Utilizing the F3App to Capture Learner Feedback About Faculty Teaching

Linda Myerholtz, PhD; Hannah M. Baker, MPH; Lisa Rollins, PhD; Cristen P. Page, MD, MPH

BACKGROUND AND OBJECTIVES: The Accreditation Council for Graduate Medical Education requires soliciting learner feedback on faculty teaching, although gathering meaningful feedback is challenging in the medical education environment. We developed the Faculty Feedback Facilitator (F3App), a mobile application that allows for real-time capture of narrative feedback by residents. The purpose of our study was to assess efficacy, usability, and acceptability of the F3App in family medicine residency programs.

METHODS: Residents, faculty, and program directors (PDs) from eight residency programs participated in a beta test of the F3App from November 2017 to May 2018; participants completed pre- and postimplementation surveys about their evaluation process and the F3App. We interviewed PDs, and analyzed responses using a thematic analysis approach.

RESULTS: Survey results showed significant postimplementation increases in faculty agreement that accessing evaluations is easy (42%), evaluations are an effective way to communicate feedback (34%), feedback is actionable and meaningful (24%), and the current system provides meaningful data for promotion (33%). Among residents, agreement that the current system allows meaningful information sharing and is easy to use increased significantly, by 17% each. The proportion of residents agreeing they were comfortable providing constructive criticism increased significantly (22%). PDs generally reported that residents were receptive to using the F3App, found it quick and easy to use, and that feedback provided was meaningful.

CONCLUSIONS: Participating programs reviewed the F3App positively as a tool to gather narrative feedback from learners on faculty teaching.

Faculty evaluation systems serve a variety of functions. They are used for annual reviews, identifying areas for faculty development, informing promotion and tenure activities, determining merit increases and/or the dispensation of rewards (eg, serving as an incentive to faculty), determining funding for individual departments, and even enhancing the visibility of a department's or institution's educational mission.[1-4] Foremost, however, faculty evaluations are used to assess the faculty member's clinical teaching abilities, the improvement of which is paramount in a medical learning environment. As such, the Accreditation Council for Graduate Medical Education (ACGME) requires assessment of faculty clinical teaching abilities during an annual performance review and "written, anonymous and confidential evaluations by the residents."[5]

Multiple approaches have been used to evaluate teaching faculty, including global rating forms, group interviews/focus groups, individual interviews, peer review, simulated teaching encounters or objective structured teaching encounters (OSTE), and creating teaching portfolios. Yet, the culture of medicine poses some challenges to developing effective and fully integrated faculty evaluation systems. While there has been a move to identify more uniform expectations of learners through specialty-based milestones[6] and entrustable professional activities,[7] no such uniform expectations are currently available for clinical instruction, although there are numerous proposals in the literature.[8-10] Beyond the difficulty of achieving valid bidirectional feedback in a hierarchical structure, the particular environment of medical education poses unique challenges; much of the educational process takes place within a service context where patient care takes priority over teaching[11] and the learning process often takes place over a series of brief and discrete encounters that may involve multiple instructors.

From the Department of Family Medicine (Dr Myerholtz), Lineberger Comprehensive Cancer Center (Ms Baker), University of North Carolina Chapel Hill (Dr Page); and the Department of Family Medicine, University of Virginia, Charlottesville, VA (Dr Rollins).

Despite these limitations, resident observations of faculty are shown to be a valid, reproducible measure of teaching quality.[12-14] Some evidence suggests that the best methods of data capture are those that do not significantly affect the flow of work.[4] Further evidence suggests that narrative feedback can enhance quantitative feedback, adding important contextual information.[15] Quantity of narrative feedback can result in a perceived increase in the quality and value of feedback.[16] With this in mind, we sought to evaluate the efficacy, usability, and acceptability of the Faculty Feedback Facilitator (F3App), a mobile application that allows for real-time capture of narrative feedback provided by medical residents in the medical education setting.

Methods

App Functioning and Setting

In 2014, we developed the Mobile Medical Milestones (M3App), which allows faculty and peers to record narrative observations of learner behavior.[16] In 2017, we developed the companion app, F3App, which allows learners and peers to record narrative observations of faculty. F3App was produced by an external contractor, then alpha tested by the Department of Family Medicine at the University of North Carolina at Chapel Hill. Figure 1 illustrates components of the app.

Residents log into the F3App and enter a narrative observation, either at point-of-observation or at a later time. By default, residents' observations are anonymous; however, they can uncheck the default in order to attach their name. Each faculty member can access her/his own F3App observations, which are listed under the date they were entered and either with an anonymous tag or the name of the person entering (Figure 1); residency leadership can also access all faculty observations. The administrative interface allows residency programs to generate summary reports that can be used for faculty evaluation, promotion, and continuous quality improvement in faculty teaching.

We offered the opportunity to beta test the F3App to eight family medicine residency programs with established use of the M3App. We gave program directors the flexibility to implement the F3App in the way they felt most appropriate for their program, but offered them the opportunity to attend bimonthly calls with the study team for idea generation and technical assistance and troubleshooting. All programs accepted the offer. The beta test occurred from November 2017 to May 2018.

Data Collection

We used a pre/post design that incorporated an anonymous online survey of faculty, residents, and program directors (PDs), and semistructured telephone interviews with PDs. We adapted surveys from a previous evaluation of M3App.[16]

Faculty and Program Director Survey

Faculty surveys (also completed by PDs) addressed accessibility, utility, and effectiveness of the current teaching evaluation system on a 5-point Likert scale (strongly disagree to strongly agree); the amount of narrative feedback currently received by faculty and the descriptive information it provides (5-point scale: far too little–far too much); and the degree of importance program leadership attached to evaluation of teaching and improvement of faculty teaching (5-point scale: not at all important–extremely important). Additional items categorized learners taught, length of time teaching, and the amount of time spent teaching or supervising residents. We asked PDs to report numbers of teaching faculty and residents in the program, the systems used to evaluate resident and faculty teaching, and faculty development needs. We also asked them to rate the utility of their current systems regarding assessment of faculty teaching skills and identifying faculty development needs.

The postimplementation survey repeated all of the above items, and was emailed to participants 6 months following implementation of the F3App. The postimplementation survey also included items regarding frequency of asking for feedback from learners and of accessing feedback received, ease of use of the F3App, desire for its continued use, likelihood of recommending it to others, and the importance of the app's anonymous feature. We analyzed survey data from PDs and faculty together, and discussed them as faculty responses.

Resident Survey

Resident preimplementation surveys included the same items assessing program leadership's attitudes about evaluating and improving teaching, and describing the ideal teaching evaluation process as found on the faculty survey. Preimplementation surveys also asked residents for their opinions of the program's current faculty teaching evaluation system regarding ease of use, ability to convey meaningful information about faculty teaching, and their comfort providing constructive criticism (5-point scale: strongly disagree–strongly agree). The survey queried residents about (1) the amount of time spent in evaluating faculty (3-point scale: not enough, about the right amount, too much), (2) the typical time lapse between observing and completing evaluation of faculty (4-point scale: same day–a month or more), and (3) the amount of training received on providing feedback (5-point scale: none at all–a great deal). The residents' postimplementation survey repeated the above items and included a similar block of items to the faculty postimplementation survey. Results from the open-ended items (eg, ideal teaching evaluation process) are reported elsewhere.[17]

Program Director Interviews

In addition to the surveys, one member of the research team (H.B.) conducted semistructured phone

Figure 1: F3App Function and Information Flow (F3App Is an Optional Add-on to the M3App)
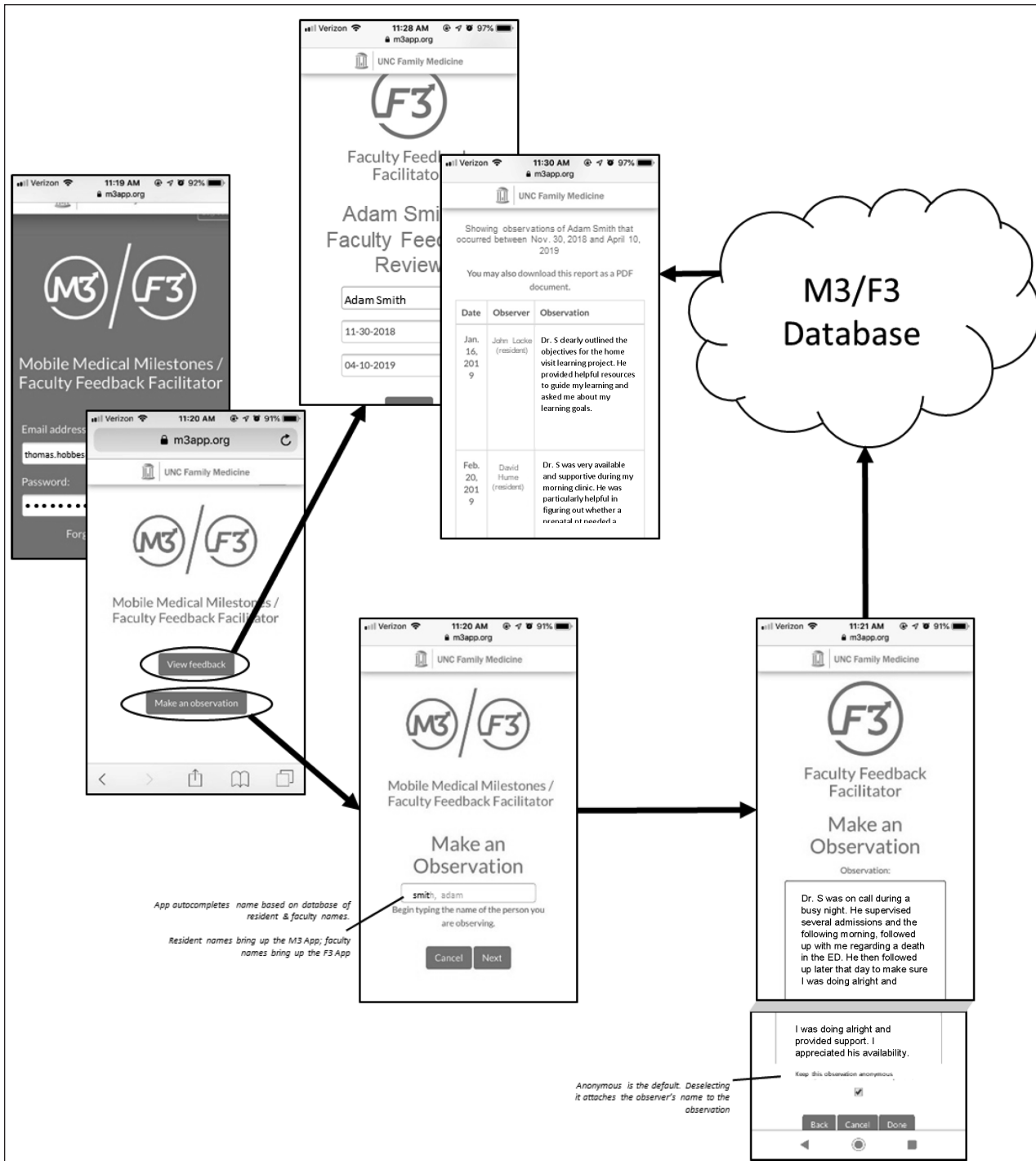
interviews with each PD, using an interview guide developed by a panel of faculty experts in the fields of graduate medical education, behavioral science, and qualitative data collection, and a former residency program director. The interview questions focused on how programs introduced the F3App to residents and faculty, facilitators and barriers to implementation, perceptions of the usefulness of the app, and the general culture of feedback within the program. The interviews were recorded and transcribed verbatim.

Data Analysis
We obtained the number of F3App observations entered by each program during the study period from the F3App database. To account for differences in program size, we calculated a volume score by dividing the total number of a program's

observations by the product of the number of residents and the number of faculty.

For analysis, we dichotomized the agree/disagree items (strongly agree-agree, and disagree-strongly disagree); we excluded neutral responses from the denominator for both the pre- and postdata aggregation. To test pre- and postimplementation differences, we entered survey items as dependent variables in a logistic regression model that included program characteristics of setting (university based or not), program size (more than median number of residents or not), and app use volume (more than median or not). We similarly analyzed items appearing only on postimplementation surveys to assess the effect of these program characteristics, and we compared faculty characteristics using the χ^2 test for independence to assess similarity in pre- and postimplementation samples. We tabulated and computed responses and descriptive statistics in Microsoft Excel 2016; inferential statistical analyses utilized Stata 10.1 (Stata Corp, College Station, TX).

We used a thematic analysis approach to analyze PD interviews.[18] The research team divided the interview transcripts among three members (A.R., L.M., and L.R.) to review. Each member identified codes and themes relevant to their transcripts and then shared those with other team members. The team then met to review emergent themes across transcripts and to discuss any discrepancies in theme identification. The intention of this qualitative approach was not to reach saturation or to quantify how often themes emerged across transcripts, but rather to identify themes that resonated with the overall purpose of the study and that provided insight into the implementation of the app overall.

This study was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill (IRB #17-2052).

Results
All eight PDs responded to both pre- and postimplementation surveys. Of 126 faculty members, 99 (79%) responded to the preimplementation survey and 82 (65%) responded postimplementation. Of 216 residents, 152 (70%) and 101 (47%) respectively responded to pre- and postimplementation surveys. Program characteristics, including setting, teaching evaluation modalities, size, and F3App usage volume are shown in Table 1.

Pre- and Postimplementation Surveys
As shown in Table 2, we found statistically significant postimplementation increases in faculty agreement that accessing evaluations is easy (42% increase), evaluations are an effective way to communicate feedback (34% increase), feedback is actionable and meaningful (24% increase), and the current system provides meaningful data for promotion (33% increase). Agreement regarding use of feedback to make changes in teaching and in the proportion of faculty characterizing the amount of narrative descriptive feedback received as "about right" increased (by 9% and 8%, respectively), although

<div align="center">Table 1: Program Characteristics</div>

	Participating Program (Setting Type)								
	Program A (University Department)	Program B (Community Hospital)	Program C (University Department)	Program D (University Department)	Program E (Community Hospital)	Program F (Community Hospital)	Program G (University Department)	Program H (University Department)	
Evaluation Modalities									
Electronic platform	X	X	X	X	X	X			
Online form		X	X			X	X		
Paper form							X	X	
Resident group discussion						X			
Program Demographics									**Median**
Number of faculty	11	18	40	25	7	15	6	4	**13**
Number of residents	30	39	32	24	18	26	27	20	**26.5**
F3App Usage Volume*									
	0.0665	0.062	0.059	0.084	0.270	0.104	0.0669	0.008	**0.0667**

* Number of observations, adjusted for number of faculty and number of residents

Table 2: Comparison of Faculty and Resident Pre- and Postintervention Survey Responses Accounting for Program Setting, Size, and F3App Usage Volume

Faculty Survey Factors	Preintervention	Postintervention	Model P Value[1]
Number of faculty responses	99	82	n/a
	Percent Who Agree		
Accessing evaluations is easy and straightforward.*	26	68	<.0001
Evaluations are an effective way to communicate feedback.*	30	64	.0011
Feedback I receive is actionable and meaningful.*	36	60	.037
I use feedback to make changes in my teaching.	77	86	.46
Current system provides meaningful data for promotion.*	39	72	.016
Percent characterizing feedback system as providing about the right amount of narrative, descriptive feedback.	20	28	.7
	Percent Responding Very or Extremely Important to Program Leadership		
How important is the faculty teaching evaluation process?	44	61	.18
How important is improving teaching among faculty?	70	78	.65
Resident Survey Factors			
Number of resident responses	152	101	
	Percent Who Agree		
Current system allows me to share meaningful information*	71	88	.013
I feel comfortable providing constructive criticism.*	55	77	.001
The evaluation platform is easy to use.*	65	82	.033
Characterizing the amount of time spent completing faculty evaluations as "about the right amount.	44	44	.33
Providing feedback to faculty in 1 week or less.*	21	38	.0042
Reporting receiving little to no training in providing meaningful feedback.*	73	49	<.0001
	Number Responding Very or Extremely Important to Program Leadership ...		
How important is the faculty teaching evaluation process?	64	72	.122
How important is improving teaching among faculty?	75	75	.628

1. Logistic regression model including program setting (university based or not), program size (> median number of residents or not), F3App usage volume (> median or not)

* Items with significant differences in pre/postintervention

these differences were not statistically significant. Likewise, faculty perception that program leadership considered the faculty teaching evaluation process and the improvement of teaching as very or extremely important also increased (by 17% and 8%, respectively), but not significantly. The proportion of very or extremely responses regarding importance of improving teaching were 70% or greater both pre and post. There was no significant independent relationship in any of the above measures with practice setting, program size, or F3App usage volume.

On the postimplementation survey, one-quarter of faculty members reported asking for feedback frequently from learners, although 88% of faculty agreed that they were comfortable asking for feedback—all independent of program setting, program size, or app usage volume (Table 3). One-quarter of faculty also reported that they accessed their own feedback.

Among residents, agreement that the current system allows meaningful information sharing and is easy to use, as well as the proportion who reported providing feedback to faculty in a week or less increased significantly (each by 17%), independent of program setting, size, and app use volume. The proportion agreeing they were comfortable providing constructive criticism increased

Table 3: Postimplementation Faculty and Resident Assessment of F3App
(Faculty Responses, n=82, Resident Responses, n=101)

Factor	Faculty Agree	Residents Agree	Model *P* Value[1]
Since using the F3App asked for feedback frequently[2]	25	n/a	n/a
Since using the F3App accessed own feedback[2]	25	n/a	n/a
I feel comfortable asking for feedback on my teaching from learners.[2]	88	n/a	n/a
I feel comfortable providing feedback to faculty members on their teaching.[3]	n/a	90	n/a
F3App is easy to set up.*	94	80	.045
Learning to operate F3App was easy.	97	92	.40
F3App has a clear, uncluttered screen design.	97	91	.18
The F3App menu items are well organized and functions are easy to find.	95	92	.49
Information presented on F3App screens is easy to comprehend.	96	94	.48
I would like my program to continue using F3App to evaluate faculty members.*	94	81	.04
I would recommend F3App to a colleague.	95	88	.13
F3App's anonymous feature is very or extremely important.*	42	74	<.0001

* Item has significant difference

1. Logistic regression model including program setting (university-based or not), program size (> median number of residents or not), F3App usage volume (> median or not)

2. Item included on faculty survey only

3. Item included on resident survey only

significantly (22%) and the proportion of residents who reported receiving little or no training in providing meaningful feedback decreased significantly (24%). Residents' perception of the importance to program leadership of the teaching evaluation process or on improving faculty teaching did not differ significantly from pre to post. Nor was there any significant difference in the proportion of residents responding that the time spent on faculty evaluations was about right.

Postimplementation Faculty-Resident Comparisons
As shown in Table 3, proportions of faculty and residents agreeing differed significantly on only two items. Although substantial majorities of both faculty and residents agreed that F3App was easy to set up, this was true for a significantly smaller proportion of residents. Fewer than half of faculty indicated that F3App's anonymous feature was important,

compared to nearly three-quarters of residents.

Independent Effects of Program Characteristics
Table 4 shows results from the logistic regression analysis indicating statistically significant relationships between program characteristics and items on faculty and resident surveys, independent of whether survey responses were pre- or postintervention. Despite an overall increase in the likelihood that residents provided feedback in a week or less (from point of observation), residents from larger and higher volume programs were less likely to do so. Residents from university-based programs were less likely to report receiving little to no training in providing meaningful feedback either pre- or postimplementation. Resident perceptions of the importance program leadership attached to faculty teaching evaluation and improving faculty teaching did not change postimplementation; however, residents from

larger and higher volume programs were more likely to indicate that the faculty teaching evaluation process was important to program leadership.

Although we found no overall difference between residents and faculty on programs' continued use of F3App, respondents from larger programs were five times as likely to favor it. Finally, although residents and faculty differed significantly regarding the importance of anonymity in F3App, in university programs and higher volume programs, both faculty and residents were significantly more likely to consider it important.

Postimplementation PD Interviews
PDs generally reported that they felt residents were receptive to using the F3App and that they felt residents found the F3App to be quick and easy to use. PDs reported being less certain about how faculty perceived the F3App, usually because they had

**Table 4: Multivariate Comparisons of Resident and Faculty Responses
Showing Significant Effects of Program Characteristics**

	Statistically Significant Odds Ratios and 95% Confidence Intervals				
Comparing Pre- to Postimplementation	**Postimplementation: All Respondents**	**Setting (University)**	**Size (> Median)**	**Volume (> Median)**	**Model *P* Value[1]**
Residents providing feedback to faculty in 1 week or less	2.4 (1.3, 4.3)	n/a	0.36 (0.16, 0.80)	0.32 (0.12, 0.87)	.0042
Residents reporting receiving little to no training in providing meaningful feedback	0.38 (0.22, 0.37)	0.31 (0.14, 0.62)	n/a	n/a	<.0001
How important is the faculty teaching evaluation process?	n/a	n/a	2.6 (1.2, 5.7)	3.1 (1.2, 7.9)	.032
How important is improving teaching among faculty?	n/a	n/a	n/a	2.8 (1.02, 7.8)	.009
Comparing Faculty to Residents, Postimplementation	**Postimplementation: Faculty Respondents Only**	**Setting (University)**	**Size (> Median)**	**Volume (> Median)**	**Model *P* Value[1]**
Faculty agreeing that the program should continue using F3App to evaluate faculty members	n/a	n/a	5.2 (1.2, 22.9)	n/a	.036
Faculty agreeing F3App's anonymous feature is very or extremely important	0.10 (0.04, 0.27)	3.0 (1.1, 8.3)	n/a	9.4 (1.7, 53.7)	<.0001

[1] Logistic regression model including program setting (university-based or not), program size (> median number of residents or not), F3App usage volume (> median or not)

not asked faculty directly. From the PD perspective, feedback provided was meaningful, eg:

> …most of the F3 apps I saw were not, just well let me put something down to get it done, but they were very thoughtful and I think when you see any kind of a personal narrative where a person has put some thought and energy into it, it's, it can be deeply moving.

In terms of implementation challenges, some PDs assumed that the F3App rollout would take less effort due to programmatic experience with the M3App. Thus, they did not follow up with residents or faculty regarding implementation. One PD commented that "…residents really like M3, maybe they'll just sort of latch onto F3 and just do that, but that plan did not work." PDs also noted a common challenge among their residents of providing constructive feedback due to both the skill needed to do this effectively and comfort in doing so. Another common barrier PDs

reported was residents feeling overburdened:

> …unless you're telling them all the time to do it, they've got more important things to do than just that. Residents are busy trying to remember all of the rules (ACGME, local hospital, Medicaid) and trying to learn medicine, and have administrative stuff to do (required to be on committees in the hospital and in clinic) and they are on stimulus overload.

Programs used a variety of strategies to introduce the F3App, including face-to-face training sessions and designating an F3 champion. Strategies that helped increase adoption of F3App included ensuring all residents had the F3App loaded on their phones, setting aside time in meetings for residents to record observations, a multifaceted reminder system for residents (ie, screen savers on shared computers, email reminders, etc), and encouraging faculty to solicit feedback. A few also attempted to give residents

expectations on the number of observations they should record, although tracking this was difficult due to the default anonymous feature.

PDs recommended keeping the anonymous feature, though concerns about the ability to maintain anonymity remained among smaller programs:

> We are such a small program that the residents didn't know that it would indeed be anonymous. Talked with the residents about it several times but residents [were] still concerned that they could be identified.

Discussion
It is widely accepted that feedback enhances learning beyond mere trial and error. Yet, feedback is far more complex than reported in the earlier literature.[19] The feedback process is widely affected by attitudes and context of both the giver and the receiver as well as the broader context and culture within the residency program. The intent of this study was to evaluate program perceptions and use of a mobile interface (F3App) to

provide narrative feedback to family medicine faculty.

The resident and faculty response to the F3App was markedly positive in regard to the various aspects of app functionality. Similarly, the majority of residents and faculty reported that they "would like their program to continue using F3" and "would recommend F3App to a colleague." In addition, among faculty we found statistically significant pre/post differences that faculty perceived the feedback they received was more actionable and meaningful and their feedback system provided more meaningful data for promotion. Residents noted that F3App allowed them to share meaningful information, and results indicate that they received training in providing meaningful feedback during F3App implementation. It is widely recognized that training improves the specificity of resident feedback,[20] and that the vast majority of feedback from learners to faculty tends to be positive rather than constructive in nature.[15] Thus, implementation of the F3App may have improved the utility of feedback given.

The greatest difference between faculty and residents pertained to the issue of anonymity, with a larger percentage of residents reporting the importance of this app feature compared to faculty. This finding is not surprising considering the traditional power differential between residents and faculty. Indeed, some research even suggests that anonymous feedback may be a more accurate reflection of teaching abilities.[21] However, such concerns could be mitigated by certain program characteristics, including programmatic emphasis on the importance of feedback, training on how to give feedback, and an overall programmatic culture of feedback-giving/receiving. These programs in turn tended to be higher utilizers of the app and were relatively less focused on anonymity, although anonymity remained important to the vast majority of residents.

Of particular note, at postimplementation, only 25% of faculty reported that they accessed their own feedback in the app even though a majority (60%) reported that their feedback was actionable and meaningful. We do not know if this is because PDs were providing this information to faculty directly, if residents were supplementing their written comments with verbal comments, or if faculty were basing this response on sources of feedback beyond the F3App. This finding is interesting when considered alongside the lack of pre/post significance across two items in the faculty survey: "I use feedback to make changes in my teaching," and "How important is improving teaching among faculty?" It is possible that a ceiling effect exists among participating faculty regarding their use of feedback to improve teaching. These findings imply a need to promote a sense of continuous improvement in teaching in addition to other areas (eg, clinical skills, continuing medical education, etc). To facilitate faculty acquisition of feedback, we enhanced our login interface to include an option for faculty to obtain their data. Additionally, we are piloting a program that distributes learners' quarterly and annual feedback, which we hope to implement for faculty as well.

Study Limitations

This study only targeted family medicine programs, thus these results may not generalize to other medical specialties. The study focused on residency education and did not include undergraduate medical education teaching. Participating programs volunteered to participate, which could lead to selection bias. Our survey was anonymous, precluding pairing of pre/post responses; however, we found no difference in characteristics of faculty from pre- and postimplementation surveys, suggesting that our samples of faculty members were comparable between the two surveys. The postimplementation response rate was lower, particularly for residents. This may have been due in part to the fact that the postimplementation surveys were distributed in June, which is a particularly busy time of the academic year in residency programs. In addition, this study involved a relatively small number of programs (eight) and while they varied in terms of location, program setting (university department or community hospital), and size, results may not be generalizable to other family medicine programs.

Overall, the F3App was reviewed positively by participating programs. Future studies should examine contexts in which residents decide to intentionally remove the anonymity feature from their feedback. Other areas for future study include assessing the quality and content of feedback from learners, as we know that it may vary depending on time of year given, level of learner, etc.[22] Finally, it may be important to better define expectations around teaching and how faculty and programs can translate feedback into improvements in teaching performance.

CONFLICT DISCLOSURE: Cristen Page, a coinvestigator on this study, serves as chief executive officer of Mission3, the educational nonprofit organization that has licensed from UNC the tool, the F3App, evaluated in this study. If the technology or approach is successful in the future, Dr Page and UNC Chapel Hill may receive financial benefits.

CORRESPONDING AUTHOR: Address correspondence to Dr Linda Myerholtz, 590 Manning Drive CB# 7595, Chapel Hill, NC 27599-7595. 919-962-4764.Fax: 919-843-3418. Linda_Myerholtz@med.unc.edu.

References

1. Bland CJ, Wersal L, VanLoy W, Jacott W. Evaluating faculty performance: a systematically designed and assessed approach. Acad Med. 2002;77(1):15-30.

2. Geraci SA, Kovach RA, Babbott SF, et al. AAIM Report on Master Teachers and Clinician Educators Part 2: faculty development and training. Am J Med. 2010;123(9):869-872.e6.

3. Mallon WT, Jones RF. How do medical schools use measurement systems to track faculty activity and productivity in teaching? Acad Med. 2002;77(2):115-123.

4. Snell L, Tallett S, Haist S, et al. A review of the evaluation of clinical teaching: new perspectives and challenges. Med Educ. 2000;34(10):862-870.

5. Accreditation Council for Graduate Medical Education. ACGME Common Program Requirements (Residency) Section I-V. http://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRs_2017-07-01.pdf. Accessed January 30, 2020

6. Accreditation Council for Graduate Medical Edu cation. Milestones. 2018. https://www.acgme.org/What-We-Do/Accreditation/Milestones/Overview. Accessed January 30, 2020.

7. Association of American Medical Colleges. The Core Entrustable Professional Activities (EPAs) for Entering Residency. 2018. https://www.aamc.org/initiatives/coreepas/. Accessed January 30, 2020.

8. Harris DL, Krause KC, Parish DC, Smith MU. Academic competencies for medical faculty. Fam Med. 2007;39(5):343-350.

9. Srinivasan M, Li STT, Meyers FJ, et al. "Teaching as a Competency": competencies for medical educators. Acad Med. 2011;86(10):1211-1220.

10. Colletti JE, Flottemesch TJ, O'Connell TA, Ankel FK, Asplin BR. Developing a standardized faculty evaluation in an emergency medicine residency. J Emerg Med. 2010;39(5):662-668.

11. Watling C, Driessen E, van der Vleuten CPM, Vanstone M, Lingard L. Beyond individualism: professional culture and its influence on feedback. Med Educ. 2013;47(6):585-594.

12. Donnelly MB, Woolliscroft JO. Evaluation of clinical instructors by third-year medical students. Acad Med. 1989;64(3):159-164.

13. Irby D, Rakestraw P. Evaluating clinical teaching in medicine. J Med Educ. 1981;56(3):181-186.

14. Benbassat J, Bachar E. Validity of students' ratings of clinical instructors. Med Educ. 1981;15(6):373-376.

15. van der Leeuw RM, Overeem K, Arah OA, Heineman MJ, Lombarts KMJMH. Frequency and determinants of residents' narrative feedback on the teaching performance of faculty: narratives in numbers. Acad Med. 2013;88(9):1324-1331.

16. Page C, Reid A, Coe CL, et al. Piloting the Mobile Medical Milestones Application (M3App): a multi-institution evaluation. Fam Med. 2017;49(1):35-41.

17. Myerholtz L, Reid A, Baker HM, Rollins L, Page CP. Residency Faculty Teaching Evaluation: What Do Faculty, Residents, and Program Directors Want? Fam Med. 2019;51(6):509-515.

18. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77-101.

19. Ende J. Feedback in clinical medical education. JAMA. 1983;250(6):777-781.

20. van der Leeuw RM, Schipper MP, Heineman MJ, Lombarts KMJMH. Residents' narrative feedback on teaching performance of clinical teachers: analysis of the content and phrasing of suggestions for improvement. Postgrad Med J. 2016;92(1085):145-151.

21. Afonso NM, Cardozo LJ, Mascarenhas OA, Aranha AN, Shah C. Are anonymous evaluations a better assessment of faculty teaching performance? A comparative analysis of open and anonymous evaluation processes. Fam Med. 2005;37(1):43-47.

22. Shea JA, Bellini LM. Evaluations of clinical faculty: the impact of level of learner and time of year. Teach Learn Med. 2002;14(2):87-91.