Published in final edited form as:
Patient Educ Couns. 2009 October ; 77(1): 128–135. doi:10.1016/j.pec.2009.03.013.

A Method to Quantify and Compare Clinicians' Assessments of Patient Understanding during Counseling of Standardized Patients

Michael H. Farrell[1], **Pramita Kuruvilla**[2], **Kerry L. Eskra**[1], **Stephanie A. Christopher**[1], and **Rebecca S. Brienza**[3]

[1]Center for Patent Care and Outcomes Research, Medical College of Wisconsin, Milwaukee, WI

[2]Contra Costa Regional Medical Center Family Practice Residency Program, Martinez, CA

[3]Southport, CT

Abstract

OBJECTIVES—to introduce a method for quantifying clinicians' use of assessment of understanding (AU) questions, and to examine medicine residents' AU usage during counseling of standardized patients about prostate or breast cancer screening.

METHODS—Explicit-criteria abstraction was done on 86 transcripts, using a data dictionary for 4 AU types. We also developed a procedure for estimating the "load" of informational content for which the clinician has not yet assessed understanding.

RESULTS—Duplicate abstraction revealed reliability κ=0.96. Definite criteria for at least one AU were found in 68/86 transcripts (79%). Of these, 2 transcripts contained a request for a teach-back ("what is your understanding of this?"), 2 contained an open-ended AU, 46 (54%) contained only a close-ended AU, and 18 (21%) only contained an "OK?" question. The load calculation identified long stretches of conversation without an AU.

CONCLUSION—Many residents' transcripts lacked AUs, and included AUs were often ineffectively phrased or inefficiently timed. Many patients may not understand clinicians, and many clinicians may be unaware of patients' confusion.

PRACTICE IMPLICATIONS—Effective AU usage is important enough to be encouraged by training programs and targeted by population-scale quality improvement programs. This quantitative method should be useful in population-scale measurement of AU usage.

Keywords

communication; physician-patient relationship; health literacy; prostate-specific antigen; mammography

Corresponding author: Michael H. Farrell, MD, Assistant Professor, Internal Medicine & Pediatrics, Center for Patient Care and Outcomes Research, 8701 Watertown Plank Road, Milwaukee, Wisconsin 53226-0509, Voice (414) 456-8381, Fax (414) 456-6689, mfarrell@mcw.edu.

1. Introduction

Effective interpersonal communication, called "the vehicle by which technical care is implemented and on which its success depends" [1], may require clinicians to adopt a wide variety of patient-centered communication behaviors [2-4]. For example, clinicians are advised to explicitly assess their patients' understanding [2-4]. Assessment of understanding questions (AUs) may be especially important for patients with limited health literacy, when the topic of conversation is complicated, or when patients' facial expressions do not correlate with emotions for cultural reasons or because of a sense of shock [4,5]. Unfortunately, many clinicians may be missing opportunities to assess patients' understanding [6-9]. In this paper we introduce a method to quantify clinicians' likely AU usage and use the method to examine internal medicine residents' use of AUs during counseling of standardized patients about prostate or breast cancer screening.

We chose to study communication about prostate and breast cancer screening because professional organizations recommend routine counseling about potential risks of screening [10-14]. Counseling is recommended prior to screening with prostate specific antigen (PSA) because there is insufficient evidence that the medical benefits of screening are greater than its medical and psychological harms [10-13]. Counseling recommendations for breast cancer screening were instituted for women in the 40-49 age group because evidence about benefit versus harm was slower to develop for this age group than for older women [13-15]. Recommendations for counseling in both situations are intended to insure that screening decisions are consistent with patient's values and underlying beliefs, and that patients will be prepared in the event of an abnormal result [10-14,16].

This analysis is part of a larger effort to develop communication assessment tools that will be inexpensive and concrete enough for use over entire geographic regions by quality improvement personnel with typical training [6,17-23]. This scope of attention is sometimes referred to as a "population" scale approach [24]. We are interested in population-scale methods because clinicians can develop communication problems long after training and far away from the training centers where many efforts to improve communication are done [25,26]. Our methods are adapted from Quality Improvement [6,17-23] because of that field's track record for changing physician behavior on a population scale [27], and because of our concerns that education-derived methods often require more resources and physician preceptor time than is available for population-scale use. Another advantage to adopting Quality Improvement methods is that their explicit definitions and concrete approach may improve on limitations of earlier communication measures such as sometimes insufficiently detailed definitions for quantification [1,25]. It is worth emphasizing that since AU usage is only one facet of communication quality, our AU analyses should be considered as part of a panel of measures that include our other measures: inclusion and timing of content messages [19,22,23], jargon and explanations [20,21] and assessment of emotional reactions [28,29]. The overall approach of locating and counting up these individual communication behaviors is based on research from cognitive psychology about the mental demands of simultaneously processing new information, experiencing emotions, and holding several unfamiliar concepts in mind [30-33]. Dispersal of AUs is important because each new misunderstanding can contribute to a backlog of confusion that could interfere with comprehension of the next concept [30-33].

We identified four types of AU questions (Table 1) in a literature search for our study of AU usage after newborn screening [6] and described a reliable, explicit-criteria method to determine whether a given transcript includes each AU type [6]. For the present study we also developed a second-generation quality indicator called "estimated mental load," which of necessity uses a more complicated algebra to calculate the amount of information presented before, between, and after each individual AU. The mental load concept is consistent with

research from cognitive psychology about the demands of holding in mind several new and unfamiliar concepts, each of which may contribute to a backlog of confusion that can interfere with comprehension of subsequent information [30-33]. The accumulated mental load can be mitigated if the clinician uses an AU to give the patient an opportunity to get questions answered [2-4]. The aims of this project are to determine the type and frequency of AU usage, and to characterize the timing of AU usage with respect to accumulating mental load.

2. Methods

2.1. Data source

This analysis used an explicit-criteria procedure to abstract transcripts of conversations between internal medicine residents and standardized patients portrayed to have a question about screening for prostate or breast cancer. Transcripts were made from tapes collected during workshops in a Primary Care Internal Medicine residency program. The workshops were part of the official curriculum, but residents gave informed consent and were free to decline the use of their tapes in research. Methods were approved by institutional review boards at Yale and the Medical College of Wisconsin.

Residents were taped in two standardized patient encounters before the teaching portion of the workshops. In one encounter, a 50-year-old man asked about prostate cancer screening. In the other encounter, a 43-year-old woman asked about breast cancer screening. The order of the two encounters was random for each resident. A handout stated that the patient had no family history of cancer and had had an unremarkable physical exam the week before, so the resident would not feel obliged to do a physical or take an extended history. The handout did not contain any suggestions about how to discuss the screening test.

Patients began with a short speech patterned after the following example:

> I'm sorry I'm back so soon after my physical, but I had to leave so quickly that I didn't get a chance to ask a question. I recently saw an advertisement about prostate [or breast] cancer screening, but I wasn't sure if it was for me. What do you think?

Patients were coached to not appear anxious or confused, to not make requests for clarification, and to adopt a neutral body language with the small, polite nod that does not necessarily denote understanding. These instructions standardized the counseling task and helped our analysis to focus on AUs rather than the resident's ability to reclarify for patients who appear mystified (an important skill, but not the subject of this study). The first author reviewed audio recordings of each standardized patient before the next workshop to ensure that the standardized patients were holding to these instructions. Little or no feedback to the patients was needed after these reviews.

Tapes were transcribed verbatim and proofread for accuracy by a board-certified internist. The proofreaders deleted personally identifying text about the residents to reduce abstractor bias. The proofreader also ensured that each transcript included all of the standardized patient's nonverbal continuers ("uh-huh") and the resident's pauses for the patient's response.

To provide a content-related unit of duration within the transcript, we used a sentence-diagramming procedure to parse transcripts into individual "statements" which were defined as having one subject and one predicate. An example of parsed text is shown in Figure 1, with each statement beginning at a number in pointed brackets { } and ending with a double slash //. Single slashes / and a quadruple slash //// are used to mark divisions or the end of compound statements.

The final sample for analysis consisted of 86 transcripts (41 for prostate cancer screening and 45 for breast cancer screening).

2.2. Abstraction procedures

The transcript abstraction procedure was adapted from explicit-criteria methods used in medical record review [34]. Abstractors evaluated each statement one at a time, comparing it to detailed definitions and examples in an explicit-criteria data dictionary derived from existing guidelines [2,35,36]. To allow for ambiguous statements, abstractors could use “definite” and “partial” designations. For a “definite” designation, the AU had to be followed by a pause for patient response, even for a simple continuer like “uh huh.” A “partial” AU designation was used for AUs that were not followed by a pause, or for AUs that used a leading syntax (e.g., “You don’t have any questions, do you?”). Discrepancies between abstractors were automatically resolved to “partial” by a spreadsheet. Half of the transcripts were duplicatively abstracted to assess inter-abstractor reliability. Half of these were discussed afterwards to ensure quality control and consistency, following the suggestion by Feinstein [37].

A second wave of abstractions was done to identify content-containing statements for the mental load procedure (see section 2.3.1), using procedures that were similar to the AU abstraction procedures and our previously described work with content message and content timing assessment [22,23].

2.3. Analyses

Statistical analyses were performed using JMP software (SAS Institute, Cary, NC). One-way ANOVA and Chi-squared tests were used as appropriate for variable type. Inter-abstractor reliability was calculated using Cohen’s method, with an adaptation for ordinal data.

2.3.1. Calculation of the “estimated mental load” ($\hat{L}$) of potential misunderstanding—

We adapted the mental load concept to represent the amount of information presented before, between, and after each individual AU. It should be noted that while assessing understanding is a patient-centered behavior for clinicians [3], the mental load calculation procedure we developed is clinician-centered because it only examines clinician speech. Mental load is therefore an indicator of communication quality regardless of whether the patient makes use of the opportunity. We include the descriptor “estimated” in the current project because there are limited data available to suggest a procedure for the quantification of the informational content present in any given statement. Following the standard convention, the estimated nature of the load variable is denoted by a circumflex accent or caret placed above the variable letter in equations ($\hat{L}$) or before the variable letter in typeface ($\hat{}L$).

The estimated mental load for any given statement *“S”* ($\hat{}L_S$) reflects both the informational content presented in the statement (C_S) and the information left over from the previous statement ($\hat{}L_{S\text{-}1}$). To standardize the present analysis, we made a provisional assumption that *Cs* can either equal 1 for any information-giving statement, or 0 for any AU or other statements about the conversation or the patient’s cognitive or emotional reaction to the conversation. We also assume C_S is zero for relationship-building, small talk, nonverbal continuers, and stalling statements (e.g., “that type of thing,” statement 54 in Figure 1).

We also assume for standardizing purposes that load before the first statement ($\hat{}L_0$) is negligible.

Thus for a simple sequence of 4 information-giving statements with no AUs, $\hat{}L$ would be calculated via cumulative addition as follows:

$$\widehat{L}_{1to4}=\sum_{S=1}^{end}\left(\widehat{L}_{S-1}+C_S\right)=(0+1)+(1+1)+(2+1)+(3+1)=1+2+3+4=10$$

If a sequence of statements includes an AU question, the patient has an opportunity to have clarification about statements presented thus far in conversation. To account for the effect of AUs on the accumulating $\hat{}L_{total}$, a "discounting function" D is applied to the $\hat{}L_S$ formula, with D_S provisionally modeled to be 1 if S is not an AU, 0 if S is a definite-criteria AU, or 0.5 if S is a partial AU or an "OK?" question. The load formula for any individual statement S is therefore

$$\widehat{L}_S=D_S\times\left(\widehat{L}_{S-1}+C_S\right)$$

and $\hat{}L_{total}$ for a set of statements is then equal to the sum of all the $\hat{}L_S$ values in the set, or

$$\widehat{L}_{total}=\sum_{S=1}^{end}\widehat{L}_S=\sum_{S=1}^{end}\left[D_S\times(\widehat{L}_{S-1}+C_S)\right]$$

$\hat{}L_S$ undergoes quadratic growth with the number of statements (see Appendix), so $\hat{}L_{total}$ may appear better for short transcripts with no AUs than for long transcripts with excellent AU coverage. We therefore developed a standardized version of the load calculation ($\hat{}L_{Std}$) to compare load regardless of transcript length or detail. $\hat{}L_{Std}$ is based on the ratio of $\hat{}L_{total}$ to the maximum possible load for that transcript if there were no AUs (i.e., D_S is removed from the formula), with a square root transform added to compensate for the quadratic distribution of $\hat{}L_{total}$ (see Appendix).

$$\widehat{L}_{Std}=\sqrt{\frac{\widehat{L}_{total}}{\text{max.possible }\widehat{L}_{total}}}\quad\text{or,}\quad\sqrt{\frac{\sum_{S=1}^{end}\left[D_S\times\left(\widehat{L}_{S-1}+C_S\right)\right]}{\sum_{S=1}^{end}\left(\widehat{L}_{S-1}+C_S\right)}}$$

When evaluating two transcripts the clinician with the higher $\hat{}L_{Std}$ assessed understanding for less of his or her content than the clinician with the lower $\hat{}L_{Std}$. A very low value could suggest repetitive assessment of understanding, as with a speech mannerism where the clinician asks "OK?" every third or fourth statement.

To demonstrate the $\hat{}L_{total}$ and $\hat{}L_{Std}$ calculations, Figure 1 shows a transcript excerpt of 35 statements that includes 30 information giving statements (C_S =1) and three AUs: a close-ended AU at statement {55}, an "OK?" question at statement {60}, and a request for a teach-back at statement {80}. As shown in the table at the right, $\hat{}L_S$ accumulates at a rate of 1 per statement until reset to zero by the AU at statement {55}. $\hat{}L_S$ then starts again at 1 for statement {56}, accumulates until reset by half for the partial AU at {60}, accumulates again until reset back to zero for the request for teach-back at the end of the excerpt. The value of $\hat{}L_{total}$ for the excerpt is then equal to the sum of the individual $\hat{}L_S$ values,

$\hat{L}_{total}$ = (1 + … + 8) + 8 + (1 + … + 4) + 2 + 3 + (3 + … + 20) = 266, and while $\hat{L}_{Std}$ is equal to the square root of $\hat{L}_{total}$ divided by the sum of each of the numbers 1 to 35,

$$\widehat{L}_{Std} = \sqrt{\frac{\widehat{L}_{total}}{(1+2+\ldots+34+35)}} = \sqrt{\frac{266}{630}} = 0.65$$

Some additional implications of the mental load calculation are listed in the Appendix. It is worth emphasizing here that while the current assumptions about D_S and C_S are provisional, we anticipate the development of mathematical functions to reflect variance of D_S with each patient's state of mind, emotion, health literacy and prior knowledge, as well as with the complexity of the topic being discussed. These factors may also influence the optimal values for $\hat{L}_{Std}$ and $\hat{L}_{total}$.

3. Results

Descriptive data for the participants (Table 2) were similar to those of the population of the residency program at the time of the study. The interviews ranged from 2 to 21.9 minutes (mean 10.1, SD 4.1), and included an average number of 147.4 statements per transcript (SD 56.7). The average number of statements that were abstracted to be information-giving (C_S = 1) for the $\hat{L}$ calculations in section 3.3 was 120.2 (SD 52.1). To assess feasibility of our methods we tracked time and expenses. The entire project was done for less than $50 per transcript, most of which was for transcription. Use of the abstraction instrument took less than 5 minutes of abstractor time per transcript.

A total of 213 definite AUs and 288 partial AUs were identified over the entire set of transcripts. Inter-abstractor agreement for these was better than 99%, and the reliability corrected for chance for overall identification of AUs was κ=0.96. Reliability for individual AU types were 0.90 for close-ended AUs and 0.98 for "OK?" questions, and lower for the more rare requests for a teach-back (0.88) and open-ended AUs (0.58).

3.1. Number and types of AUs included

At least one definite AU was seen in 68/86 transcripts (79%), leaving about one-fifth of transcripts with no evidence of a definite attempt to assess patient understanding. Among those transcripts with AUs, the average number of AUs included was 3.0 per transcript (SD 2.6, range 1 to 18).

The transcripts where prostate cancer screening was discussed were more likely than breast cancer transcripts to include at least one definite-criteria AU (90% versus 69%, p = 0.017). There were no statistically significant associations between AU presence or AU number and the resident's gender or year in residency, or the duration of counseling (power was 80% to detect a difference of 1.6 AUs).

The percentages of transcripts meeting definite criteria for each AU type are shown in Table 3. The two most effective AU types (requests for teach-backs and open-ended AUs) were rarely included in the residents' counseling. Transcripts more commonly included the less effective close-ended and "OK?" question AUs. The "OK?" question was the sole type of AU in 18/86 transcripts (20.9%).

One transcript included a statement that met criteria for an advance request for questions.

3.2. Effect of adding partial abstraction criteria

As mentioned in section 2.3.1, the definite/partial/absent scheme helped to differentiate between AUs and statements that may have been intended to assess understanding but were less likely to be effective. When abstractors' partial ratings were also considered, at least one AU "attempt" was seen in an additional 5 transcripts, leaving 15% of transcripts that did not even meet partial criteria for AU inclusion. Among those transcripts with AUs, the average number of AUs increased to 6.7 per transcript (SD=8.4, range 1 to 54). The AU type with the most new inclusions from the addition of partial-criteria was the "OK?" question, which was seen with no pause for response in 40 transcripts (46.5%) with a range of up to 40 instances in a single transcript.

3.3. The estimated mental load ($\hat{L}$) of potential misunderstanding

The average $\hat{L}_S$ value across the entire project was 43.2 (SD 38.4, skew 1.2). This figure suggests that on average, patients were asked **to process** about 43 concepts in their minds between clinician attempts to assess understanding. The $\hat{L}_{total}$ scores, which represent the accumulated potential for misunderstanding with no adjustment for transcript length, averaged 6341.4 (SD=8405.5, range 420.75 to 29,556.5). The $\hat{L}_{total}$ scores have no units, but are proportionate to mental workload and are comparable across transcripts (See Appendix for further discussion about the quadratic growth curve for $\hat{L}_{total}$.).

The $\hat{L}_{Std}$ score, which adjusts $\hat{L}_{total}$ for transcript length and its quadratic curve, averaged 0.80 (SD 0.16) with a range from 0.46 to a total lack of AUs at 1.0. This suggests that some transcripts contain dispersal of AU usage, but that there are many other transcripts with longer stretches of content without an AU. Also as shown in Figure 2, $\hat{L}_{Std}$ scores were significantly lower for transcripts in which prostate cancer was being discussed than for mammography screening transcripts (0.76 versus 0.84, 1-way ANOVA, p=0.016). This difference is consistent with both a lower usage of AUs for mammography explanations and a separate tendency to bunch up those AUs that are used.

A histogram of the $\hat{L}_{Std}$ scores (Figure 2, total bar height) shows wide variety with a peak of transcripts in the ranges from 0.4 to 0.59. The differences for prostate versus mammography transcripts lead us despite the small sample size to investigate the possibility of a bimodal distribution for $\hat{L}_{Std}$. The divided histogram (Figure 2) suggests that the midrange peak in the main distribution corresponds with the mean of a roughly symmetrical $\hat{L}_{Std}$ distribution for prostate cancer screening transcripts (skew -0.1, kurtosis -0.8), while the mammography transcripts' $\hat{L}_{Std}$ scores may be consistent with an L-shaped distribution with most transcripts scoring at the high range.

4. Discussion and conclusion

4.1. Discussion

Assessments of understanding help clinicians to know if their communication efforts with patients have been effective [2-4]. The two purposes of this study were to introduce a new method of quantifying AU behaviors that will be suitable for use on a population-wide scale, and to use the new method to investigate internal medicine residents' use of AU questions during counseling of standardized patients about prostate and breast cancer screening. We aimed to characterize AU usage both at a transcript level and at a level of individual statements.

Many of the residents' transcripts included definite criteria for at least one AU. Unfortunately, almost all of the AUs identified were close-ended, used a leading syntax, or were not followed by a pause for patient response. This finding parallels our earlier study of AU usage after newborn screening [6], but for this study we also found that most of the residents' AUs were

either bunched inefficiently together or rare enough to leave much of the conversation unassessed. Generalizability for these analyses is limited by the use of a small sample of residents from a single program, so further research at other programs and with physicians working outside of academic settings should be conducted before more specific conclusions can be made. The field will also need research about communication topics other than cancer screening, and scenarios in which patients do not initiate the topic of conversation and could therefore be less engaged (see section 2.1).

For this study we coached standardized patients to adopt neutral expressions, but for future research it may be useful to evaluate explicit AU usage in datasets including video of actual patients' facial expressions. Clinicians can use facial expressions to verify understanding [2-5], so including facial expressions in analyses may help to explain to researchers why some conversations do not include explicit AUs. For everyday practice and for the primary quality measure, however, clinicians and researchers should continue to focus on explicit AUs because patients' facial expressions may not correlate with understanding [4,5].

The most important limitation of the current state of our methods is that our quantitative approach of counting up individual AUs has not been validated against a measure of patient comprehension. Previous AU research has been done with ordinal scales, which are largely unable to account for AU type or timing [8,9]. Until research such as our ongoing validation study is complete [38], questions will remain about the number of AUs that should be necessary for a "high quality" rating. Further research can also be done to tailor values of D_S and C_S for patients' health literacy, patients' engagement, and emotional state as well as for the complexity of the topic and proximity of other AUs in conversation. Many of these influences may interact with patient factors such as health literacy or environmental factors like distracting noise.

Despite these cautions, our research into patterns of AU usage raises concern about the quality of communication likely to be experienced by many patients. Poor communication will be a problem for prostate and breast cancer screening, since patients without a clear understanding of screening may opt for or against screening when the opposite plan would have been more consistent with their values [10-14,16]. Fortunately, the finding that many residents were at least attempting to assess the patients' understanding suggests to us that some clinicians may simply need feedback about how to do it effectively.

We see this project as contributing to both the existing scholarship on patient-centered communication and to a new dimension of quality improvement. Population-scale improvements in clinicians' communication are needed and part of this need is to improve the usage of AUs. While many clinicians' training programs may have encouraged them to use AUs [36,39,40], without periodic reinforcement it is unclear whether AU behaviors will persist for long after training. Reinforcement for practicing clinicians can be done by faculty development and refresher programs [41,42], but low levels of enrollment may reflect the programs' high requirements for time, effort, and money. With Quality Improvement methods come several advantages over education-derived methods, including a track record for improving other clinician behaviors on population scales, high quantitative reliability, the ability to function on a lean budget, and concrete methods that are transparent to clinicians and easy to implement by quality improvement professionals with typical training [1,27].

It is worth commenting that the advantages of population-scale Communication Quality Assurance may be accompanied by some further limitations for study of communication. Qualitative methods would provide a richer description of communication, but qualitative methods have limited reliability and are very resource-intensive for use in Quality Improvement. Using simulated patient encounters instead of actual patient encounters may reduce generalizability because of artificial circumstances or because a sense of observation

prompts clinicians to greater efforts. On the other hand, we included patient simulation in our data collection methods because we are aiming to develop techniques that can be immediately, affordably generalized for widespread use in settings where communication improvement has not previously been feasible. Simulation avoids logistical, privacy, and consent issues that would pose tremendous challenges for routine quality improvement projects. Finally, the standardization that is possible with simulation allows for comparisons across clinicians to be made fairly and equally without confounding effects of variation in patients' medical risks and communication challenges.

4.2. Conclusion

The data from this project are preliminary, but they suggest that problems with AU usage may exist in some clinician-patient encounters. This study offers part of a low-cost, quantitatively reliable method to measure processes of communication and clinicians' adherence to recommendations for counseling prior to offering complex screening tests and other instances of shared decision-making. Further research and development needs to be done, but methodological innovations hold significant promise for a nascent field of population-scale Communication Quality Assurance.

4.3. Practice implications

We hope that our observations of an unfortunately high number of missed opportunities to assess patients' understanding will expand awareness in clinicians and medical educators about AUs and how to improve their use. Improvements in AU usage and effectiveness may improve the informed decision-making that guidelines recommend.

Appendix

Appendix

In this paper we have proposed a new method to quantify mental load of potential misunderstanding in a conversation, or what we believe to be the next methodological step in an innovative approach to communication process measurement. Our goal in developing this new method was to have a group of single numbers that can be used with a large number of clinicians to compare and rank performance and also to track changes after an intervention. Our development of the load calculations (Appendix Table) is the result of several cycles of trial and evaluation. To achieve our single number goal for AU timing, we found it necessary to incorporate a quantitative, algebraic procedure that is more elaborate than many previously described methods in the field of health care communication. During the development process, we identified several methodological points that will be important to colleagues and to the field in general, but which were not immediately relevant to the aims of our analysis. The purpose of this appendix is to provide brief commentary on each of these points.

1. More about the provisional values of mental load calculation variables

We envision several possibilities for the refinement of C_S, D_S, and L_0. The current study used standardized patients in order to evaluate feasibility for Communication Quality Assurance projects. As a result, the current load calculation procedure can currently provide descriptive data of communication process, and has no ability to infer outcomes like patient understanding. Further research with both communication processes and outcomes will be needed to validate and refine the method. Some answers may be derived from our ongoing statewide study of counseling after cystic fibrosis and sickle cell hemoglobinopathy [23]. We plan to tailor the provisional assumptions mentioned in section 2.3.2 for the effects of content message complexity (C_S), varying effectiveness of the different AU types (D_S). Finally, since the amount

and content of underlying beliefs about a topic is critical [16], further research is needed to evaluate the influence of prior mental load (L_0) on load calculations.

2. Quadratic growth of $\hat{L}_{total}$

Under the provisional assumptions for values of C_S, D_S, and $\hat{L}_0$, the $\hat{L}_{total}$ function grows via cumulative addition following a quadratic curve that can be described by the equation,

$$\widehat{L}_{total}=0.5(x^2+x)$$

where x is equal to the current statements' number in the transcript. Appendix Figure shows a plot of the resulting curve, called the "ceiling curve" because it describes the maximum possible $\hat{L}_{total}$ for a set of statements if none of the statements is an AU. For comparison, Appendix Figure also shows the $\hat{L}_{total}$ values for each of the transcripts in this study.

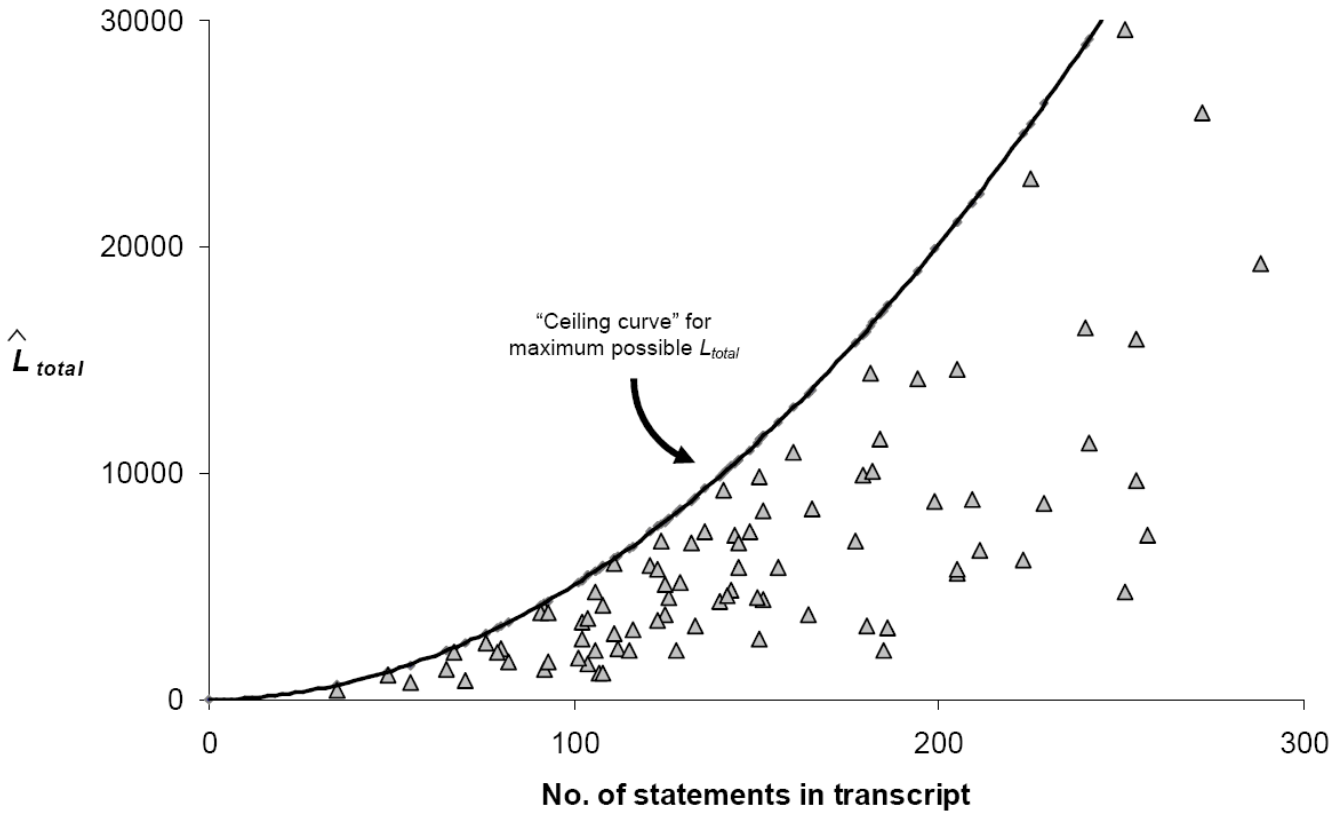
The quadratic growth of $\hat{L}_{total}$ is an important issue to recognize because of the resulting implication that differences between clinicians in AU usage and load may have different significances at longer versus shorter conversations. Further research and validating efforts need to be completed, but the current implications are consistent with the decreasing likelihood of overall understanding as the number of concepts increases.

3. The square root transform

The purpose of the square root transform function in the $\hat{L}_{Std}$ formula is to optimize mental load estimates data so that they can be more comparable across clinicians, such that differences between transcripts will be similar at different ends of the $\hat{L}_{Std}$ spectrum. The square root enables the $\hat{L}_{total}$ functions in the numerator and denominator of the $\hat{L}_{Std}$ formula to approach linearity instead of following a quadratic curve. As a result, the absolute difference in $\hat{L}_{total}$ from an $\hat{L}_{Std}$ of 0.9 to 0.8 is the same as the difference from an $\hat{L}_{Std}$ of 0.5 to 0.4. Without the square root, the relative difference in $\hat{L}_{total}$ would be significantly greater as $\hat{L}_{Std}$ approaches zero.

4. A potential ceiling for $\hat{L}_S$

To standardize comparisons across residents, the current version of the load calculation assumes that that there is no maximum value for $\hat{L}_S$. Put another way, load operationalizes the amount of information that the patient is asked to consider, rather than the amount of information that he or she actually does consider. Load will therefore accumulate long after the patient's cognitive capacity has been overwhelmed. In a subsequent version of the load calculation, it may become useful to acknowledge that $\hat{L}_S$ reaches a ceiling, such that patients can only think about a small number of new or unfamiliar concepts at any one time. Such a ceiling is consistent with research surrounding what is now called "Miller's Magic Number" of 7 ± 2, the number of data that the human mind appears to be capable of actively considering or remembering at any one time [31]. If such a ceiling for patient comprehension is observed, however, our unlimited cumulative addition approach to mental load would still be valuable for purposes of standardized comparison across transcripts.



Appendix Figure.
Comparison of total load for residents' transcripts versus the "Ceiling curve"

Appendix Table

Legend to key terms and abbreviations

Term	Meaning
AU	Assessment of understanding
statement	String of text containing one subject and one predicate
C_S	Quantitative value of informational content present in a statement.*
D_S	Discounting function applied to cumulative load calculations to reflect the resetting effect of an AU.*
$\hat{L}_S$ (or $\hat{L}_S$)	Estimated** mental load present in a statement 'S', corresponding with the number of concepts the patient is asked to keep in mind during the statement.
$\hat{L}_0$ (or $\hat{L}_0$)	Mental load about the topic prior to initiation of conversation. Corresponds with the patient's prior misunderstanding or questions.
$\hat{L}_{total}$ (or $\hat{L}_{total}$)	Sum of *LS* terms over an entire transcript (i.e. S=1 to S=last statement)
$\hat{L}_{Std}$ (or $\hat{L}_{Std}$)	Standardized mental load ($\hat{L}_{total}$ adjusted for transcript length to facilitate comparison of clinicians)

* In the current analysis, C_S and D_S have been assigned provisional values.

** A caret ^ denotes that L_S is an estimate of true mental workload. The estimate descriptor may be removed when the C_S and D_S functions have been assigned nonprovisional values.

Acknowledgments

MF is supported in part by grants K01HL072530 and R01HL086691 from the National Heart, Lung, and Blood Institute. When the project was begun, PK was a medical student in the Yale University School of Medicine in New Haven, CT.

References

1. Donabedian A. The quality of care. How can it be assessed? JAMA 1988;260:1743–8. [PubMed: 3045356]
2. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med 2001;76:390–3. [PubMed: 11299158]
3. Epstein, RM.; Street, RL, Jr. Patient-Centered Communication in Cancer Care: Promoting Healing and Reducing Suffering. Monograph from the National Cancer Institute; Bethesda, MD, USA: 2007. [cited 3/11/2009]; Available from: http://outcomes.cancer.gov/areas/pcc/communication/monograph.html
4. Nielsen-Bohlman, L.; Panzer, A.; Kindig, D., editors. Health Literacy: A Prescription to End Confusion. National Academies Press; Washington, D.C: 2004.
5. Bensing JM, Kerssens JJ, van der Pasch M. Patient-directed gaze as a tool for discovering and handling psychosocial problems in general practice. Journal of Nonverbal Behavior 1995;19:223–242.
6. Farrell MH, Kuruvilla P. Assessment of parental understanding by pediatric residents during counseling after newborn genetic screening. Arch Pediatr Adolesc Med 2008;162:199–204. [PubMed: 18316655]
7. Lerman C, Daly M, Walsh WP, Resch N, Seay J, Barsevick A, Birenbaum L, Heggan T, Martin G. Communication between patients with breast cancer and health care providers. Determinants and implications. Cancer 1993;72:2612–20. [PubMed: 8402483]
8. Braddock CH 3rd, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. JAMA 1999;282:2313–20. [PubMed: 10612318]
9. Stillman PL, Sabers DL, Redfield DL. The use of paraprofessionals to teach interviewing skills. Pediatrics 1976;57:769–74. [PubMed: 940718]
10. Lim LS, Sherin K. Screening for prostate cancer in U.S. men ACPM position statement on preventive practice. Am J Prev Med 2008;34:164–70. [PubMed: 18201648]
11. Board of Directors of the American Urological Association. AUA Policy Statement on Early Detection of Prostate Cancer. 2006. [cited 3/11/2009]; Available from: http://www.auanet.org/content/guidelines-and-quality-care/policy-statements/e/early-detection-of-prostate-cancer.cfm
12. Screening for prostate cancer: recommendation and rationale. Ann Intern Med 2002;137:915–6. [PubMed: 12458992]
13. Smith RA, Cokkinides V, Eyre HJ. Cancer screening in the United States, 2007: a review of current guidelines, practices, and prospects. CA Cancer J Clin 2007;57:90–104. [PubMed: 17392386]
14. Qaseem A, Snow V, Sherif K, Aronson M, Weiss KB, Owens DK. Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2007;146:511–5. [PubMed: 17404353]
15. Ransohoff DF, Harris RP. Lessons from the mammography screening controversy: can we improve the debate? Ann Intern Med 1997;127:1029–34. [PubMed: 9412285]
16. Farrell MH, Murphy MA, Schneider CE. How underlying patient beliefs can affect physician-patient communication about prostate-specific antigen testing. Eff Clin Pract 2002;5:120–9. [PubMed: 12088291]
17. Farrell MH, Makoul GT, Christopher SA. A Regression-Based Approach to Combining Concepts from Disparate Theories for Communication Quality Assurance. 2008manuscript in preparation
18. Farrell MH, Makoul GT. Communication Quality Assurance: Parameters for Assessing and Improving Clinician-Patient Communication on a Population Scale. 2008manuscript in preparation
19. Farrell MH, Chan ECY, Ladouceur L, Stein JM. Comparison of Expert Recommendations and Medicine Residents' Efforts for Informed Consent before Prostate-Specific Antigen Screening. 2008submitted manuscript

20. Farrell M, Deuster L, Donovan J, Christopher S. Pediatric residents' use of jargon during counseling about newborn genetic screening results. Pediatrics 2008;122:243–9. [PubMed: 18676539]
21. Deuster L, Christopher S, Donovan J, Farrell M. A Method to Quantify Residents' Jargon Use During Counseling of Standardized Patients About Cancer Screening. J Gen Intern Med 2008;23:1947–52. [PubMed: 18670828]
22. La Pean A, Farrell MH. Initially misleading communication of carrier results after newborn genetic screening. Pediatrics 2005;116:1499–505. [PubMed: 16322177]
23. Farrell MH, La Pean A, Ladouceur L. Content of communication by pediatric residents after newborn genetic screening. Pediatrics 2005;116:1492–8. [PubMed: 16322176]
24. Kindig D, Stoddart G. What is population health? Am J Public Health 2003;93:380–3. [PubMed: 12604476]
25. Cegala DJ, Lenzmeier Broz S. Physician communication skills training: a review of theoretical backgrounds, objectives and skills. Med Educ 2002;36:1004–16. [PubMed: 12406260]
26. Hulsman RL, Ros WJ, Winnubst JA, Bensing JM. Teaching clinically experienced physicians communication skills. A review of evaluation studies. Med Educ 1999;33:655–68. [PubMed: 10476016]
27. Jencks SF, Huff ED, Cuerdon T. Change in the Quality of Care Delivered to Medicare Beneficiaries, 1998-1999 to 2000-2001. JAMA 2003;289:305–12. [PubMed: 12525231]
28. Donovan, JJ.; Farrell, MH.; Deuster, L.; Christopher, SA. "Precautionary empathy" by child health providers after newborn screening. Poster at International Conference on Communication and Health Care; Charleston, SC. 2007.
29. Donovan J, Deuster L, Christopher SA, Farrell MH. Residents' precautionary discussion of emotions during communication about cancer screening. Poster at 2007 annual meeting of the Society for General Internal Medicine. 2007
30. Graber, DA. The Theoretical Base: Schema Theory. In: Graber, DA., editor. Processing the News: How People Tame the Information Tide. Longman; New York: 1988. p. 27-31.
31. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 1994;101:343–52. [PubMed: 8022966]
32. Seel N. Mental Models in Learning Situations. Advances in Psychology 2006;138:85–110.
33. White JD, Carlston DE. Consequences of Schemata for Attention, Impressions, and Recall in Complex Social Interactions. Journal of Personality & Social Psychology 1983;45:538–49. [PubMed: 6620125]
34. Allison J, Wall T, Spettell C, Calhoun J, Fargason CJ, Kobylinski R, Farmer R, Kiefe C. The art and science of chart review. Jt Comm J Qual Improv 2000;26:115–136. [PubMed: 10709146]
35. Silverman, J.; Kurtz, S.; Draper, J., editors. Skills for Communicating with Patients. Radcliffe Medical Press; Abingdon, Oxon, UK: 1998.
36. Smith, RC., editor. Patient-Centered Interviewing: An Evidence-Based Method. Vol. second. Lippincott, Williams and Wilkins; Philadelphia: 2002.
37. Feinstein, A. Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia: WB Saunders; 1985.
38. Farrell, MH. R01 HL086691, Improvement of communication process and outcomes after newborn genetic screening. National Heart, Lung, and Blood Institute; Medical College of Wisconsin: 2008.
39. Coulehan, JL.; Block, MR. The medical interview : mastering skills for clinical practice. Vol. 5. Philadelphia: F.A. Davis Co; 2006. p. xixp. 409
40. Kurtz, S.; Silverman, J.; Draper, J. Teaching and Learning Communication Skills in Medicine. Abingdon, Oxon, U.K.: Radcliffe Medical Press; 1998.
41. Bylund CL, Brown RF, di Ciccone BL, Levin TT, Gueguen JA, Hill C, Kissane DW. Training faculty to facilitate communication skills training: Development and evaluation of a workshop. Patient Educ Couns 2008;70:430–6. [PubMed: 18201858]
42. Lang F, Everett K, McGowen R, Bennard B. Faculty development in communication skills instruction: insights from a longitudinal program with "real-time feedback". Acad Med 2000;75:1222–8. [PubMed: 11112727]

Doc: {46}{47}{48}{49} Um, with the mammogram, um, you go in /{50} -it's a pretty quick procedure/{46 cont} -you go and, um, disrobe,/{47 cont} you-they-they actually use a machine which actually squeezes your breast tissue between two plates /{48 cont}{49 cont} and they /{48 cont} shoot x-rays /{49 cont} and do it bi-laterally.////{51} They actually do it from-they can do it from multiple sides as well.//{52} After that, they send off to radiologists //{53} and it probably takes maybe five minutes to do,/{54} that type of thing.////{55} Is that clear? //

Pt: Sure.

Doc: {56}{57} I mean, some women /{56 cont} say it's uncomfortable/{57 cont} they don't like getting it /{58} but they do it /{59} because they know it's, uh, important in their life.////{60} OK?//

Pt: Well, I think I basically understand the-the procedure-itself.

Doc: {61} It's more-it's more of the why you should be screened.//{62} So, to start off, um,/{63} there is a lot of controversy between forty-in forty to forty-nine-//{64} and what has been recommended is screening of women fifty and above-fifty to like sixty-nine.//{65} That's basically from some studies of a large group of women //{66}{67} and they looked at,/{66 cont} um, mammography /{67 cont} and the risk of developing breast cancer after-after screening /{68} to see how many women developed it.//// {69} And it seemed from that study that /{70}{71} women /{70 cont} between fifty and sixty-nine had -/{71 cont} if they were screened -/{70 cont}{71 cont} had a reduction - in developing, um, breast cancer, um, that was incurable.////

Pt: Ok.

Doc: {72}{73} Um, now the forty to forty-nine, um,/{72 cont} it wasn't as clear /{73 cont} and there was a big controversy for a while there.////{74} Some people think that it isn't as worthwhile,//{75} and some people just recommend it.////{76} A lot of my patients in their forties actually ask about this,//{77} and I tell them its fine to get it.//{78}{79} I just want to make sure that they know that /{78 cont} there's not as much evidence for a benefit,/{79 cont} and that there might be a good enough result to know what it means.////{80} Um, so when you hear all this, what is it saying to you?//

Stmt	C_S	$\hat{L}_S$
46	1	1
47	1	2
48	1	3
49	1	4
50	1	5
51	1	6
52	1	7
53	1	8
54	0	8
55 (close-ended AU)		0
56	1	1
57	1	2
58	1	3
59	1	4
60 ("OK?" question)		2
61	1	3
62	0	3
63	1	4
64	1	5
65	1	6
66	1	7
67	1	8
68	1	9
69	1	10
70	1	11
71	1	12
72	1	13
73	1	14
74	1	15
75	1	16
76	1	17
77	1	18
78	1	19
79	1	20
80 (request for teachback)		0
L_{total}		266

Figure 1.
Transcript excerpt to illustrate AUs and load (ˆLS) calculation

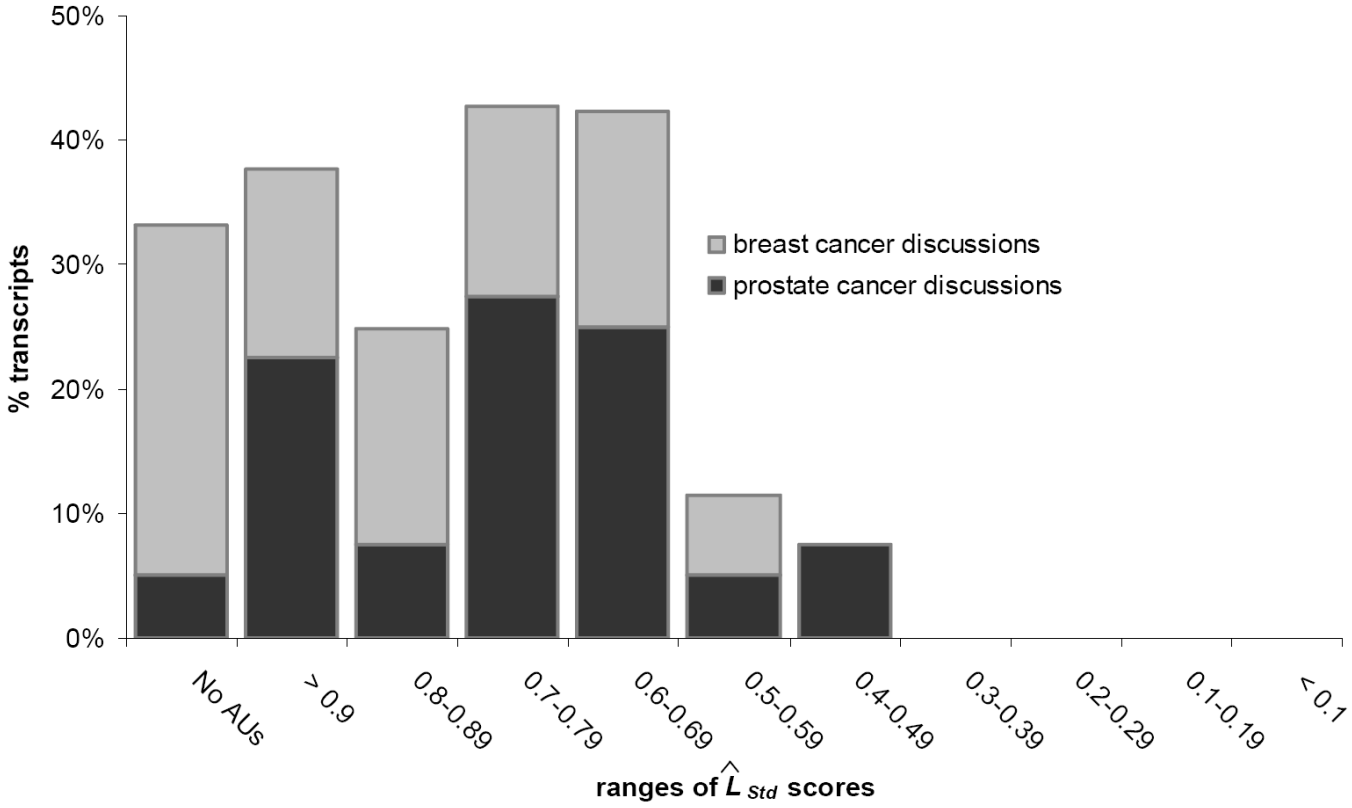


Figure 2.
Histogram of standardized estimated mental load ($\hat{L}_{Std}$) scores for total sample of transcripts, subdivided into two subset samples of prostate cancer discussions and breast cancer discussions

Table 1

Types of Assessments of Understanding (AU) Questions

AU type and definition	Examples
Request for teach-back	
A question whose natural answer is the patient's own version of the information covered in conversation. Some clinicians include an explanation of why he or she is requesting the teach-back.	*To make sure that I was being clear enough, could you please share with me the main points you got from our discussion?* *You may have to discuss this issue with your wife, so lets practice now how you will explain this to her.*
Open-ended AU (effective)	
An AU question that places few or no restrictions on response.	*What questions do you have?*
In many cases the natural answer will be an actual question.	*What parts of this are hard to understand?*
Close-ended AU (less effective)	
AUs to which the natural answer is restricted, such as a "yes," a "no," or some other brief, limited answer.	*Do you have any questions?* *Does that make sense?*
"OK?" question (ambiguous- could be interpreted in many ways)	
Close-ended AU that does not specifically mention understanding or a question. The patient may interpret this AU type as a question about emotions or as a request for permission to proceed. Must include a rising voice pitch signifying an interrogative.	*OK?* *Alright?*

Patient Educ Couns. Author manuscript; available in PMC 2010 October 1.

Table 2

Participant Characteristics

	No. responding	(%)
Gender		
Male	20	(42)
Female	28	(58)
Age*		
25-29 years	26	(55)
30-37 years	21	(45)
Year in residency		
1st	10	(20)
2nd	19	(40)
3rd (or 4th year med-peds)	19	(40)

* one resident did not answer the age question

Table 3

Number of transcripts containing Definite Criteria for at Least One Assessment of Understanding*

AU inclusion	Transcripts No.	*(% total)*
No AUs at all	18	*(20.9)*
At least one AU	68	*(79.1)*
AU type		
Request for a teach-back	2	*(2.3)*
Open-ended	2	*(2.3)*
Close-ended	48	*(55.8)*
"OK?" question	48	*(55.8)*

* Many transcripts contain more than one type of AU