Benefit of Report Card Feedback After Point-of-Care Assessment of Communication Quality Indicators

Michael H. Farrell  
Clair R. Sprenger  
Shelbie L. Sullivan  
Bree A. Trisler  
Jessica J F Kram  
Erin K. Ruppel

Follow this and additional works at: https://aurora.org/jpcrr

Part of the Health and Medical Administration Commons

Recommended Citation

Journal of Patient-Centered Research and Reviews (JPCRR) is a peer-reviewed scientific journal whose mission is to communicate clinical and bench research findings, with the goal of improving the quality of human health, the care of the individual patient, and the care of populations.
Benefit of Report Card Feedback After Point-of-Care Assessment of Communication Quality Indicators


1Aurora University of Wisconsin Medical Group, Aurora Health Care, Milwaukee, WI
2Center for Urban Population Health, Milwaukee, WI
3Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee, WI
4Diederich College of Communication, Marquette University, Milwaukee, WI
5Department of Communication, University of Wisconsin-Milwaukee, Milwaukee, WI

Purpose
Communication is crucial for patient experience and biomedical outcomes. Training programs improve communication but are too resource-intensive for sustained use across an entire health care organization. This study demonstrates in a heterogeneous set of encounters the efficacy of quantitative feedback on two groups of physician communication behaviors: 1) jargon explanation, and 2) assessment of patient understanding.

Methods
We used a secure Internet application to audio-record conversations between primary care physicians and 54 patients. Transcripts were quantitatively abstracted using explicit-criteria definitions for assessments of understanding and jargon explanations. These data were conveyed to physicians using a previously tested report card. Finally, physicians were audio-recorded with 48 other patients and compared against their baseline.

Results
Baseline transcripts included an average of 15.5 unique jargon words. Many words were spoken more than once so the total jargon count averaged 25.1. Jargon explanations were infrequent (median of 2.6/transcript). The jargon explanation ratio (fraction of jargon words spoken after or alongside a jargon explanation for that word) averaged 0.26 out of 1.0. Assessments of understanding were found in 61.1% of transcripts, but most were "OK?" questions (median of 2.22/transcript) or close-ended assessments of understanding (median of 0.59/transcript). After the report card, use of jargon explanations improved to a median of 4.8/transcript (P<0.001), and the jargon explanation ratio improved to 0.37 (P<0.02). Assessments of understanding improved to 81.3% of transcripts (P<0.03), largely due to increased use of close-ended assessments of understanding to 1.08/transcript (P<0.006).

Conclusions
It is feasible to audio-record at the point of care, abstract transcripts at a central office and improve physician-to-patient communication quality via a report card. A larger, multifaceted program may improve patient experience and biomedical outcomes. (J Patient Cent Res Rev. 2017;4:7-17.)

Keywords
health communication; patient satisfaction; physician-patient relations; health care quality

Communication is a crucial contributor to the patient experience and is said to be the “main ingredient” of health care. Effective communication improves adherence and patient satisfaction and has been associated with improvements in symptoms, physiological measures and biomedical outcomes. Unfortunately, there are many problems with health care communication and with the patient experience, as evidenced by many research studies and also by the patient complaints that organizations routinely receive.

Schools and residency programs teach communication skills, but training alone does not seem to have solved the problems seen in the health care system. As a result, payors and health care organizations are attempting to improve patient experience by surveying patients and conveying results to providers and clinical sites. The best known of these programs use data from the Consumer

Correspondence: Michael H. Farrell, MD, Aurora Sinai Medical Center, 1020 N. 12th Street, Milwaukee, WI, 53233, T: 414-219-7914, Email: dr.mike.farrell@aurora.org
Assessment of Healthcare Providers and Systems (CAHPS) surveys. The hospital version of CAHPS (H-CAHPS) is especially motivating for organizations because payments for some hospitalizations are decreased for hospitals with poor H-CAHPS scores and increased for hospitals with higher H-CAHPS scores. Public reporting of survey data also may motivate organizations if the publicity affects contracts with payors or prompts patients to switch to another health care organization. Consumer surveys also may provide motivation for individual health care providers, perhaps from a sense of duty or embarrassment or because the results may affect their salaries.

Unfortunately, patient surveys can have limitations. Validity of data may be uncertain because patients’ responses to a survey question can be influenced by factors other than the question’s topic. For example, a patient could be annoyed by a prolonged wait, an unpleasant experience at the front desk or a provider’s refusal to prescribe a desired medication. It also may be difficult to compare responses of patients from differing cultural backgrounds or who have different calibrations for responses such as “always” versus “usually.” Surveys also increase the burden of paperwork on patients, especially for persons with limited health literacy.

The utility of patient-reported data is also uncertain. Most questions on patient surveys do not supply a concrete target for improvement. For example, there is no specific strategy to improve patients’ responses to the H-CAHPS question, “How often did nurses explain things in a way you could understand?” Some organizations address this question by teaching their employees to include comments such as, “I want to explain things in a way that is easy to understand.” These types of remarks may increase CAHPS scores but do not improve the explanations themselves. Without true improvement in communication, it seems unlikely that survey data will translate to better satisfaction, outcomes or value.

Even if a communication improvement program is valid and successful, the health care organization must consider issues such as cost, sustainability and employee relations. For example, the widespread use of periodic assessment by simulated patients would be too expensive and labor-intensive to be sustainable in most organizations. Busy clinicians also might object to such a program’s interference with their time and productivity.

To fill the methodological gap between early training and patient surveys, we continue to adapt effective techniques from quality improvement (QI) for the needs of communication. The result is an approach that the first author has coined “Communication Quality Assurance” (Comm QA). Comm QA methods are designed to be objective, quantitatively reliable, transparent and unobtrusive enough to be acceptable to busy clinicians as well as straightforward enough to be implemented by existing personnel on a lean budget. Communication behaviors are operationalized using an objective “quality indicator” approach instead of subjective rating scales. As with indicators in traditional QI, each indicator in Comm QA uses explicit-criteria definitions and represents a small but important domain within the overall concept of health care quality. In a previous randomized controlled trial, the first author demonstrated the effectiveness of similar techniques used after a telephone counseling conversation about newborn screening results. In that study, feedback was conveyed by a report card adapted from so-called “dashboard” graphic designs that are often used in traditional QI projects.

The research design for the study reported here implemented the next necessary innovation: to embed these methods at the point of care for actual patients. To take this step, we reprised the previously studied methods in an intervention-only pilot project, focusing on two groups of communication quality indicators — assessment of patient understanding and explanation of jargon (Table 1).

**METHODS**

**Design**

We conducted a Comm QA project with physicians in an urban medical group. For regulatory reasons the project was considered to have two components: a QI component based on earlier research, and a secondary research analysis that would use the data from the QI effort to develop new measures. The QI component did not require informed consent from physicians, but the secondary research analysis required informed consent from both physicians and patients.
This paper presents results from the QI component, which was intended to improve physicians’ use of jargon explanations and assessments of understanding. The QI component followed a pre/post design with a single cohort of physicians. Physician scores on a panel of communication quality indicators were measured at baseline, and these scores were used to populate a report card for feedback. We then audio-recorded physicians with new patients to obtain follow-up scores for comparison. Materials and procedures were approved by our institutional review board.

Participants and Recruitment
Both physicians and patients were considered research participants but with important differences.

Physicians: An introductory email was sent by the principal investigator to 45 family physicians and 40 internal medicine physicians from six clinics. The email, which was cosigned by the director of the medical group, briefly described the overall project, how participation in the QI component would be expected by the medical group, and how the first author would not be able to identify their data by name. The introductory email asked physicians to complete a short online survey using Qualtrics software (Qualtrics LLC, Provo, UT). The survey was designed to obtain some contextual data about respondents’ previous experience with communication assessment and whether or not they were “nervous” about the upcoming recordings. Relevant questions are further described in the Results section.

The researchers then met with physicians to ensure they understood the project. During this meeting, signed informed consent was obtained for the secondary research analysis portion of the project. Each physician was slated to be audio-recorded on two separate clinic days, have one or more “baseline” encounters and have one or more “follow-up” encounters. Research staff kept in contact with the physicians to work within their schedules.

Patients: Informed consent was needed from patients for the entire project, since for them the audio-recording was entirely voluntary. There were up to three opportunities for outreach to patients. Approximately 2 weeks before each physician’s anticipated day for recording, a project staff person reviewed the clinic schedule for patients who were age 18 years or older, did not require a language interpreter and did not have an entry in the electronic medical record that suggested they had intellectual disability or a similar challenge for informed consent. A list of patients without these exclusion criteria was sent to the physician scheduled to see the patient to determine if there might be some other contraindication to participation (e.g. fragile emotional status). The clinic then mailed the patients a letter and information sheet that contained required

<table>
<thead>
<tr>
<th>Quality indicator</th>
<th>Description/Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use of jargon</strong></td>
<td></td>
</tr>
<tr>
<td>Total jargon words</td>
<td>Count of every jargon word* in the transcript.</td>
</tr>
<tr>
<td>Unique jargon words</td>
<td>Count of jargon words used at least once in the transcript (each unique word only counted once)</td>
</tr>
<tr>
<td>Jargon explanation</td>
<td>Any text string that explains a word or concept</td>
</tr>
<tr>
<td>Jargon explanation ratio</td>
<td>Percentage of total jargon words that either follow an explanation of the word, or precede an explanation by two text strings (possible range, 0–1.0)</td>
</tr>
<tr>
<td><strong>Assessments of understanding</strong></td>
<td>Example**</td>
</tr>
<tr>
<td>Close-ended</td>
<td>&quot;Do you have any questions?&quot;</td>
</tr>
<tr>
<td>Open-ended</td>
<td>&quot;What questions do you have for me?&quot;</td>
</tr>
<tr>
<td>Request for teach-back</td>
<td>&quot;It would be helpful to me if you could repeat back that last point in your own words.&quot;</td>
</tr>
<tr>
<td>“OK?” question**</td>
<td>&quot;OK?&quot; or &quot;All right?&quot;</td>
</tr>
</tbody>
</table>

*Note that a compound term (e.g. cystic fibrosis) counts as a single jargon word.

**Denotes a negative indicator, or a test indicator whose value for communication is unclear.
messages for informed consent, such as an assurance that participation was voluntary.

The contact protocol also included provisions for telephoning patients within 1–5 days of the anticipated recording day because patients can be added to the schedule at the last moment. Oral content of the telephone call was the same as that in the letter and information sheet.

On the day of the appointments, we excluded patients who were slated to be seen by a medical student. The front desk person gave a copy of the letter and information sheet to the patient. A project staff person approached the patient in the examination room before the physician’s entry and offered the patient a chance to ask questions and decline the recording. Written documentation of patients’ informed consent had been waived by the institutional review board, but consent was confirmed aloud by asking, “The recorder is on, is that OK?” and waiting for a verbal response.

Physicians were instructed how to cancel the recording if necessary, as when a conversation grew emotional enough that the recorder might be troubling to the patient. To our knowledge, this occurred once during the project.

Procedures

Recording and Data Management: Encounters were audio-recorded using AVA software (People Designs, Durham, NC), an encrypted online application that runs in an Internet browser on exam room computers or over a wirelessly connected iOS device (Apple Inc., Cupertino, CA). When a network browser or wireless was not available, then a handheld digital audio-recorder was used and carefully tracked by the project staff person to ensure that devices would not be misplaced or cause a confidentiality breach.

Audio-recordings were transcribed without identifying words, then transcripts parsed into individual sentences (this was a simplification of our usual procedure of using strings of text with one subject and one predicate). Subsequently, the sentences were abstracted for two groups of communication quality indicators — jargon explanations and assessments of understanding. Abstraction process used previously demonstrated explicit-criteria procedures. Abstractors were blind to whether transcripts were from baseline or follow-up encounters.

Report Card: Communication quality indicator data from each transcript were used to generate personalized report cards (see Figure 1 for a de-identified example). When a physician had more than one transcript for abstraction, the report card presented medians for each communication quality indicator.

We presented jargon data in comparison with peer data by quartiles and a comparative adjective (Figure 1, at left). We presented data about assessments of understanding in a table (Figure 1, upper right) and a histogram that compared the physician with his/her peers (Figure 1, lower right). The histogram incorporated previously described ordinal “feedback categories,”16–19 which we have used to provide feedback that would be less esoteric than our more technical measures. Definitions for the categories are provided in the Results section and Figure 4.

Researchers emailed report cards to physicians via the institution’s email system. Physicians were offered the opportunity to listen to their recordings and read their transcripts.

Analysis

Statistical procedures included $\chi^2$ test, t-test, ANOVA and Wilcoxon signed-rank test depending on variables’ respective characteristics. The Wilcoxon test is important for nonparametric statistics, i.e. when distribution of data does not resemble the normal distribution that is necessary for parametric tests such as t-test. For many tests, comparisons were made with matched (pre/post) data rather than simple means or medians. Analyses were done with JMP software (SAS Institute Inc., Cary, NC).

RESULTS

Survey

We used the baseline survey to understand the scope of medical group physicians’ perspectives about communication assessment. We received 41 responses (48.2% of physicians contacted). A portion of the survey asked about previous experiences with communication assessment. A total of 38 respondents
Original Research

(92.7%) indicated that their communication had previously been assessed by a simulated patient; 28 (68.3%) indicated that a teacher had directly observed them with an actual patient and subsequently provided feedback; and 33 (80.5%) indicated that they had been video-recorded or audio-recorded with an actual patient and later received feedback from a teacher about the recording. For each of these assessment methods, the survey asked physicians about the types of communication that had been assessed (Table 2).

Another survey question asked, “How nervous are you about having your communication audiotaped with real patients?” A slider was supplied for response on a Likert scale with a 1 to 5 range (5 being “very nervous” and 1 being “totally relaxed, not nervous at all”). This was important because participation by the medical group was expected by its leadership. None of the physicians responded in the slider range around “very nervous” (Table 3). About half responded in the slider range around “somewhat nervous.”

**Participation**
Of the 85 (45 family, 40 internal medicine) physicians contacted, only 30 (35%) completed both the baseline and follow-up encounters. Participant characteristics for the final sample of physicians are shown in Table 4. Six of the family physicians were excused for maternity or medical leave; another 8 family physicians were not able to complete the project because of graduation from residency. For the internists, plans for inclusion were dropped when only 6 physicians (4 faculty, 2 residents) responded to several recruitment emails for the baseline surveys or for the introductory meetings. The reason for such a significant disparity in response to emails was never clear, since the email made clear that participation...
was expected by the medical group and was cosigned by the medical group’s director (a respected and widely liked internist). Resources and logistical challenges limited our ability to work in the internists’ clinics, however by the time we had made the decision to exclude this subgroup, we had recorded 2 internists.

By the end of the project, there were 54 baseline transcripts and 48 follow-up transcripts. The average amount of time spent by the patient in the exam room was 47.3 minutes (standard deviation: 20.1). The average duration of the provider-patient conversations was 22.9 minutes (standard deviation: 8.9), or about 48% of the time spent in the exam room.

### Jargon Usage and Explanation

Results for the jargon analyses are listed in Table 5. The total and unique counts of jargon words were far greater than the count of jargon explanations, regardless of whether the conversation was recorded at baseline or follow-up. The counts of unique and total jargon words did not appear to differ from baseline to follow-up. However, the number of jargon explanations increased significantly from baseline to follow-up ($P=0.007$ on matched Wilcoxon). Jargon explanation data are nonparametric and therefore best represented with a box-and-whisker plot (Figure 2) in which boxes represent the groups’ medians and interquartile ranges.

The timing of jargon explanation also improved. At baseline, the average jargon explanation ratio was 0.26 out of a possible range of zero to 1.0, which
means that 26% of physician-used jargon words either followed an explanation or were followed within two sentences by an explanation. In follow-up transcripts, the average jargon explanation ratio improved to 0.36 (P=0.008 on matched t-test). Figure 3 depicts these data graphically.

A second analysis was done with the word counts standardized for duration of conversation, which should be proportional to the total number of words in the encounter. In this analysis the beneficial effect of the report card persisted for jargon explanations (medians of 1.3 at baseline to 3.9 at follow-up, P=0.002 on Wilcoxon). The variance in the adjusted count of jargon words increased enough that the significance was lost for the jargon explanation ratio. This loss of significance suggests that the timing of jargon explanation was not distributed evenly across the range of duration-adjusted transcripts, presumably because of sample size.

### Assessment of Understanding

The quality indicator approach allows assessment of understanding to be analyzed by transcript, by participant or by individual behavior. In the transcript-level analysis, 33 baseline transcripts (61.1%) met criteria for at least one assessment of understanding. After the report card, 39 transcripts (81.3%) met criteria for at least one assessment of understanding (P<0.03 on \( \chi^2 \)).

In the participant-level analysis, comparison was done for the feedback categories received by each physician on their respective report cards (Figure 1, lower right). Figure 4 shows the number of participants who received each of the feedback categories (“expert,” “good,” etc.), along with a definition for each category. When the feedback categories were analyzed by their

---

**Table 5. Participant Usage of Jargon and Jargon Explanations**

<table>
<thead>
<tr>
<th>Quality indicator data</th>
<th>Baseline, mean (SD)</th>
<th>Follow-up, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique jargon words (each word counted once)</td>
<td>15.5 (11.2)</td>
<td>15.8 (8.2)</td>
</tr>
<tr>
<td>Total number of jargon words (all words)</td>
<td>25.1 (17.9)</td>
<td>27.2 (16.2)</td>
</tr>
<tr>
<td>Jargon explanation</td>
<td>2.6 (3.6)</td>
<td>4.8 (4.1)</td>
</tr>
<tr>
<td>Jargon explanation ratio</td>
<td>0.26 (0.16)</td>
<td>0.37 (0.17)</td>
</tr>
</tbody>
</table>

SD, standard deviation.
ordinal numbers (possible range of 0–4), there was an improvement from 1.25 at baseline to 1.68 at follow-up (P<0.03 on matched t-test).

The behavior-level analysis counts each individual assessment of understanding behavior in conversation (Table 6). This more detailed analysis confirmed the report card’s effect on close-ended assessments of understanding, with medians of zero at baseline and 1 at follow-up (P=0.04 on Wilcoxon) (Figure 5). Variance was too great within this sample to detect a difference for open-ended assessments of understanding or the “OK?” question. Only one request for teach-back was seen in the entire project, so no difference for that assessment of understanding behavior could be measured.

DISCUSSION
A positive patient experience is necessary for health care providers to be effective and for health care organizations to be successful. Communication and satisfaction may be improved by continuing education or reminders, but there is no evidence that the effects of these methods last for long. To fill the gap between initial training and long-term practice, we explored the applicability of the Comm QA approach, which uses methods that are straightforward enough to be implemented by existing personnel on a lean budget so that health care organizations can afford to continue them as part of ongoing QI efforts.16-28

The results of this pilot project confirmed our earlier findings27 that a modest Comm QA intervention can produce quantifiable improvements in specific communication behaviors. We observed nearly a doubling of jargon explanations and a 38% relative improvement in the timing of explanations. For assessments of understanding, there was a 32% relative improvement. For this pilot we added some brief explanatory remarks to the report card (Figure 1), but we suspect that Comm QA may be even more successful if participants are given more persuasive materials.

Methodologically, the pilot demonstrated that design principles from our earlier research can be generalized to a busy clinical setting.16-28 For example, it is feasible to use communication quality indicators with a heterogeneous set of conversational topics. Feedback was limited to only two groups of quality indicators, reducing the chance that participants would be overwhelmed by a large number of details. Feedback data for the physicians

![Figure 4. Number of transcripts for each assessment of understanding (AU) feedback category.](image-url)
were compared with those for peers, so participants would know how their communication skills compared.

Most importantly, communication quality indicators suggest a concrete target for improvement. For example, physicians’ assessment of understanding scores could have been improved by using open-ended questions instead of close-ended questions. Better yet, a request for teach-back could have been included in more challenging conversations. The finding that the increase in assessments of understanding was mostly due to close-ended questions suggests that physicians may lack awareness of the most effective ways to assess understanding.

### Study Limitations

There are several limitations to consider. Resource challenges for this pilot led to a low sample size, but we still had enough power to show benefit from a modest intervention. Generalizability also may be limited to nonacademic settings, but data from the baseline survey help to show that the sample included physicians with varying degrees of comfort and experience with communication assessment. The pilot was resourced as a pre/post study, thus theoretically the improvement could have been a test/retest phenomenon or a result of awareness of the project’s goals. On the other hand, the same methods were found to be effective in a previous randomized trial. Furthermore, a Hawthorne effect from awareness of being recorded should facilitate efforts to assess competence because participants are likely to be on their best behavior when feeling observed.

### CONCLUSIONS

CAHPS and other initiatives are giving new prominence to provider-patient communication and the patient experience. Research efforts thus far are promising, but the next necessary research goal is to determine whether broader-scale implementation will result in improved CAHPS scores, biomedical outcomes and communication outcomes (e.g. understanding, psychological states, satisfaction, etc.). This project demonstrated how Communication Quality Assurance methods can be used at the point of care to improve specific communication behaviors. Ideally, the next step will be part of a mandatory, comprehensive Comm QA program with a budget for outreach, explanatory materials and access to outcomes databases. Broader demonstration of Comm QA’s effect on outcomes should help health care organizations understand the value of its affordable, acceptable approach to service across the entire workforce.

### Table 6. Mean Number of Individual Assessment of Understanding Behaviors

<table>
<thead>
<tr>
<th>Assessment of understanding</th>
<th>Baseline, mean (SD)</th>
<th>Follow-up, mean (SD)</th>
<th>Sigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close-ended</td>
<td>0.59 (1.30)</td>
<td>1.08 (1.27)</td>
<td>P&lt;0.006</td>
</tr>
<tr>
<td>Open-ended</td>
<td>0.48 (0.89)</td>
<td>0.21 (0.45)</td>
<td>NS</td>
</tr>
<tr>
<td>Request for teach-back</td>
<td>0.02* (0.14)</td>
<td>0 (0)</td>
<td>NS</td>
</tr>
<tr>
<td>“OK?”</td>
<td>2.22 (4.16)</td>
<td>2.02 (2.54)</td>
<td>NS</td>
</tr>
</tbody>
</table>

*Request for teach-back was seen in a single transcript at baseline only.

SD, standard deviation; Sig, statistical significance.

### Figure 5. Median number of close-ended assessments of understanding (AUs), with range and interquartile range. Box-and-whisker plots for Wilcoxon comparison of matched, nonparametric data. Boxes represent medians and interquartile ranges. Whisker lines represent the range of data excepting outliers (i.e. greater than 150% of the interquartile range). Means for each group are represented in each column for purposes of visual comparison. Jittering of data points is solely for visual clarity.
Patient-Friendly Recap

- Effective communication between physicians and patients is crucial for quality health care so researchers are looking at means of improvement that can be sustained across a large health system’s workforce.
- The authors studied one such approach, called Communication Quality Assurance, which is more objective, unobtrusive and transparent than traditional training methods.
- In addition to determining physician/patient conversations can be audio-recorded at the point of care, the authors found that physicians, when given a report card judging their past conversations in clinic, increased their explanations of potentially confusing jargon and efforts to ensure that patients truly understood what was said.

Conflicts of Interest
None.

Funding Sources
The project was funded by the Aurora Health Care Foundation and a pilot project award from the University of Wisconsin Foundation Milwaukee Clinical Faculty Fund.

Acknowledgments
The authors are very grateful to Karen Pankowski of the Center for Urban Population Health for her excellent assistance with transcription.

References


© 2017 Aurora Health Care, Inc.