Development and Field Pretest of the Feedback Quality Index: A Tool for Assessing Feedback Frequency and Quality in a Clinical Setting

Michael FitzGerald*,[1] and Corinne Lehmann[2]

[1]*Cincinnati Children's Hospital Medical Center, Division of Emergency Medicine, Cincinnati, Ohio, USA*

[2]*Cincinnati Children's Hospital Medical Center, Division of Adolescent Medicine, Cincinnati, Ohio, USA*

Abstract: *Objectives*: This paper describes the systematic development, design, and field pretest of the Feedback Quality Index (FQI) - a brief survey for assessing feedback regularity and quality in a clinical setting. Medical educators need this type of tool so they can identify specific problems related to feedback provision and evaluate the impact of efforts designed to address those problems. The purposes of this study are to provide evidence for the feasibility and validity of the tool and to provide guidance and suggestions regarding its use.

Methods: Decisions regarding the design and implementation of the FQI were guided by a systematic approach to survey development to ensure that common sources of non-sampling error were identified and addressed. A field pretest implementation was conducted to gather evidence regarding the feasibility of the tool and the validity of the data. Feasibility was assessed based on the effort needed to administer, complete, and analyze the FQI while evidence for validity was based on an analysis of question quality.

Results: Field pretest results indicated that the FQI can be feasibly administered, completed, and analyzed. An analysis of question quality revealed that most questions were understood correctly by respondents and the level of detail they provide in describing the feedback received provides evidence for the accuracy of their recollections.

Conclusions: The current version of the FQI is a useful tool that programs could use to assess feedback frequency and quality, identify specific problems, and evaluate efforts to address those problems. Additional studies need to be conducted to further assess and improve the validity of the FQI. Such efforts will not only improve the accuracy of the FQI but could also enhance our conceptual understanding of what constitutes effective feedback in the clinical setting.

Keywords: Feedback, educational assessment, education, graduate medical, clinical teaching, ACGME.

BACKGROUND AND RATIONALE

Numerous commentaries on the importance of feedback [1] and suggestions for improving its frequency and quality [2, 3] exist in the medical education literature. But effective improvement depends upon the accurate identification of specific problems and the accurate assessment of any changes in desired outcomes. The existing literature on feedback in medical education might provide some helpful guidance to medical educators, but there is currently no existing method or tool that can be used to feasibly and accurately assess feedback frequency and quality in the clinical setting. Consequently, medical educators are currently unable to effectively target or evaluate their improvement efforts [4].

Methods used in the handful of existing studies tend to be too impractical or error-prone for use in improvement efforts. Salerno, Jackson, and O'Malley [5] used live audio recordings of teaching encounters to assess changes in the frequency and content of feedback "utterances" before and after a faculty development workshop. They were able to capture accurate and detailed information about feedback provided in those instances but this, and any live recording method, is too resource intensive for most institutions and difficult to implement because of privacy protection issues.

Surveys, a more feasible alternative to live recording, have been used in a few studies but most have required the learner to recall feedback received over a long period of time (e.g., an entire rotation). As a result, the data from these surveys are prone to a number of recall-related errors including the failure to encode the information in the first place (i.e., encoding error), the failure to completely recall information (i.e., recency or primacy error), and recollection of erroneous information (i.e., halo or devil effect).

Many of the surveys used in prior studies also lack the data precision necessary for drawing conclusions about feedback regularity and quality because they require respondents to report on these in ambiguous terms. A good example comes from a study by Gil, Heins, and Jones [6] that is frequently cited as evidence for the poor state of feedback in medical education. In their study respondents were asked to recall the frequency of feedback received using response options ranging from *never* to *always* and asked to describe the *overall* quality of that feedback. But respondents likely differ in their perceptions of what it means to *always* receive feedback and what constitutes the *overall* quality of feedback. Consequently, one can't draw any reasonably specific

*Address correspondence to this author at the Cincinnati Children's Hospital Medical Center, Division of Emergency Medicine, 3333 Burnet Avenue, MLC 2008, Cincinnati, OH, 45229-3039, USA; Tel: (513) 636-1373; Fax: (513) 636-7967; E-mail: mike.fitzgerald@cchmc.org

or confident conclusions about the state of feedback regularity and quality from their data.

Kogan and Hauer [7] made a good attempt to address these recall and precision problems when they asked medical students to recall the actual number of episodes of feedback received *that day* as well their satisfaction with its *specificity*. But recall inaccuracy is still a problem in their study because it would be difficult for anyone to remember – even over the course of a day - the actual number feedback of episodes they received. Their data also lack precision because feedback quality is only assessed in terms of its *specificity* and neglects other important elements of effective feedback (e.g., whether or not it was based on direct observation).

This paper describes the systematic development and field pretest of the Feedback Quality Index (FQI): a new tool designed to provide medical educators with a feasible means for obtaining a reasonably accurate and precise measure of both feedback regularity and quality in a clinical setting. This initial field pretest was conducted to answer the following questions:

Question 1 – Is the FQI feasible to implement, complete, and evaluate?

Question 2 – Is there any evidence for the validity of the data captured with the FQI?

In addition, the field pretest data captured with the FQI are provided to serve as an illustration of the information that can be obtained and reported.

This field pretest is an important first step toward developing a tool that will enable medical educators to more precisely target and evaluate efforts to improve feedback provision in their programs.

DEVELOPMENT AND IMPLEMENTATION OF THE FQI

The Feedback Quality Index (FQI) and implementation process were developed using a stepwise survey design approach described by Biemer and Lyberg [8] to ensure that common sources of non-sampling error were proactively addressed. Such sources of error include specification errors, frame errors, nonresponse bias, and measurement errors which are among the various threats to validity that medical education researchers have been urged to address [9].

Specification of FQI Objectives and Corresponding Questions

A critical first - and frequently neglected - step in the survey design process is to specify the full range of conceptual objectives the survey should address [8]. This reduces the risk of *specification errors* that occur when the questions included in a survey fail to provide the information needed for its intended purpose or use. This section specifies each objective of the FQI, describes the underlying concept, and provides the rationale for including it. A summary of these objectives, including the specific question(s) that correspond to each, are provided in Table **1**.

The initial questions were developed by the primary author who has a background in survey methodology and experience in the development and implementation of surveys in organizational settings. Four other faculty members,

including the co-author and two residents, reviewed iterative drafts to ensure that the questions were clearly stated and fully addressed the specified objectives. After a number of revisions, the FQI consisted of 18 closed-ended and 6 open-ended questions. (See the Appendix for the full version).

The first objective of the FQI was to obtain some clear and precise indication regarding the regularity of feedback residents receive from faculty. Starting with the widely shared view that - at its core - feedback is *any information about performance*, the FQI is designed to assess whether or not residents are receiving information about their own performance from attending physicians or fellows during a given shift. Since a survey can't capture every instance of feedback, respondents are asked whether they received *any* instance of corrective or positive feedback during their most recent shift (see Q's 1.1 and 2.1 in Table **1**).

The FQI is designed to ensure that respondents consider and describe instances of both *corrective feedback* and *positive feedback* separately. Corrective feedback targets ineffective behaviors for the purpose changing those behaviors, while positive feedback targets effective behaviors for the purpose of ensuring that those behaviors continue. Residents need both but evidence suggests that positive feedback tends to occur more frequently than corrective feedback [7] and faculty have reported a greater reluctance to provide corrective feedback [6]. While the conceptual elements underlying the effectiveness of corrective and positive feedback are similar (e.g., specificity), it is important – from an assessment and improvement standpoint – to ask about the occurrence of each separately to ensure that the frequency estimates are as accurate as possible and so that interventions to improve feedback frequency and quality can be appropriately targeted to one, the other, or both.

In terms of an outcome variable, data from these questions serve as an indicator of feedback regularity by providing an estimated proportion of cases in which residents received *at least one instance of corrective feedback* during their most recent shift and an estimated proportion of cases in which residents received *at least one instance of positive feedback* during their most recent shift.

The second objective of the FQI was to assess the quality of the feedback being provided. The following is a description of the various elements of effective feedback that the FQI is designed to assess.

The first element is whether or not the feedback was *based on direct observation*; individuals are more likely to accept and respond positively to feedback when it comes directly from the person who observed the individual [10]. Feedback based on indirect or second-hand information can be problematic because it tends to draw the receiver's immediate attention to distracting thoughts about the accuracy of the information and to the identity, credibility and motives of the source [11]. In the case of corrective feedback, these thoughts can quickly become doubts that provoke a defensive reaction and a rejection of the feedback. Such defensiveness might not occur when the feedback is positive, but the receiver might still be distracted by similar thoughts about the identity and motives of the second-hand source. Questions 1.5a and 2.5a of the FQI (see Table **1**) were in-

Table 1.   **FQI Objectives with Corresponding Feedback Elements and Questions**

FQI Objective	Corresponding feedback element	Corresponding FQI question (Q#)
Objective #1 - Determine if residents are regularly receiving feedback from faculty.	Frequent	During your last shift did you receive any feedback from an attending or fellow on something you could improve? (1.1) During your last shift did you receive any feedback from an attending or fellow on something you did well? (2.1)
Objective #2-Assess the quality of the feedback residents are receiving.	Based on direct observation	Was the feedback based on something they observed directly? (1.5a & 2.5a)
	Focused on specific behaviors	Did they describe SPECIFICALLY what you did that needs improvement? (1.5b) Did they describe SPECIFICALLY what you did well? (2.5b) Please briefly describe what they said you did. (1.5c & 2.5c)
	Clarity regarding the importance (motivational factor)	Did they describe the potential impact of what you did? (1.5d) Did they describe the positive impact of what you did well? (2.5d)
	Opportunity to respond	Did they provide an opportunity for you to respond to the feedback? (1.5e)
	Includes improvement guidance	Did their feedback include any guidance or discussion on how to improve? (1.5f)
Objective #3-Assess residents' perceptions of the utility of the feedback they received	NA	Overall, how would you describe the usefulness of the feedback they provided? (1.6 & 2.6) "What about the feedback was particularly useful to you?" (1.7 & 2.7) "What, if anything, would have made the feedback more useful?" (1.8 & 2.8)
Supplemental FQI questions	Solicited vs. unsolicited	Did you actively solicit this feedback? (1.2 & 2.2)
	Verbal vs. written	How did they provide the feedback? (1.3 & 2.3)
	Timeliness	When did they provide the feedback? (1.4 & 2.4)

cluded to determine if the feedback was based on direct observation.

The next, and perhaps most critical element of effective feedback, is that it should *focus on specific observable behaviors* and avoid any mention of personal traits or characteristics. In their frequently cited meta-analysis, Kluger and Denisi [12] concluded that feedback on specific behaviors has a more positive effect on performance because it directs the receiver's attention outward to the task and not inward to the self. Much like reactions to feedback from indirect sources, feedback that draws the learner's focus inward will be distracting and possibly give rise to defensive thoughts and actions. Questions 1.5b and 2.5b were included to assess whether or not the feedback included a description of specific behaviors.

In addition to specificity, the behaviors targeted for feedback should also be *within the learner's control, limited in scope* [13], and *directly related to clear learning objectives or performance expectations* [10]. Instead of adding new questions to assess each of these elements separately, open-ended questions 1.5c and 2.5c were included. These ques-

tions ask respondents to recount the specific behaviors targeted in the feedback they received. Then, from these open-ended responses, a determination could be made about whether or not the behavior was something controllable, limited in scope, or related to learning objectives. While this does increase the effort required for data analysis, it also minimizes the number of questions thereby reducing respondent burden which is a common source of inaccurate or incomplete data [14]. In addition, the open-ended responses also serve as an important check on the validity of the claim that specific feedback was received.

The next element of effective feedback targeted in the FQI is the supported view that feedback is more likely to be accepted and used if the recipient *understands and agrees that it will result in important positive outcomes* [13]. Consequently, feedback messages should include some description of the positive impact of the desired behavior. In the case of corrective feedback, it might also be important to specify the negative consequences of the observed behavior since people are often unaware of the consequences of their

actions [15]. Questions 1.5d and 2.5d were included to assess this element of effective feedback.

To best ensure that the feedback message adequately addresses each of the previously mentioned elements it is important for the deliverer to give the recipient *an opportunity to respond*. Corrective feedback, no matter how well delivered, is likely to illicit some defensiveness. Learners will be more likely to accept and use it when they have the opportunity to clarify – or even challenge – the observations and judgments of the feedback provider [10, 11]. Question 1.5e was included to assess this element.

A few supplemental or optional elements that have an unclear relationship to feedback effectiveness were also included in the FQI. Questions 1.2 and 2.2 were included to determine if the feedback being reported was *solicited or unsolicited*. Research on feedback seeking suggests that a number of intrapersonal and interpersonal factors prevent learners from actively seeking feedback - particularly corrective feedback [16] - but the advantages of one over the other aren't clear. Perhaps residents find feedback more useful when they request it themselves. Another possibility is that attending physicians might be more likely to provide thoughtful and specific feedback when it's solicited. Including these questions might help shed light on these possibilities. In addition, these questions would also enable medical educators to assess the effectiveness of interventions designed to encourage residents to solicit feedback rather than wait to receive it.

Questions 1.3 and 2.3 were included to determine if the feedback was provided in *verbal or written* form. A number of institutions have developed processes that encourage faculty to provide regular feedback to residents in written form, but little is known about the relative advantages or disadvantages of this kind of intervention in terms of its impact on feedback regularity and quality.

Question 1.5f was included to gain a better understanding of the role or importance of *guidance in corrective feedback*. Learners might appreciate guidance as a follow-up to corrective feedback [10] but there is no compelling evidence that its omission undermines feedback effectiveness. Some evidence even suggests that including guidance might be risky because the feedback provider might not understand the underlying causes of the performance problem (e.g., mental health issues) and might provide guidance that is ineffective at best [12].

A few elements that have a clear relationship to feedback effectiveness were excluded from the FQI. Factors including *trust* [17], the recipient's perceptions of the *source's intentions* [18], and *cultural norms and expectations* [19] have been shown to have a mediating impact on feedback effectiveness. However, it was felt that many residents would either be too uncomfortable or unable to provide accurate responses to these questions.

It should be noted that the FQI is designed as an *index* that provides information about the presence or absence of various elements of feedback rather than a *scale* that combines items into an overall score of *feedback quality*. A better understanding of the full range of elements of effective feedback would be needed to produce a valid scale. Consequently, the third objective for the FQI was to include questions to assess residents' overall perceptions of the utility of the feedback that could serve as a dependent variable for analyzing the relationships of the various elements to feedback effectiveness. Open-ended questions 1.7, 1.8 and 2.7, 2.8 were included for this purpose and for the purpose of identifying any additional elements that should be incorporated into the FQI.

TARGET POPULATION

Since the FQI is designed to estimate the proportion of cases in which residents received feedback during their most recent shift, the target population for the FQI is every resident/shift pairing over a specified period of time. The specific target population for the field pretest was the 910 resident/shift pairings of categorical pediatric residents in the Emergency Department at Cincinnati Children's Hospital Medical Center (CCHMC) over a five-month period of time. PGY1 residents were excluded because they were participating in a new curriculum that included observation-based assessments in the department; we wanted to test the tool with a population that was not receiving feedback as a formal part of the curriculum. The five-month time span was chosen since patient volumes-which might affect feedback regularity and quality-tend to be higher in certain months than others.

SAMPLING APPROACH

For the field pretest, a two-stage random sampling procedure was used to ensure the data represented the entire five-month period from September into February and included responses from a broad range of individual residents. For the first-stage sampling unit, the five months were divided into 10 two-week blocks. From those, 10 day and evening shifts were randomly selected within each block for a total of 70 sampled shifts over the five-month period. It should be noted that this sample size was determined by the researchers to be sufficient for assessing question quality and was not based on any desired margin of error or confidence level. For the second-stage sampling unit, one resident was randomly selected within each of the randomly selected shifts. Residents within a given shift who had already completed the FQI were not included so that question quality and response rate data were not biased by factors related to individual residents and prior participation.

This two-stage sampling procedure was considered the best option for the purpose of the field pretest (i.e., testing feasibility and question quality). The downside of such cluster sampling procedures, however, is the complexity involved in determining appropriate sample sizes to obtain estimates based on desired levels of error and confidence.

Since the FQI is intended to be a feasible method for medical educators to obtain good estimates of feedback regularity and quality, systematic sampling should be used as an alternative in practice. Systematic sampling involves the selection of every k^{th} unit from an ordered population where k (i.e., the sampling interval) is equal to the total population N divided by the sample size n. Fig. (**1**) depicts how systematic sampling would have been used in the context of administering the FQI for the field pretest. Its simplicity and potential for greater accuracy are the main advantages over simple random sampling, but there are a number of cautions that
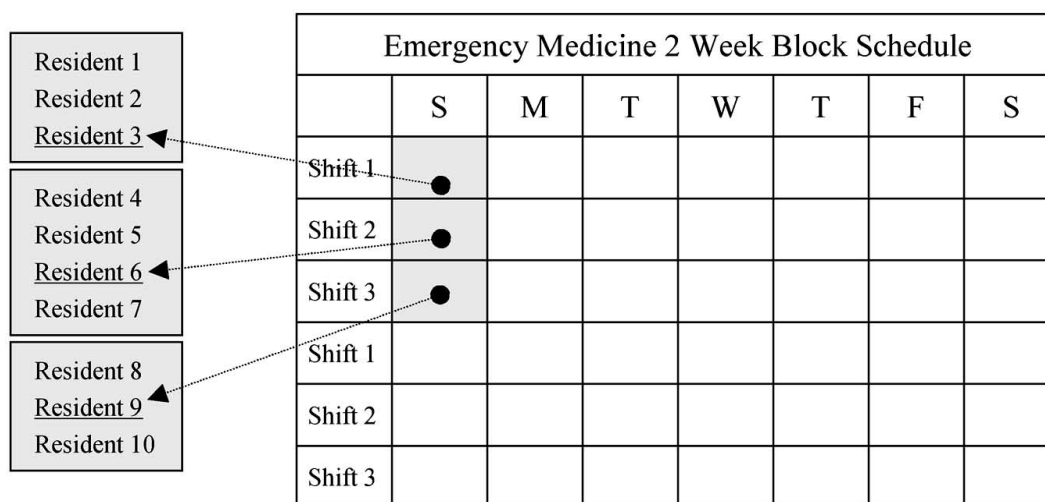
Emergency Medicine 2 Week Block Schedule							
	S	M	T	W	T	F	S
Shift 1	●						
Shift 2	●						
Shift 3	●						
Shift 1							
Shift 2							
Shift 3							

Resident 1
Resident 2
Resident 3

Resident 4
Resident 5
Resident 6
Resident 7

Resident 8
Resident 9
Resident 10

Fig. (1). Hypothetical illustration of systematic sampling in the context of the field pretest population – Assuming the same population (i.e., 910 shifts over a five-month period) a sample size of 270 would have been needed to obtain an estimated proportion with a 5% margin of error at a 90% confidence level. The sampling interval would've been 3.37 (910/270) which could be round down to 3 to keep it simple and ensure the required sample size is obtained. Then – as illustrated above – every third resident would've been selected for participation over the five-month period.

users should consider when using systematic random sampling [20].

MODE OF ADMINISTRATION

The FQI was administered using SurveyMonkey [21], a web-based survey application. Email requests containing a direct link to the survey were automatically delivered to each sampled resident at the end of the shift. To enhance response rates residents were offered a token incentive worth five dollars for completing the survey within 48 hours. In practice, the FQI should be administered using a computer or web-based survey application to take advantage of various functions that minimize respondent burden (e.g., question routing) and maximize data integrity (e.g., response validation).

FIELD PRETEST MEASURES

The feasibility of the FQI was evaluated by assessing the time required for respondents to complete the form and the time required to administer the form, analyze the data, and generate reports. The validity of the FQI was evaluated through an assessment of question quality using a method commonly used in the pretesting phase of survey development to identify questions that are unclear, prone to misinterpretation, or frequently unanswered [22].

FIELD PRETEST RESULTS

Feasibility of the FQI

Tasks related to sampling and participant notification required an average of 50 minutes per week over the five month period. Tasks related to data analysis, coding, and report generation took approximately three hours. Residents took an average of 2:27 to complete the survey with minimum and maximum times of 0:26 and 6:46 respectively.

Validity of the FQI

The field pretest response rate for the FQI was 63% with 44 of the 70 residents responding. The adequate response (AR) percentages for the FQI questions - our measure of

question quality - are presented in Table **2**. A response was determined to be inadequate if the respondent selected or wrote "didn't know" or "can't recall" (coded DK), gave an irrelevant answer (coded IA), or refused to answer (coded RF). As a general standard, questions with an adequate response (AR) rate of less than 85% should be revised or eliminated [22]. All other responses were coded as an adequate response (AR).

All closed-ended questions except Q 2.5d resulted in an adequate response (AR) rate above 85% with over half at 100%. The majority of responses to both open-ended question 1.5c (92%) and 2.5c (94%) included a detailed description of the specific behavior that was targeted for feedback. The open-ended follow-up questions 1.7 and 2.7, designed to identify aspects of feedback that were particularly useful to respondents, resulted in AR rates of only 35% and 15% respectively. Many of the respondents misinterpreted the question and described the benefits of feedback in general instead of describing what was useful about the particular feedback they received. The responses to open-ended questions 1.8 and 2.8 - designed to identify elements that would have made the feedback more useful - resulted in AR rates of 61% and 70% respectively. A fully modified and updated version of the FQI based on these results is provided in the Appendix.

Report of FQI Data

Frequencies were calculated for each closed-ended question to illustrate the kind of information that can be gathered and reported with the FQI. Table **3** shows the response frequencies for each core FQI question and respondent suggestions for improving feedback. Questions 1.2 to 1.4 and 2.2 to 2.4 were included as supplemental – and optional - questions to assess whether the feedback was solicited or unsolicited, if it was received verbally or in written form, and when the feedback was delivered. 72% of the corrective and 92% of the positive feedback instances were reported as *unsolicited*. All reported feedback instances, both corrective and positive,

were delivered *verbally*. In terms of timing, 72% of corrective and 80% of positive feedback instances were reported as being provided *immediately* with the rest reported as provided *before end of shift*. Table **3** also shows a summary of the qualitative responses to Questions 1.8 and 2.8 that were included to understand residents' views for improving the feedback they received. It should be noted that there were too few adequate responses to report in response to Questions 1.7 and 2.7 which were designed to determine what residents liked about the feedback they received.

Table **4** shows a summary of responses to Questions 1.5c and 2.5c which were designed both as a check on the accuracy of each resident's recollection about receiving and feedback and as a way to determine the types of skills that are being targeted for feedback.

DISCUSSION

Feasibility, Validity, and Limitations of the FQI

The field pretest results indicated that the FQI can be administered, completed, and analyzed efficiently if a web-based survey application (e.g., SurveyMonkey) is used. The response rate of 63% was well within the acceptable rate for the behavioral sciences [23] but could probably be improved by sending follow-up reminder notices to participants. This process could easily be automated within the web-based survey application itself or within a separate email application.

In terms of data validity, the careful decisions made in the design and implementation of the FQI proactively addressed many common sources of survey error and the analysis of question quality indicated that most FQI ques-

Table 2. Percentage of Adequate Responses (AR) for Each FQI Question

Corrective Feedback Questions		Response Code Percentage**					
		n*	DK	IA	RF	AR	Action*
1.1	During your last shift did you receive any feedback from an attending or fellow on something you could improve?	44				100%	-
1.2	Did you actively solicit this feedback?	18				100%	-
1.3	How did they provide the feedback?	18				100%	-
1.4	When did they provide the feedback?	18				100%	-
1.5a	Was the feedback based on something they observed directly?	18	11%			89%	-
1.5b	Did they describe SPECIFICALLY what you did that needs improvement?	18	6%			94%	-
1.5c	Please briefly describe what they said you did.	13	8%			92%	-
1.5d	Did they describe the impact of what you did?	18				100%	-
1.5e	Did they provide an opportunity for you to respond to the feedback?	18	6%			94%	-
1.5f	Did their feedback include any guidance or discussion on how to improve?	18				100%	-
1.6	Overall, how would you describe the usefulness of the feedback they provided?	18				100%	-
1.7	What about the feedback was particularly useful to you?*	17		41%	24%	35%	Modified
1.8	What, if anything, would have made the feedback more useful?*	18	6%		33%	61%	Modified
Positive Feedback Questions							
2.1	During your last shift did you receive any feedback from an attending or fellow on something you did well?	44	3%			97%	-
2.2	Did you actively solicit this feedback?	26				100%	-
2.3	How did they provide the feedback?	26				100%	-
2.4	When did they provide the feedback?	26				100%	-
2.5a	Was the feedback based on something they observed directly?	26	3%			97%	-
2.5b	Did they describe SPECIFICALLY what you did well?	26	3%			97%	-
2.5c	Please briefly describe what they said you did well.	14	6%			94%	-
2.5d	Did they describe the positive impact of what you did?	26	20%			80%	Reassess
2.6	Overall, how would you describe the usefulness of the feedback they provided?	26				100%	-
2.7	What about the feedback was particularly useful to you?	26		73%	12%	15%	Modified
2.8	What, if anything, would have made the feedback more useful?	26		12%	18%	70%	Modified

*n = All 44 were presented with Q's 1.1 and 2.1 but the remaining questions were only presented to those who answered "yes" to the question preceding it.

**Explanation of codes. DK = respondent "didn't know" or "couldn't recall", IA = respondent gave an irrelevant answer, RF = respondent didn't answer the question, AR = all other responses.

Table 3.   Percent Responses Frequencies to FQI Questions

	n	Corrective Feedback	n	Positive Feedback
Percentage responding "Yes" to each of the following questions:				
During your last shift did you receive any feedback?	44	41%	44	59%
Was the feedback based on something they observed directly?	18	56%	26	77%
Did they describe specifically what you did?	18	72%	26	54%
Did they describe the potential impact of what you did?	18	78%	26	38%
Did they provide an opportunity for you to respond?	18	83%	-	Not asked
Did the feedback include guidance or discussion on how to improve?	18	83%	-	Not asked
Percentage responding "Very useful" to the following question:				
Overall, how would you describe the usefulness of the feedback?	18	67%	26	37%
Respondent suggestions for improving feedback				
More of this type of feedback	18	50%	26	42%
More specific about the importance of the desired behavior	18	17%	26	4%
More specific about what I did	18	0%	26	19%
Provide it more immediately	18	11%	26	0%
Other	18	11%	26	12%

tions were consistently understood and could be answered adequately. Unfortunately, the quality of data from questions 1.7 and 2.7 - designed to determine the aspects of feedback that residents saw as particularly effective - was too poor to be useful. Those questions have been redesigned.

Most encouraging - regarding evidence for validity - was the fact that a large majority of residents were able to provide relatively detailed descriptions of the feedback they received. This provides some good evidence that the FQI is capturing true-positive, as opposed to false-positive, recollections. On the other hand, it's impossible to determine the prevalence of false-negatives (i.e., instances in which a resident actually did receive feedback but reported not receiving

it) from this field pretest. Consequently, the results obtained from the FQI might actually be an underestimation of feedback regularity.

The next step toward validating the FQI will be a controlled experimental study to assess residents' abilities to accurately recall and describe feedback from a prior shift. If unacceptable levels of false negatives are found, design changes (e.g., require a response to the FQI immediately after the shift) will be needed to improve accuracy. Even with this evidence, however, the ability to establish construct validity for the FQI is limited because some elements of effective feedback (e.g., recipient's level of respect toward the provider) simply can't be assessed with a recall-based tool.

Table 4.   Types of Behaviors or Skills on which Residents Received Feedback by ACGME Competency*

Comments grouped by ACGME Core Competency	Q1.5c – Corrective Feedback (n=18)		Q2.5c - Positive Feedback (n=19)	
	Freq	**Example**	**Freq**	**Example**
Patient Care	5	"use the duck walk in a sports PE"	5	"correct technique in the lab"
Medical Knowledge	5	"guidelines for CXR in bronchiolitis"	3	"asthma management"
Communication and Interpersonal Skills	5	"how to present a simple case in the clinic quickly"	5	"explaining the diagnosis to the family"
Systems and Practice Based Improvement	3	"triage and taking care of 2 pts in ER"	5	"correct documentation for billing"
Professionalism	0		1	"patience with difficult patient"

*Responses to open-ended questions 1.5c and 2.5c were grouped by ACGME competencies.

Furthermore, the elements included in the FQI might not represent the full range of elements that could, and should, be included in a recall-based measure.

A few additional limitations regarding the interpretation of FQI data should be mentioned. First, the FQI doesn't capture instances of feedback that might occur days or even weeks after a given shift. This kind of delayed feedback might not be ideal, but it's probably better than no feedback at all. Second, the data from the FQI say little about whether or not a given resident is receiving a sufficient amount of feedback for his or her level of development or need. Some residents might actually need more than one instance of feedback during a given shift. Third, the quality of the feedback being reported might be biased toward higher quality since those instances might be more readily recalled while an indeterminable number of lower quality feedback instances go unreported. These limitations, however, would be difficult to fully eliminate without significantly undermining the feasibility of the FQI.

Potential Uses for the FQI

Despite the limitations noted above, the FQI - in its present form - could certainly provide medical educators with useful information regarding the frequency and quality of feedback being provided to residents in their programs because it addresses some of the recall and data precision issues that are problematic in the methods and tools used in prior studies. More specifically, the FQI only requires the learner to consider feedback instances over a 24-48 hour period and asks the learner to describe the quality of feedback in terms of elements known – or strongly believed – to be essential to the effectiveness of the feedback. With this tool educators could gain at least some sense for the frequency and quality of feedback being provided in their programs and could use the data to identify targeted areas for improvement (e.g., increasing the specificity of feedback being provided) and to gauge the success of improvement efforts.

In addition to its use as an assessment tool for improvement efforts the FQI could also be used effectively as an educational tool to help learners and teachers see and understand what effective feedback entails. For example, it might be used in conjunction with a workshop on improving feedback skills where participants engage in role-play exercises or as a tool to formatively evaluate teachers in real or simulated context.

Customizability of the FQI

While the FQI was designed to assess feedback between teaching physicians and residents, the form could be easily modified to assess feedback from others in the clinical setting including peers, nurses, and others on the care team who might actually be in a better position to directly observe and give feedback to residents. In fact, the form could be easily modified to assess feedback frequency and quality in any setting (e.g., organizational) and between a variety of other roles (e.g., manager/employee).

In addition, the FQI form could be shortened, or slightly lengthened, to suit particular needs. While the core questions of the FQI should always be included, other questions could, and should, be left off if the information will not be used. For example, some users might not be interested in knowing whether or not the feedback was solicited or unsolicited.

Finally, some users might want to focus their assessment of feedback by targeting a specific skill or competency area that is of particular concern (e.g., interpersonal communication with patients). The FQI could be modified by specifying a particular area in questions 1.1 and 2.1. For example, question 1.1 could be reworded to say, "During your last shift did you receive any feedback from an attending on something you could improve related to your interpersonal communication with patients?"

SUMMARY CONCLUSIONS

The FQI was designed to address the need for a feasible tool and method that medical educators can use to identify and assess specific problems related to feedback regularity and quality in the clinical setting. The current version of the FQI, the guidelines for administering the tool, and the examples for reporting and using the data contained in this paper will be useful to medical educators who want to improve feedback in their particular settings. With further evidence for the validity of the FQI - perhaps after additional modifications - the tool could be used with even greater confidence in improvement efforts to enhance our conceptual and theoretical understanding of factors that contribute to effective feedback in the clinical setting.

ACKNOWLEDGEMENTS

APPENDIX

Feedback Quality Index

Corrective Feedback Questions

Q1.1   During your last shift/session did you receive any feedback from an attending or fellow on something you could improve?

Yes, from an attending .... ☐
Yes, from a fellow .... ☐
No .... ☐ Go to Q2.1
Can't recall .... ☐ Go to Q2.1

Q1.2   Did you actively solicit this feedback?

Yes .... ☐

No .... ☐

Q1.3   How did they provide the feedback?

Verbally .... ☐

In writing .... ☐

Both .... ☐

Q1.4   When did they provide the feedback?

Immediately .... ☐

After some delay, but before end of shift .... ☐

After end of shift .... ☐

Q1.5a Was the feedback based on something they observed directly?

Yes .... ☐

No .... ☐

Not sure .... ☐

Q1.5b Did they describe SPECIFICALLY what you did that needs improvement?

Yes .... ☐

No .... ☐ Go to Q1.5d

Can't recall .... ☐ Go to Q1.5d

Q1.5c Please briefly describe what they said you did. *(This will help us better understand the general areas in which residents are receiving feedback)*

☐ (text box)

Q1.5d Did they describe the impact of what you did?

Yes .... ☐

No .... ☐

Can't recall .... ☐

Q1.5e Did they provide an opportunity for you to respond to the feedback?

Yes .... ☐

No .... ☐

Can't recall .... ☐

Q1.5f Did their feedback include any guidance or discussion on how to improve?

Yes .... ☐

No .... ☐

Can't recall .... ☐

Q1.6   Overall, how would you describe the usefulness of the feedback they provided?

Very useful .... ☐

Somewhat useful .... ☐

Not at all useful .... ☐ Go to Q1.8

Q1.7   What did they say or do that made the feedback useful to you?

(text box)

Q1.8   What, if anything, could they have said or done differently that would have made the feedback more useful to you?

(text box)

Positive Feedback Questions

Q2.1   During your last shift/session did you receive any feedback from an attending or fellow on something you did well?

Yes, from an attending .... ☐

Yes, from a fellow .... ☐

No .... ☐ Go to End

Can't recall .... ☐ Go to End

Q2.2   Did you actively solicit this feedback?

Yes .... ☐

No .... ☐

Q2.3   How did they provide the feedback?

Verbally .... ☐
In writing .... ☐
Both .... ☐

Q2.4   When did they provide the feedback?

Immediately .... ☐
After some delay, but before end of shift .... ☐
After end of shift .... ☐

Q2.5a   Was the feedback based on something they observed directly?

Yes .... ☐
No .... ☐
Not sure .... ☐

Q2.5b   Did they describe SPECIFICALLY what you did well?

Yes .... ☐
No .... ☐ Go to Q2.5d
Can't recall .... ☐ Go to Q2.5d

Q2.5c   Please briefly describe what they said you did. *(This will help us better understand the general areas in which residents are receiving feedback)*

☐

Q2.5d   Did they describe the positive impact of what you did?

Yes .... ☐
No .... ☐
Can't recall .... ☐

Q2.6   Overall, how would you describe the usefulness of the feedback they provided?

Very useful .... ☐
Somewhat useful .... ☐
Not at all useful .... ☐ Go to Q2.8

Q2.7   What did they say or do that made the feedback useful to you?

☐

Q2.8   What, if anything, could they have said or done differently that would have made the feedback more useful to you?

☐

REFERENCES

[1]   Wood BP. Feedback: A key feature of medical training. Radiology 2000; 215: 17-9.
[2]   Ende J. Feedback in clinical medical education. J Am Med Assoc 1983; 250: 777-81.
[3]   Irby DM. What clinical teachers in medicine need to know. Acad Med 1994; 69: 333-42.
[4]   McIlwrick J, Nair B, Montgomery G. "How am I doing?": Many problems but few solutions related to feedback delivery in undergraduate psychiatry education. Acad Psychiatry 2006; 30: 130-5.
[5]   Salerno SM, Jackson JL, O'Malley PG. Interactive faculty development seminars improve the quality of written feedback in ambulatory teaching. J Gen Intern Med 2003; 18: 831-4.
[6]   Gil DH, Heins DM, Jones PB. Perceptions of medical school faculty members and students on clinical clerkship feedback. J Med Educ 1984; 59: 856-64.
[7]   Kogan JR, Hauer KE. Brief report: Use of the mini-clinical evaluation exercise in internal medicine core clerkships. J Gen Intern Med 2006; 21: 501-2.
[8]   Biemer PP, Lyberg LE. The survey process and data quality. Introduction to survey quality Hoboken, NJ: Wiley; 2003; pp. 26-62.
[9]   Colliver JA, McGaghie WC. The reputation of medical education research: Quasi-experimentation and unresolved threats to validity. Teach Learn Med 2008; 20: 101-3.
[10] Hewson MG, Little ML. Giving feedback in medical education: Verification of recommended techniques. J Gen Intern Med 1998; 13: 111-6.
[11] Bing-You RG. Feedback falling on deaf ears: Residents' receptivity to feedback tempered by sender credibility. Med Teacher 1997; 19: 40-4.
[12] Kluger AN, DeNisi A. The effects of feedback interventions on performance feedback: A historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996; 119: 254-84.
[13] Kulhavy RW, White MT, Topp BW, Chan AL, Adams J. Feedback complexity and corrective efficiency. Contemp Educ Psychol 1985; 10: 285-91.
[14] Crawford SD, Couper MP, Lamias MJ. Web Surveys: Perceptions of Burden. Soc Sci Comput Rev 2001; 19: 146.
[15] Nicol D, MacFarlane-Dick D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Stud High Educ 2006; 31: 199-218.
[16] London M. Feedback dynamics. Job feedback: giving, seeking, and using feedback for performance improvement. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 2003; pp. 38-51.
[17] Peterson RS, Behfar KJ. The dynamic relationship between performance feedback, trust, and conflict in groups: A longitudinal study. Organ Behav Hum Decis Process 2003; 92: 102-12.
[18] Fedor DB, Eder RW, Buckley MR. The contributory effects of supervisor intentions on subordinate feedback responses. Organ Behav Hum Decis Process 1989; 44: 396-414.
[19] de Luque MFS, Sommer SM. The impact of culture on feedback-seeking behavior: An integrated model and propositions. Acad Manage Rev 2000; 25: 829-49.
[20] Cochran WG. Systematic sampling. Sampling techniques. 3rd ed. New York: John Wiley & Sons; 1977; pp. 205-29.

[21]   Finley R. SurveyMonkey. Portland, OR: SurveyMonkey.com; 2010.

[22]   van der Zouwen J, Smit JH. Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach. In: Presser S, Rothgeb JM, Couper MP, Lessler JT, Martin E, Martin J, *et al*., Eds. Methods for testing and evaluating survey questionnaires Hoboken, N.J.: Wiley-Interscience; 2004; pp. 109-30.

[23]   Baruch Y. Response Rate in Academic Studies—A Comparative Analysis. Hum Relat 1999; 52: 421-38.