Physicians in all disciplines are called upon to demonstrate leadership skills even if they do not hold a formal leadership role. Additionally, organizations led by physicians perform better than organizations with nonphysician leaders.1 For these reasons, leadership skill development must become a routine part of medical education at all levels and in all specialties.1-20 National organizations recognize the need for effective leadership training. The Interprofessional Education Collaborative defined competencies in interprofessional collaboration for a variety of medical professionals, including physicians, that include the use of leadership skills to support collaborative practice and team effectiveness.21 The Institute of Medicine’s July 2014 report Graduate Medical Education that Meets the Nation’s Health Needs recommended “production of a physician workforce better prepared to work in, help lead, and continually improve an evolving health care delivery system that can provide better individual care, better population health, and lower cost.”19 The Accreditation Council for Graduate Medical Education (ACGME) Milestones for Family Medicine specify competencies in team leadership, collaboration with public health and community agencies, advocacy, and leadership of systems and organizational strategies.22 The ACGME Milestones for Internal Medicine and Pediatrics identify competencies in team management, communication, practice improvement, and advocacy.23,24

From the Department of Family and Preventive Medicine (Drs Van Hala, Cochella, Kiraly, Pohl, and Gren, and Ms Jaggi), and College of Social Work (Dr Frost), University of Utah, Salt Lake City, UT.
While many innovative curricula have been developed, few target all learners, use a validated assessment, or show meaningful impact. Validated assessment is needed to demonstrate impact and compare curriculum. A systematic review of leadership training in health care teams found that defining best practices is difficult due to lack of a standard definition of leadership, supporting frameworks, and robust assessments.

Though assessments do exist for health care administrators, practicing physicians, and surgeons’ non-technical skills in the operating room (including leadership), there are no published assessment instruments of leadership skills for physicians in residency training programs. Knowing that leadership development covers a broad spectrum, we specifically wanted to describe the foundational skills that early-career physicians, specifically residents, need to learn to be effective team leaders. Residents routinely participate in interprofessional teams focused on clinical care and quality improvement. Theoretically though, foundational leadership skills should translate across a variety of settings.

The purpose of this study was to develop a health care leadership self-assessment of foundational competencies and validate the tool among family medicine residents.

Methods
Faculty at the University of Utah Family Medicine Residency Program developed the initial leadership self-assessment from a compilation of materials on general leadership skills, including those for health care executives. Our intent was to identify leadership competencies for early-career physicians, specifically resident physicians, to be effective team leaders.

To validate the self-assessment, we used a sequential exploratory study design, which utilizes “an initial phase of qualitative data collection and analysis. The overall purpose of this approach is “to explore a phenomenon.” This strategy may also be useful when developing and testing a new instrument. The Institutional Review Board at the University of Utah approved this study.

Qualitative Methods
To refine the self-assessment, we conducted two focus groups. The primary focus group question was: “What are the foundational leadership skills that a family medicine physician needs to be an effective leader of interprofessional teams?”

The focus group participants discussed the family medicine physician’s role in interprofessional teams and described the foundational skills that family medicine physicians need to be effective leaders. Participants then completed the self-assessment, and discussed its comprehensiveness, relevance, clarity, and ease of use. Each focus group lasted approximately 1 hour, was audio recorded, and detailed notes were taken by the facilitators. This qualitative data was analyzed using open coding to identify emerging themes from the data as they linked to the assessment.

Face Validity. In March 2015, we conducted the first focus group with leaders in medicine to elicit expert opinion on the skills needed for leadership in family medicine and to identify whether the devised items captured the noted leadership skills. This cross-disciplinary group included experts in community medicine, academic medicine, family medicine residencies, and team practice. The insights from this focus group were used to guide revisions of the self-assessment.

Content Validity. The purpose of the second focus group (June 2015) was to further refine the self-assessment with eight third-year residents from our home institution. We followed a focus group process, similar to the expert focus group, to explore resident perspective on leadership competencies. The residents reviewed the self-assessment, which had been revised based on information from the first focus group. We elicited their views on the clarity and comprehensiveness of the self-assessment, and used their responses to clarify and simplify the self-assessment items.

Quantitative Methods
For the second phase of this study, we contacted 30 family medicine residency directors to distribute the pilot self-assessment to their residents. The directors were selected by purposive sample, based on geographic region and type of program (community versus university), in order to increase diversity of representation. The author (SVH) contacted the directors by email with an invitation to participate in a validation study of a leadership self-assessment. We sent an email survey link to the participating directors who were instructed to forward the survey to their entire resident cohort. Two weeks later we sent a second email to improve participation. REDCap, a secure web application, was the survey platform.

We collected basic demographic information and asked residents to complete the 33-item self-assessment on a scale from beginner (1) to expert (5). We compared the demographics of our sample to those of US family medicine residents (as reported by the Accreditation Council for Graduate Medical Education, 2013-2014) using chi-square analysis and Fisher’s exact computation for P value.

Construct Validity. We conducted a principal component analysis with varimax rotation, to reduce the number of assessment items, and group these items into domains using Stata (v 13.1 College Station, TX). We retained individual items with eigenvalues over 1.0. Factor loadings of at least 0.30 were used to group items into domains. Once the domains were established, we calculated and compared mean scores overall and for each domain across demographic variables using linear regression analysis.
Domain Development
Principal component analysis grouped the self-assessment statements into domains. Once the domains were established, we named the domains. Our research team completed a comprehensive review of the scholarly and academic literature to determine the vocabulary and categorization of leadership behaviors.\textsuperscript{38,45,47,66,67} We drew from our collective experience in teaching medical, public health, and social work students about leadership. We named the leadership domains to represent the statements contained within each domain.

Results
Initially, 30 program directors were contacted to inquire whether they would be interested in participating in this validation study. Of these, 22 programs agreed to send the survey link out to their residents. Of these, 13 program directors confirmed via email that the link was sent out to residents. For the remaining programs for which no confirmation was obtained, it is not possible to determine which programs did distribute the link to their residents, as survey responses were anonymous and respondents were asked only to identify their state, but not their program. Of the programs that initially agreed but did not confirm sending the link, three programs must have sent the link, as there were responses from that state and their program was the only one contacted in the state. Together, these 16 programs are considered verified participants. The participation of four programs cannot be determined, as there were multiple programs contacted in the same state, preventing researchers from delineating whether these programs did in fact send the link. Finally, two programs likely did not send the link out to residents, as the program was the only program contacted in that state and no responses from that state were received (Figure 1). Overall, responses were received from between 16 and 20 programs with a broad geographic distribution across 12 states (West: California, Colorado, Utah, Washington; Midwest: Illinois, Michigan, Minnesota; Northeast: New York, Pennsylvania; South: Missouri, North Carolina, Texas).

The maximum number of residents potentially included in the sample across the 20 programs that most likely sent out the link was 564, while the number of residents in the 16 programs verified to have circulated the assessment was 468. We received responses from 163 residents, giving a response rate of 28.9% across all 20 programs, or 34.8% across the 16 programs with verified participation. The majority of respondents were female (64.4%) and white (65.6%). Demographics for respondents, US family medicine residents, and all US medical residents are presented in Table 1.\textsuperscript{64} We compared the demographics of our sample to all US family medicine residents, and found significant differences by sex ($P=0.015$), race/ethnicity ($P=0.001$) and residency year ($P=0.000$).

The principal component analysis reduced the number of items from 33 on the pilot assessment to 21 on the final assessment, which was titled the Foundational Healthcare Leadership Self-assessment (FHLS). The

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{participating_residency_programs.png}
\caption{Participating Residency Programs}
\end{figure}
21 items of the FHLS were assigned to five leadership domains: accountability, collaboration, communication, team management, and self-management (Table 2).

An average score for each domain was calculated by summing the self-assessment values (from 1=beginner to 5=expert) from the appropriate items, and then dividing by the number of items in the domain. The average scores from our survey are presented in Table 3. Regression analysis of average scores found no significant differences by sex, but did find important differences by residency year, with more advanced residents scoring higher ($P<0.01$). Additionally, there was suggestive evidence that nonwhite residents reported higher leadership scores than their white peers ($P=0.055$, Table 4).

Regression analysis for each of the five domains revealed similar patterns, with residency year remaining the strongest predictor of average score (results not shown).

**Discussion**

We created and validated the FHLS as a self-assessment of foundational leadership skills for early-career physicians. We validated this...
assessment using two groups—experts and users—using qualitative and quantitative analysis for face, content, and construct validity. A self-assessment can be used longitudinally as a formative tool during residency to help residents engage in personalized leadership development. The FHLS is the first step in the development of educational and evaluative assessments for training clinician leaders.

The FHLS was developed with a focus on foundational skills for leading interprofessional teams. During residency, residents work in and lead teams, in both formal and informal roles, providing clinical care in ambulatory and inpatient settings, and conducting quality improvement. We imagine these foundational skills are generalizable across different team environments.

The scores from this tool showed discernable differences by year of residency, and no gender or age differences. Scores increased by year of residency, which is an expected progression with advancing training. As residents develop more experience, they feel more skilled.

Table 3: Average FHLS Scores* (Standard Deviation) Overall and by Domain

<table>
<thead>
<tr>
<th>Overall Score</th>
<th>Accountability</th>
<th>Collaboration</th>
<th>Communication</th>
<th>Team Management</th>
<th>Self-management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sample</td>
<td>3.4 (0.6)</td>
<td>3.7 (0.6)</td>
<td>3.2 (0.8)</td>
<td>3.7 (0.6)</td>
<td>3.0 (0.8)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>3.4 (0.5)</td>
<td>3.7 (0.4)</td>
<td>3.3 (0.8)</td>
<td>3.7 (0.5)</td>
<td>3.0 (0.8)</td>
</tr>
<tr>
<td>Female</td>
<td>3.4 (0.5)</td>
<td>3.7 (0.6)</td>
<td>3.2 (0.8)</td>
<td>3.7 (0.5)</td>
<td>3.1 (0.8)</td>
</tr>
<tr>
<td>Minority Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>3.3 (0.6)</td>
<td>3.6 (0.6)</td>
<td>3.2 (0.8)</td>
<td>3.7 (0.5)</td>
<td>3.0 (0.8)</td>
</tr>
<tr>
<td>Nonwhite</td>
<td>3.5 (0.5)</td>
<td>3.8 (0.5)</td>
<td>3.3 (0.9)</td>
<td>3.8 (0.6)</td>
<td>3.2 (0.8)</td>
</tr>
<tr>
<td>Residency Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3.1 (0.6)</td>
<td>3.5 (0.6)</td>
<td>2.9 (0.8)</td>
<td>3.5 (0.6)</td>
<td>2.6 (0.8)</td>
</tr>
<tr>
<td>2</td>
<td>3.5 (0.5)</td>
<td>3.8 (0.5)</td>
<td>3.2 (0.8)</td>
<td>3.8 (0.5)</td>
<td>3.2 (0.7)</td>
</tr>
<tr>
<td>3</td>
<td>3.7 (0.3)</td>
<td>3.9 (0.4)</td>
<td>3.7 (0.6)</td>
<td>3.9 (0.5)</td>
<td>3.5 (0.5)</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>3.5 (0.5)</td>
<td>3.8 (0.5)</td>
<td>3.3 (0.8)</td>
<td>3.8 (0.5)</td>
<td>3.2 (0.8)</td>
</tr>
<tr>
<td>South</td>
<td>3.3 (0.6)</td>
<td>3.7 (0.6)</td>
<td>3.2 (0.9)</td>
<td>3.6 (0.6)</td>
<td>2.9 (0.9)</td>
</tr>
<tr>
<td>Midwest</td>
<td>3.4 (0.6)</td>
<td>3.7 (0.7)</td>
<td>3.2 (0.8)</td>
<td>3.7 (0.8)</td>
<td>3.1 (0.7)</td>
</tr>
<tr>
<td>Northeast</td>
<td>3.3 (0.6)</td>
<td>3.6 (0.6)</td>
<td>3.1 (0.9)</td>
<td>3.6 (0.5)</td>
<td>2.9 (1.0)</td>
</tr>
</tbody>
</table>

*Scores: 1=beginner to 5=expert

Table 4: Regression Analysis of Average FHLS Score

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>(ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>-0.052</td>
<td>-0.213-0.109</td>
<td>0.526</td>
</tr>
<tr>
<td>Age</td>
<td>0.010</td>
<td>-0.019-0.038</td>
<td>0.503</td>
</tr>
<tr>
<td>Minority Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>(ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonwhite</td>
<td>0.159</td>
<td>-0.003-0.322</td>
<td>0.055</td>
</tr>
<tr>
<td>Residency Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.297</td>
<td>0.114-0.481</td>
<td>0.002</td>
</tr>
<tr>
<td>3</td>
<td>0.591</td>
<td>0.399-0.783</td>
<td>0.000</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>(ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>-0.177</td>
<td>-0.391-0.037</td>
<td>0.105</td>
</tr>
<tr>
<td>Midwest</td>
<td>0.019</td>
<td>-0.187-0.226</td>
<td>0.854</td>
</tr>
<tr>
<td>Northeast</td>
<td>-0.059</td>
<td>-0.284-0.166</td>
<td>0.603</td>
</tr>
</tbody>
</table>

The majority of respondents to this survey were female and white. Compared to family medicine and all residents nationally, respondents to the survey were disproportionately female. However, our analysis showed no difference in response by gender. Nonwhite respondents scored higher in the total score than white respondents (P=0.055). This phenomenon should be studied further. While respondents represented
both academic and community-based residency programs, we did not ask them to identify the type of program, in order to maintain anonymity. Type of program could be evaluated in future studies with the FHLS.

The FHLS is a brief self-assessment that takes less than 5 minutes to complete and quantifies skill within five domains of foundational leadership (Table 2).

Implications
To cultivate leadership skills in medical residents, residency faculty need validated instruments to measure progress and to target educational interventions. Such a tool now exists. The FHLS is intended to be a formative tool for resident leadership development within a leadership curriculum. It can inform individualized learning plans and tailored educational experiences. The FHLS can be used to guide residency curricular development for educational needs in a resident cohort.

Strengths
The strength of our study is in the use of both qualitative and quantitative methods to validate the FHLS. In our qualitative analysis, we involved experts and the users of the tool, with input from both groups. This process provided face and content validity for the items presented in the initial tool, as we used the collective expertise of the leaders, and feedback from the intended users. Construct validity was assessed in our quantitative analysis, and resulted in a streamlined tool containing five domains. For construct validity, we had broad geographic representation. This purposive sample targeted residents in all US geographic regions and included respondents in both academic and community-based residencies.

Limitations
The FHLS was initially developed using qualitative methods at a single geographic location. The initial focus group included physician leaders in various leadership positions in the region. Therefore, a bias may have occurred due to the geographic location and practices of the experts who designed and gave feedback on the tool. However, the tool was designed using published materials from a variety of locations, and the group of experts had a collectively broad set of life and leadership experiences. If there were geographic differences, we would expect to see differences in response by geography, which we did not.

References


33. Gillam S. Teaching doctors in training about management and leadership. BMJ. 2011;343;iep313 (3:i5672).


58. Taking You and Your Organization to the Next Level with Quint Studer @ 2005. The Studer Group, LLC.


