6-2017

# An Item-Response Theory Approach to Safety Climate Measurement: The Liberty Mutual Safety Climate Short Scales

Yueng-hsiang Huanga
*Liberty Mutual Research Institute for Safety*

Jin Lee
*Kansas State University*

Zhuo Chen
*University of Connecticut*

MacKenna Laine Perry
*Portland State University*, mackenna.perry@gmail.com

Janelle H. Chung
*Oregon Health & Science University*

***See next page for additional authors***



Citation Details

Huang, Y., Lee, J., Chen, Z., Perry, M., Cheung, J. H., & Wang, M. (2017). An item-response theory approach to safety climate measurement: The Liberty Mutual Safety Climate Short Scales. Accident; Analysis And Prevention, 10396-104. doi:10.1016/j.aap.2017.03.015

Accident Analysis and Prevention 103 (2017) 96–104

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

# An item-response theory approach to safety climate measurement: The Liberty Mutual Safety Climate Short Scales

Yueng-hsiang Huang[a,*], Jin Lee[a,b], Zhuo Chen[a,c], MacKenna Perry[a,d], Janelle H. Cheung[a,e], Mo Wang[f]

[a] Liberty Mutual Research Institute for Safety, Hopkinton, MA, USA
[b] Kansas State University, Manhattan, KS, USA
[c] University of Connecticut, Storrs, CT, USA
[d] Portland State University, Portland, OR, USA
[e] Oregon Health & Science University, Portland, OR, USA
[f] University of Florida, Gainesville, FL, USA

ARTICLE INFO

*Keywords:*
Safety climate
Item response theory
Shortened scales

ABSTRACT

Zohar and Luria's (2005) safety climate (SC) scale, measuring organization- and group- level SC each with 16 items, is widely used in research and practice. To improve the utility of the SC scale, we shortened the original full-length SC scales. Item response theory (IRT) analysis was conducted using a sample of 29,179 frontline workers from various industries. Based on graded response models, we shortened the original scales in two ways: (1) selecting items with above-average discriminating ability (i.e. offering more than 6.25% of the original total scale information), resulting in 8-item organization-level and 11-item group-level SC scales; and (2) selecting the most informative items that together retain at least 30% of original scale information, resulting in 4-item organization-level and 4-item group-level SC scales. All four shortened scales had acceptable reliability ($\geq 0.89$) and high correlations ($\geq 0.95$) with the original scale scores. The shortened scales will be valuable for academic research and practical survey implementation in improving occupational safety.

## 1. Introduction

### 1.1. Safety climate

Safety climate research has been ongoing for more than 35 years, since Zohar published his seminal work in 1980 defining this construct as workers' shared perceptions regarding their organization's policies, procedures, and practices in relation to the value and importance of safety within that organization (Zohar, 1980; Griffin and Neal, 2000; Zohar, 2000, 2002, 2003). The study of safety climate is based on perceptions of workers, with the major factors relating to (a) management commitment to safety and (b) communication pertaining to safety as a true priority from top management and direct supervisors (Dejoy et al., 2004). Prior research has stated that safety climate is a multilevel construct encompassing two managerial levels: (1) organization-level safety climate, which refers to employees' perceptions of the company's or top management's commitment to and prioritization of safety, and (2) group-level safety climate, meaning employees' perceptions of their direct supervisors' commitment to and prioritization of safety (e.g., Zohar and Luria, 2005; Huang et al., 2013a,b). Several meta-analyses have provided robust evidence that safety climate is one of the best leading indicators of organizational safety outcomes, such as frequency or severity of injury incidents (Christian et al., 2009; Beus et al., 2010; Nahrgang et al., 2011). Overall, safety climate influences employees' motivation and knowledge to act in a safe manner, which in turn lead to safer behaviors and fewer accidents and injuries (Griffin and Neal, 2000; Christian et al., 2009).

Since the inception of safety climate research, many safety climate scales have been developed and validated in the scientific literature. One of the most widely used safety climate scales published in the field, which has robust evidence of reliability and validity, is a generic safety climate scale developed by Zohar and Luria (2005). Their scale includes 32 total items: 16 items to measure organization-level safety climate and 16 items to measure group-level safety climate. In Zohar and Luria's (2005) study, the Cronbach's alpha of the scale was 0.92 for organizational-level safety climate (OSC) and 0.95 for group-level safety climate (GSC). In terms of criterion-related validity, OSC was correlated with safety audit/observation scores at 0.46, and GSC was correlated with safety behavior observations at 0.38. According to Google Scholar (retrieved January, 2017), their paper has been cited by

* Corresponding author at: Center for Behavioral Sciences, Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748, USA. (Y.H. Huang)
*E-mail address:* Yueng-hsiang.Huang@Libertymutual.com (Y.-h. Huang).

http://dx.doi.org/10.1016/j.aap.2017.03.015
Received 8 December 2016; Received in revised form 15 February 2017; Accepted 19 March 2017

nearly 800 publications, many of which use their measure. For example, one of the heavily cited papers (Johnson, 2007) found that GSC was significantly correlated with injury frequency at −0.50 and safety behaviors at 0.78. Examining OSC, Martínez-Córcoles et al. (2011) found a correlation with safety behaviors at 0.43, while Brondino et al. (2012) found correlations with safety compliance and safety participation ranging from 0.27 to 0.36. Due to its increasingly high usage in research and practice, the current study focuses on increasing the utility of this scale by shortening the number of items required while maximizing information provided.

### 1.2. Length of safety climate scales

Safety researchers are frequently faced with a dilemma in field research: whether to use brief measures or longer, more exhaustive and thorough measures. A longer measure can capture a fuller range of construct content and variance of interest, whereas a brief measure can boost both participant engagement and the efficiency of data collection. There are times when a longer scale is preferable, but shorter scales may be more effective in other cases.

Overall, a survey instrument should not overwhelm respondents with too many questions. Previous research has demonstrated that survey length can negatively impact response rates (e.g., Crawford et al., 2001). By shortening the length of a survey, individuals may be more likely to perceive that they have time to participate in survey research, even when they do not feel participation will directly benefit themselves (Woods and Hampson, 2005). Furthermore, in cases where measures contain many items focused on a very similar topic, many participants may interpret items as redundant and may have negative reactions toward the overall survey assessment (Wanous et al., 1997).

An additional issue with longer measures is that their use can limit the nature of models that can be tested to explore relations among various constructs (Fisher et al., 2016). Zohar and Luria's (2005) generic safety climate scale includes 32 items, which is a fairly long measurement scale. Despite the existence of this psychometrically solid and widely accepted scale, Zohar (2010) stated that more work is needed to explore how safety climate emerges and how safety climate is influenced or changed (i.e., which factors contribute to the development of safety climate perceptions). In order to fill this gap, researchers need to collect additional data on many other variables simultaneously with safety climate. With the current length of the safety climate scale, it is challenging to achieve this goal within realistic limitations that researchers face. In order to further explore potential factors influencing safety climate, a shorter and valid generic safety climate scale is needed.

### 1.3. Item response theory (IRT)

We propose an Item Response Theory (IRT) approach because it assesses multiple psychometric features of individual scale items. In comparison, Classical Test Theory (CTT) places more emphasis on the scale's composite score. IRT is a probabilistic non-linear modeling technique for developing and evaluating psychological measurement scales. For example, it can be posited that items of a scale are designed to assess a certain psychological attribute (e.g., safety perception) such that endorsing higher values on the items suggests a stronger underlying psychological attribute (e.g., stronger safety perception). If respondents give undiscriminating endorsements to an item when they indeed differ in terms of the underlying psychological attribute, the item should be deemed improper as a measure of the psychological attribute. To this end, IRT calculates the respondents' probability of endorsing particular response options of each scale item and estimates each item's ability to differentiate respondents, which can be used for strategic tailoring of lengthy psychological scales.

It needs to be noted that even though IRT has been frequently used with educational and psychological tests which have correct or wrong answers, it can be applied to Likert scale-based measures (i.e., item with ordered categorical – polytonomous – response options) of psychological trait/attribute such as perceived job security (Probst, 2003) and personality (e.g., Reise and Henson, 2000). Likewise, higher levels of underlying trait/attribute are assumed to lead to higher probabilities of stronger endorsement (e.g., choosing the category 'strongly agree' on a 5-point Likert scale). IRT is free from limitations faced by conventional linear regression-based development and validation techniques such as circular sample dependency of item/person statistics (Fan, 1998). Furthermore, IRT considers the differentiating/discrimination ability and difficulty of each item as information to be incorporated in the scale. It allows researchers to more efficiently assemble the items that offer the most information for measuring the targeted underlying trait/attribute.

The unique parameters offered by IRT, such as slope and difficulty parameters, can be derived based on the probability of responses, which is illustrated by the item option response functions (ORFs). For a five-point Likert scale, each item has five response options. In the polytomous IRT model, ORFs are used to describe participants' response patterns. Each option has an ORF curve, with the $x$-axis representing the trait being measured (θ) and the $y$-axis representing the probability of endorsing this particular option; an ORF thus depicts the relationship between the participants' trait and their responses to an item.

The slope, discrimination, or differentiation parameter determines the slope of the option response functions (ORF) for each item. Every item will have one slope parameter. If all other difficulty parameters are equal, items with high slope parameters will have smaller overlap of θ values between the option response functions, representing better differentiation. In the current study, the slope parameter represents each item's sensitivity to the overall level of safety climate.

The difficulty parameter determines the location of the ORF along the θ axis and indicates on which part of the range of θ the item is most informative, or the θ value at which people have a 50% chance of selecting specified responses (i.e., the cutoff points that separate the response option categories). In the current study, each item was rated on a 5-point Likert scale. Therefore, each item has four ORFs and four difficulty parameters (i.e., the cutoff points that separate response 1 from responses 2–5, responses 1–2 from 3 to 5, responses 1–3 from 4 to 5, and, finally, responses 1–4 from 5). These four difficulty parameters jointly indicate the overall difficulty of an item. In the current study, the item's difficulty represents whether an item is more informative (i.e., sensitive in differentiating the level/strength of estimated target trait) at lower or higher ranges of safety climate scores.

The item information curve (IIC) for each item is a function of both the slope and difficulty parameters. The amount of information that a particular item provides depends on both the size of the slope parameter and the spread of the category thresholds. An IIC represents the amount of information provided by a specific item across the entire continuum of the latent construct of interest. The area of the IIC above the $x$-axis (θ) equals the item information. If an item has a larger amount of item information, the item has higher discriminating ability to differentiate respondents along the θ axis. Depending on the slope and difficulty parameters, the amounts of information offered by items will differ. By aggregating the IICs of items in a measure, the test information function (TIF) for a scale can be generated. Similar to IICs, the area of the TIF above the θ axis equals the total test information. If a scale has a larger amount of total test information, the scale score has higher discriminating ability along the latent θ value.

Overall, the current study aims to utilize IRT to shorten Zohar and Luria's (2005) 32-item safety climate scale. Both slope and difficulty parameters for each item in the existing scale were calculated, and all information available was carefully considered to decide on the best items to include in the final shortened scales and the ideal number of items to include. The new, shortened scales are expected to benefit future safety climate research and practice by allowing for more diverse data collection opportunities and addressing concerns that organiza-

tions and participants may have with implementation of a longer scale, while maintaining the usefulness of the existing measure.

## 2. Method

### 2.1. Participants and data collection procedure

Safety climate survey data were collected online as part of an evaluation package for customers of a safety consulting group. The service consultants invited their corporate customers to participate in the survey. After an organization agreed to participate, all employees of the company were invited to participate in the online safety climate survey administered by the research team. Example items include: "Top management at this company tries to continually improve safety levels in each department," and "My direct supervisor discusses how to improve safety with us." The items were all on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Raw data were handled by only the research team, and the lead consultant received only a report with analyzed, aggregated data to share with the customer. No identifiable personal information was collected from participants.

Survey data were collected from 29,185 frontline employees of 46 companies from various industries (e.g., manufacturing, construction, and transportation). Six respondents did not answer more than 50% of the scale questions, so they were excluded from the analysis, leaving a final sample for analysis of 29,179 participants. Company size ranged from 45 to 12,000, with an average of 1274 employees. The within-company response rate ranged from 30.16% to 98.83%, with an average of 62.39%.

### 2.2. Data analysis procedure

#### 2.2.1. IRT analysis

IRT analyses were performed with the R open source package LTM (Latent Trait Modeling) developed by Rizopoulos (2006). IRT assumes the scale items are measuring a single construct, representing the target trait. Hence, unidimensionality of the OSC and GSC scales were individually examined before running IRT analyses. Both discrimination and difficulty parameters for every item of the safety climate scale were calculated. The discrimination (or differentiating) parameter represents the slope of the ORFs that capture the relationship between the latent construct (i.e., overall safety climate perception) and the probability of endorsing a particular response option for each item's response options. The standardized discrimination parameter (e.g., *z*-score) can be used to judge the statistical significance of the item's trait-differentiating capacity such that if it is greater than 1.96, it is significant at $p < 0.05$. The difficulty parameters determine the location of the ORF along the axis of θ (i.e., latent trait; representing overall level of safety climate perception).

Based on the discrimination and difficulty parameters, the Item Information Curve (IIC) for each of the 32 items can be generated. The IIC shows the distribution of information an item provides on a continuum of the estimated level of the latent trait, θ. The area of IIC above the θ axis represents the amount of information provided by a specific item across the entire continuum of the latent trait of interest. An item typically offers a larger amount of item information if it has a greater discriminating parameter (i.e., steeper slopes) and a broader range of difficulty parameters along the θ axis.

The Test Information Function (TIF) for a scale can be generated by aggregating all the IICs of the items included in the scale. Similar to IIC, the area of TIF above the θ axis equals the total test information. Our aim was to shorten the original scales by selecting the items that provided the most information, while also ensuring the TIFs of the shortened scales maintained a shape that was similar to those of the original scales. Two approaches were used to determine how many items should be included in the shortened scales, as described below.

*2.2.1.1. Shortening via item information criteria.* We first shortened the original OSC and GSC scales by selecting items that offered above-average information because an item with more information can more precisely differentiate the overall level of OSC or GSC based on respondents' ratings on the item. The amount of the information is indicated by the area under the item information function curve across the θ axis. For a 16-item scale, if each item is assumed to differentiate the level of OSC or GSC by an equal amount, each item should provide 6.25% of the total test information (i.e., 100% divided by 16 items). In reality, some items have better discriminating ability than others. In other words, they provide more than 6.25% of the total test information. Therefore, we shortened the original OSC and GSC scales by selecting items that had better than average discriminating ability (i.e. providing more than 6.25% of total test information).

*2.2.1.2. Shortening via total test information.* At the same time, in order to give companies more flexibility in the scale length they select, the original OSC and GSC scales were further shortened and made more concise by selecting the most discriminating items that, in total, retained at least 30% of the original total scale information (c.f., 100% information by entire 16 items, respectively for OSC and GSC scales). Put differently, we retained items with the highest percentages of information until the sum of item information was equal to or greater than 30%. It should be noted that the 30% criterion was chosen in consideration of the minimum number of items (i.e., over three; Kenny, 2016) needed to ensure model identification in a confirmatory factor analysis (CFA) and acceptable reliability of the scale (Cortina, 1993). We tested the correlations between scores of these more concise scales and the original scales to examine the representativeness of the shorter versions and to justify the appropriateness of using the criterion of retaining at least 30% of total scale information (see Section 2.2.3).

#### 2.2.2. Reliability test

We calculated the Cronbach's alpha of all shortened scales to determine the reliability of the shortened versions of the safety climate scales. The generally accepted criterion for good internal consistency (i.e., Cronbach's alpha = 0.70) was used (Nunnally and Bernstein, 1994).

#### 2.2.3. Validity test

After we created and calculated the mean scores of the shortened versions of the safety climate scales, we then examined the convergent validity of the shortened and original scales by calculating the correlation between the scales' mean scores. Generally, a correlation between two variables of greater than 0.80 (Brown, 2006) or 0.85 (Kenny, 1979) indicates the two variables are measuring the same construct. Because the validity of the original Zohar and Luria (2005) safety climate scale has been demonstrated in various previously-published scientific articles (e.g., Zohar and Luria, 2005), if these two scales are demonstrated to measure the same construct (i.e., correlation coefficients between scores on the shortened and original versions fall above the recommended values), we are able to infer the validity of the IRT-based shortened version of the safety climate scale.

#### 2.2.4. Supplemental test for robustness

We further cross-validated the results by running analyses with 50% of the dataset (Davison and Hinkley, 1997) to examine the consistency of results regarding which items are most discriminating. We randomly selected 50% of respondents in each company to create two company-level stratified split-half samples. We ran the IRT analyses using the two split-half samples and compared the discrimination and difficulty parameters. When results are consistent and robust across the split-half samples, we report the results using only the whole sample.

**Table 1**
Basic descriptive information for OSC scale.

| Item | Option1 | Option2 | Option3 | Option4 | Option5 | Mean | SD |
|---|---|---|---|---|---|---|---|
| OSC1 | 2.49% | 4.36% | 13.97% | 42.78% | 36.40% | 4.06 | 0.95 |
| OSC2 | 1.79% | 3.62% | 15.61% | 45.49% | 33.50% | 4.05 | 0.89 |
| OSC3 | 1.70% | 3.27% | 13.76% | 44.75% | 36.52% | 4.11 | 0.88 |
| OSC4 | 2.57% | 5.64% | 14.75% | 41.47% | 35.57% | 4.02 | 0.98 |
| OSC5 | 2.87% | 6.29% | 18.84% | 40.73% | 31.27% | 3.91 | 1.00 |
| OSC6 | 2.87% | 6.88% | 19.66% | 40.36% | 30.23% | 3.88 | 1.01 |
| OSC7 | 2.36% | 5.34% | 17.54% | 41.76% | 32.99% | 3.98 | 0.96 |
| OSC8 | 2.98% | 5.32% | 29.41% | 37.57% | 24.72% | 3.76 | 0.98 |
| OSC9 | 1.75% | 3.39% | 18.85% | 45.15% | 30.87% | 4.00 | 0.89 |
| OSC10 | 2.57% | 6.52% | 21.12% | 39.33% | 30.47% | 3.89 | 1.00 |
| OSC11 | 1.56% | 3.23% | 18.56% | 46.84% | 29.82% | 4.00 | 0.87 |
| OSC12 | 3.23% | 5.98% | 20.16% | 41.81% | 28.82% | 3.87 | 1.00 |
| OSC13 | 3.52% | 7.71% | 23.56% | 39.84% | 25.37% | 3.76 | 1.03 |
| OSC14 | 1.83% | 3.95% | 16.35% | 44.12% | 33.75% | 4.04 | 0.91 |
| OSC15 | 3.27% | 7.02% | 19.49% | 39.47% | 30.75% | 3.87 | 1.03 |
| OSC16 | 2.15% | 3.39% | 19.46% | 42.78% | 32.22% | 4.00 | 0.92 |

*Note:* Percentage represents the percentage of respondents who endorsed specified options 1–5 on a 5-point Likert scale for each item (Option1 = completely disagree, Option5 = completely agree). OSC = organizational-level safety climate. OSC1-OSC16 refer to the original 16 items in Zohar and Luria (2005).

## 3. Results

### 3.1. Basic descriptive

The mean OSC and GSC scores for the original, full-length scales were 3.95 (SD = 0.76) and 3.97 (SD = 0.79), respectively. Tables 1 and 2 list the option endorsement percentages (percentage of respondents who endorsed specified options 1–5 on a 5-point Likert scale for each item), mean score, and standard deviation for each item of the OSC and GSC scales, respectively.

### 3.2. Unidimensionality

We tested the unidimensionality of the OSC and GSC scales using Mplus 6.1 (Muthén and Muthén, 2010). Results of a confirmatory factor analysis (CFA) showed good model fit for a one-factor model of the OSC scale, $\chi^2$ (N = 29,179, df = 104) = 18648.91, $p < 0.001$, CFI = 0.95, TLI = 0.94, RMSEA = 0.078, 90% C.I. = [0.077, 0.079], and SRMR = 0.027. A one-factor model also fit well for the GSC scale, $\chi^2$ (N = 29,179, df = 104) = 18750.93, $p < 0.001$, CFI = 0.96, TLI = 0.95, RMSEA = 0.078, 90% C.I. = [0.077, 0.079], and SRMR = 0.023.

**Table 2**
Basic descriptive information for GSC scale.

| Item | Option1 | Option2 | Option3 | Option4 | Option5 | Mean | SD |
|---|---|---|---|---|---|---|---|
| GSC1 | 1.94% | 3.96% | 13.09% | 44.12% | 36.89% | 4.10 | 0.91 |
| GSC2 | 1.88% | 4.53% | 16.07% | 44.00% | 33.52% | 4.03 | 0.92 |
| GSC3 | 2.02% | 4.50% | 17.14% | 43.32% | 33.02% | 4.01 | 0.93 |
| GSC4 | 1.96% | 4.43% | 18.23% | 43.78% | 31.60% | 3.99 | 0.92 |
| GSC5 | 2.53% | 5.96% | 19.36% | 41.44% | 30.72% | 3.92 | 0.98 |
| GSC6 | 2.24% | 5.31% | 18.97% | 41.97% | 31.51% | 3.95 | 0.96 |
| GSC7 | 6.38% | 7.94% | 18.82% | 37.92% | 28.94% | 3.75 | 1.14 |
| GSC8 | 2.73% | 5.51% | 20.41% | 41.11% | 30.24% | 3.91 | 0.98 |
| GSC9 | 1.94% | 3.64% | 18.08% | 43.99% | 32.34% | 4.01 | 0.91 |
| GSC10 | 1.80% | 3.56% | 15.48% | 44.05% | 35.11% | 4.07 | 0.90 |
| GSC11 | 1.72% | 2.44% | 17.52% | 41.95% | 36.37% | 4.09 | 0.89 |
| GSC12 | 4.44% | 6.53% | 22.57% | 36.90% | 29.56% | 3.81 | 1.07 |
| GSC13 | 2.54% | 4.73% | 24.81% | 39.42% | 28.50% | 3.87 | 0.97 |
| GSC14 | 3.04% | 6.09% | 22.30% | 40.57% | 28.00% | 3.84 | 1.00 |
| GSC15 | 2.46% | 5.82% | 20.65% | 40.28% | 30.79% | 3.91 | 0.98 |
| GSC16 | 1.78% | 2.28% | 13.27% | 39.94% | 42.73% | 4.20 | 0.88 |

*Note:* Percentage represents the percentage of respondents who endorsed specified options 1–5 on a 5-point Likert scale for each item (Option1 = completely disagree, Option5 = completely agree). GSC = group-level safety climate. GSC1-GSC16 refer to the original 16 items in Zohar and Luria (2005).

### 3.3. IRT model results

#### 3.3.1. IRT model testing

We fit the items using graded response models (GRM; Samejima, 1997) because the OSC and GSC were all based on polytomous responses (i.e., five response options). GRM estimates one slope parameter and four difficulty parameters for each five-option item of the original scales. Two GRM models were estimated and compared for each scale: (1) a parsimonious GRM that specified an equal discrimination parameter for all of the items; and (2) a full GRM that freely estimated a discrimination parameter for each item. The first model was nested within the second model. Therefore, comparison of the change in −2*loglikelihood (−2*LL, which is based on a Chi-square distribution) can evaluate which model fit better.

For the OSC scale, the parsimonious GRM yielded a −2*LL value of −414488.3, AIC = 829106.6, BIC = 829644.8, whereas the full GRM resulted in a −2*LL value of −412756.4, AIC = 825672.9, BIC = 826335.4. The likelihood ratio test yielded a LRT = 3463.71, df = 15, $p < 0.001$. This indicates that the full GRM was significantly better than the parsimonious GRM, and the sixteen OSC items had significantly different discrimination parameters.

The GSC scale had similar results. The parsimonious GRM yielded a −2*LL value of −374430.7, AIC = 748991.3, BIC = 749529.6, whereas the full GRM resulted in a −2*LL value of −369942.4, AIC = 740044.8, BIC = 740707.3. The likelihood ratio test yielded a LRT = 8976.51, df = 15, $p < 0.001$. This means the full GRM fit better and the sixteen items had significantly different discrimination parameters.

For the two full GRM models, we also examined the model-data fit. The value of $\chi^2$/df for all possible item pairs and item triples of both OSC and GSC scales were less than 1, which indicates the two full GRM models fit well to the data (Chernyshenko et al., 2001).

#### 3.3.2. IRT parameters and information

Tables 3 and 4 list the parameter and information results of the full GRM models for OSC and GSC items, respectively. Fig. 1a and b depict the item information curve for each item of the OSC and GSC scales, respectively. Fig. 2a and b solid lines show the total test information function for the OSC and GSC scales, respectively.

For the 16 OSC items, the discrimination parameters ranged from 1.98 to 3.35, and the percentage of total test information each item provided ranged from 4.28% to 8.81%. This is consistent with previous model comparison results and indicates considerable variation in the OSC items' discrimination ability. The difficulty parameters reflected a sizeable range of the underlying construct, OSC (−2.74 to 0.92), indicating that the OSC scale was generally more useful in identifying companies with poor to average OSC safety climate scores than very high OSC scores (i.e., approximately 1SD+ mean range).

Results for the 16 GSC items were quite similar. The discrimination parameters ranged from 1.70 to 3.77, and the percentage of total test information each item provided ranged from 2.74% to 8.31%. This is consistent with previous model comparison results and indicates considerable variation in the GSC items' discrimination ability. The difficulty parameters reflected a sizeable range of the underlying construct, GSC (−2.86 to 0.86), indicating that the GSC scale was generally more useful in identifying companies with poor and average GSC scores than very high level of safety climate (i.e., approximately 1SD + mean range).

#### 3.3.3. Item selection for the shortened scales

*3.3.3.1. Item information criteria method.* First, we shortened the scales by selecting items that had above-average discriminating ability (i.e. provided more than 6.25% of total test information), as described

**Table 3**
Results of parameters and information of the full GRM for OSC items.

| Item | Slope (Discrimination) | Difficulty | | | | Information Total Test = 131.92 | |
|---|---|---|---|---|---|---|---|
| | | Diff1 | Diff2 | Diff3 | Diff4 | Value (Rank) | Percentage |
| OSC1 | 2.51 | −2.53 | −1.85 | −0.99 | 0.45 | 7.51 (11) | 5.69% |
| OSC2 | 2.55 | −2.74 | −1.99 | −0.97 | 0.55 | 8.05 (10) | 6.10% |
| ***OSC3*** | ***2.99*** | ***−2.65*** | ***−1.95*** | ***−1.03*** | ***0.41*** | ***9.73(2)*** | ***7.38%*** |
| OSC4 | 2.30 | −2.57 | −1.76 | −0.93 | 0.49 | 6.75 (14) | 5.12% |
| OSC5 | 2.61 | −2.39 | −1.63 | −0.71 | 0.62 | 8.07 (9) | 6.12% |
| **OSC6** | **2.75** | **−2.35** | **−1.56** | **−0.66** | **0.65** | **8.71 (7)** | **6.60%** |
| OSC7 | 2.21 | −2.66 | −1.82 | −0.85 | 0.59 | 6.59 (15) | 5.00% |
| OSC8 | 2.28 | −2.48 | −1.78 | −0.42 | 0.92 | 6.97 (13) | 5.28% |
| ***OSC9*** | ***2.84*** | ***−2.65*** | ***−1.94*** | ***−0.83*** | ***0.62*** | ***9.31(3)*** | ***7.06%*** |
| OSC10 | 2.35 | −2.55 | −1.68 | −0.65 | 0.68 | 7.23 (12) | 5.48% |
| ***OSC11*** | ***3.35*** | ***−2.63*** | ***−1.91*** | ***−0.83*** | ***0.63*** | ***11.62(1)*** | ***8.81*** |
| **OSC12** | **2.79** | **−2.27** | **−1.60** | **−0.65** | **0.70** | **8.79 (6)** | **6.66%** |
| **OSC13** | **2.67** | **−2.23** | **−1.48** | **−0.50** | **0.85** | **8.44 (8)** | **6.40%** |
| ***OSC14*** | ***2.87*** | ***−2.61*** | ***−1.85*** | ***−0.89*** | ***0.52*** | ***9.30(4)*** | ***7.05%*** |
| OSC15 | 1.98 | −2.56 | −1.70 | −0.71 | 0.70 | 5.64 (16) | 4.28% |
| **OSC16** | **2.88** | **−2.50** | **−1.89** | **−0.79** | **0.57** | **9.21 (5)** | **6.98%** |

*Note:* Bold indicates that the item was selected for the shortened 8-item scale; Italics (rank 1-4 in Value column) indicate that the item was selected for the more concise 4-item scale. GRM = graded response models; OSC = organization-level safety climate; OSC1-OSC16 refer to the original 16 items in Zohar and Luria (2005).

above. The shortened OSC scale included eight items: items 11, 3, 9, 14, 16, 12, 6, and 13 (descending order of information provided). This shortened OSC scale retained 56.94% of the total test information of the original scale. Reliability of the shortened 8-item OSC scale was 0.94. The difficulty parameters ranged from −2.65 to 0.85.The shortened GSC scale included 11 items: items 10, 4, 3, 9, 5, 13, 6, 2, 14, 11, and 15 (descending order of information provided). This shortened GSC scale retained 77.71% of the total test information of the original scale. Reliability of the shortened 11-item GSC was 0.97. The difficulty parameters ranged from −2.70 to 0.80.

The dashed lines in Fig. 2a and b demonstrate the test information functions of the shortened OSC and GSC scales, respectively. More specifically, they show how well the ratings on given sets of safety climate scale items are capable of precisely differentiating respondents with different levels of overall safety climate perceptions. According to the figures, the test information function curves of the two shortened scales are similar to those of the original scales in both shape and coverage across the safety climate continuum, which indicates that they are representative of the original scales. Although the shortening of the scale inevitably results in the shrinkage of area under the curves, which is the amount of scale information, the shrinkage was relatively less substantial considering the sizeable number of items that were removed. Also, general trends of the estimated safety climate level and scale information relationship were similar (see 3.3.4), suggesting that item reduction did not distort the original scales.

*3.3.3.2. Total test information method.* Because the shortened OSC and GSC scales together have 19 items, which may still be too long for some applications, we further shortened the original OSC and GSC scales. To provide more scale length options, we selected the most discriminating items that, in total, retained at least 30% of the original total test information. Based on this criterion, the more concise OSC scale included four items: items 11, 3, 9, and 14, which together retained 30.29% of the total test information of the original scale. Reliability of the four-item OSC scale was 0.89. The difficulty parameters ranged from −2.65 to 0.63.The more concise GSC scale included four items: items 10, 4, 3, and 9, which together retained 30.88% of the total test information of the original scale. Reliability of the four-item GSC scale was 0.92. The difficulty parameters ranged from −2.57 to 0.63.

The dotted lines in Fig. 2a and b depict the test information

**Table 4**
Results of parameters and information of the full GRM for GSC items.

| Item | Slope (Discrimination) | Difficulty | | | | Information Total Test = 159.28 | |
|---|---|---|---|---|---|---|---|
| | | Diff1 | Diff2 | Diff3 | Diff4 | Value (Rank) | Percentage |
| GSC1 | 2.43 | −2.79 | −2.01 | −1.11 | 0.49 | 7.46 (13) | 4.68% |
| **GSC2** | **3.17** | **−2.63** | **−1.83** | **−0.89** | **0.57** | **10.79 (8)** | **6.77%** |
| ***GSC3*** | ***3.41*** | ***−2.54*** | ***−1.78*** | ***−0.83*** | ***0.58*** | ***11.75(3)*** | ***7.38%*** |
| ***GSC4*** | ***3.60*** | ***−2.53*** | ***−1.79*** | ***−0.79*** | ***0.63*** | ***12.65(2)*** | ***7.94%*** |
| **GSC5** | **3.33** | **−2.40** | **−1.64** | **−0.69** | **0.67** | **11.38 (5)** | **7.14%** |
| **GSC6** | **3.19** | **−2.50** | **−1.72** | **−0.73** | **0.65** | **10.83 (7)** | **6.80%** |
| GSC7 | 1.70 | −2.32 | −1.69 | −0.77 | 0.86 | 4.37 (16) | 2.74% |
| GSC8 | 2.86 | −2.44 | −1.73 | −0.70 | 0.71 | 9.30 (12) | 5.84% |
| **GSC9** | ***3.37*** | ***−2.57*** | ***−1.89*** | ***−0.84*** | ***0.60*** | ***11.56(4)*** | ***7.26%*** |
| **GSC10** | ***3.77*** | ***−2.56*** | ***−1.88*** | ***−0.93*** | ***0.49*** | ***13.23(1)*** | ***8.31%*** |
| **GSC11** | **3.04** | **−2.70** | **−2.10** | **−0.92** | **0.47** | **10.01 (10)** | **6.28%** |
| GSC12 | 2.46 | −2.23 | −1.59 | −0.54 | 0.77 | 7.36 (14) | 4.62% |
| **GSC13** | **3.24** | **−2.42** | **−1.76** | **−0.55** | **0.76** | **11.03 (6)** | **6.92%** |
| **GSC14** | **3.15** | **−2.32** | **−1.62** | **−0.57** | **0.80** | **10.57 (9)** | **6.64%** |
| **GSC15** | **2.98** | **−2.48** | **−1.68** | **−0.65** | **0.69** | **9.97 (11)** | **6.26%** |
| GSC16 | 2.36 | −2.86 | −2.27 | −1.17 | 0.29 | 7.02 (15) | 4.41% |

*Note:* Bold indicates that the item was selected for the shortened 11-item scale; Italics (rank 1–4 in Value column) indicate that the item was selected for the more concise 4-item scale. GRM = graded response models; GSC = group-level safety climate; GSC1-GSC16 refer to the original 16 items in Zohar and Luria (2005).

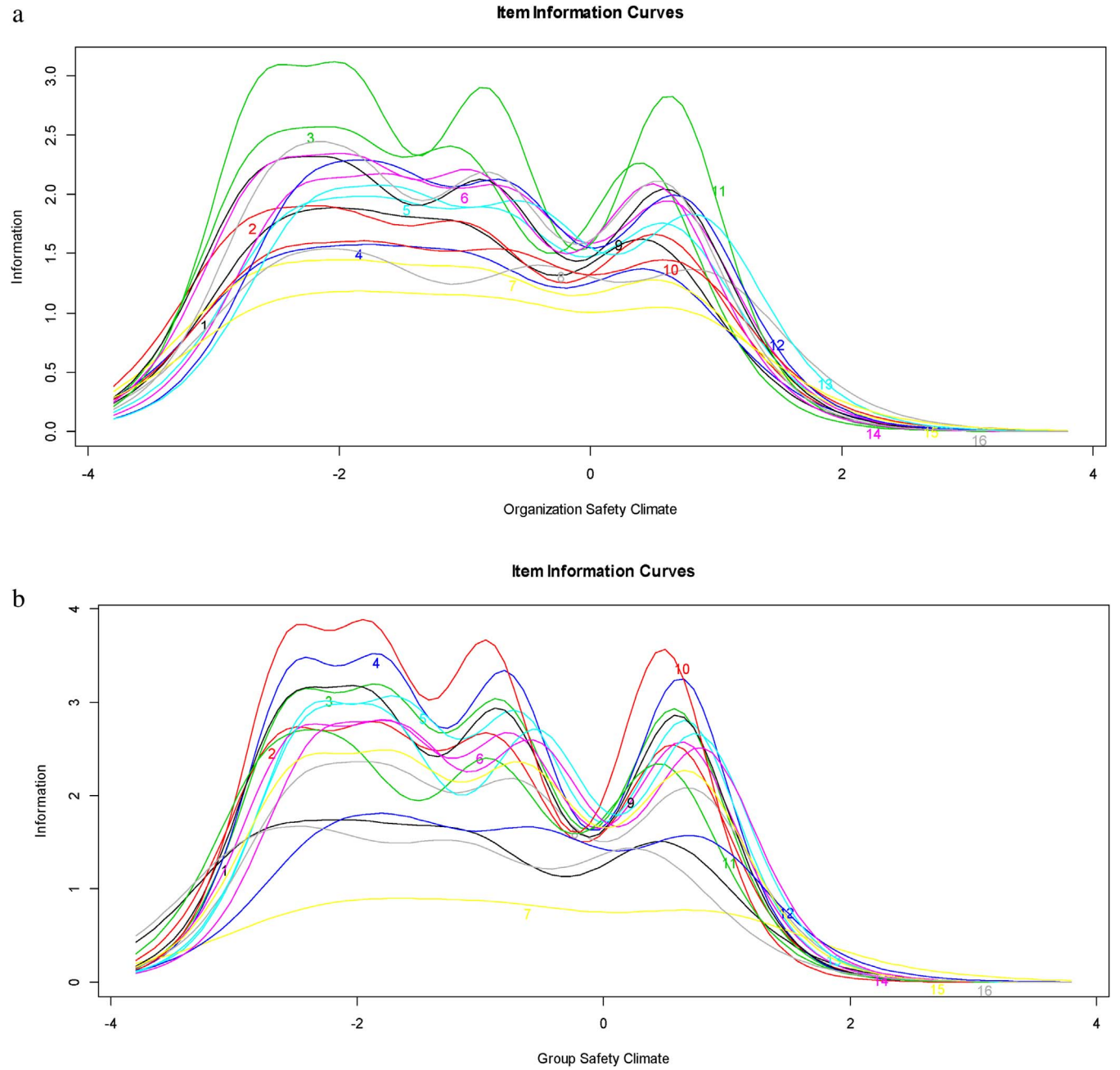


**Fig. 1.** (a) Item Information Curves for OSC items. (b) Item Information Curves for GSC items.*Note:* Each line represents an Item Information Curve of each SC scale item. Numbers indicate the item number in the original 16-item SC scale in Zohar and Luria (2005).

functions of these more concise OSC and GSC scales, respectively. The figures show that the test information function curves of the two more concise scales had shapes and coverage across the safety climate continuum similar to the original scales.

### 3.3.4. Preliminary validity evidence of the shortened scales

Results of the bivariate Pearson correlations between the original full-length scales and shortened scales, using their mean scores, are listed in Table 5. All the correlations were greater than 0.95 and significant ($p < 0.01$). Given that Zohar's original scales were predictive of important safety outcomes, the shortened scale scores should also be significantly related to those outcomes.

### 3.3.5. Supplemental analyses – split-half test for robustness

We further cross-validated the results by comparing IRT results of two split-half samples. Split-half sample A randomly selected 50% of the respondents from each company for a total number of 14589. Sample B consisted of the unselected 50% of respondents from each company for a total number of 14590. Results of the IRT analyses using the two split-half samples were consistent and robust across the two samples: the Pearson correlation coefficients of the slope and difficulty parameters for each item were all significantly correlated between sample A and sample B, $p < 0.05$. Furthermore, when using the two shortening scale methods described, the selected items remained the same for the two split-half samples. Therefore, we report the results using only the whole sample.

## 4. Discussion

The primary goal of the current study was to shorten Zohar and Luria's (2005) 32-item safety climate scale, which includes 16 items for organization-level safety climate (OSC) and 16 items for group-level

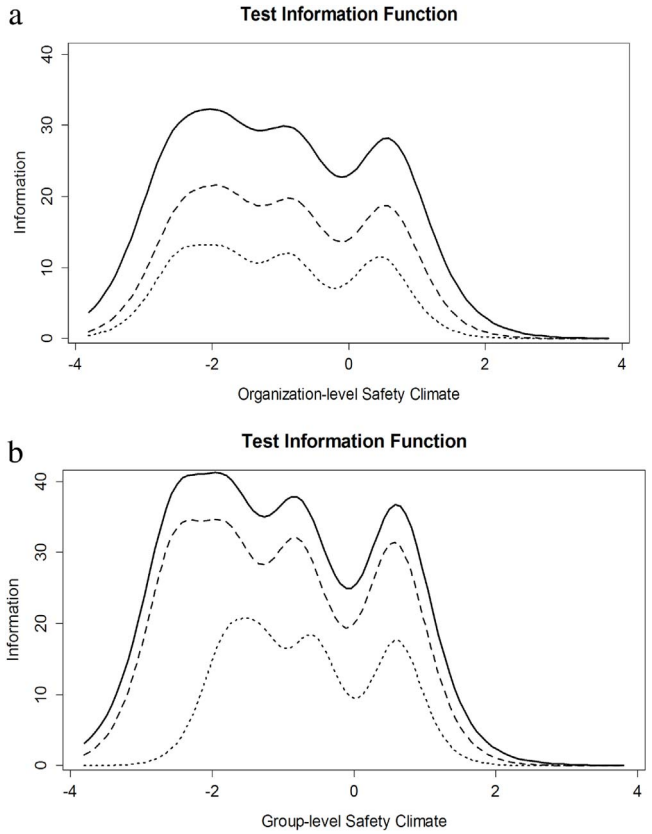


**Fig. 2.** (a) Total test information function (aggregation of all the item information curves) for the OSC scale. (b) Total test information function (aggregation of all the item information curves) for the GSC scale.*Note:* Solid lines = original 16-item scale for both OSC & GSC; Dashed lines = 8-item scale for OSC & 11-item scale for GSC; Dotted lines = 4-item scale for both OSC & GSC.

**Table 5**
Pearson correlations between original and shortened scale scores.

| Scale | | Score of Original 16-item scale |
|---|---|---|
| OSC | Score of Shortened 8-item scale | 0.98** |
| | Score of More Concise 4-item scale | 0.95** |
| GSC | Score of Shortened 11-item scale | 0.99** |
| | Score of More Concise 4-item scale | 0.96** |

*Note:* ** $p < 0.01$. OSC = organization-level safety climate; GSC = group-level safety climate.

safety climate (GSC), using an item response theory (IRT) analytical approach. We expect that a shortened safety climate scale will increase the practical utility of safety climate assessments by reducing respondent burden and increasing face validity, especially for users who are concerned with the amount of time needed for survey administration and the measurement integrity (e.g., reliability and validity). Moreover, a shortened safety climate scale would more likely allow researchers and practitioners to incorporate additional constructs into their survey assessment to advance the literature by, for example, examining and expanding the nomological network of safety climate.

Based on a series of IRT analyses using survey responses gathered from nearly 30,000 employees representing 46 companies in various industries, the discrimination parameters revealed that all OSC and GSC items in Zohar and Luria's (2005) original scale were able to effectively discriminate (or differentiate) between high and low levels of safety climate. However, the difficulty parameters indicated that, overall, the OSC and GSC items were more useful in identifying companies with poor and average safety climate scores than those with high safety climate scores.

Item information for each item was then computed as a function of both discrimination and difficulty parameters. We adopted two different procedures in shortening the OSC and GSC by 1) identifying items with above-average discriminating ability (i.e., items providing more than 6.25% of total test information) and 2) developing more concise scales that in total retained at least 30% of the original total test information, thus creating two shortened versions of the OSC scale and two shortened versions of the GSC scale.

The first procedure resulted in eight OSC items and eleven GSC items that each had above-average discriminating ability (i.e., over 6.25%) and, respectively, retained 56.94% and 77.71% of total test information (see Tables 3 and 4). In addition, these 8-item OSC and 11-item GSC scales both had acceptable Cronbach's alpha estimates (0.94 and 0.97 respectively) and significant correlations with the original scale scores, thus supporting the reliability of these shortened OSC and GSC scales.

The second procedure identified four OSC and four GSC items from the original scale that are needed to retain at least 30% of the original total test information. These 4-item OSC and 4-item GSC scales also had acceptable reliability estimates (0.89 and 0.92, respectively) and significant correlations with the original scale scores.

Depending on measurement needs and objectives, some users may prefer the 8-item OSC and 11-item GSC shortened versions, while others may prefer the 4-item OSC and 4-item GSC shortened versions. It is important to note that we are not arguing that one length is superior to the other; we adopted two lengths to provide researchers and practitioners with two different shortened scale options that they can choose from based on measurement purposes/objectives, study design, and available resources (e.g., time).

The current study makes important contributions to the literature, organizations, and safety professional communities in several ways. First, Zohar (2010) highlighted that gaps exist in our understanding of how safety climate emerges and how it is influenced. The shortened versions of OSC and GSC scales identified in the current study would allow researchers and practitioners to incorporate additional constructs into their survey instruments which could potentially explain the emergence or changes in safety climate. In other words, the use of shortened safety climate scales has the potential to increase the chances of expanding our understanding of the relationships between safety climate and other constructs.

Second, the shortened OSC and GSC scales identified in the current study are expected to broaden the usage of safety climate assessment in field settings while retaining acceptable levels of scale information. For example, a company would more likely be able to incorporate the shortened OSC and GSC scales into their existing employee assessments (e.g., employee opinion surveys) and, thus, increase understanding of safety climate in their organization.

The current study also has limitations that highlight directions for future research. First, even though we used a relatively large sample representing a number of companies, biases may exist in the survey responses because it is typically more common for organizations that prioritize safety to participate in safety climate assessments. For example, our results showed that the study participants scores were around 4 (mean OSC = 3.95, mean GSC = 3.97) out of a 5-point Likert scale, suggesting the possibility of that the sample was biased toward people who perceived a positive SC. However, as mentioned earlier, IRT parameters are not dependent on the level of target trait (i.e., SC) of the sample (Fan, 1998; Baker, 2001). In other words, the sample does not impact the estimate of the IRT parameters.

Second, we were not able to collect data on safety outcomes to validate our shortened scales. However, Zohar and Luria's original scale (2005) has been demonstrated to have good validity with quite a few safety outcomes. Because our four shortened OSC and GSC scale scores were strongly related to the original scale scores ($r > .95$), we believe that our shortened scales have good validity and can be used to predict safety outcomes. Future studies can consider collecting responses on

safety outcomes (e.g., self-reported safety behaviors and objective workers' compensation data) in order to establish criterion-related validity of the shortened scales.

Third, the range of the difficulty parameters of our shortened scales focuses on the low end of safety climate, which is similar to the range of difficulty parameters from the original scale items (see Tables 3 and 4). The low end difficulty range shows that our selected items are more useful in differentiating companies with poor, average, and better than average safety climate (less than +1 standard deviation), which is where safety improvement is most needed. However, these items were less efficient in differentiating the companies with highest safety climate (top 20%). Although this might be a minor issue, given safety climate assessment is commonly used for identifying companies with low safety climate for safety promotion, future studies may consider adding items with difficulty parameters in the higher end.

In conclusion, using an IRT analytical approach, the current study developed shortened versions of Zohar and Luria's (2005) 16-item OSC and 16-item GSC scales. Specifically, we identified 8 OSC and 11 GSC items with above-average discriminating ability, and further selected 4 OSC and 4 GSC items that retained at least 30% of the original total test information. It is our expectation that these shortened safety climate scales will increase the utility of safety climate assessments in both research and practice.

## Acknowledgements

The authors wish to thank the following team members for their invaluable assistance: Marvin Dainoff, Susan Jeffries and Peg Rothwell (Liberty Mutual Research Institute for Safety) for data collection, analysis and general assistance; Don Tolbert and Julie Thompson (Liberty Mutual Insurance) for technical consulting.

## Appendix A

A. Organization-Level Safety Climate Scales.

| OSC | Top management at this company: | 8-item | 4-item |
|---|---|---|---|
| 1. | Reacts quickly to solve the problem when told about safety hazards. | | |
| 2. | Insists on thorough and regular safety audits and inspections. | | |
| 3. | Tries to continually improve safety levels in each department. | X | X |
| 4. | Provides all the equipment needed to do the job safely. | | |
| 5. | Is strict about working safely when work falls behind schedule. | | |
| 6. | Quickly corrects any safety hazard (even if it's costly). | X | |
| 7. | Provides detailed safety reports to workers (e.g., injuries, near accidents). | | |
| 8. | Considers a person's safety behavior when moving–promoting people. | | |
| 9. | Requires each manager to help improve safety in his or her department. | X | X |
| 10. | Invests a lot of time and money in safety training for workers. | | |
| 11. | Uses any available information to improve existing safety rules. | X | X |
| 12. | Listens carefully to workers' ideas about improving safety. | X | |
| 13. | Considers safety when setting production speed and schedules. | X | |
| 14. | Provides workers with a lot of information on safety issues. | X | X |
| 15. | Regularly holds safety-awareness events (e.g., presentations, ceremonies). | | |
| 16. | Gives safety personnel the power they need to do their job. | X | |

*Note:* The original 16 items are from Zohar and Luria (2005). The 8-item and 4-item shortened scales are referred to as the Liberty Mutual Safety Climate Short Scales.

B. Group-Level Safety Climate Scales.

| GSC | My direct supervisor: | 11-item | 4-item |
|---|---|---|---|
| 1. | Makes sure we receive all the equipment needed to do the job safely. | | |
| 2. | Frequently checks to see if we are all obeying the safety rules. | X | |
| 3. | Discusses how to improve safety with us. | X | X |
| 4. | Uses explanations (not just compliance) to get us to act safely. | X | X |
| 5. | Emphasizes safety procedures when we are working under pressure. | X | |
| 6. | Frequently tells us about the hazards in our work. | X | |
| 7. | Refuses to ignore safety rules when work falls behind schedule. | | |
| 8. | Is strict about working safely when we are tired or stressed. | | |
| 9. | Reminds workers who need reminders to work safely. | X | X |
| 10. | Makes sure we follow all the safety rules (not just the most important ones). | X | X |
| 11. | Insists that we obey safety rules when fixing equipment or machines. | X | |
| 12. | Says a "good word" to workers who pay special attention to safety. | | |
| 13. | Is strict about safety at the end of the shift, when we want to go home. | X | |
| 14. | Spends time helping us learn to see problems before they arise. | X | |
| 15. | Frequently talks about safety issues throughout the work week. | X | |
| 16. | Insists we wear our protective equipment even if it is uncomfortable. | | |

*Note:* The original 16 items are from Zohar and Luria (2005). The 11-item and 4-item shortened scales are referred to as the Liberty Mutual Safety Climate Short Scales.

## References


Baker, F.B., The Basics of Item Response Theory, 2001, ERIC Clearinghouse on Assessment and Evaluation; College Park, MD. Retrieved from http://echo.edres.org:8080/irt/baker/. Original work was published by Heinemann in 1985.

Beus, J.M., Payne, S.C., Bergman, M.E., Arthur, W., 2010. Safety climate and injuries: an examination of theoretical and empirical relationships. J. Appl. Psychol. 95 (4), 713–727.

Brondino, M., Silva, S.A., Pasini, M., 2012. Multilevel approach to organizational and group safety climate and safety performance: co-workers as the missing link. Saf. Sci. 50 (9), 1847–1856.

Brown, T.A., 2006. Confirmatory Factor Analysis for Applied Research. Guilford Press, New York.

Chernyshenko, O.S., Stark, S., Chan, K.Y., Drasgow, F., Williams, B., 2001. Fitting item response theory models to two personality inventories: issues and insights. Multivariate Behav. Res. 36 (4), 523–562.

Christian, M.S., Bradley, J.C., Wallace, J.C., Burke, M.J., 2009. Workplace safety: a meta-analysis of the roles of person and situation factors. J. Appl. Psychol. 94, 1103–1127.

Cortina, J.M., 1993. What is coefficient alpha? An examination of theory and applications. J. Appl. Psychol. 78, 98–104.

Crawford, S.D., Couper, M.P., Lamias, M.J., 2001. Web surveys perceptions of burden. Soc. Sci. Comput. Rev. 19 (2), 146–162.

Davison, A.C., Hinkley, D.V., 1997. Bootstrap Methods and Their Application, vol. 1 Cambridge University Press, Cambridge, UK.

Dejoy, D.M., Schaffer, B.S., Wilson, M.G., Vandenbert, R.J., Butts, M.M., 2004. Creating safer workplaces: assessing the determinants and role of safety climate. J. Saf. Res. 35, 81–90.

Fan, X., 1998. Item response theory and classical test theory: an empirical comparison of their item/person statistics. Educ. Psychol. Meas. 58 (3), 357–381.

Fisher, G.G., Matthews, R.A., Gibbons, A.M., 2016. Developing and investigating the use of single-item measures in organizational research. J. Occup. Health Psychol. 21 (1), 3–23.

Griffin, M.A., Neal, A., 2000. Perceptions of safety at work: a framework for linking safety climate to safety performance, knowledge, and motivation. J. Occup. Health Psychol. 5 (3), 347–358.

Huang, Y.H., Zohar, D., Robertson, M.M., Garabet, A., Lee, J., Murphy, L.A., 2013a. Development and validation of safety climate scales for lone workers using truck drivers as exemplar. Transp. Res. F: Traffic Psychol. Behav. 17, 5–19.

Huang, Y.H., Zohar, D., Robertson, M.M., Garabet, A., Murphy, L.A., Lee, J., 2013b. Development and validation of safety climate scales for mobile remote workers using utility/electrical workers as exemplar. Accid. Anal. Prev. 59, 76–86.

Johnson, S.E., 2007. The predictive validity of safety climate. J. Saf. Res. 38, 511–521.

Kenny, D.A., 1979. Correlation and Causality. Wiley-Interscience, New York.

Kenny, D.A., 2016. Multiple Latent Variable Models: Confirmatory Factor Analysis. Lecture Notes Online Web site. Retrieved from http://davidakenny.net/cm/mfactor.htm.

Martínez-Córcoles, M., Gracia, F., Tomás, I., Peiró, J.M., 2011. Leadership and employees' perceived safety behaviours in a nuclear power plant: a structural equation model. Saf. Sci. 49 (8), 1118–1129.

Muthén, L.K., Muthén, B.O., 2010. Mplus: the comprehensive modeling program for applied researchers. User's Guide V 6. 1. Muthén & Muthén, Los Angeles.

Nahrgang, J.D., Morgeson, F.P., Hofmann, D.A., 2011. Safety at work: a meta-analytic investigation of the link between job demands, job resources, burnout, engagement, and safety outcomes. J. Appl. Psychol. 96 (1), 71–94.

Nunnally, J.C., Bernstein, I.H., 1994. Psychometric Theory, 3rd edition. McGraw-Hill, New York.

Probst, T.M., 2003. Development and validation of the Job Security Index and the Job Security Satisfaction Scale: A classical test theory and IRT approach. J. Occup. Organiz. Psychol. 76, 451–467.

Rizopoulos, D., 2006. LTM: An R package for latent variable modeling and item response theory analyses. J. Stat. Software 17 (5), 1–25.

Samejima, F., 1997. Graded response model. In: van der Linden, W.J., Hambleton, R.K. (Eds.), Handbook of Modern Item Response Theory. Springer, New York, pp. 85–100.

Wanous, J.P., Reichers, A.E., Hudy, M.J., 1997. Overall job satisfaction: how good are single-item measures? J. Appl. Psychol. 82 (2), 247–252.

Woods, S.A., Hampson, S.E., 2005. Measuring the Big Five with single items using a bipolar response scale. Eur. J. Pers. 19 (5), 373–390.

Zohar, D., Luria, G., 2005. A multilevel model of safety climate: cross-level relationships between organization and group-level climates. J. Appl. Psychol. 90, 616–628.

Zohar, D., 1980. Safety climate in industrial organizations: theoretical and applied implications. J. Appl. Psychol. 65, 96–102.

Zohar, D., 2000. A group-level model of safety climate: testing the effects of group climate on microaccidents in manufacturing jobs. J. Appl. Psychol. 85, 587–596.

Zohar, D., 2002. The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. J. Org. Behav. 23, 75–92.

Zohar, D., 2003. Safety climate: conceptual and measurement issues. In: Quick, J.C., Tetrick, L.E. (Eds.), Handbook of Occupational Health Psychology. American Psychological Association, Washington, DC, pp. 123–142.

Zohar, D., 2010. Thirty years of safety climate research: reflections and future directions. Accid. Anal. Prev. 42 (5), 1517–1522.