GUIDELINES FOR A CULTURE OF SAFETY SURVEY

THAT DEMONSTRATES VALIDITY, CONSISTENCY, AND RELIABILITY

For the purposes of reporting on Safe Practice 2 Culture Measurement, Feedback, and Intervention, hospitals must conduct a culture of safety survey of their employees (the units surveyed must account for at least 50% of the aggregated care delivered to patients within the facility, and includes the high patient safety risk units or departments) using a "nationally recognized tool that has demonstrated validity, consistency and reliability" within the past 24 months. There are numerous nationally recognized culture of safety survey tools that have demonstrated validity, consistency, and reliability in peer-reviewed literature. However, some hospitals have stated these culture surveys do not meet their needs. These hospitals often choose to create their own surveys. In consultation with Leapfrog's Culture of Safety Expert Panel, we have developed a set of guidelines for what constitutes a valid, consistent, and reliable survey tool for the purposes of reporting on Safe Practice 2 on the Leapfrog Hospital Survey.

The Leapfrog Culture of Safety Expert Panel has identified a set of options for survey content:

- **Option 1:** Unaltered, nationally recognized tool with information on the validity, consistency, and reliability of the tool published in peer-reviewed literature. Examples of Option 1 surveys are:
    - AHRQ Hospital Survey on Patient Safety Culture
    - Patient Safety Climate for Healthcare Organizations
    - Safety Attitudes Questionnaire
    - SCORE
    - Zohar
- **Option 2:** Minor modifications to an Option 1 tool, such as simple word changes that do not alter the intent or tone of questions (e.g., changing "unit" to "department"). However, the survey is utilized in its entirety.
- **Option 3:** Major modifications to an established tool or development of a self-created tool (e.g., re-wording questions from an Option 1 tool from negative to positive, employee engagement surveys that include some questions related to culture of safety, etc.). Examples of Option 3 surveys that have already been tested and have demonstrated the criteria outlined below are:
    - Press Ganey Safety Solution
    - Glint Patient Safety Pulse (GPS Pulse)

Hospitals that use an **Option 3** survey (as described above), other than the Press Ganey Safety Solution, must ensure that the tool demonstrates similar psychometric properties to an **Option 1** survey for the purposes of

reporting on Safe Practice 2 Culture Measurement, Feedback, and Intervention on the Leapfrog Hospital Survey. Hospitals that use an **Option 3** survey may be asked to provide documentation that the tool has been assessed for validity, consistency, and reliability.

While Leapfrog encourages hospitals to use an Option 1 survey over Option 2 or Option 3 surveys as standardized tools help the understanding of broader trends, Leapfrog scores hospitals in Safe Practice 2 equivalently regardless of which Option survey they use, as long as the tool has been demonstrated to be valid, consistent, and reliable.

Criteria for an Option 3 Survey

The Expert Panel did not identify specific domains or questions that need to be included to be considered a "culture of safety survey," however, Option 3 survey tools need to be grounded in an underlying theory of the organization's culture. The tool developers must establish why they included the questions and domains in the survey to support the theory of culture. In addition to establishing a theory of culture, hospitals must document that the *tool demonstrates three of the four following qualities:

The testing done to demonstrate the validity, consistency, and reliability of an Option 3 survey tool needs to include ALL the questions that make-up the Culture of Safety survey (i.e., the entire survey tool), not just new questions or new domains. This is done, as to understand how the survey as a whole demonstrates the desired qualities.

> **1. Scale reliability**.
>
> One common assessment of reliability is Cronbach's Alpha, a statistical index of internal consistency that also provides an estimate of the ratio of true score to error in Classical Test Theory.
>
> **2. Exploratory/confirmatory factor analysis** (this quality is **required** if creating new domain or composite scores).
>
> Exploratory Factor Analysis (EFA) is a technique whose overarching goal is to identify the underlying relationships between measured variables. EFA is commonly used by researchers when developing a scale and serves to identify a set of latent constructs underlying a battery of measured variables.
>
> Confirmatory Factor Analysis (CFA) is used to test whether measures of a construct are consistent with a researcher's understanding of the nature of that construct. The objective of CFA is to test whether the data fit a hypothesized measurement model.
>
> **3. The tool was assessed and shown to be valid**, which could include:

Convergent/divergent validity. To establish convergent validity, one needs to show that measures that should be related are in reality related. To establish discriminant validity, one needs to show that measures that should not be related are in reality not related. Both are commonly demonstrated by examining the item inter-correlations for all item pairings.

Predictive validity. Predictive validity is the extent to which a score on a scale predicts scores on some criterion. As an example, the predictive validity of a culture of safety scale would be demonstrated by examining the correlation between a unit's score on the scale (at a point in time) and rates of preventable harm in the unit (at a future point in time).

Face validity. Face validity is a measure of how representative a scale or tool is 'at face value,' and whether it appears to be a good scale/tool. On a measure of happiness, for example, the test would be said to have face validity if it appeared to actually measure levels of happiness. Face validity only means that the test looks like it works; it does not mean that the test has been proven to work.

4. The tool underwent cognitive testing.

Data from cognitive interviews are qualitative, and analysis of this data can identify sources of response error as well as various interpretations of questions. By conducting a comparative analysis of the cognitive interviews, it is possible to identify patterns of error and misinterpretation across groups of people.

For a more complete understanding of these concepts, please see the following articles:

Sexton JB, Helmreich RL, Neilands TB, Rowan K, Vella K, Boyden J, et al. The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Services Research*. 2006;6:44.

Singer SJ, Meterko M, Baker L, Gaba G, Falwell A, and Rosen A. Workforce perceptions of hospital safety culture: Development and validation of the Patient Safety Climate in Healthcare Organizations survey. *Health Services Research.* 2007;42(5):1999.

Sorra J and Dyer N. Multilevel psychometric properties of the AHRQ hospital survey on patient safety culture. *BMC Health Services Research.* 2010;10:199

The following guidelines should be considered for survey administration:
- Employees completing the survey should have familiarity with the facility (i.e., not employees who rarely work in the facility).
- The survey should be administered to, at least, the frontline clinical staff and leadership.
- The survey should be administered in a way that avoids coercion.

- The survey's sample population should be representative of the facility's clinical and administrative workforce.
- High response rates are considered more important than high survey scores.