

Short-Form Measures of Well-Being Centered Leadership for Matrixed Healthcare Systems

Anthony C Waddimba¹⁻³, Rahul R Gunukula^{2,4}, Megan E Douglas⁵, Tait D Shanafelt⁶, J Michael DiMaio^{2,7,8}, Jamile A Ashmore^{3,9}

¹Department of Surgery, Baylor University Medical Center, Dallas, TX, USA; ²Research Development & Analytics Core, Baylor Scott and White Research Institute, Dallas, TX, USA; ³Department of Medical Education, College of Medicine, Texas A&M University, Dallas, TX, USA; ⁴Academic Research Team, Baylor Scott & White-The Heart Hospital, Plano, TX, USA; ⁵Trauma Research Consortium, Baylor Scott and White Research Institute, Dallas, TX, USA; ⁶Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; ⁷Division of Cardiothoracic Surgery, Baylor Scott & White-The Heart Hospital, Plano, TX, USA; ⁸Department of Biomedical Engineering, College of Medicine, Texas A&M University, College Station, TX, USA; ⁹Office of Professionalism and Well-Being, Baylor Scott & White-The Heart Hospital, Plano, TX, USA

Correspondence: Anthony C Waddimba, Department of Surgery, Baylor University Medical Center, 3500 Gaston Avenue, Dallas, Texas, 75246, USA, Tel +1-214-820-0291, Email Anthony.Waddimba@BSWHealth.org; waddimba@gmail.com

Purpose: This study extracted the empirically strongest short-form abbreviations of the adapted Mayo Leadership Impact Index (MLII) and evaluated their comparative validity and reliability versus the full-length reference scale.

Participants and Methods: Data were sourced from physician-respondents to the 2023 (n=158) and 2024 (n=112) annual “well-being” surveys in a tri-hospital Cardiovascular healthcare system in Texas. Priority ranking of the full-length scale’s items based on factor loadings from single-factor confirmatory factor analysis and information function plots from unidimensional graded response model analysis yielded single-, 3-, and 4-item short-form scales. Convergent validity of short-form versus full-length scales was compared using Spearman correlations with professional fulfillment, self-valuation, autonomy support, peer connectedness, and peer respect measures. Divergent validity was tested via correlations with occupational burnout. Internal consistency was based on Cronbach’s and ordinal reliability coefficients. Inter-rater reliability was based on areas under the ROC curve, accuracy, precision, sensitivity, specificity, and Cohen’s kappa coefficient.

Results: Single-item, 3-item, and 4-item short-form scales, from the four highest ranked items, had similar convergent and divergent validity indexes plus inter-rater reliability to the 9-item adapted MLII. Internal consistency reliabilities for the 9-item, 4-item, and 3-item versions of the adapted MLII were >0.900. Sensitivity/specificity ratios and kappa coefficients for MLII-4, MLII-3, and MLII-1 extracts versus the adapted MLII-9 were all >0.700.

Conclusion: Findings verify the reliability and validity of short-form abbreviations of the adapted MLII as lower-burden options for assessing well-being centered leadership within matrixed healthcare systems.

Plain Language Summary: Question: What are the most valid and reliable short-form versions of the revised Mayo Leadership Impact Index adapted for healthcare systems with matrixed reporting structures?

Findings: Classical test theory and item response theory analyses of cross-sectional survey data from 2023 and 2024 physician-respondents supported the comparative validity and reliability of 4-item, 3-item, and single-item short-form scales. The parent scale and its short-form proxies had similar positive correlations with professional fulfillment, autonomy support, self-valuation, plus peer connectedness/respect, and negative correlations with burnout.

Meaning: Findings affirm the short-form scales’ validity and reliability, supporting use of these abbreviated measures to capture well-being centered leadership within matrixed healthcare systems.

Keywords: physicians, occupational well-being, well-being centered leadership, abbreviated measures, psychometric validity, reliability

Introduction

Burnout afflicts 45.2% of United States physicians, and only 36% report high professional fulfillment.¹ Physicians' ratings of the support received from their leader are one of the strongest predictors of occupational well-being.^{2–13} Leadership behavior is associated with physicians' burnout levels,^{3–6,8–13} job satisfaction,^{3–6,10,11,13} professional fulfillment,^{8,10–12} and intent to leave their healthcare organization.^{8,10,11} Emerging evidence indicates that quality improvements in well-being-centered leadership behaviors practiced in clinical workplaces effectively enhance occupational well-being among physicians,^{14–22} and other healthcare professionals.⁴ Hence, the formulation of the construct of *well-being centered leadership*, integrating elements from transformational, relational, situational, and values-based leadership models aimed at empowering individuals, cultivating relationships, and aligning values.²³ This form of leadership centers well-being and professional fulfillment as core drivers, not peripheral by-products, of organizational effectiveness, and reframes leader effectiveness to include sustenance of human capacity, meaning, plus engagement, not just performance or change attainment.²³ Given the resulting imperative and incentive for healthcare organizations to promote the practice of well-being centered leadership behaviors,²³ valid and reliable measures for quantification, assessment, and tracking of these behaviors must be implemented.^{24,25} Measurements of leader behaviors could enable healthcare entities to set well-being-centered leadership improvement goals, implement training programs, and subsequently monitor the success of such initiatives.^{25,26}

Effective assessment of well-being-centered leadership behaviors is complicated by the fact that healthcare organizations are not uniform in terms of their leadership structures.²⁷ Some health systems have a hierarchical structure that assigns a primary, direct supervisor to each physician, while other networks have a matrixed structure, ie., one wherein a physician can source leadership support from any of multiple individuals.¹¹ Guidance, mentorship, support, or accountability is obtainable from department heads, service-line leaders, administrative leaders, quality control managers, medical education leaders, informal physician leaders/peers, depending on role, task, or context. Resonant leadership relationships can, thus, overlap or even vary across time. Loosely integrated healthcare networks with matrixed structures are increasing as more physician practices consolidate with hospital systems.²⁸ Complexity science considers matrixed structures to be better at resolving non-linear, dynamic, self-organizing, or emergent occurrences.²⁹ Many large healthcare organizations, including those in which a single direct supervisor is assigned for each physician, are nowadays adopting a matrixed leadership structure for aspects of their operational management.

The Mayo Leadership Impact Index (MLII), a well-researched measure of well-being centered leadership behaviors,³ was originally designed and validated in hospital systems whereby individuals had a direct supervisor assigned.³⁰ Its validity in systems where direct-report supervisors are not the norm is unknown as it can prompt physicians to evaluate a leader who does not have the deepest influence on their well-being. With permission from the MLII copyright holder, we recently validated an adapted version of this measure for implementation in settings with matrixed, flexible, or multi-form leadership structures.¹¹ Due to the time pressure under which clinicians work,³¹ plus the need to minimize respondent burden, and optimize response rates, organizations often have to shorten their multi-construct questionnaires.^{32–34} Long-form parent scales can be replaced with short-form abbreviations or single-item measures that assess the construct briefly and conserve questionnaire space. A trade-off is made between *depth* (ie., in-depth information on few constructs) versus *breadth* (ie., abbreviated feedback on a broader range of constructs),³⁵ based on an organization's needs and on the short-form scales' validity and reliability. A prior investigation of 1-, 2-, and 3-item proxies of the traditional 9-item MLII was limited to contexts with "direct-report" supervisors.³⁶ There has been no validation of abbreviations of the MLII as adapted for settings with matrixed leadership structures. We addressed this evidence gap by assessing comparative validity and reliability of abbreviated versions versus the full-length 9-item version of the revised MLII adapted for matrixed structures. This study extends prior research by focusing on matrixed rather than hierarchical structures, by validating multiple abbreviations of the revised, adapted MLII.

Materials and Methods

Study Setting

The study occurred within a multi-specialty integrated cardiovascular healthcare delivery system that comprised three enterprise hospitals in North and Central Texas (United States) during 2023–2024. Collectively, these three hospitals perform the highest number of carotid artery bypass grafts and third highest number of heart valve surgeries annually in the U.S.

Study Design

This psychometric validation study is based on data from a cross-sectional anonymized “well-being quality improvement” survey distributed in two consecutive years. Baylor Scott and White Research Institute’s Institutional Review Board assessed the anonymized survey as satisfying minimal risk standards, waived written informed consent requirements, and approved the study (# 023–171).

Study Participants

The study included credentialed physicians providing patient care within the host institution who responded to the annual “Physician Well-being Survey” in 2023 and/or 2024. Affiliated physicians work under a service volume-based payment model that links financial remuneration to clinical productivity measured in relative value units. Physicians in residency or fellowship training were excluded from the sample in both years.

Data Collection

Survey data were collected using the Research Electronic Data Capture (REDCapTM), a secure, web-based electronic data management platform^{37,38} hosted by our institution. A weekly recurring e-mail containing a hyperlink to the online questionnaire was sent out to all eligible and currently practicing physicians with the organization. Posters with scannable QR codes were also openly displayed in physician workspaces.

Study Measures

Reference Scale

The full-length 9-item Mayo Leadership Impact Index (MLIITM) version adapted for utilization in contexts with matrixed, flexible, and multi-form leadership structures¹¹ was our reference scale. Each of the items rates a leader’s supportive behavior on a 5-point Likert spectrum ranging from 1 (“Strongly Disagree”) to 5 (“Strongly Agree”). The MLIITM is scored by summing up constituent items so that higher total scores (minimum = 9; maximum = 45) indicate more supportive leadership behavior, and vice versa. Scale scores ≥ 36 were deemed to indicate a “high” wellness-centered leadership rating overall. The MLII was used with permission from the copyright holder and all derivatives of the MLII remain proprietary to Mayo Clinic.

Index Scales

We extracted the strongest four-item, three-item, and single-item abbreviations of the adapted MLII to serve as index measures for comparative validation with the reference scale. Score thresholds for “high” well-being centered leadership ratings on the 4-, 3-, and single-item abbreviated scales were set at ≥ 16 , ≥ 12 , and ≥ 4 , respectively.

Perceived Autonomy Support

Perceived autonomy support was assessed using the six-item Physician Perceptions of Autonomy Support (PPAS-6) scale,³⁹ framed with the work organization as the referent. The PPAS-6 is scored by summing items (after reverse coding one negatively worded “interference” item) such that higher scores (minimum=6; maximum=30) indicate greater autonomy support.

Self-Valuation

Self-valuation was assessed using the four-item Stanford Self-Valuation Scale (SVS).⁴⁰ The scale combines two items assessing deferment of self-care to prioritize work demands (eg. “I put off taking care of my own health due to time pressure”), with two items assessing harsh responses to personal imperfections or errors (eg. “When I made a mistake, I felt more self-condemnation than self-encouragement to learn from the experience”). Item response options are on a five-point Likert spectrum ranging from 0 (“Never”) to 4 (“Always”). The SVS is scored by summing individual items, with higher scale scores indicating greater self-valuation and vice versa. SVS scores ≥ 9 indicated “moderate to high” self-valuation, and SVS scores < 9 “low” self-valuation.

Burnout

Burnout was assessed using the ten-item Overall Burnout Subscale (OBS) of the Stanford Professional Fulfillment Index (PFI).⁴¹ The OBS of the PFI rates burnout experienced in the preceding two weeks using four *Work Exhaustion* items and six *Interpersonal Disengagement* items. Item responses range from 0 (“not at all”) to 4 (“extremely”). The *Work Exhaustion* and *Interpersonal Disengagement* subscales of the PFI have been found to correspond strongly with the emotional exhaustion and depersonalization scales of the Maslach Burnout Inventory.⁴² The OBS was scored by summing then averaging constituent item scores (minimum score=0; maximum score=4).

Professional Fulfillment

Professional fulfillment was assessed with the six-item professional fulfillment subscale (PFS) of the Stanford PFI.⁴¹ The PFS of the PFI rates perceptions of one’s work in the preceding two weeks on a 0 (“not at all true”) to 4 (“completely true”) spectrum. The PFS was scored by summing up then averaging constituent item scores (minimum score=0; maximum score=4).

Peer Connectedness

Peer connectedness was captured via a single questionnaire item (“I feel connected to my peers at work”) scored on a five-point, bi-polar rating spectrum from “strongly disagree” to “strongly agree”.¹¹ Higher scores indicate greater perceptions of social connectedness with physician peers.

Peer Respect

Peer respect among physicians was captured via a single questionnaire item (“I feel respected by my peers at work”) scored on a five-point, bi-polar “strongly disagree” to “strongly agree” rating spectrum.¹¹ Higher scores indicate greater perceptions of social connectedness with physician peers.

Analytic Strategy

Constituent items of the reference scale were ranked in terms of individual item validity based on two empirical criteria. First, we fitted a single-factor diagonally weighted least squares (DWLS) confirmatory factor analysis (CFA)^{43,44} model of the full-length scale and compared individual items’ factor loadings. Next, we conducted a unidimensional graded response model (GRM) item response theory (IRT) analysis⁴⁵ of the full-length scale and compared item-level information function plots, which indicate each item’s precision in capturing well-being centered leadership. The highest-performing item, on these two criteria, was extracted as the single-item abbreviated scale. The three and four highest-performing items were constituted into 3- and 4-item abbreviated scales, respectively.

Divergent validity of the reference versus index scales was compared via Spearman rank correlations (ρ) with the OBS. Convergent validity was evaluated via correlations (ρ) with the PFS, PPAS-6, SVS, Peer Connectedness, and Peer Respect measures. Internal consistency reliability of the full-length adapted MLII versus its multi-item abbreviations was compared using Cronbach’s alpha coefficient⁴⁶ plus ordinal alpha and theta coefficients.^{47,48}

Logistic regressions were fitted to test the predictive validity of continuous index scale scores in discriminating high well-being centered leadership ratings on the reference scale. Short-form scales were compared based on odds ($\pm$ 95% confidence intervals) of discriminating highly supportive leadership behavior on the full-length scale and on area(s) under receiver operating characteristic (ROC) curves (c-statistic). Logistic regressions of high reference scale scores on

elevated index scale scores were also fitted to compare index scales' inter-rater reliability based on accuracy, precision, sensitivity, specificity,⁴⁹ and Cohen's Kappa (κ) coefficients of agreement.⁵⁰ Raw ordinal scale scores were interpreted as defined in their original validation studies and were not normed to a linear 0–10 spectrum as is the practice in some studies. Analyses were performed using SAS[®] software version 9.4 (SAS Inc., Cary, North Carolina, USA).

Results

Study Sample Attributes

The 2023 survey targeted 500 eligible physicians of whom 158 (31.6%) submitted responses, whereas the 2024 survey targeted 815 physicians and elicited responses from 112 (14.95%). Respondents were 75.95% male in 2023, and 75.76% male in 2024; as well as 43.04% non-Hispanic whites in 2023, and 44.44% non-Hispanic whites in 2024. Cardiologists (both non-invasive and non-interventional) comprised 32.28% of respondents in 2023 and 35.35% in 2024, while cardiovascular surgeons were 9.49% of respondents in 2023 and 11.11% in 2024. More than one third of respondents (36.71% in 2023; 35.35% in 2024) had accumulated >20 years of clinical practice experience. Figure 1 (below) summarizes the demographic attributes of the respondents in both 2023 and 2024.

Item-Level Validity and Reliability

In the single-factor CFA of the adapted MLII, items with the highest, second, third, and fourth highest standardized factor loadings both in the 2023 ($\lambda_{\text{standardized}} = 0.9048, 0.8943, 0.8693, \text{ and } 0.890$, respectively) and 2024 samples ($\lambda_{\text{standardized}} =$

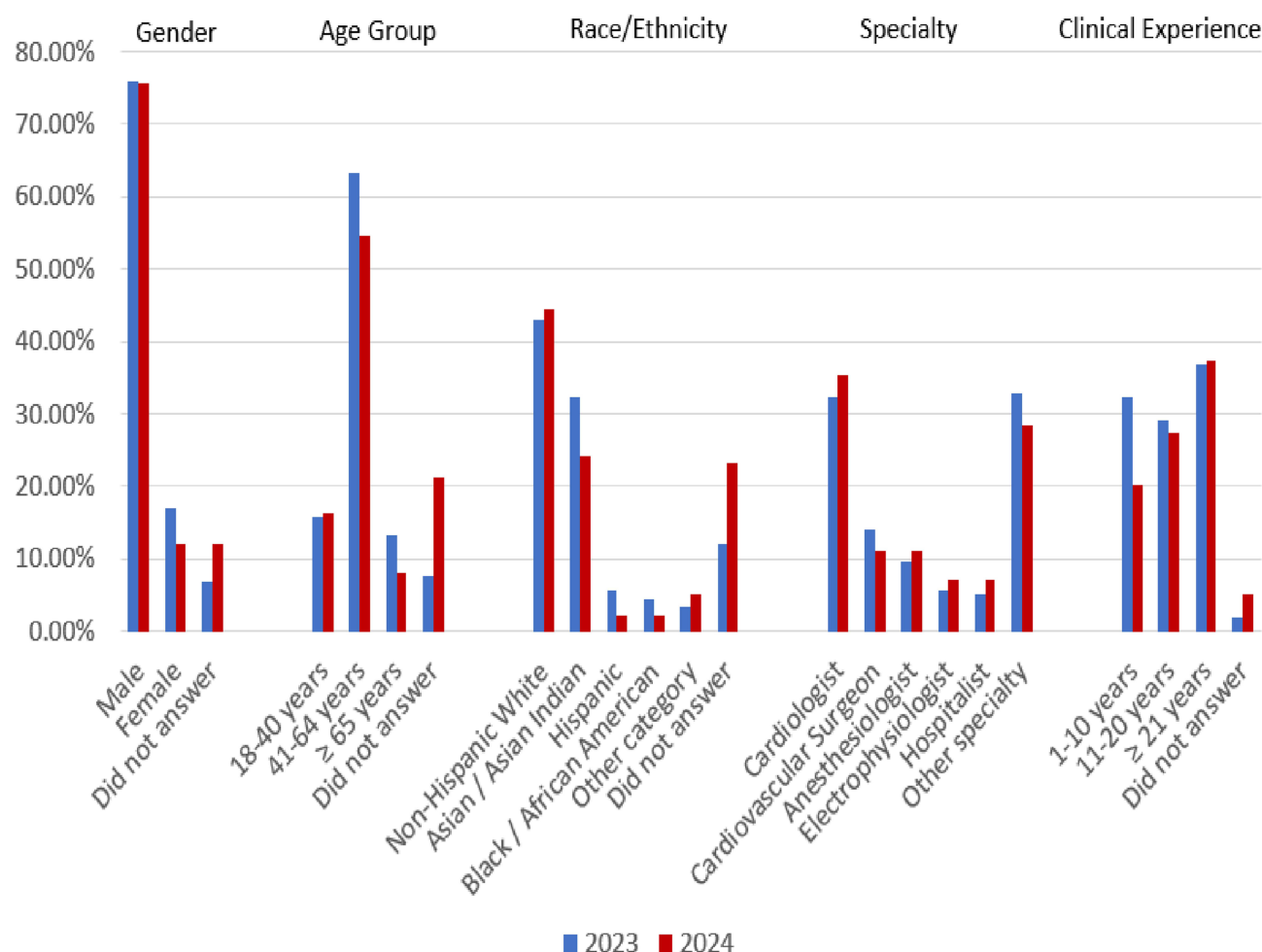


Figure 1 Characteristics of Survey Respondents.

0.9191, 0.9016, 0.8879, and 0.8557, respectively) were: items **5** “provides helpful feedback and coaching”, **6** “recognizes me for a job well done”, **8** “encourages me to develop my talents.skills”, and **2** “empowers me to do my job”. Items 5, 6, 8, and 2 also had the highest proportions of variance in item scores accounted for by the latent factor (well-being centered leadership) in the 2023 ($R^2 = 0.8187, 0.7998, 0.7912, \text{ and } 0.7556$, respectively) and 2024 samples ($R^2 = 0.8448, 0.8128, 0.7884, \text{ and } 0.7322$, respectively). The single-factor CFA model showed excellent fit to the overall sample in both 2023 (SRMR = 0.0354; CFI = 0.9999; TLI = 0.9972) and 2024 (SRMR = 0.0384; CFI = 0.9998; TLI = 0.9974). [Table 1](#) summarizes findings from the CFA model. [Figure 2](#) depicts item-level information function curve plots from a unidimensional GRM analysis of the adapted MLII. The highest, second, third, and fourth highest item-level amount of psychometric information across the breadth of variability in well-being centered leadership was captured by items 5, 6, 2, and 8, respectively, among 2023 respondents; and items 5, 6, 8, and 2, respectively, among 2024 respondents. Based on these CFA and GRM findings, the four-item abbreviation of the adapted MLII combined items 5, 6, 8, and 2 from the full-length parent scale; the three-item abbreviation combined items 5, 6 and 8; while item 5 alone became the single-item abbreviation.

Table 1 Confirmatory Factor Analysis of the Adapted Mayo Leadership Impact Index

#	Item	Sample	Factor Loading (β)	R ²	Error Variance of R ²	Item's Validity Rank
1	Holds career development conversations with me	2023 survey	0.8273	0.6844	0.31564	6th
		2024 survey	0.7593	0.5765	0.42349	9th
2	Empowers me to do my job	2023 survey	0.8693	0.7556	0.24437	4th
		2024 survey	0.8557	0.7322	0.26776	4th
3	Encourages employees to suggest ideas for improvement	2023 survey	0.8499	0.7223	0.27769	5th
		2024 survey	0.8541	0.7294	0.27059	5th
4	Ensures that I am treated with respect and dignity	2023 survey	0.7785	0.6060	0.39401	8th
		2024 survey	0.8138	0.6623	0.33771	7th
5	Provides helpful feedback and coaching on my performance	2023 survey	0.9048	0.8187	0.18128	1st
		2024 survey	0.9191	0.8448	0.15519	1st
6	Recognizes me for a job well done	2023 survey	0.8943	0.7998	0.20019	2nd
		2024 survey	0.9016	0.8128	0.18716	2nd
7	Keeps me informed about changes... in my clinic, work unit, or hospital	2023 survey	0.8061	0.6498	0.35018	7th
		2024 survey	0.7843	0.6151	0.38490	8th
8	Encourages me to develop my talents and skills	2023 survey	0.8895	0.7912	0.20875	3rd
		2024 survey	0.8879	0.7884	0.21162	3rd
9	Overall, I am satisfied with the person(s) who provide(s) ...the support listed above	2023 survey	0.7774	0.6044	0.39561	9th
		2024 survey	0.8541	0.7294	0.27055	5th
Indexes of Overall Goodness-of-fit		Sample	N	SRMR	GFI	TLI
		2023 survey	135	0.0354	0.9999	0.9972
		2024 survey	112	0.0384	0.9998	0.9974

Abbreviations: β , standardized beta coefficient; GFI, goodness of fit index; N, number of respondents; R^2 , squared multiple correlation; SRMR, standardized root mean square residual; TLI, Tucker-Lewis index.

Criterion Validity and Internal Consistency Reliability

The 2023 and 2024 respondent samples yielded similar Spearman rank correlations of the full-length adapted MLII with its single-item ($\rho=0.8903$ vs 0.9198), three-item ($\rho=0.9548$ vs 0.9531), and four-item abbreviations ($\rho=0.9649$ vs 0.9642). Likewise, we observed similar correlations of the four-item abbreviation with its three-item ($\rho=0.9805$ vs 0.9756) and single-item ($\rho=0.9132$ vs 0.9422) counterparts, and between three- and single-item abbreviations ($\rho=0.9242$ vs 0.9628), in 2023 versus 2024.

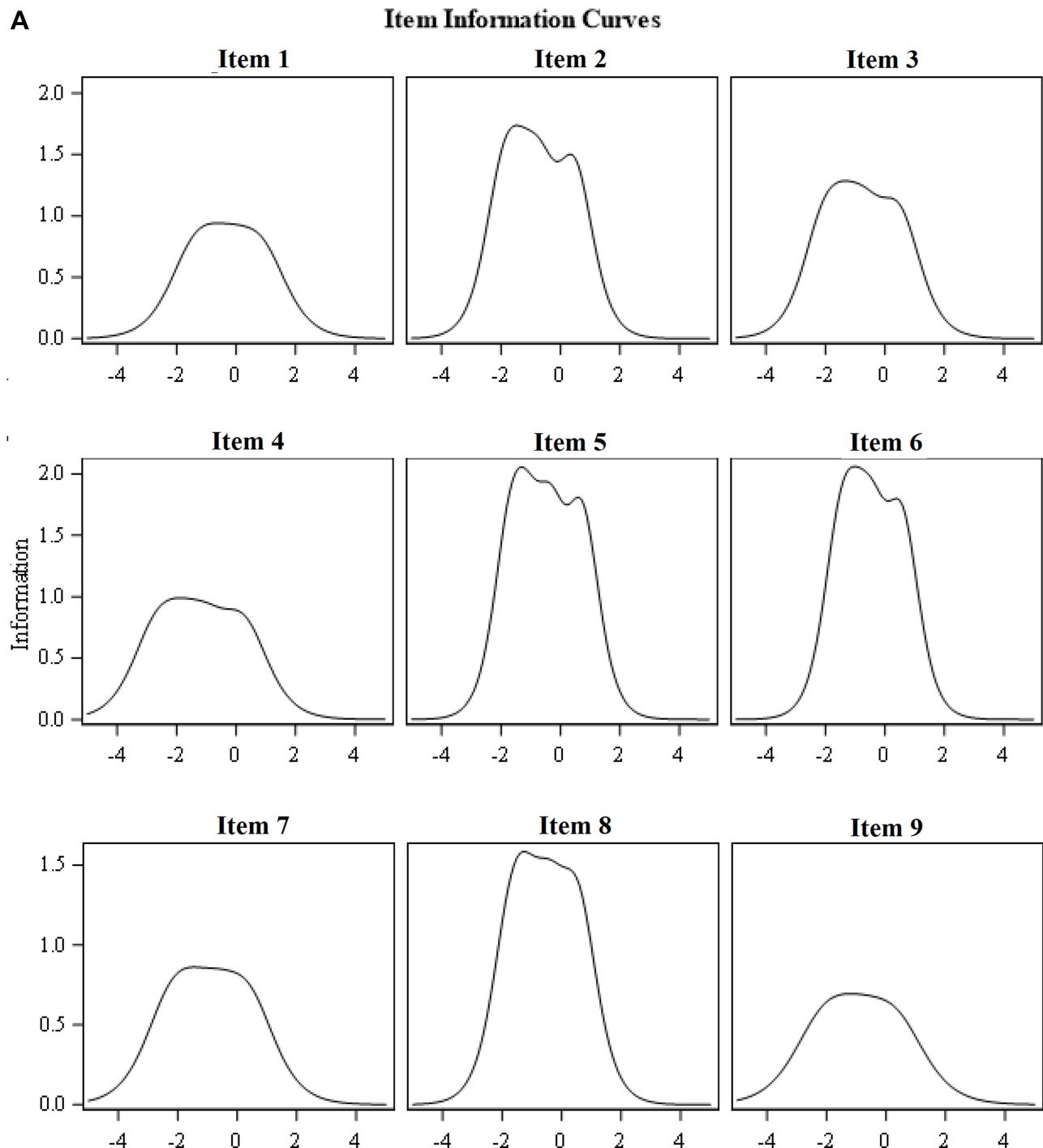


Figure 2 continued.

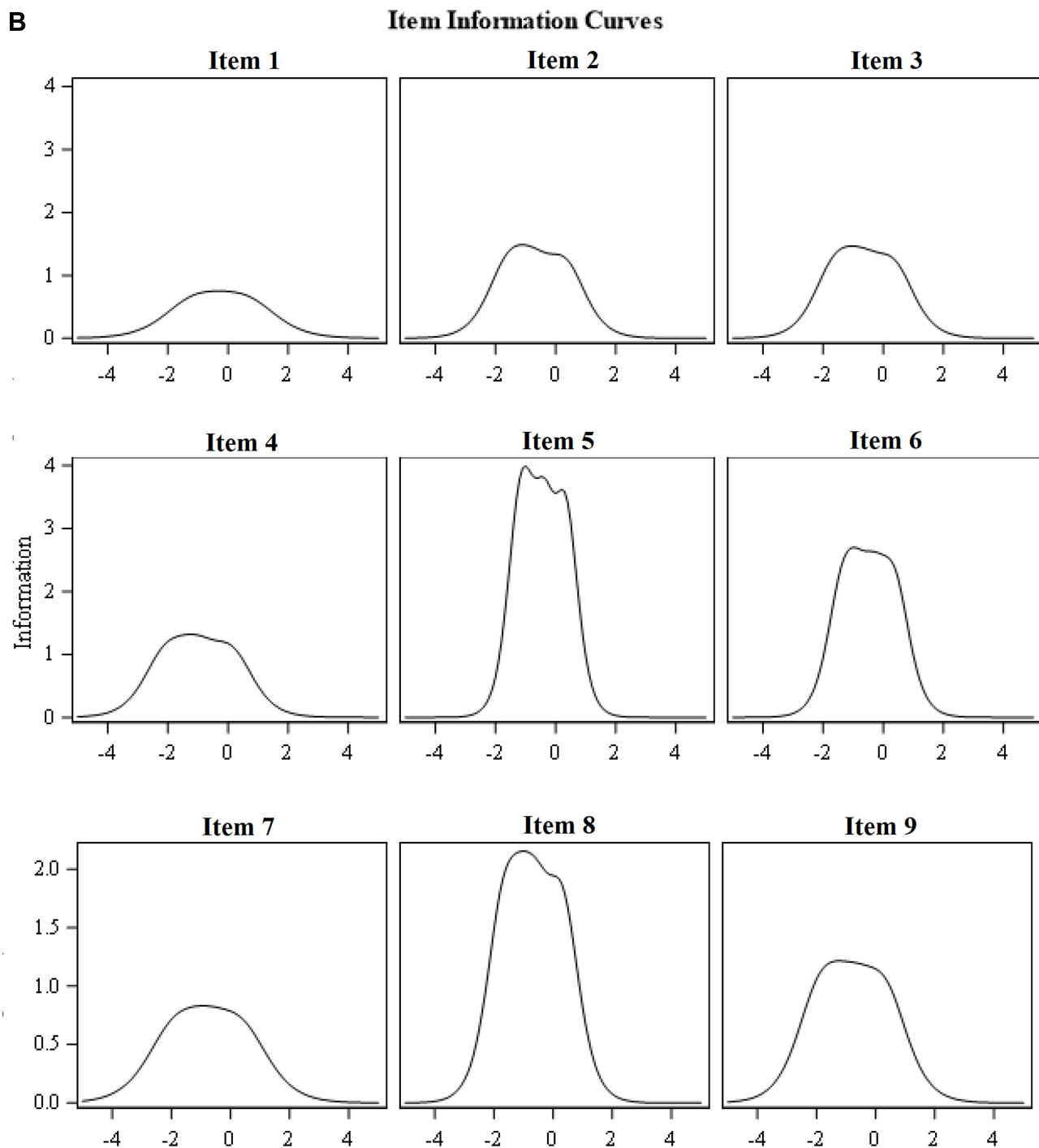


Figure 2 (A) Item Information Curve Plots from a Graded Response Model Analysis on the Full 2023 Respondents' Sample **(B)** Item Information Curve Plots from a Graded Response Model Analysis on the Full 2024 Respondents' Sample.

Correlations of abbreviated MLII scales with well-being measures were all in expected directions (ie., negative correlations with overall burnout but positive correlations with professional fulfillment, autonomy support, self-valuation, peer connectedness, and peer respect).

Abbreviations of the adapted MLII evaluated had virtually identical divergent and convergent validity to that of the full-length, parent scale among both 2023 and 2024 respondents. Table 2 outlines the aforementioned correlation indexes. Four-item and three-item abbreviations were identical to the full-length scale in their internal consistency reliability across 2023 and 2024 respondent populations (all reliability coefficients > 0.900; see [eSupplementary Table 1](#) for details).

Table 2 Distribution of Scores on the Full-Length and Short-Form Versions of the Adapted Mayo Leadership Impact Index and Correlations with Physician Well-Being Measures

	Single-Item Abbreviated Index		3-Item Abbreviated Index		4-Item Abbreviated Index		9-Item Full-Length Index	
	2023	2024	2023	2024	2023	2024	2023	2024
<i>Survey Sample</i>								
<i>Distribution of scale scores</i>								
Range: minimum – maximum	1 – 5	1 – 5	3 – 15	3 – 15	4 – 20	4 – 20	9 – 45	9 – 45
Mean ± standard deviation	3.7 (1.1)	3.7 (1.3)	11.1 (3.1)	11.2 (3.4)	14.6 (4.2)	14.6 (4.6)	34.6 (8.2)	34.3 (9.1)
<i>Spearman correlation coefficient (ρ)</i>								
Single-Item Abbreviated Index	1.0000	1.0000						
3-Item Abbreviated Index	0.9242	0.9628	1.0000	1.0000				
4-Item Abbreviated Index	0.9132	0.9422	0.9805	0.9756	1.0000	1.0000		
9-Item Full-length Index	0.8903	0.9198	0.9548	0.9531	0.9649	0.9622	1.0000	1.0000
Perceived Autonomy Support Scale	0.4721	0.4964	0.4994	0.5007	0.5212	0.5086	0.5670	0.5597
Self-valuation Scale	0.2488	0.3246	0.2833	0.3232	0.2901	0.3403	0.2862	0.3546
Overall Burnout Scale	–0.2423	–0.2884	–0.2544	–0.3244	–0.2700	–0.3560	–0.3032	–0.3822
Professional Fulfillment Index	0.4020	0.4766	0.4398	0.4827	0.4717	0.4884	0.4696	0.4825
Connectedness with Peers (item)	0.3513	0.4817	0.3767	0.5069	0.3877	0.5230	0.4106	0.5249
Respect from Peers (item)	0.4081	0.5344	0.4824	0.5209	0.4826	0.4969	0.4957	0.5274

Predictive Validity and Inter-Rater Reliability

Logistic regression analysis showed that scores on four-, three-, and single-item abbreviations of the adapted MLII had strong predictive associations with the odds of high well-being centered leadership ratings by 2023 and 2024 respondents on the full-length scale (areas under the ROC curve > 0.900). This provides evidence of the abbreviated scales' predictive validity. We also observed high coefficients of agreement (see Table 3) between the abbreviated scales and full-length adapted MLII in classifying highly supportive well-being centered leadership behavior (κ coefficients > 0.700, accuracy > 85%, sensitivity > 0.850, specificity > 0.700). This evidence supports the abbreviated scales' inter-rater reliability vis-à-vis the full-length, parent scale.

Table 3 How the Short-Form Abbreviations Compare with the 9-Item Full-Length Version of the Adapted Mayo Leadership Index in Identifying Highly Supportive Leadership Behaviors

	Single-Item Abbreviated Index		3-Item Abbreviated Index		4-Item Abbreviated Index	
	2023	2024	2023	2024	2023	2024
<i>Survey sample</i>						
<i>Association of scores on short-form scale With high scores on 9-item Adapted Index^a</i>						
Odds ratio	29.59	33.40	6.02	4.34	5.14	3.00
95% CI of odds ratio	(8.86, 98.84)	(8.00, 139.51)	(2.92, 12.42)	(2.31, 8.17)	(2.45, 10.80)	(1.91, 4.71)
Significance (p value)	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Area under the ROC curve (c statistic)	0.918	0.946	0.963	0.970	0.974	0.971

(Continued)

Table 3 (Continued).

	Single-Item Abbreviated Index		3-Item Abbreviated Index		4-Item Abbreviated Index	
<i>Agreement with 9-item Adapted Index In classifying high leadership support^b</i>						
Kappa (κ) coefficient (standard error)	0.703 (0.060)	0.764 (0.065)	0.807 (0.051)	0.846 (0.507)	0.808 (0.051)	0.776 (0.066)
95% CI of κ coefficient	0.586–0.821	0.636–0.891	0.708–0.907	0.735–0.957	0.708–0.907	0.646–0.906
Z statistic	8.353	8.426	9.390	10.501	9.408	9.975
Significance (p value)	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
<i>Comparative ability of short-form scale to Discriminate high leadership support^b</i>						
Accuracy (%)	85.185	89.286	90.370	94.326	90.370	92.199
Precision ratio	0.793	0.841	0.887	0.966	0.937	0.944
Recall (or Sensitivity) ratio	0.956	0.983	0.926	0.918	0.868	0.864
Specificity ratio	0.746	0.793	0.881	0.963	0.940	0.963
F ₁ score	0.867	0.906	0.906	0.941	0.901	0.903

Notes: a = logistic regression of high leadership support according to the 9-item MCLII on scores across the entire range of short-form abbreviations of the MCLII; b = logistic regression of high leadership support according to the 9-item MCLII on high ratings by the abbreviated scales.

Discussion

The present study investigated the comparative validity of short-form abbreviations of the revised MLII adapted for matrixed/multiform leadership structures versus the full-length 9-item scale as the reference. We elicited evidence that the psychometrically strongest 4-item, 3-item, and single-item abbreviations of the adapted MLII are effective proxies of the parent scale. The findings were mostly consistent across two consecutive years of survey data from physicians practicing in a tri-hospital multi-specialty cardiovascular health system. The abbreviations of the adapted MLII provide validated options for accurate surveillance of well-being centered leadership in matrixed healthcare networks aiming for shorter, less costly multi-construct questionnaires.

Physicians in our study rated providing helpful feedback/coaching on one's performance, captured by item 5 of the adapted MLII, as the most salient well-being centered leadership trait. The second, third, and fourth most salient traits were: recognizing a physician for a job well done (item 6), encouraging talents/skills development (item 8), and empowering one to do one's job (item 2). A prior study found encouragement of talents/skills development to be the strongest item of the traditional MLII, with the second and third strongest being the items on helpful feedback/coaching plus empowering one to do one's job.³⁶ Differences in item(s) rated as the most salient by the two studies must be interpreted with caution. As noted before, the Dyrbye et al (2024) study sampled only physicians in traditional hierarchical structures with "direct-report" supervisors and predated the COVID-19 pandemic,³⁶ which substantially affected physicians' well-being.⁵¹ In multiform settings, rather than constrain respondents to rate a direct-report supervisor, each individual ought to evaluate the leader(s) with sufficient influence to empower well-being in the matrix, making the adapted MLII more aligned with such contexts than the traditional MLII. Future studies should assess differences in the salience of specific leadership traits for physicians in healthcare settings with matrixed/multiform leadership structures versus those in more hierarchical settings that assign a single "direct report" supervisor to each physician.

This study yields three short-form proxies of the adapted MLII that are psychometrically strong but less burdensome to incorporate within multi-construct questionnaires by health systems with matrixed, multiform leadership structures that aim to track well-being centered leadership behaviors. The single-item abbreviation of the adapted MLII might be attractive to organizations that can only afford space for one additional item on their multi-construct survey-questionnaire. This choice would share the known advantages and disadvantages of other single-item measures.^{52–55} While less time-consuming, less burdensome for respondents, and less costly to incorporate within multi-construct

questionnaires than long-form scales, single items inadequately capture complex constructs due to a narrow conceptual depth, and their internal consistency is hard to confirm.^{35,54,55} Abbreviated scales with at least three items are short enough to maintain low-burden advantages of single items while adding conceptual nuance.⁵⁶ Compared to their single-item counterpart, 3- and 4-item abbreviations of the adapted MLII have the advantage of enabling latent variable models to be fitted as these typically require three or more indicator items for each latent construct.⁵⁷ Internal consistency reliability is also easier to confirm for the multi-item versus single-item abbreviations.⁵⁵

From an implementation science perspective, the abbreviated short-form versions of the adapted MLII are suitable for scaling across healthcare systems. Within the RE-AIM framework, abbreviated measures enhance “Reach” by reducing respondent burden, foster “Adoption” by lowering dissemination costs, enhance “Implementation” feasibility by fitting into routine survey workflows, and aid “Maintenance” by facilitating repeated, low-burden longitudinal assessment.⁵⁸ From a CFIR viewpoint, these tools align with favorable *intervention characteristics* such as simplicity and adaptability, fit well within the *inner setting* of existing measurement infrastructures, and strengthen the *process* domain by enabling standardized feedback and iterative learning across units and institutions.^{59,60} These attributes position short-form MLII measures as pragmatic tools for implementing and sustaining well-being-centered leadership initiatives in complex, matrixed healthcare systems.

Limitations of this study are acknowledged. Respondents were from a single healthcare system. Generalizability and external validity of our findings needs to be confirmed by future studies in matrixed networks situated in diverse geolocations and cultures (eg., other countries) whose physicians include a diverse mix of specialties, experience levels, and demographics (eg., non-cardiovascular specialties and trainees). Our survey response rates also fell below published averages for online surveys of specialist physicians.^{61,62} While this can limit the generalizability of estimated point prevalence of the domains assessed, it is less likely to influence which MLII items have the highest factor loading on reference scales and/or the correlations between MLII scores and other measures. We performed a sensitivity analysis to assess the extent to which the modest sample size challenged statistical conclusion validity. The sensitivity analysis duplicated our methods on an expanded, simulated dataset generated via 100 multiple imputations⁶³ of the study sample and obtained identical psychometric indexes, suggesting that the findings are robust to sample size limitations. Sahin et al show that, in a unidimensional IRT model testing a 10-item scale, a sample of 150 persons (ie., 15 persons per item) is sufficient to yield accurate item parameters.⁶⁴ The strict anonymity of survey responses did not permit linking of each respondent’s cross-sectional data across both survey years, which precluded the calculation of test-retest reliability. The study also did not test acquiescence response bias. The main strengths of the study are the robust reliability and validity indexes, which are similar to those obtained from larger studies. Consistency of findings across two consecutive years is another strength.

Conclusion

This study successfully evaluated the psychometric reliability, validity, and utility of 4-item, 3-item, and 1-item abbreviated versions of the revised MLII adapted for organizational settings with matrixed, multiform leadership structures. These three short-form abbreviations provide less burdensome but strong proxies of the 9-item full-length scale from which organizations can select a most appropriate option based on specific needs for brevity and parsimony. The single-item abbreviation can act as an ultra-brief, limited screening measure of well-being centered leadership to rapidly identify under-served physician subgroups for subsequent deeper analysis/intervention. The 3- and 4-item abbreviations provide more nuance, maintaining low-burden advantages of single items, but with less detailed feedback, especially for organizations seeking to identify specific leadership behavior domains for improvement, than the full-length scale. Future studies should evaluate short-form and full-length versions of the adapted MLII within a broader range of matrix-structured healthcare networks.

Acknowledgments

The authors acknowledge the assistance of Maris Adams, MS, and Colleen Parro, BS, of Baylor Scott and White Health Research Institute, as research coordinators on the parent project. The Physicians’ advisory panel at Baylor Scott and

White-The Heart Hospital helped to champion the survey among fellow physicians. We are also indebted to all the respondents to our survey in 2023 and 2024.

Disclosure

Dr. Tait Shanafelt, a co-inventor of the Mayo Leadership Impact Index (MLII), shares a portion of the royalties with Mayo Clinic, owners of the proprietary copyright and license for this measure. As an internationally renowned expert on clinician well-being, Dr. Shanafelt often presents grand rounds or keynote lectures and advises healthcare organizations on improving their practice environments. He receives honorarium for some of these engagements. Other authors have no potential conflicts of interest to disclose regarding this research.

References

- Shanafelt TD, West CP, Sinsky C, et al. Changes in burnout and satisfaction with work-life integration in physicians and the general us working population between 2011–2023. *Mayo Clin Proc.* **2025**;100(7):1142–1158. doi:10.1016/j.mayocp.2024.11.031
- Demmy TL, Kivlahan C, Stone TT, Teague L, Sapienza P. Physicians' perceptions of institutional and leadership factors influencing their job satisfaction at one academic medical center. *Acad. Med.* **2002**;77(12 Pt 1):1235–1240. doi:10.1097/00001888-200212000-00020
- Shanafelt TD, Gorringer G, Menaker R, et al. Impact of organizational leadership on physician burnout and satisfaction. *Mayo Clin Proc.* **2015**;90(4):432–440. doi:10.1016/j.mayocp.2015.01.012
- Dyrbye LN, Major-Elechi B, Hays JT, Fraser CH, Buskirk SJ, West CP. Relationship between organizational leadership and health care employee burnout and satisfaction. *Mayo Clin Proc.* **2020**;95(4):698–708. doi:10.1016/j.mayocp.2019.10.041
- Dyrbye LN, Leep Hunderfund AN, Winters RC, et al. The relationship between residents' perceptions of residency program leadership team behaviors and resident burnout and satisfaction. *Academic Med.* **2020**;95(9):1428–1434. doi:10.1097/acm.00000000000003538
- Dyrbye LN, Major-Elechi B, Hays JT, Fraser CH, Buskirk SJ, West CP. Physicians' ratings of their supervisor's leadership behaviors and their subsequent burnout and satisfaction: a longitudinal study. *Mayo Clin Proc.* **2021**;96(10):2598–2605. doi:10.1016/j.mayocp.2021.01.035
- Shanafelt TD, Wang H, Leonard M, et al. Assessment of the association of leadership behaviors of supervising physicians with personal-organizational values alignment among staff physicians. *JAMA Netw Open.* **2021**;4(2):e2035622. doi:10.1001/jamanetworkopen.2020.35622
- Mete M, Goldman C, Shanafelt T, Marchalik D. Impact of leadership behaviour on physician well-being, burnout, professional fulfilment and intent to leave: a multicentre cross-sectional survey study. *BMJ Open.* **2022**;12(6):e057554. doi:10.1136/bmjopen-2021-057554
- Meredith LS, Bouskill K, Chang J, Larkin J, Motala A, Hempel S. Predictors of burnout among US healthcare providers: a systematic review. *BMJ Open.* **2022**;12(8):e054243. doi:10.1136/bmjopen-2021-054243
- Tawfik DS, Adair KC, Palassoff S, et al. Leadership behavior associations with domains of safety culture, engagement, and health care worker well-being. *Jt Comm J Qual Saf.* **2023**;49(3):156–165. doi:10.1016/j.jcjq.2022.12.006
- Ashmore JA, Waddimba AC, Douglas ME, Coombes SV, Shanafelt TD, DiMaio JM. The mayo leadership impact index adapted for matrix leadership structures: initial validity evidence. *J. Healthc. Leadersh.* **2024**;16:315–327. doi:10.2147/jhl.S465170
- Waddimba AC, Ashmore J, Douglas ME, et al. Association of well-being-centered leadership with burnout and professional fulfillment among physicians: mediating effects of autonomy support and self-valuation. *Leadership Health Services.* **2025**;38(5):65–81. doi:10.1108/lhs-01-2025-0001
- Spilg EG, McNeill K, Dodd-Moher M, et al. Physician leadership and its effect on physician burnout and satisfaction during the COVID-19 pandemic. *J. Healthc. Leadersh.* **2025**;17:49–61. doi:10.2147/JHL.S487849
- Gilin DA, Anderson GG, Etezad S, Lee-Baggle D, Cooper AM, Preston RJ. Impact of a wellness leadership intervention on the empathy, burnout, and resting heart rate of medical faculty. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes.* **2023**;7(6):545–555. doi:10.1016/j.mayocpiqo.2023.09.005
- Calderón V, Mogul Wyman A, Miller G. Preliminary findings from a pilot professional coaching program on the components of burnout in a diverse group of physician leaders. *Glob Adv in Integr Med Health.* **2024**;13:27536130241296088. doi:10.1177/27536130241296088
- Briggs SE, Heman-Ackah SM, Hamilton F. The impact of leadership training on burnout and fulfillment among direct reports. *J. Healthc. Manag.* **2024**;69(6):402–413. doi:10.1097/jhm-d-23-00209
- Kiser SB, Sterns JD, Lai PY, Horick NK, Palamara K. Physician coaching by professionally trained peers for burnout and well-being: a randomized clinical trial. *JAMA Netw Open.* **2024**;7(4):e245645–e245645. doi:10.1001/jamanetworkopen.2024.5645
- Peter KA, Voirol C, Kunz S, et al. Reducing work-related stress among health professionals by using a training-based intervention programme for leaders in a cluster randomised controlled trial. *Sci. Rep.* **2024**;14(1):23502. doi:10.1038/s41598-024-73939-y
- Smith S, Goldhaber N, Maysent K, Lang U, Daniel M, Longhurst C. Impact of a virtual coaching program for women physicians on burnout, fulfillment, and self-valuation. *BMC Psychol.* **2024**;12(1):331. doi:10.1186/s40359-024-01763-0
- Hartung K, Swann-Thomsen H, Schneider K. Wellness-centered leadership: a key differentiator for successfully reducing burnout and building a culture of well-being among physicians and apPs. *J. Healthc. Leadersh.* **2025**;17:145–157. doi:10.2147/JHL.S513209
- Sears DM, Bejcek A, Kilpatrick L, et al. Leadership development as a novel strategy to mitigate burnout among female physicians. *PLoS One.* **2025**;20(3):e0319895. doi:10.1371/journal.pone.0319895
- James TT, Nayak AC, Houff AM, et al. Well-being leadership training to reduce clinician burnout in a metropolitan community health system. *Healthcare.* **2025**;13(3177):3177. doi:10.3390/healthcare13233177
- Shanafelt TD, Trockel M, Rodriguez A, Logan D. Wellness-centered leadership: equipping health care leaders to cultivate physician well-being and professional fulfillment. *Academic Med.* **2021**;96(5):641–651. doi:10.1097/acm.0000000000003907
- Swensen SJ, Shanafelt TD. Agency action: measuring leader behaviors. in: *Mayo Clinic strategies to reduce burnout: 12 actions to create the ideal workplace*. New York, NY, USA; Oxford University Press, Inc.; **2020**. 105–120. doi:10.1093/med/9780190848965.003.0015

25. Swensen S, Shanafelt TD. *Cultivating Leadership: Measure and Assess Leader Behaviors to Improve Professional Well-Being*. American Medical Association (AMA), Professional Satisfaction and Practice Sustainability Group; 2021. <https://edhub.ama-assn.org/steps-forward/module/2774089>.
26. American Medical Association. *Wellness-Centered Leadership Playbook: Cultivating a Culture of Wellness Within Your Organization*. Chicago, Illinois, USA: American Medical Association (AMA) STEPS Forward 2024.
27. Heeringa J, Mutti A, Furukawa MF, Lechner A, Maurer KA, Rich E. Horizontal and vertical integration of health care providers: a framework for understanding various provider organizational structures. *Int. J. Integr. Care*. 2020;20(1):1–10. doi:10.5334/ijic.4635
28. Lyu PF, Chernen ME, McWilliams JM. Soft consolidation in Medicare ACOs: potential for higher prices without mergers or acquisitions. *Health Affairs*. 2021;40(6):979–988. doi:10.1377/hlthaff.2020.02449
29. Khan S, Vandermorris A, Shepherd J, et al. Embracing uncertainty, managing complexity: applying complexity thinking principles to transformation efforts in healthcare systems. *BMC Health Serv. Res.* 2018;18(192):1–8. doi:10.1186/s12913-018-2994-0
30. American Hospital, American Medical Association. Integrated leadership for hospitals and health Systems: principles for success; 2015. https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/about-ama/ama-aha-integrated-leadership-principles_0.pdf. Accessed April 29, 2026.
31. Prasad K, Poplau S, Brown R, et al. Time pressure during primary care office visits: a prospective evaluation of data from the healthy work place study. *J. Gen. Intern. Med.* 2020;35(2):465–472. doi:10.1007/s11606-019-05343-6
32. Maloney P, Grawitch MJ, Barber LK. Strategic item selection to reduce survey length: reduction in validity? *Consult. Psychol. J: Pract Res.* 2011;63(3):162–175. doi:10.1037/a0025604
33. Kost RG, de Rosa JC. Impact of survey length and compensation on validity, reliability, and sample characteristics for ultrashort-, short-, and long-research participant perception surveys. *J Clin Transl Res.* 2018;2(1):31–37. doi:10.1017/cts.2018.18
34. Kato T, Miura T. The impact of questionnaire length on the accuracy rate of online surveys. *J. Mark. Anal.* 2021;9(2):83–98. doi:10.1057/s41270-021-00105-y
35. Sibley CG, Stronge S, Lilly KJ, et al. Comparative reliability of 108 scales and their short-form counterparts. *New Zealand J Psychol.* 2024;53(2):57–76. doi:10.63146/001c.138416
36. Dyrbye LN, Satele DV, West CP. A pragmatic approach to assessing supervisor leadership capability to support healthcare worker well-being. *J. Healthc. Manag.* 2024;69(4):280–295. doi:10.1097/jhm-d-23-00137
37. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 2009;42(2):377–381. doi:10.1016/j.jbi.2008.08.010
38. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* 2019;95:103208. doi:10.1016/j.jbi.2019.103208
39. Waddimba AC, Mohr DC, Beckman HB, Meterko MM. Physicians' perceptions of autonomy support during transition to value-based reimbursement: a multi-center psychometric evaluation of six-item and three-item measures. *PLoS One*. 2020;15(4):e0230907. doi:10.1371/journal.pone.0230907
40. Trockel MT, Hamidi MS, Menon NK, et al. Self-valuation: attending to the most important instrument in the practice of medicine. *Mayo Clin Proc.* 2019;94(10):2022–2031. doi:10.1016/j.mayocp.2019.04.040
41. Trockel M, Bohman B, Lesure E, et al. A brief instrument to assess both burnout and professional fulfillment in physicians: reliability and validity, including correlation with self-reported medical errors, in a sample of resident and practicing physicians. *Academic Psychiatry*. 2018;42(1):11–24. doi:10.1007/s40596-017-0849-3
42. Brady KJS, Ni P, Carlasare L, et al. Establishing crosswalks between common measures of burnout in US physicians. *J. Gen. Intern. Med.* 2022;37(4):777–784. doi:10.1007/s11606-021-06661-4
43. Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychol. Methods*. 2004;9(4):466–491. doi:10.1037/1082-989x.9.4.466
44. DiStefano C, Morgan GB. A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling*. 2014;21(3):425–438. doi:10.1080/10705511.2014.915373
45. Samejima F. Graded Response Models. In: van der Linden WJ, editor. *Handbook of Item Response Theory, Volume One: Models*. Boca Raton, FL: Chapman & Hall/CRC Press, Taylor & Francis Group; 2016:95–108. doi:10.1201/9781315374512
46. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297–334. doi:10.1007/BF02310555
47. Zumbo B, Gadermann A, Zeisser C. Ordinal versions of coefficients alpha and theta for likert rating scales. *J Mod. Appl. Stat. Method.* 2007;6(1):21–29. doi:10.22237/jmasm/1177992180
48. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for likert-type and ordinal item response data: a conceptual, empirical, and practical guide. *Practical Assessment, Research and Evaluation*. 2012;17(3). doi:10.7275/n560-j767
49. Gwet KL. *Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters*. Fifth ed. Gaithersburg, Maryland, USA: AgreeStat Analytics; 2021. <https://agreestat.com/books/cac5>. Accessed April 25, 2026.
50. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20(1):37–46. doi:10.1177/001316446002000104
51. Sarode AL, Hu X, Dill MJ. COVID –19 and physician burnout in the United States: cross-sectional and longitudinal evidence from a national survey. *Health Serv. Res.* 2025;60:0e70003. doi:10.1111/1475-6773.70003
52. Loo R. A caveat on using single-item versus multiple-item scales. *J. Manag. Psychol.* 2002;17(1):68–75. doi:10.1108/02683940210415933
53. Bowling A. Just one question: if one question works, why ask several? *J Epidemiol Community Health*. 2005;59(5):342–345. doi:10.1136/jech.2004.021204
54. Hays RD, Reise S, Calderón JL. How much is lost in using single items? *J. Gen. Intern. Med.* 2012;27(11):1402–1403. doi:10.1007/s11606-012-2182-6
55. Allen MS, Iliescu D, Greiff S. Single item measures in psychological science: a call to action. *Eur. J. Psychol. Assess.* 2022;38(1):1–5. doi:10.1027/1015-5759/a000699
56. Riley MR, Mohr DC, Waddimba AC. The reliability and validity of three-item screening measures for burnout: evidence from group-employed health care practitioners in upstate New York. *Stress and Health*. 2018;34(1):187–193. doi:10.1002/smi.2762
57. Kline RB. *Principles and Practice of Structural Equation Modeling*. Fifth ed. New York, NY, USA: Guilford Press.; 2023.

58. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the re-aim framework. *Am. J. Public Health.* 1999;89(9):1322–1327. doi:10.2105/ajph.89.9.1322
59. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement. Sci.* 2009;4(50). doi:10.1186/1748-5908-4-50
60. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated consolidated framework for implementation research based on user feedback. *Implement. Sci.* 2022;17(75). doi:10.1186/s13012-022-01245-0
61. Cunningham CT, Quan H, Hemmelgarn B, et al. Exploring physician specialist response rates to web-based surveys. *BMC Med Res Methodol.* 2015;15(32). doi:10.1186/s12874-015-0016-z
62. Meyer VM, Benjamins S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. *Ann Surg.* 2022;275(1):e75–e81. doi:10.1097/sla.0000000000004078
63. Graham JW, Schafer JL. On the performance of multiple imputation for multivariate data with small sample size. In: Hoyle RH, editor. *Statistical Strategies for Small Sample Research*. 1st ed. Thousand Oaks, California, USA: SAGE Publications, Inc; 1999:1–29.
64. Şahin A, Anıl D. The effects of test length and sample size on item parameters in item response theory. *Educ. Sci.: Theory Pract.* 2017;17(1):321–335. doi:10.12738/estp.2017.1.0270

Journal of Healthcare Leadership

Publish your work in this journal

The Journal of Healthcare Leadership is an international, peer-reviewed, open access journal focusing on leadership for the health profession. The journal is committed to the rapid publication of research focusing on but not limited to: Healthcare policy and law; Theoretical and practical aspects healthcare delivery; Interactions between healthcare and society and evidence-based practices; Interdisciplinary decision-making; Philosophical and ethical issues; Hazard management; Research and opinion for health leadership; Leadership assessment. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/journal-of-healthcare-leadership-journal>

Dovepress
Taylor & Francis Group