Safety Culture Survey: Measuring Your Organization's Safety Engagement

As an operator's safety culture matures, engagement and commitment to continuous safety improvement is a long-term goal.

January 31, 2022

Table of Contents

2021 Safety Culture Annual Report

Summary

Baldwin Safety & Compliance's 2021 Safety Culture Survey (Baldwin's SCS) reports that operators participating in a Safety Performance Monitoring Program (SPMP) received higher scores in safety culture assessment than non-participants.

The survey confirms quantifiable results: as an operator's safety culture matures, engagement and commitment to continuous safety improvement is a long-term goal. The results reflect higher average scores by SPMP operators within the safety culture segment. The scores further suggest the successful use of Baldwin's Safety Management System (SMS) is attributable to the hands-on mentoring by the Baldwin safety team of experts as part of the SPMP program.

The current Baldwin's SCS was compiled in December 2021 with data from more than 800 respondents: 111 of them SPMP participants. Higher average scores by SPMP operators were found within the safety culture segment.

About the SPMP Program

Baldwin's Safety Performance Monitoring Program (SPMP) has been available to eligible Baldwin clients since 2019. The program utilizes Baldwin's algorithm-based Safety Monitor which tracks key components of safety management to provide definitive information that an auditor, client, and most importantly, an operator can use to assess its SMS performance. The standout part of the program is the mentoring as well as continuous monitoring by Baldwin's safety professionals.

Baldwin's SPMP begins with an initial safety culture survey and an overall safety assessment as part of the entrance into the program. The survey is then administered periodically to provide an analysis and indicate the maturation of an organization's safety culture.

The results can help operators attain an integrated safety culture like the Stage 3 IS-BAO/IS-BAH registration requires, which is the highest international standard level an aviation organization can achieve.

Baldwin Safety & Compliance's 2021 Safety Culture Survey (Baldwin's SCS) is designed to gather information on employee perceptions of the current safety culture by measuring four dimensions. The annual safety culture report aggregates and analyzes the survey responses for both offerings over the calendar year.

Baldwin offers its Safety Culture Survey to its clients requesting this service. The survey offerings are designated as the lite survey, providing only quantitative feedback on the survey questions, and the full survey, which provides survey feedback and professional assessment. The lite offering is reserved for only Safety Performance Monitoring Program (SPMP) members, while the full offering is available to all.

Background:

When it comes to safety management systems (SMS) - or more accurately implementing an effective SMS - safety culture is a term that almost becomes a mantra in an organization. Often it is heard that an organization desires to *get* a safety culture or *obtain* a just culture. However, safety culture is not something that can be acquired, bought, or otherwise obtained (Skybrary, 2019). It is in this context that some organizations feel that a safety culture needs to be obtained to have a complete SMS. While this is not too far from fact, it is essential to see the actual benefits of fostering a culture that embraces safety and safety practices. However, to do this, it is also necessary to understand what organizational culture is, how it relates to safety culture, and ultimately how measuring safety culture can be the foundation to understanding and improving the organization.

Safety culture, as a concept, has been studied extensively in academia. The findings of this research generally all support the notion that positive safety culture and positive worker safety behaviors are linked (Sellers, 2014; Zohar, 2010). In other words, an

organization with a strong safety culture is more likely to have fewer accidents. In the context of SMS, it has been stated that the policies, processes, and procedures of an SMS combined with safety culture help to define the overall safety performance of an organization. It is important to note that every organization has a safety culture that can loosely be defined in terms of weak, neutral, or strong (Robbins & Judge, 2017; Skybrary, 2019). Therefore, a strong safety culture is what organizations strive after, in conjunction with safety management, to realize enhanced safety performance. Understanding safety culture, however, as a trait of organizational culture is vital in understanding that the shared values and beliefs are the underlying foundation of what is seen day-to-day regarding safety behavior in the organization.

Organizational culture - like safety culture - is also widely studied in the context of organizational performance. Leaders are always looking for that environment that fosters highly driven, innovative, and entrepreneurial employees. Indeed, the culture of an organization has a profound impact on workplace behaviors (Sellers, 2014). Sellers argues that an individual's behavior in the organization is the result of many interactions within the organization, and organizational culture is one of the main drivers. Organizational culture finds its roots in the sociological, anthropological, and psychological literature. It can be defined broadly as a set of basic assumptions and behavior learned within the organization that has shown to be effective in dealing with problems of external adaptation and internal integration (Abubakar & Dogoji, 2015; Schein, 1984). According to Schein, there are three paradigms of an organizational culture, which include basic assumptions, values, and artifacts. The basic assumptions are underlying and invisible, while the values and artifacts manifest in day-to-day interactions. In this view, the underlying basic assumptions are intangible, while values such as organizational citizenship behavior, and artifacts such as teamwork, creativity, collaboration, and goal setting are visible and measurable. Organizational culture, however, is a broad term and concept. In the definition, basic assumptions, values, and artifacts are mentioned, but the question remains *about what?* It is here that the modifier *safety* can be inserted to come to a concept of shared basic assumptions, values, and artifacts regarding safety.

Organizational culture has been expressed as the colloquialism "*The way we do things 'round here*." In the case of safety culture, it would be "*The way we do things 'round here about safety*." Borrowing from Schein's (1984) definition of organizational culture and modifying it with safety, safety culture then becomes basic assumptions, values, and artifacts regarding safety in the organization. Antonsen (2009) views culture as the personal frames of reference through which information, rituals, and values are interpreted, and the conventions for behavior, interaction, and communication are generated. It is this latter view that brings an experiential component, which is essential when it comes to attempting to measure safety culture. Through the popular survey approach, questions are directed towards an individual's perceptions of the artifacts and values regarding safety in the organization.

Measuring safety culture can be a challenging prospect because safety culture itself tends to be an abstract concept (IATA, 2019). As such a concept, Sellers (2014) and IATA (2019) recommend that measurement of safety culture utilizes both quantitative and qualitative data. In other words, interviews with organizational members can provide more insight into the perceptions and narratives regarding safety culture. However, the use of a survey is foundational for performing quantitative analysis that allows for statistical testing to discern differences or similarities. Regardless, organizations should engage in safety culture measurement and analysis periodically. One important reason to measure safety culture is to uncover underlying issues such as employee reluctance to report problems (Skybrary, 2019). As is known, safety reporting is critical for the success of safety management and uncovering any inhibitors to it would allow the organization an excellent opportunity for improvement. Other reasons to measure safety culture include (but not limited to): (a) showing improvement, (b) understanding the organization better, (c) identifying areas of improvement, (d) improve processes and management broadly, and (e) help redefine operational philosophies. Measuring safety culture also enables management to track progress and evaluate the effectiveness of safety management efforts, set goals, and develop leading indicators.

Developing an instrument (survey) for measuring safety culture can be challenging, but this does not preclude an organization from being able to create one in-house. It could be as simple as determining the significant facets of what safety culture means to the organization and asking questions to obtain perceptions of each facet (Khaneman, 2011). In reality, creating a framework to describe safety culture and asking questions around each facet - or dimension - of the framework is how most safety culture instruments are constructed. Reason (1997) developed a framework of four dimensions to describe safety culture, which includes reporting, just, flexible and learning cultures. Baldwin Safety and Compliance's legacy survey (still in production for benchmarking) utilizes four dimensions as well: (a) top management, (b) just culture, (c) continuous improvement, and (d) full participation. In 2019, Baldwin updated its safety culture survey and now measures safety culture across five dimensions: (a) safety leadership, (b) reporting, (c) justness, (d) trust and accountability, and (e) learning ability. This framework allows for a more refined focus on specific elements and amplification of potential issues.

Survey Demographics

Baldwin Safety & Compliance survey team administered the safety culture survey to 31 different organizations in 2021. Of these 31, one organization represented a different transportation mode outside of aviation operations and therefore was excluded. The remaining 30 valid organizations represent aviation companies from:

- private business aviation
- unscheduled commercial business aviation
- aeromedical aviation
- public use segments

837 individuals took Baldwin’s Safety Culture Survey from the 30 different organizations. However, any survey that contained missing responses across any of the five dimensions (see Safety Culture Model Description) was removed from the analysis to prevent skewing the results. After removing those surveys, the total number of

deidentified responses for analysis was 778. The table below provides the organization demographics and number of valid versus total responses.

		Cases					
		Valid		Missing		Total	
	Org Type	N	Percent	N	Percent	N	Percent
Total_ave	Private	298	95.5%	14	4.5%	312	100.0%
	Commercial	301	90.7%	31	9.3%	332	100.0%
	Air Medical	136	93.8%	9	6.2%	145	100.0%
	Public Use	43	89.6%	5	10.4%	48	100.0%

Table 1: Demographics

Each of the 30 organizations was coded for size based on the number of people registered in its respective Baldwin Safety & Compliance web portal at the time of survey administration. The size categories used are derived from the Bureau of Labor Statistics (BLS) firm size categories. The BLS has eight size categories based on the number of individuals employed by the company. The Baldwin Safety Culture Survey utilized six of these categories, which are: 1-4 employees (Class 1), 5-9 employees (Class 2), 10 -19 employees (Class 3), 20-49 employees (Class 4), 50-99 employees (Class 5), and 100-499 employees (Class 6). The table below provides the breakdown of valid responses per size category.

		Cases					
		Valid		Missing		Total	
	Org Size	N	Percent	N	Percent	N	Percent
Total_ave	Class 1 (1-4)	3	100.0%	0	0.0%	3	100.0%
	Class 2 (5-9)	36	97.3%	1	2.7%	37	100.0%
	Class 3 (10-19)	144	93.5%	10	6.5%	154	100.0%
	Class 4 (20-49)	137	94.5%	8	5.5%	145	100.0%
	Class 5 (50-99)	219	90.9%	22	9.1%	241	100.0%
	Class 6 (100-249)	239	93.0%	18	7.0%	257	100.0%

Table 2: Organization Size

The 30 organizations were also coded for the aircraft category most predominately operated. The two categories are fixed-wing (F/W) and rotary-wing (R/W). The table below provides a breakdown of valid responses by aircraft category.

		Cases					
		Valid		Missing		Total	
	FW/RW	N	Percent	N	Percent	N	Percent
Total_ave	FW	573	92.7%	45	7.3%	618	100.0%
	RW	205	93.6%	14	6.4%	219	100.0%

Table 3: Aircraft by Category

Finally, the organizations were also coded as to whether they participated in Baldwin's Safety Performance Monitoring Program (SPMP). Baldwin's SPMP program is built around the notion that a highly experienced safety mentor, part of the Baldwin team, monitors and encourages participating organizations to engage with and complete safety management activities in the Baldwin portal. A rigorous audit is also part of this program which the organization must successfully complete to remain in the program. More information about Baldwin's SPMP can be found here. The table below provides a breakdown of valid responses by participating organizations.

		Cases					
		Valid		Missing		Total	
	SPMP Program	N	Percent	N	Percent	N	Percent
Total_ave	No	667	93.3%	48	6.7%	715	100.0%
	Yes	111	91.0%	11	9.0%	122	100.0%

Table 4: Valid Responses by Organization

The inclusion of SPMP program members for this report is to allow for the ability to analyze the safety culture scores of these organizations against non-participating organizations.

Safety Culture Survey Description

The Baldwin Safety Culture Survey is designed to assess perceptions of the organization's safety culture through interaction between the leadership, safety policy, procedures, and processes in the safety climate (Zohar, 2010). Since culture is described as shared values, norms, rituals, and language (Schein, 1984), everyday

shared perceptions evaluated in the climate will provide a robust indication of the underlying, persistent safety culture of the organization.

Participants in the 2021 survey were asked to rate items across six dimensions of safety culture using a Likert-type scale (e.g., 1 = completely disagree to 5 = completely agree). The survey consists of 36 questions across six factors of safety culture. These factors were adopted from Reason's (1997) four dimensions of safety culture, which include: reporting culture, just culture, flexible culture, and learning culture. Additionally, based on research from Díaz-Cabrera, Hernández-Fernaud, and Isla-Díaz (2007), the fifth factor of safety leadership was added. The sixth factor is safety citizenship behavior. Many of the questions in this survey were adapted from the Díaz-Cabrera et al. (2007) instrument for the aviation and transportation contexts.

The combined score of all the questions represented the aggregate of the scores across the six factors of safety culture used in the Baldwin Safety & Compliance Safety Culture Survey. Again, these factors are safety leadership, reporting culture, the justness of the culture, trust and accountability, learning ability, and safety citizenship behaviors.

Safety Leadership

The importance of leadership in developing and sustaining a robust safety culture cannot be overstated. Leadership – specifically executive leadership - in an organization has a profound effect on the culture of the organization (Zohar, 2003). A leader's approach to leading has been shown to induce a climate that mirrors that leadership style (Liden, Wayne, Liao, & Meuser, 2014). In other words, if a leader's approach to leadership is laissez-faire or absent, then this will create a perception of distrust and no organizational support (Clarke, 2013). However, if the leader's style of leadership is transformational and engaged, then the culture in the organization will thrive (Mullen, Kelloway, & Teed, 2011).

Concerning safety, if a leader prioritizes the policies, procedures, and infrastructure to maintain safe operations while promoting the importance of safety, then this will help

induce a robust safety culture. On the other hand, if the leader is considered absent or inaccessible about safety, then the safety culture will flounder (Mullen, Kelloway, & Teed, 2011). Therefore, this factor of safety culture demonstrates how supportive and active leadership is in fostering positive safety behavior, including engagement in the SMS.

Reporting Culture

Safety reporting - or rather engaging employees in the act of reporting - is a difficult task in most organizations. First, an employee can have anxiety about reporting an error, slip, or lapse of judgment. This is especially true if she does not feel confident that no retribution will be taken against her. Second, if an employee doesn't perceive that any benefit will manifest from his effort to take the extra time to report, the engagement rate in reporting will be meager (Reason, 1997). Therefore, organizations need to foster a culture of shared values around and beliefs regarding the importance of employee reporting.

Visible indicators of a healthy reporting culture include ease of access to reporting, anonymous or confidential reporting ability, feedback provided from the reports, and indemnification against disciplinary proceedings (Reason, 1997). Concerning shared values and beliefs regarding reporting, these tenets of a reporting program need to be pervasive throughout the organization. This factor of safety culture demonstrates how strong these shared values and beliefs are in the organization.

Justness of Culture

In the last decade, the notion of a culture that is just has become an overarching indicator for a healthy safety culture. However, the shared values and beliefs around being just are only a factor - or trait - of an overall healthy safety culture. Regardless, a culture that is just as important in facilitating engagement in safety reporting. Justness is especially important if an organization wishes to learn from the slips, lapses, and mistakes of its employees on the frontline.

The justness of the culture describes the organization's belief and practice of discerning slips, lapses, and mistakes from reckless, willful acts. In reality, it would not make sense for an organization to punish all errors; however, it would also not make sense to dismiss all erroneous acts either. Therefore, an organization with a robust just culture will have developed an agreed-upon set of formal guidelines to determine the motive in erroneous acts (Reason, 1997). These guidelines will provide clarity between acts that could be considered reckless, negligent, or even criminal from those that are not.
This one factor of overall safety culture, therefore, describes the organization's shared values and beliefs regarding restoring individuals who have committed errors without bad intent. However, the company should also take disciplinary action against those whose acts are determined to be against the greater good of the organization (Reason, 1997).

Trust and Accountability

The safety culture factor of trust and accountability is, in essence, the trust leadership has in its employees - including managers - to make appropriate decisions regarding safety. Additionally, this factor describes the transparency leadership maintains regarding the safety status of the organization.

Organizational cultures that are high in trust and accountability have leaders that empower frontline employees to make informed decisions regarding risk. Also, cultures high in trust and accountability foster open scrutiny and corrective input to improve safety in operations.

This factor is closely related to Reason's (1997) factor of a flexible culture. Reason argues that flexible cultures are built on trust so that during times of high operations tempo, the authority for work and risk decision-making is decentralized and placed on the frontline workers and managers.

Learning Ability

Knowledge dissemination is vital for the survival of any organization in the world today. Whether it is tacit knowledge shared through employee interaction or formal knowledge through training, organizations that are unable to learn will have a very short life (Reason, 1997). The learning ability of an organization is an important factor in safety culture for this reason. We have seen examples of organizations that had poor knowledge capital, such as Kodak and Blockbuster movies, fail. Imagine how much shorter an aviation organization can sustain without dissemination of safety knowledge.

This survey examines this factor through assessing training, sharing of lessons learned, and open discussion of safety-related events. According to Reason (1997), the learning ability factor of an organization is the easiest to establish but the hardest to make effective. If any of the elements within this factor are deficient, then the organization likely has a knowledge dissemination problem.

Safety Citizenship

Safety citizenship behavior, as a recognized component of organizational citizenship behavior, is defined as voluntary personal behavior taken on by employees to ensure the safety performance of coworkers and to help achieve the safety performance objectives of the organization (Curcuruto & Griffin, 2018; Hofmann & Morgeson, 1999). This stems from the definition of organizational citizenship behavior, which is voluntary employee behavior that is not formally or directly compensated for by the organization for the betterment of that organization (Bhatla, 2016).

Safety citizenship behavior, like organization citizenship behavior, are behaviors controlled by factors external to economic exchange that control the relationship between individuals and their organization (Curcuruto & Griffin, 2018).

Results and Analysis

The 778 data points from the 30 valid organizations were used for the analysis of safety culture across the group as well as for comparisons by differing characteristics within the group. The sections below will describe the analyses and results of the analyses.

All Organizations

The 30 organizations that engaged with the Baldwin Safety & Compliance's 2021 Safety Culture Survey represent different segments in the business aviation industry (see Table 1: Demographics). The aggregate safety culture score across these segments is described in the table below.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Total_ave	778	1.75	5.00	4.2669	.62469
Valid N (listwise)	778				

Table 5: Average Safety Culture Score

The table shows that the average safety culture score across the 30 organizations is 4.27, which is above the benchmark score of 4.0, indicating a healthy safety culture. However, it is interesting to note the range in scores as depicted by the "minimum" and "maximum" statistics. The range in average scores by the respondent is 3.25 (the difference between "maximum" and "minimum"), which indicates a wide variability in responses. This variability – or range - in scores is also depicted in the boxplot below.

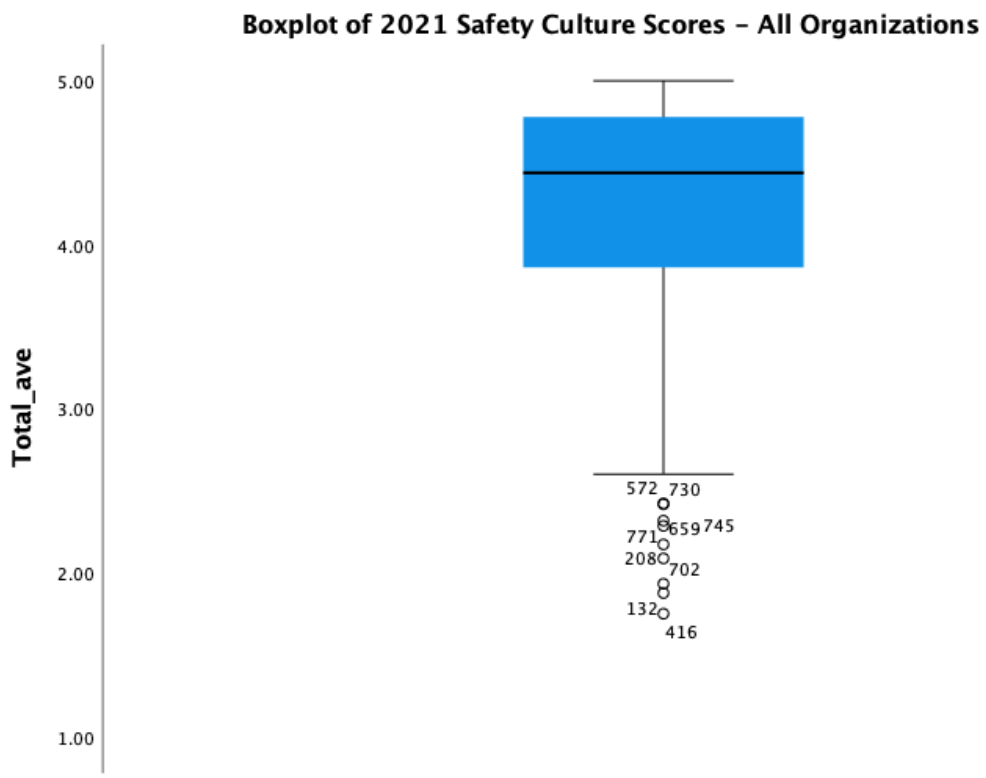

Table 6: Safety Culture Scores - All Organizations

In the boxplot above, the blue box represents the inter-quartile range (IRQ) which contains the middle 50% of the scores. These are the scores that lie between the 75th percentile and the 25th percentile. The bottom tail, or "T," depicts the bottom 25%, while the top tail, or "T," represents the top 25% of scores. The thick black line represents the median.

As shown in the boxplot, the median score across the 30 organizations is ~4.4, which is higher than the mean score of 4.27. The difference between the median and mean scores is because the median is less susceptible to outliers, which are depicted in the boxplot as the points plotted below the bottom 25% tail. The boxplot also shows that 75% of the scores range between 3.9-5.0 (IRQ + top 25%) while the bottom 25% (not including outliers) range between 2.6-3.9. This demonstrates the wide variability in the data and shows that the dataset is negatively skewed. It is normal to see the variance in the perception of safety culture as reported by the respondents; however, when aggregating across the group, the variance appears to be amplified, which may be due

to the nature of the different segments themselves. Therefore, the next section analyzes the average safety culture scores by the different business aviation segments represented in the data.

Analysis by Business Aviation Segments

To determine if safety culture may be impacted by organization type and operating context, the average safety culture scores from each business aviation segment represented in the safety culture survey data were determined (see table below).

Average Safety Culture Scores by Business Aviation Segment

		Total_ave				
		Mean	Maximum	Minimum	Range	Standard Deviation
Org Type	Private	4.57	5.00	1.75	3.25	.44
	Commercial	4.15	5.00	2.09	2.91	.60
	Air Medical	3.81	4.98	1.93	3.05	.70
	Public Use	4.40	5.00	3.34	1.66	.50

Table 7: Average Scores by Segment (data)

In this table, the average safety culture score is shown for each business aviation segment represented in the survey data. Similar to the 2019 Safety Culture Review results, the Private category scored the highest, with Air Medical scoring the lowest. What is also interesting are the standard deviation statistics also shown in the table to the far right. The standard deviation is basically described as the average amount of variation from the mean in each sample. A small standard deviation is indicative of a more tightly clustered sample around the mean and implies less variance in responses. In other words, in the context of the safety culture scores, a small standard deviation implies a more unified perception of safety culture in an organization. The Private category boasts the smallest standard deviation statistic while Air Medical the largest.

The chart below depicts boxplots for each of the business aviation segments and gives a better visual representation of the data, including the variance and any skewness.

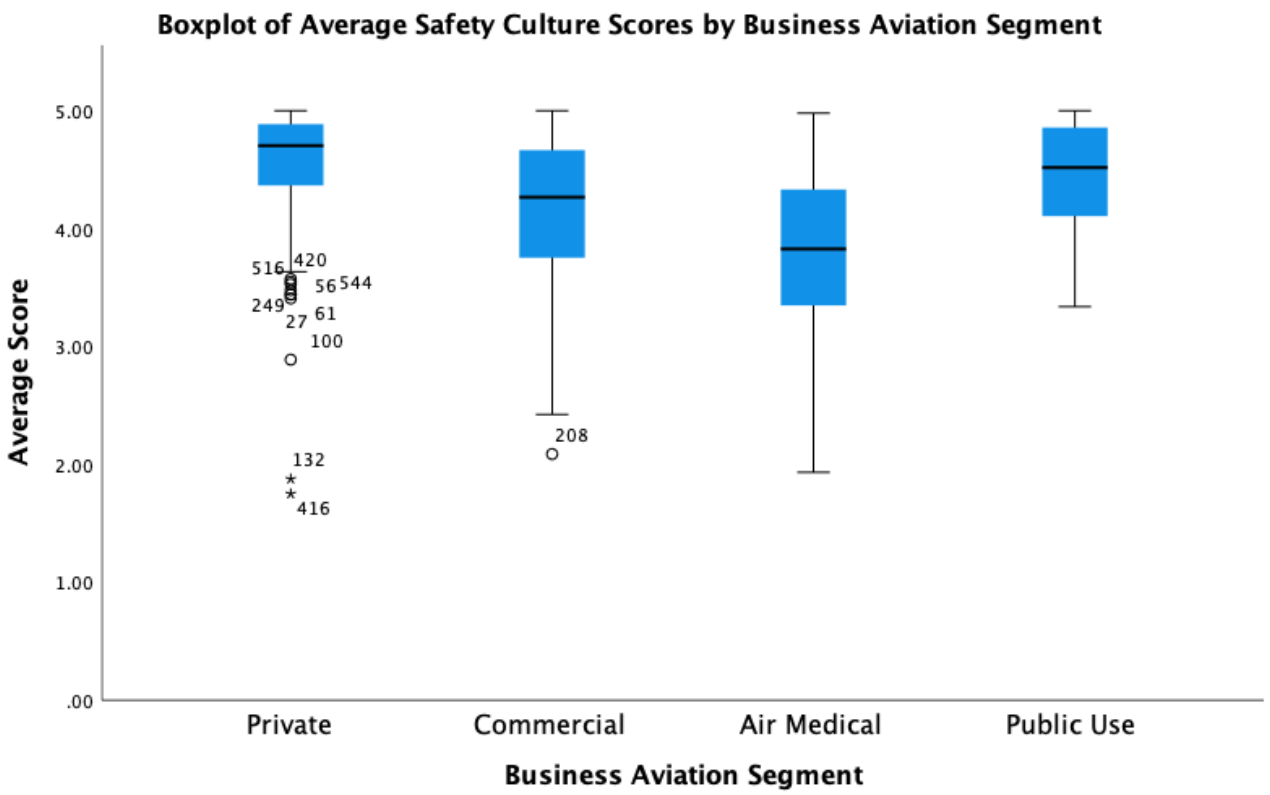

Table 8: Average Scores by Segment (boxplot)

In the boxplots, the median of each segment follows the same pattern as displayed in the table where the Private category has the highest average, followed by Public Use, then Commercial, with Air Medical yielding the lowest. One interesting aspect shown in the graph is the differences in the range across quartiles of the data. The Air Medical category appears to have the widest range spanning from 2.0-5.0, while Private has the smallest (3.7-5.). This is also represented by the standard deviation statistics from the table above. The ranges in each segment infer the shared perception of safety culture at a given organization in the representative segments. In the Private category, it appears that perceptions of safety culture are more strongly shared than in Air Medical.

Another interesting aspect shown in the graph is the number of outliers in the Private category. To evaluate this, it is important to understand how outliers are determined in boxplots. The formula for determining outliers (in either direction) is Q1 – 1.5*IQR and Q3 + 1.5*IQR. Just looking at the lower outliers, the IQR for the Private category is 4.9-4.4 (Q3 – Q1), which equals .5. Using the formula 1.5 * IQR, the resulting value is .75.

Subtracting this from 4.4 yields a lower bound of 3.65. Therefore, any average score below 3.65 is considered an outlier. This is purely a function of the range of the IQR. As stated, since the range is so low in the Private category, a seemingly small variation in the average response score will be seen as an outlier. In other words, the perception of the organization's safety culture is so strong and unified in the Private category that more moderate perceptions are considered outliers. This speaks a lot to the organizational safety cultures in this segment.

When looking at the different average scores by segment, it is readily apparent that some scores are higher and some lower. We cannot state that these differences – which infer stronger or weaker safety cultures – are indeed significant or due to chance alone without statistically testing the differences. A statistical test called an Analysis of Variance (ANOVA) was performed to test if the difference in average safety culture scores by segment is statistically significant. The test showed that there are significant differences in the different average safety culture scores between the different segments, $F(3,774)=65.3$, $p<.001$. This test just indicates that there are significant differences between the average scores, so another test was performed to determine where those differences are. The results of the test are indicated in the table below.

Multiple Comparisons

Dependent Variable: Total_ave
Tukey HSD

(I) Org Type	(J) Org Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Private	Commercial	.41772*	.04569	<.001	.3001	.5354
	Air Medical	.76585*	.05786	<.001	.6169	.9148
	Public Use	.16816	.09121	.254	-.0667	.4030
Commercial	Private	-.41772*	.04569	<.001	-.5354	-.3001
	Air Medical	.34814*	.05777	<.001	.1994	.4969
	Public Use	-.24956*	.09115	.032	-.4842	-.0149
Air Medical	Private	-.76585*	.05786	<.001	-.9148	-.6169
	Commercial	-.34814*	.05777	<.001	-.4969	-.1994
	Public Use	-.59770*	.09782	<.001	-.8496	-.3458
Public Use	Private	-.16816	.09121	.254	-.4030	.0667
	Commercial	.24956*	.09115	.032	.0149	.4842
	Air Medical	.59770*	.09782	<.001	.3458	.8496

*. The mean difference is significant at the 0.05 level.

What the table shows is that the differences in means between the Private and Commercial categories (.42) and between the Private and Air Medical categories (.77)

are significant at $p<.001$. In other words, there is a less than 0.1% chance the average score differences between Private and Commercial, and Private and Air Medical are due to chance alone or sampling error. This indicates that the populations are different, and the differences in the scores are indicative of some other influencing factors. As shown in the table, the only non-statistically significant differences in averages occur between the Private and Public Use segments. In other words, there may be a common perception of an organization's safety culture between these two segments.

Analysis by Aircraft Category

The next analysis was to evaluate the data by operating category (FW vs. RW) across the different segments and organization sizes. The table below shows the average safety culture scores, score ranges, and standard deviation by aircraft category.

Average Safety Culture Scores by Aircraft Category

		Total_ave			
		Mean	Maximum	Minimum	Standard Deviation
FW/RW	FW	4.42	5.00	1.75	.52
	RW	3.83	5.00	1.93	.68

Table 10: SCS by Flight Category (data)

As seen in the table, the average score for the FW category is higher than for the RW category. The standard deviation statistics reveal that there is a greater central tendency of the data in the FW group than in the RW. The boxplots below help to visualize the data better.

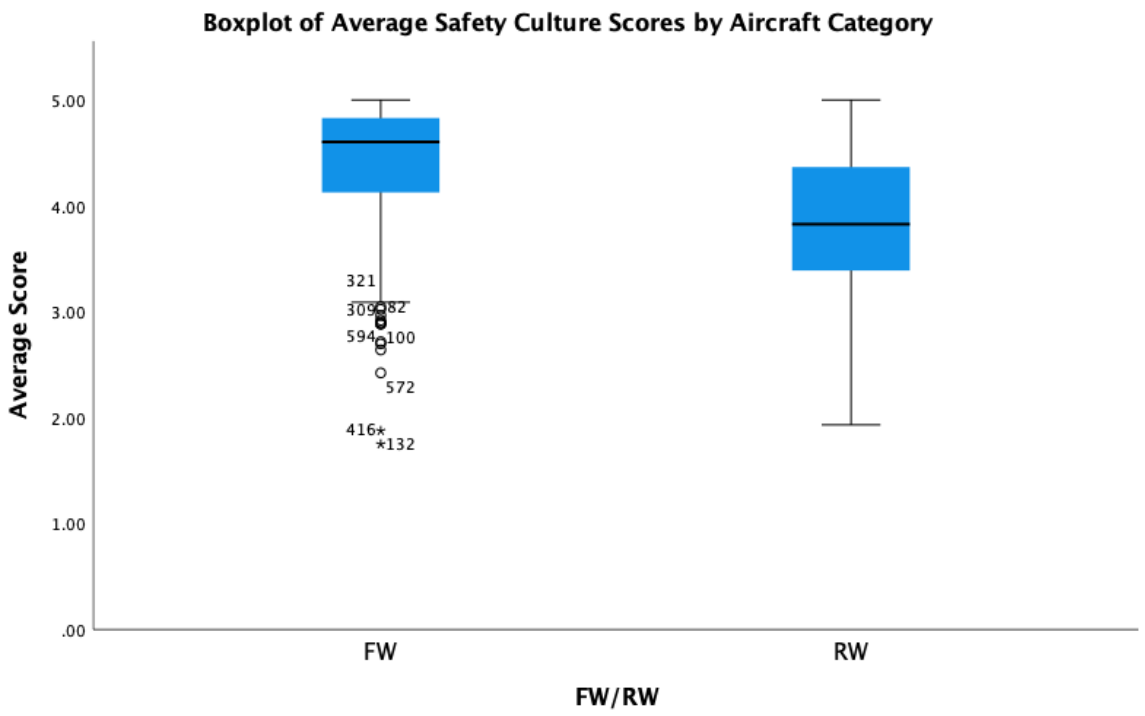

Table 11: SCS by Flight Category (boxplot)

In this chart, we see some similarities to the boxplot by organization type. The number of outliers for the FW category again is due to the tight tolerance as a result of the small interquartile range (IRQ). Again, to determine if the differences in averages between the two categories are significant, a statistical test was performed. This time, a *t*-Test was used for the test, and the results show that the means are significantly different, $t(776)=12.72$, $p<.001$. In other words, there is less than a .1% chance that the difference between the FW and RW safety culture score averages (.59) is due to chance or sampling error alone. However, this result is not surprising considering that the majority of RW responses are from the Air Medical organization type (see below).

Number of Responses by Organization Type and Aircraft Category

				Count
Org Type	Private	FW/RW	FW	312
			RW	0
	Commercial	FW/RW	FW	274
			RW	58
	Air Medical	FW/RW	FW	0
			RW	145
	Public Use	FW/RW	FW	32
			RW	16

Table 12: Response by Organization & Aircraft Type (data)

This fact more than likely skews the data when looking at the differences between FW and RW.

Analysis by Organization Size

The next analysis performed looks at the size of the organization rather than the operating context or equipment used. The table below shows the average safety culture scores, score range, and standard deviation by the different size categories.

Average Safety Culture Scores by Organization Size

		Total_ave			
		Mean	Maximum	Minimum	Standard Deviation
Org Size	Class 1 (1-4)	4.95	5.00	4.88	.06
	Class 2 (5-9)	4.59	5.00	1.88	.61
	Class 3 (10-19)	4.44	5.00	1.75	.52
	Class 4 (20-49)	4.51	5.00	2.89	.42
	Class 5 (50-99)	4.29	5.00	2.09	.61
	Class 6 (100-249)	3.94	5.00	1.93	.66

Table 13: Safety Culture Scores by Organization Size (data)

In the table, the different average safety culture scores across the different Bureau of Labor Statistics organization size categories are depicted. All categories except Class 6 reported averages above 4.0. The standard deviations (which speak to the variance in

the scores) are all similar except for the Class 1 organizations. The very low standard deviation for this class is likely due to the fact there were only three responses in this category and were from a private organization. A visualization is presented below using a series of boxplots.

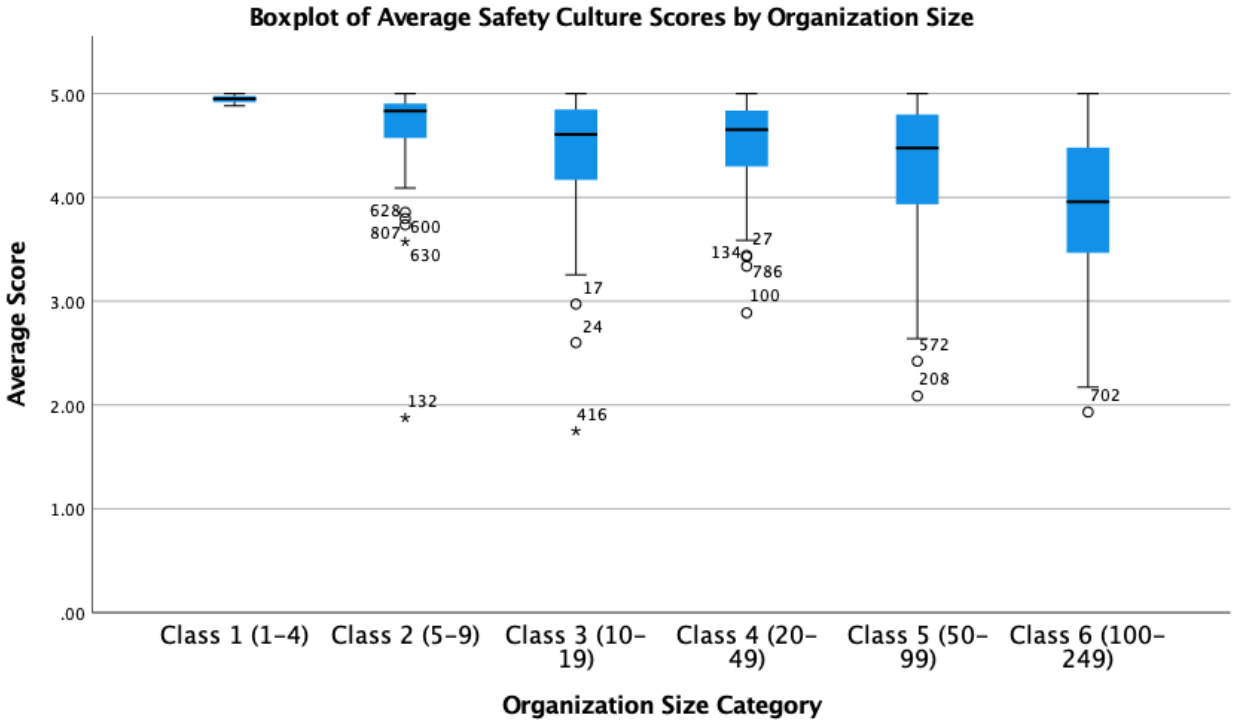

Table 14: Safety Culture Scores by Organization Size (boxplot)

In the above chart, average safety culture scores for the six size categories are plotted. Again, the numbered points outside of the quartiles represent outliers determined using the aforementioned formula. The table below shows the breakdown of respondents by the organization size category and type.

Number of Respondents by Organization Size Category and Type

				Count
Org Size	Class 1 (1-4)	Org Type	Private	3
			Commercial	0
			Air Medical	0
			Public Use	0
	Class 2 (5-9)	Org Type	Private	34
			Commercial	0
			Air Medical	0
			Public Use	3
	Class 3 (10-19)	Org Type	Private	114
			Commercial	11
			Air Medical	13
			Public Use	16
	Class 4 (20-49)	Org Type	Private	55
			Commercial	61
			Air Medical	0
			Public Use	29
	Class 5 (50-99)	Org Type	Private	106
			Commercial	135
			Air Medical	0
			Public Use	0
	Class 6 (100-249)	Org Type	Private	0
			Commercial	125
			Air Medical	132
			Public Use	0

Table 15: Safety Culture Scores by Size, Type, Aircraft (data)

As shown in the above table, the bulk of the respondents from the Private category are in organization size classes 2-5, whereas class 2 is predominately comprised of private operators. This explains some of the traits we see in the above boxplot by size class, including narrow IQRs and, as a result, multiple outliers.

The initial table providing the average scores by size class as well as the boxplot chart do reveal that there are different safety culture averages by class. Again, to determine if these differences are significant, an ANOVA was run. The initial results do show that there are statistically significant differences between the safety culture means amongst the different size classes at $F(5,772)=25.87$, $p<.001$. A *post hoc* test was performed to see which safety culture averages were different, and the results are in the table below.

Multiple Comparisons

Dependent Variable: Total_ave
Tukey HSD

(I) Org Size	(J) Org Size	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Class 1 (1-4)	Class 2 (5-9)	.35747	.34854	.909	-.6383	1.3532
	Class 3 (10-19)	.50196	.33833	.675	-.4646	1.4685
	Class 4 (20-49)	.43488	.33851	.793	-.5322	1.4020
	Class 5 (50-99)	.65180	.33715	.383	-.3114	1.6150
	Class 6 (100-249)	1.00608*	.33696	.035	.0434	1.9687
Class 2 (5-9)	Class 1 (1-4)	-.35747	.34854	.909	-1.3532	.6383
	Class 3 (10-19)	.14448	.10808	.764	-.1643	.4532
	Class 4 (20-49)	.07741	.10863	.980	-.2329	.3877
	Class 5 (50-99)	.29433	.10431	.055	-.0037	.5923
	Class 6 (100-249)	.64860*	.10369	<.001	.3524	.9448
Class 3 (10-19)	Class 1 (1-4)	-.50196	.33833	.675	-1.4685	.4646
	Class 2 (5-9)	-.14448	.10808	.764	-.4532	.1643
	Class 4 (20-49)	-.06708	.06922	.928	-.2648	.1307
	Class 5 (50-99)	.14985	.06223	.155	-.0279	.3276
	Class 6 (100-249)	.50412*	.06119	<.001	.3293	.6789
Class 4 (20-49)	Class 1 (1-4)	-.43488	.33851	.793	-1.4020	.5322
	Class 2 (5-9)	-.07741	.10863	.980	-.3877	.2329
	Class 3 (10-19)	.06708	.06922	.928	-.1307	.2648
	Class 5 (50-99)	.21693*	.06318	.008	.0364	.3974
	Class 6 (100-249)	.57120*	.06215	<.001	.3936	.7488
Class 5 (50-99)	Class 1 (1-4)	-.65180	.33715	.383	-1.6150	.3114
	Class 2 (5-9)	-.29433	.10431	.055	-.5923	.0037
	Class 3 (10-19)	-.14985	.06223	.155	-.3276	.0279
	Class 4 (20-49)	-.21693*	.06318	.008	-.3974	-.0364
	Class 6 (100-249)	.35427*	.05425	<.001	.1993	.5093
Class 6 (100-249)	Class 1 (1-4)	-1.00608*	.33696	.035	-1.9687	-.0434
	Class 2 (5-9)	-.64860*	.10369	<.001	-.9448	-.3524
	Class 3 (10-19)	-.50412*	.06119	<.001	-.6789	-.3293
	Class 4 (20-49)	-.57120*	.06215	<.001	-.7488	-.3936
	Class 5 (50-99)	-.35427*	.05425	<.001	-.5093	-.1993

*. The mean difference is significant at the 0.05 level.

Table 16: Multiple Comparisons

In the table above, the far-left column gives the first size class and the second and third columns give the comparisons to the other size classes. The last row is the comparison of class 6 organizations to other classes. In the third column, the value is the difference between the Class 6 safety culture average and the average of the other respective classes. The small asterisk next to each value indicates that the difference is statistically significant. According to the above table, the Class 6 organization average safety culture score is statistically significantly less than all the other classes. In other words, the organizations with 100-249 employees that participated in the Baldwin safety culture survey had a statistically significant lower average safety culture score than the smaller organizations. Again, since no private organizations were represented in the Class 6

category, this could have influenced the data. It is interesting to note, however, that the difference between Class 5 and Class 4 organizations (-.22) is statistically significant. Both classes have representation by private organizations: Class 4 had 55 respondents from Class 4 organizations, and Class 5 had 106. This may be a small signal that organizational size does impact the perception of safety culture.

Analysis by SPMP Participation

Baldwin's Safety Performance Monitoring Program (SPMP) is a service that pairs an experienced safety professional from the Baldwin team with an organization to help guide and encourage that organization in the SMS journey. The Baldwin safety professional interacts with the organization through their portal, provides feedback on safety management activities performed in the portal, and motivates the organization when activity begins to wane. Additionally, to be a member of the SPMP, each organization must pass a qualifying safety audit initially and recurrently on a timeframe determined by the previous audit's performance. In this analysis, the safety culture score averages for participating and non-participating organizations are compared. The bar chart below provides an overview of the average safety culture scores of participating and non-participating organizations.

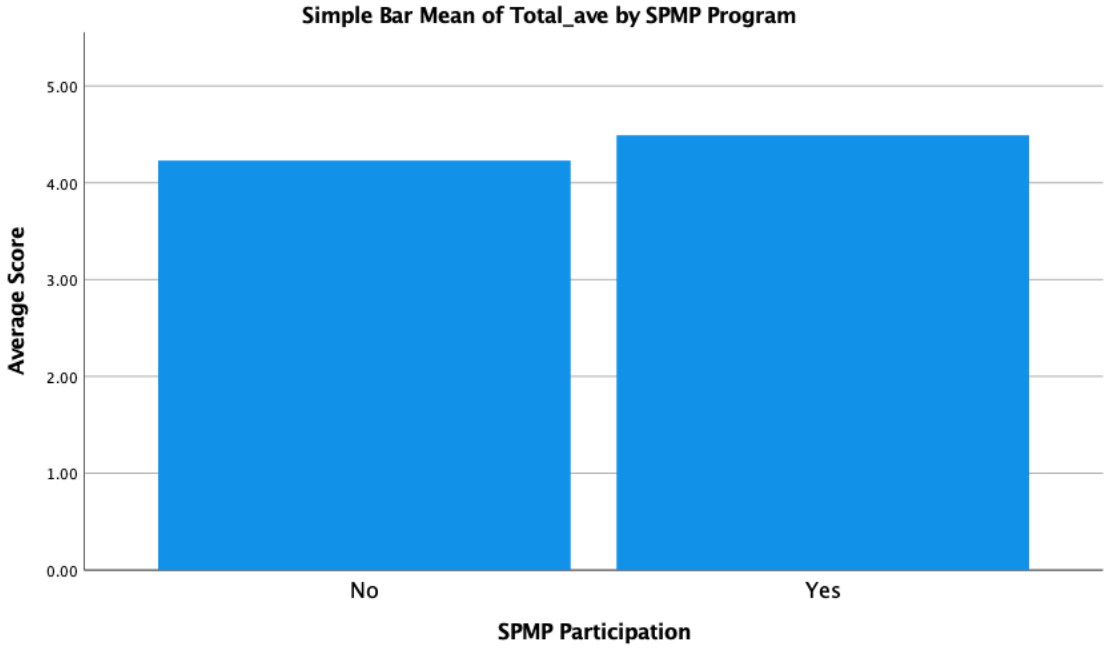

Table 17: Safety Culture Scores by SPMP participation

This chart shows that the average safety culture scores of organizations that participate in the SPMP (4.49) are slightly higher than non-participating organizations (4.23). In order to determine if this difference in average scores is significant, a *t*-Test was performed on the data. The results of the test indicate that the difference in means between participants and non-participants (.262) is statistically significant at (776)=-4.147, $p<.001$. Again, this means that the difference in the average safety culture scores between the two groups has a less than .1% chance of being due to randomness or sampling error alone. Cast in a different way; there is a 99.9% chance that other factors are influencing this difference. Since this is not an analysis of causation, reasons for this difference are suggested for future studies.

Summary

The data from the 837 respondents across 30 different organizations were analyzed to determine if there were differences in safety culture scores based on business aviation segment, operating aircraft category, organization size, and participation in Baldwin's Safety Performance Monitoring Program. A summary of the findings in each of these areas is below.

Business Aviation Segment

The four business aviation segments that were represented in the data and that we analyzed were Private, Commercial, Air Medical, and Public Use. As in the 2019 analysis, the Private category yielded the highest average safety culture score at 4.57 out of a possible 5.0. The Air Medical category had the lowest average safety culture at 3.81, followed by Commercial at 4.15 and Public Use taking the second-place spot at 4.40.

The results of statistical testing reveal that the differences in safety culture score averages are significant and meaningful. While the categories were adjusted slightly from the 2019 report, the two categories that did remain the same were the Private and Commercial categories. In the available data pool, this translates into 14 CFR 91 and

135 operations (excluding Air Medical). In 2019, the Private average safety culture score was higher than the Commercial category as it is again in 2021. Also, this difference was as significant in this analysis as it was in 2019. As also stated in the 2019 analysis, the tests only provide an indication of the significance of the data and not causation. The reasons for the differences in safety culture strength between the different organization types should be a topic of research in the future.

Operating Aircraft Category

The two different aircraft categories operated by the participating organizations are Fixed-wing (F/W) and Rotary-wing (R/W). An analysis was done comparing the safety culture scores across these two categories. The results showed that the F/W category boasted a higher safety culture score (4.42) than the R/W category (3.83). This large difference in average safety culture scores between F/W and R/W categories of .59 is significant and represents a difference in safety culture strength between the two categories.

Further analysis, however, determined that the different safety culture scores between F/W and R/W could possibly be attributed to the signal from the business aviation segment. The data show that 67% of the R/W operations are performed by Air Medical, which also scored lowest amongst the business aviation segments. In other words, it could be the fact that most R/W operations are contained in the business aviation segment that scored the lowest average safety culture score, thus suppressing the overall average score for the R/W group. Again, this is speculation, and more research should be done into this to determine if the reduced scores are due to the business aviation segment or if possibly R/W operations contribute to the lower business aviation segment score.

Organization Size

Organization size was divided into six different classes based on the categorization scheme of the Bureau of Labor Statistics. This analysis was performed to

determine – if any – the relationship between the size of the organization and safety culture strength. The analysis showed that Class 6 (100 – 249 people) organizations had – on average – a lower safety culture score than the other five classes. Furthermore, the differences in averages of the Class 6 organizations and the other classes were significant. The analysis also showed that the Class 5 (50-99 people) average safety culture score was significantly less than the Class 4 (20-49 people) safety culture score average. While safety culture average scores did generally decrease as the class increased, the differences between Classes 1 – 4 were not statistically significant.

Again, the differences in safety culture scores may be a signal from the business aviation segments represented in each class. The majority of respondents in Classes 1-4 are from the Private Business Aviation segment, while Classes 5-6 are predominately Commercial and Air Medical. The lower safety culture score averages yielded as class size increased may be a function of the organization types represented. However, it is interesting to note that 34% of the Private category respondents are represented in Class 5, which did score lower than Classes 1-4. This does open the possibility that organization size may impact safety culture strength and should be considered for future research.

SPMP Participation

An analysis of average safety culture scores by participation in Baldwin's SPMP was performed to see if any relationships do exist between safety culture strength and participation in this program. The results of the analysis do show that the difference between the average safety culture scores of participants (4.49) and non-participants (4.43) is significant. Again, the SPMP is described as a mentoring relationship between Baldwin and the organization to help increase the efficacy of and participation in the SMS. What cannot be stated because of this analysis is whether the safety culture scores of participants are higher due to the SPMP or if the SPMP is biased towards organizations with a more robust safety culture. Further analysis and research – which

should include a longitudinal study – should be performed to determine if the SPMP does have a positive impact on safety culture.

COVID-19 Implications

The Baldwin Safety Culture Survey does not have questions that related directly to the impact of COVID-19 on safety culture; however, the open-ended responses did provide information into perceptions of safety during this period. At the end of the survey, the participants are asked a question related to what they perceive safety challenges in the organization may be. This is an open-ended question, meaning the participant can respond as s/he wishes using a text field. In the early period of the pandemic, when flight operations began to wane across many organizations, anecdotally, about 10-15% of the responses indicated a concern for the loss of proficiency and complacency due to the lack of flying. Towards the tail end of the operational slow-down in mid-2021, comments related to lack of proficiency and complacency increased to 75-80%. This demonstrated that as the period of reduced operations prolonged, concern for skill maintenance and proficiency increased.

The question which cannot be answered is if the uncertainty and angst associated with the pandemic impacted overall safety culture perceptions. It may be the case that perceptions were weakened due to these concerns. It also may be the case that they could have been strengthened as organizations started investing the time now available into safety management efforts. This is an unknown at this point but certainly an interesting aspect during this period. Regardless, this may highlight the fact that large-scale societal events can impact – to some degree- the level of perceived safety in an organization.

Acknowledgments

Baldwin Safety & Compliance would like to express its profound gratitude to all the organizations that participated in the safety culture survey. While this is one of the many offerings Baldwin has that a customer may choose from, Baldwin feels that organizations which participate in this survey are very conscious about safety.

Furthermore, participation allows for summary results such as what is contained in this report to help increase awareness and understanding in the industry. Again, Baldwin thanks its participants and hopes other organizations will decide to participate in the safety culture survey in order to not only become more aware of their own culture but at the same time give even more back to our industry.

References

Clarke, S. (2013). Safety leadership: A meta-analytic review of transformational and transactional leadership styles as antecedents of safety behaviours. *Journal of Occupational and Organizational Psychology*, *86*(1), 22–49. https://doi.org/10.1111/j.2044-8325.2012.02064.x

Díaz-Cabrera, D., Hernández-Fernaud, E., & Isla-Díaz, R. (2007). An evaluation of a new instrument to measure organisational safety culture values and practices. *Accident Analysis and Prevention*, *39*(6), 1202–1211. https://doi.org/10.1016/j.aap.2007.03.005

Liden, R. C., Wayne, S. J., Liao, C., & Meuser, J. D. (2014). Servant leadership and serving culture: Influence on individual and unit performance. *Academy of Management Journal*, *57*(5), 1434–1452. https://doi.org/10.5465/amj.2013.0034

Mullen, J., Kelloway, E. K., & Teed, M. (2011). Inconsistent style of leadership as a predictor of safety behaviour. *Work and Stress*, *25*(1), 41–54. https://doi.org/10.1080/02678373.2011.569200

Reason, J. (1997). *Managing the risks of organizational accidents*. Burlington: Ashgate.

Schein, E. H. (1984). schein Comming to a new awareness. *Sloan Management Review*, *2*(25), 3–16.

Zohar, D. (2003). The effects of leadership dimensions, safety climate, and assigned priorities on minor injuries in work groups. *Journal of Organizational Behavior*, *23*(1), 75–92. https://doi.org/10.1002/job.130

Zohar, D. (2010). Thirty years of safety climate research: Reflections and future directions. *Accident Analysis and Prevention*, *42*(5), 1517–1522. https://doi.org/10.1016/j.aap.2009