received a PhD in Behavior Analysis from Queens College and the Graduate Center, CUNY. His research interests include issues related to behavioral safety, human factors, behavioral economics, education, and patient safety.

Oliver Wirth, PhD is a Research Psychologist in the Health Effects Laboratory Division of the National Institute for Occupational Safety and Health (NIOSH) at the Centers for Disease Control and Prevention (CDC). He has degrees in clinical psychology and behavior analysis. He has a background in psychological assessment and basic and applied research in learning and reinforcement theory with both human and nonhuman subjects. His research interests have included the application of behavioral technologies for the study of work-related musculoskeletal injuries, vibration disorders, safety climate and culture, safety-related decision-making, and at-risk driver behaviors.

Marc Olvina received a BS in Psychology from Queens College, CUNY. His academic interests include behavioral safety and applied behavior analysis.

Alicia Alvero is an Associate Professor at Queens College, CUNY. She received her BA in Psychology from Florida International University, and both her MA and PhD from Western Michigan University. Dr. Alvero was awarded the prestigious Ford Foundation Dissertation Fellowship and her research in the areas of performance feedback, behavioral safety, and training have been published in peer-reviewed journals such as *Professional Safety*, the *Journal of Organizational Behavior Management*, the *Journal of Agricultural Safety and Health,* etc. She has authored book chapters, sits on several editorial boards, and has been an invited speaker at numerous conferences across the country.

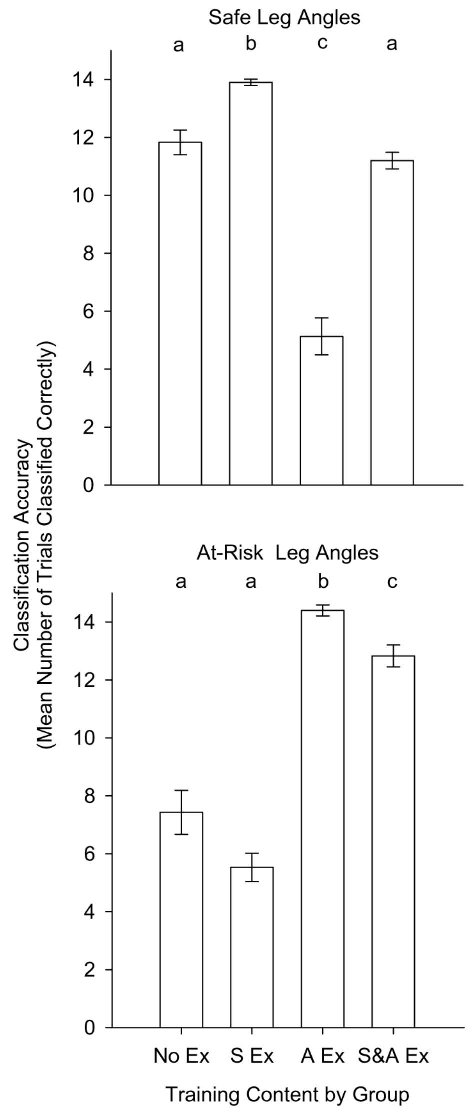

Fig. 1.

Results of the classification assessment showing the mean number of video trials with safe leg angles (top panel) and at-risk leg angles (bottom panel) classified correctly. Above the bars, dissimilar lowercase letters (a, b, and c) across training groups denote significant differences among the means. A comparison of means across top and bottom panels shows that the mean number of safe and at-risk trials classified correctly was significantly different within each training group. Each group was shown 15 trials of safe leg angles and 15 trials of at-risk leg angles. All tests used a .05 alpha level; error bars denote SEM.

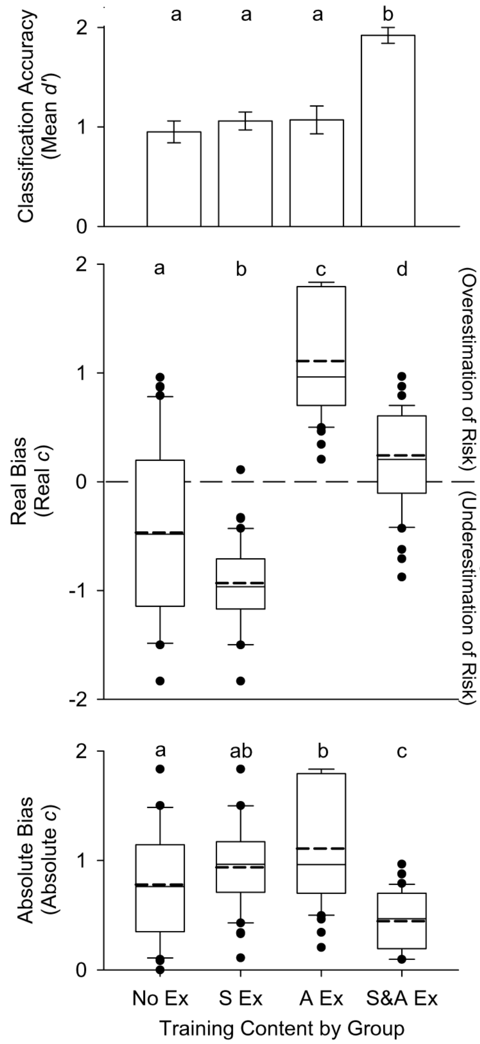

Fig. 2.

Results of the classification assessment and signal detection analysis. The top panel shows mean *d′*; higher values indicate greater classification accuracy. The middle panel presents mean real *c*; positive values indicate bias towards an at-risk classification (i.e., overestimation of risk) and negative values indicate bias towards a safe classification (i.e., underestimation of risk). The bottom panel shows mean absolute *c* or the magnitude of bias independent of direction. The boxplots depict the mean (dashed line), median (solid line), inter-quartile range (box), 10th and 90th percentiles (error bars), and values beyond 10th and 90th percentiles (black circles). Means labeled with the dissimilar lowercase letters (a, b, c, and d) above the plots indicate significant differences. Means in all panels are significantly different from zero. All tests used a .05 alpha level.

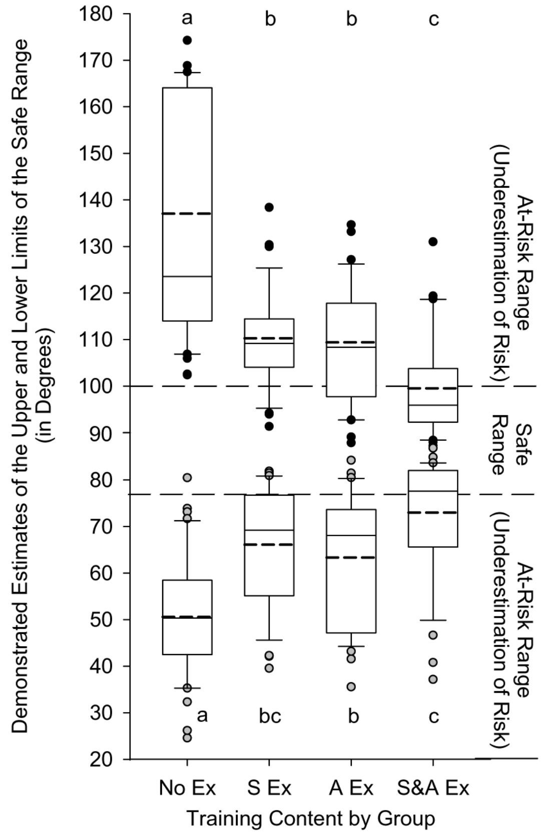

Fig. 3.

Results of the demonstration assessment showing the participants' estimates of the upper limit (100°) and lower limit (77°) of the range of safe leg angles. The boxplots show the mean (dashed line), median (solid line), inter-quartile range (box), 10th and 90th percentiles (error bars), and estimates beyond 10th and 90th percentiles (black circles for upper limit and gray circles for lower limit). Means labeled with dissimilar lowercase letters (a, b, and c) next to the plots indicate significant differences (only compare labels of plots with the same limit; either lower or upper limit). All means are significantly different from the closest limit, except the estimates for the S&A Ex group. All tests used a .05 alpha level.

Table 1

Exposure to components and content of the training and assessment phases by group.

Phase/component	Format/content	Group			
		No examples	Only safe examples	Only at-risk examples	Both safe/at-risk examples
Training					
Information (examples)	Read definition of *safe* leg angles	No	Yes	No	Yes
Information (non-examples)	Read definition of *at-risk* leg angles	No	No	Yes	Yes
Presentation (examples)	View pictures and 5-s videos of *safe* leg angles	No	Yes	No	Yes
Presentation (non-examples)	View pictures and 5-s videos of *at-risk* leg angles	No	No	Yes	Yes
Classification (examples)	Match-to-sample with pictures and 5-s videos of *safe* leg angles	No	Yes	No	Yes
Classification (non-examples)	Match-to-sample with pictures and 5-s videos of *at-risk* leg angles	No	No	Yes	Yes
Assessment					
Classification (examples and non-examples)	Match-to-sample test with 30-s videos of *safe* and *at-risk* leg angles	Yes	Yes	Yes	Yes
Demonstration	Move right leg to estimated upper and lower *safe* limit	Yes	Yes	Yes	Yes