ELSEVIER

Patient Education and Counseling 45 (2001) 23–34

Patient Education and Counseling

www.elsevier.com/locate/pateducou

The SEGUE Framework for teaching and assessing communication skills☆

Gregory Makoul*

Program in Communication & Medicine, Northwestern University Medical School, 750 North Lake Shore Drive (ABA 625), Chicago, IL 60611, USA

Abstract

This article examines uses and characteristics of the SEGUE Framework, a research-based checklist of medical communication tasks. A recent survey of US and Canadian medical schools indicates that the SEGUE Framework is the most widely used structure for communication skills teaching and assessment in North America. Student and faculty response to the SEGUE Framework as a teaching tool has been positive. Data drawn from clinical skills assessments with standardized patients provide evidence of concurrent and construct validity. Analysis of visits between general internists and their patients reinforces validity of the SEGUE Framework in an actual practice setting. Interrater reliability is high when standardized patients are recording student performance immediately after a live encounter, and when coders are evaluating videotaped or audiotaped encounters; intrarater reliability is strong as well. The SEGUE Framework has a high degree of acceptability, can be used reliably, has evidence of validity, and is applicable to a variety of contexts. Studies of predictive validity are needed.

Keywords: Communication skills; Clinical skills; Physician–patient communication; Medical education; Educational measurement; Evaluation

1. Introduction

In the 1980s, calls for improving physician–patient communication came from clinical, academic, professional, and popular circles [1–5]. While most US medical schools responded to these calls by implementing or enhancing relevant training for their students, a 1992 survey of medical schools documented ''wide variability in the quality and intensity of medical interviewing and interpersonal skills teaching'' [6]. Authors of the survey report suggested that employing checklists in the evaluation of communication skills would be an important step toward improving the quality of training. A follow-up survey indicated that by 1998, less than a third of the medical schools in North America had incorporated a checklist or similar tool into their communication skills teaching or assessment activities [7]. Instruments such as checklists, frameworks, guides, or models that are designed to be reliably used in multiple contexts can provide a structural foundation for efforts to teach, assess, study, and improve the communication skills of both physicians-in-training and physicians-in-practice.

This report details 7 years of experience with the SEGUE Framework, a checklist of medical communication tasks that is designed to facilitate the teaching and assessment of communication skills, as well as research on doctor–patient communication (see Appendix A). The checklist employs a nominal (i.e. yes/no) scale to allow both participants and observers to evaluate whether or not a physician or medical student accomplishes critical communication tasks. In addition to serving as a reminder of the general areas on which to focus (i.e. *S*et the stage, *E*licit information, *G*ive information, *U*nderstand the patient's perspective, *E*nd the encounter), the SEGUE acronym connotes the general transition or flow of the medical encounter: from beginning to end, from problems to solutions.

1.1. The task approach

The SEGUE Framework targets specific communication tasks, focusing on whether or not they were accomplished during a medical encounter. As noted by Makoul and Schofield [8]

☆ Segments of this paper have been presented to Division I of the American Educational Research Association (San Francisco, 1995), the American Academy on Physician and Patient (San Diego, 1995), the Teaching about Communication in Medicine Conference (Oxford, 1996), the Association of American Medical Colleges' Central Group on Educational Affairs (Chicago, 1998), the Eighth International Ottawa Conference on Medical Education and Assessment (Philadelphia, 1998), and the Communication in Medicine Conference (Chicago, 1999).

* Tel.: +1-312-503-6728; fax: +1-312-503-0574.
E-mail address: makoul@northwestern.edu (G. Makoul).

0738-3991/01/$ – see front matter
PII: S0738-3991(01)00136-7

Focusing on tasks provides a sense of purpose for learning communication skills [1]. The task approach also preserves the individuality of students by encouraging them to develop a repertoire of strategies and skills, and respond to patients in a flexible way.

For instance, if one task is defined as ''make a personal connection with the patient'', students and physicians can proceed in a variety of equally effective ways, choosing one that fits their style, the patient, and the situation. This built-in flexibility with respect to the skills and strategies required to accomplish each task reflects the reality and individuality of human communication. The task approach also directs attention toward communication content and process, rather than bedside manner per se. While it focuses on observable behavior, this approach can facilitate discussion and exploration of attitudes relevant to each task as well.

1.2. Uses and characteristics of the framework

The following discussion outlines how the SEGUE Framework can be used in teaching, assessment, and research. Psychometric characteristics (i.e. validity and reliability) are presented in a manner similar to that followed by Stillman et al. when describing the Arizona Clinical Interview Rating Scale [9], as their report provides an excellent model. After describing checklist development, the paper presents methods and results for several tests of the framework's utility as: a structure for communication skills education; a teaching tool; an assessment tool; and a research tool. Throughout these sections, tasks are identified in words (e.g. maintain a respectful tone) and/or labels that indicate the relevant SEGUE section and task number (e.g. *understand 23*).

2. Checklist development

Nunnally [10] emphasized that the plan and procedure of constructing measures is a primary determinant of content validity. Accordingly, the most important step in adopting the task approach is identifying the tasks themselves. Selection of the tasks and response scale included in the SEGUE Framework was informed by the research and education literature, as well as work conducted in two very different contexts; a study of general practitioner–patient communication and decision making conducted in Oxford, UK [11] and a communication skills training program the author developed for resident physicians at the Rehabilitation Institute of Chicago [12]. In both contexts, patients were either surveyed or interviewed to ensure that the framework would encompass their perspectives as well.

2.1. Item generation

General practitioners in Oxford and physicians at the Rehabilitation Institute of Chicago completed a Consultation and Patient Appraisal, one section of which had them rate the importance of communication tasks [11]. The format of this section was adapted from a survey used in the Oxford region to evaluate the degree of attitude change associated with a communication skills training course [1,13]. Fourteen tasks that would apply to multiple clinical contexts emerged as the core group of items; each of these was rated as either very important or essential by the physicians. Two additional items (i.e. make a personal connection during the visit — go beyond the medical issues at hand; discuss antecedent treatments) were rated as moderately important by physicians. The ''personal connection'' item was included because it emerged as the primary issue during in-depth interviews with patients [12]. The ''antecedent treatment'' item was incorporated because it is considered an essential component of patients' past medical history [14].

The wording, order, and number of items were refined during the first 2 years of using the checklist to teach and assess communication between resident physicians and their patients, as well as that between medical students and patient instructors. Nine items were added to the initial set of 16 during this time period; some were obviously essential (e.g. explore physical/physiological factors) and, thus, did not need to be included in the original questionnaire; others repeatedly emerged as critical factors during observations of videotaped and live encounters (e.g. avoid directive/leading questions). Each of the tasks falls under one of the SEGUE Framework's sections. For instance, the two examples given earlier in this paragraph are part of the *E*licit information section. While an additional set of seven tasks was developed for application to visits that involve a new or modified treatment plan (see Appendix A), this article focuses on the 25 basic tasks, since they are the focus of teaching, assessment, and research efforts reported herein. Representational validity [15] of the items was ensured by systematically reviewing the checklists with resident, attending and community physicians, medical school students and faculty, standardized patients (SPs) and patient instructors, and research colleagues at various institutions in several countries. The review process indicated that people in teaching, assessment, and research contexts understood the tasks as intended.

2.2. Coding rules

A guide that describes each task has been prepared for the teaching context, and a detailed codebook is used for assessment and research. General coding rules are outlined herein to provide a sense of the logic and use of the framework. Seventeen of the items focus on content. These tasks include topics to be covered or behaviors to be enacted at least once during the encounter. While some require very little time and need not happen more than once (e.g. greet patient appropriately), others might take more time to enact (e.g. discuss how health problem affects patient's life) or could happen multiple times during an

encounter (e.g. explain rationale for diagnostic procedures). Regardless, each of these items is coded as ''yes'' if the topic is covered or the behavior is enacted *at least one time* during the encounter. It is coded as ''no'' if it does not happen at all. In order to increase the utility of the instrument in different clinical and skills assessment contexts, some of the content items may be coded as ''n/a'' (i.e. not applicable).

The other eight items focus on form, or process, and are marked with arrows (i.e. → in Appendix A) on the checklist form itself to distinguish them from the content items. These tasks are best described as communication behaviors that should be maintained throughout the encounter (e.g. maintain a respectful tone). Thus, each of the process-oriented items is coded as ''no'' if there is at least one time during the encounter when the behavior is *not* enacted (e.g. a physician makes just one disrespectful comment); otherwise it is coded as ''yes''. Note that there is no ''n/a'' option, since all of these tasks are considered applicable to all contexts.

Previous research experience indicated that a nominal response scale, in combination with clear coding rules and rigorous coder training, contributes to very high interrater reliability coefficients among a large pool of coders [11]. Thus, the ''yes'' or ''no'' format was retained in the SEGUE Framework because the objective was to produce a comprehensive instrument that yields reliable data, whether the assessment is being conducted in the context of a research project or clinical skills assessment. While the identical form is used in teaching contexts, learners receive detailed feedback rather than marks of ''yes'' or ''no''.

2.3. Summary scores

The SEGUE Framework may be used both to provide feedback to a particular individual (e.g. reviewing the checklist and comments) and to summarize communication in medical encounters (e.g. determining total scores and sub-scores). While it is easy to generate total scores and sub-scores, in practice this should be done only in an effort to evaluate psychometric qualities of the instrument, to gauge performance of an individual at a macro level (i.e. to ''raise flags''), or to summarize the degree to which the communication tasks are accomplished in groups of encounters. In other words, the SEGUE Framework was not designed to provide fine detail about communication skills via summary scores. Focusing on the scores themselves can be misleading in regard to individual performance; for instance, if a student or physician accomplished all the tasks except *understand 23* (maintain a respectful tone), his or her ''high'' score might be mistakenly interpreted as an indication of highly effective communication. Since fine points regarding communication skills and strategies are better served by a more qualitative evaluation, there is ample space for narrative comments at the end of the checklist.

3. The SEGUE Framework as a structure for communication skills education

As part of the Medical Schools Objectives Project, the Association of American Medical Colleges (AAMC) conducted a survey to determine the status of communication skills teaching and assessment in North America [7]. The second stage of the survey, which was distributed to associate deans at the 144 North American medical schools in 1998, included sections on basic communication skills, end-of-life care, family violence, cultural sensitivity, and spirituality. The majority of questionnaire items asked for detailed information regarding the timing and methods of both teaching and assessment, including questions about which, if any, instruments (e.g. model, framework, checklist, guide) are used for each purpose.

This survey was completed and returned by associate deans at 99 (68.8%) of the 144 medical schools. Of the 99 schools represented, 95 teach basic communication skills at some point within their curricula. As noted in the AAMC summary report [7], 30 (31.6%) of those 95 schools reported integrating an instrument into teaching, and 30 reported using such a structure for assessment. Only 22 schools reported using an instrument for both teaching and assessment. Of those schools employing an instrument, most used the SEGUE Framework (10 for teaching and 16 for assessment) or Kurtz, Silverman, and Draper's [16] Calgary–Cambridge Observation Guide (seven for teaching and five for assessment).

4. The SEGUE Framework as a teaching tool

The SEGUE Framework is the touchstone for communication skills teaching at Northwestern University Medical School. Northwestern's Communication Skills Unit, which is situated within a comprehensive, 2-year course entitled Patient, Physician & Society (PPS), begins 5 weeks into the first-year and runs for 12 weeks [17,18]. After reviewing basic concepts in interpersonal communication, exploring the world of the sick, and explicating the roles and goals of physicians and patients, students focus on the SEGUE Framework. They work through a section each week (i.e. first S, then SE, then SEG, etc.), such that the weeks build upon one another. Additional weeks of practice allow students to apply what they have learned to smoking cessation counseling, nutritional counseling, and other complex situations.

In terms of logistics, students meet in seminar groups of 10–12 for an hour-long session with a faculty preceptor, during which they consider the importance of communication tasks outlined in the SEGUE Framework, and discuss skills and strategies to achieve them. They then practice skills for an additional hour in groups of 3–4 with a trained patient instructor. To offer a diverse experience, the small groups of students interact with a different patient instructor

every week. Further, within any given week, the patient instructors portray a different demeanor and convey a different psychosocial context for each student. Students receive immediate feedback from the patient instructor, as well as their peers and faculty preceptor.

The unit is evaluated in several ways, each of which provides some insight into use of the SEGUE Framework within a teaching context. First, the PPS course directors hold regular curriculum review sessions with students. Second, the communication skills unit director and staff speak directly with patient instructors each week as the unit progresses. Third, on their last day of the unit, all students complete an evaluation of the unit and another of their preceptor. Fourth, the faculty preceptors complete written evaluations of the unit, and evaluate each student in their seminar group. The unit evaluation forms distributed to both students and faculty include several open-ended questions. Each evaluation is reviewed by the PPS course directors, the PPS course coordinator, and the communication skills unit director.

A total of more than 1200 medical student and 30 faculty unit evaluations collected over 7 years at Northwestern University Medical School indicate that the SEGUE Framework has been a valuable tool for teaching communication skills. It is important to highlight one aspect of the feedback that has been particularly helpful; comments from the first-year of use indicated that some students perceived the SEGUE Framework as a rigid, step-by-step formulation. Since then, every effort has been made to explicitly reinforce the flexibility of the framework, both verbally and in the written guide (e.g. ''SEGUE is a flexible framework, and the order of items is not meant to be interpreted as a script. Numbering the items simply facilitates reporting of data.''). Subsequent student comments, in written unit evaluations as well as the rather informal curriculum review sessions and discussions with patient instructors, have reflected this understanding.

5. The SEGUE Framework as an assessment tool

Annual clinical skills assessments have been carried out at Northwestern University Medical School since 1994, allowing different tests of the SEGUE Framework's validity and reliability in a real-time examination context. Toward the end of their second-year of medical school, all Northwestern students are required to participate in the clinical skills assessment, a constant feature of which over the years has been an encounter with a SP who ''has'' abdominal pain. In general, students are instructed to talk with the SP and to perform a focused physical examination; they leave the room and record their findings when finished. Immediately after the visit, SPs use a computer workstation to enter student performance on the SEGUE Framework and any other items included in the assessment (e.g. physical examination). All SPs participate in three half-day training sessions before the assessment begins, and must demonstrate the ability to accurately enact the case scenario and use the checklists. The following sections offer information about interrater reliability, concurrent validity, and construct validity drawn from these clinical skills assessments.

5.1. Interrater reliability

The clinical skills assessments provided a venue for determining interrater reliability, as well as evaluating whether reliability changes with increasing memory demand on SPs. Reliability analyses were run using *Kn* [19], the appropriate index of interrater agreement in this case because it uses $1/n$ as a definition of chance agreement, in contrast to the more common reliability index, Cohen's *K* [20], which incorporates marginal proportions in its definition of chance agreement and assumes both marginals are fixed a priori. Thus, *Kn* offers a more conservative index than percent agreement, but does not carry the penalties associated with Cohen's *K* for situations in which there is a predominance of agreement in one cell (e.g. coders agree that most students accomplished a task) and a small number of ''off-diagonal'' disagreements.

5.1.1. Method

For the initial 3 years of the clinical skills assessment, interrater reliability was gauged by having an assistant spot-check the SP coding. In 1994, the first-year, this was done in real-time; one SP worked with a student and completed the SEGUE checklist immediately after the encounter while another SP monitored the interaction and completed the checklist as the encounter progressed. In 1995 and 1996, spot checks were conducted via both real-time monitoring during the assessments and videotape review after the assessment had ended. In the subsequent years, reliability has been emphasized and ensured during training sessions and reinforced during the assessment itself. All encounters during this period have been videotaped and reviewed after the assessments to thoroughly evaluate interrater reliability.

Since accuracy can be affected by the examination context (e.g. type and number of behaviors to record), we conducted a study during the 1997 assessment to examine interrater reliability for real-time use of the SEGUE Framework under conditions of increasing memory demand on SPs. Seven SPs were randomly assigned to one of three conditions: two SPs completed the SEGUE checklist and had students complete a 12-item physical examination checklist (i.e. the SPs focused only on the SEGUE Framework); two SPs completed the SEGUE checklist and a 12-item physical examination checklist; three SPs completed the SEGUE checklist and a 28-item physical examination checklist. All seven of the SPs completed paper SEGUE forms — with narrative comments for students — then entered the data at computer workstations. As mentioned above, all interactions were videotaped. An experienced research assistant, blinded to experimental condition, used the SEGUE checklist to code

Table 1
Interrater reliability for SEGUE items

SEGUE item	Real-time SP encounters[a] (*Kn*[c])			Videotapes[b] (*Kn*[c])	
	+0 Items ($n = 44$)	+12 Items ($n = 48$)	+28 Items ($n = 72$)	Criterion ($n = 16$)	Test run ($n = 52$)
Set 01	0.91	0.96	0.97	1.00	1.00
Set 02	0.95	1.00	0.97	1.00	1.00
Set 03	0.59	0.79	0.33	1.00	0.96
Set 04	0.68	0.79	0.86	0.88	0.85
Set 05	1.00	1.00	1.00	1.00	1.00
Elicit 06	0.64	0.79	0.72	1.00	1.00
Elicit 07	1.00	1.00	1.00	1.00	1.00
Elicit 08	0.95	0.88	0.81	0.91	0.92
Elicit 09	1.00	1.00	0.94	1.00	0.89
Elicit 10	0.77	0.71	0.81	1.00	0.85
Elicit 11	1.00	1.00	1.00	1.00	0.92
Elicit 12	0.68	0.50	0.42	1.00	0.81
Elicit 13	0.82	0.88	0.72	1.00	0.96
Elicit 14	0.95	0.96	0.86	1.00	0.96
Elicit 15	0.86	0.96	0.86	1.00	0.81
Give 16	1.00	1.00	0.97	1.00	1.00
Give 17	0.82	0.88	0.89	1.00	0.85
Give 18	0.91	0.88	0.78	1.00	0.96
Give 19	0.73	0.75	0.58	1.00	0.85
Understand 20	0.59	0.92	0.89	1.00	0.83
Understand 21	n/a[d]	n/a[d]	n/a[d]	1.00	1.00
Understand 22	0.77	0.92	0.81	1.00	0.81
Understand 23	0.95	0.96	0.89	1.00	1.00
End 24	0.45	0.67	0.75	1.00	1.00
End 25	1.00	0.96	0.94	1.00	1.00
Average	0.84	0.88	0.82	0.99	0.93

[a] Data from 1997 clinical skills assessment for second-year students. There were three conditions in the test on memory demand (i.e. SEGUE + 0 physical examination items, SEGUE + 12 physical examination items, SEGUE + 28 physical examination items).

[b] Data from videotaped doctor–patient encounters. Coders were instructed to view each tape no more than twice.

[c] $Kn = \frac{\sum Pii - (1/n)}{1-(1/n)}$, n = number of categories.

[d] This item was marked not applicable by the clinical skills assessment committee (i.e. it was not coded by SPs).

every videotape. We compared checklists completed by the research assistant and the SPs to determine interrater reliability for each SEGUE item.

5.1.2. Results

The spot-checks conducted in 1994–1996 indicated that most SPs were able to use the SEGUE Framework reliably. One SP was clearly erratic in her application of the coding rules and was dismissed from the assessment. The early spot-checks also highlighted problems in coding two of the items: *set 03* (outline agenda for visit) and *elicit 12* (avoid directive/leading questions). Discussions with the SPs indicated that difficulties with these items arose from different sources. First, several of the SPs worked as patient instructors during the Communication Unit for first-year students, and found it hard to reconcile the very specific coding rule for *set 03* used in the assessment (i.e. student should ask about the patient's agenda) with the more expansive definition used in teaching (i.e. student should ask about the patient's agenda *and* provide his or her own agenda). Second, the SPs had trouble recognizing and remembering directive/leading questions. Both of these issues were addressed in subsequent training sessions.

Results of the test on memory demand that was run during the 1997 assessment are illustrated in Table 1, indicating high interrater reliability in a real-time examination context, even when the SPs had to complete up to 28 physical examination items in addition to the SEGUE checklist; the mean *Kn* for each condition (i.e. SEGUE + 0 physical examination items, SEGUE + 12 physical examination items, SEGUE + 28 physical examination items) was >0.80. Coefficients of this magnitude are quite acceptable in both educational and research contexts [21,22]. There was no consistent relationship between reliability and condition at the item level. Taken together, the reliability checks conducted in 1997–1999 indicated acceptable interrater reliability in the clinical skills assessments, with a mean *Kn* of 0.80 (1997: $Kn = 0.84$, $n = 164$; 1998: $Kn = 0.79$, $n = 160$; 1999; $Kn = 0.77$, $n = 122$). While interrater reliability for coding the *set 03* and *elicit 12* items had improved, *Kn* was still at the level of 0.50 for both tasks, signaling the need for continued vigilance.

5.2. Sensitivity

While the nominal response scale lends reliability to the SEGUE Framework in a clinical skills assessment, it is important to determine if it also offers a sensitive measure of student performance. If not, the framework would be a reliable but ''blunt'' instrument, and of little practical use in differentiating students on the basis of communication skills proficiency.

5.2.1. Method

The SEGUE summary scores were generated by assigning a value of 1 to ''yes'' and a value of 0 to ''no'' for each item, and then summing the values across items for each student. We pooled data for the clinical skills assessments conducted for students at the end of their second-year in 1997–1999. For these three assessments, the maximum possible SEGUE summary score is 24, as one item (*understand 21*) was marked ''not applicable'' by the assessment team (i.e. SPs were not given the option to record whether or not students accomplished this task).

5.2.2. Results

As illustrated in Fig. 1, the SEGUE scores ranged from 11 to 24, with a mean of 21.1 (S.D. = 2.2) and median of 22. The majority of students accomplished at least 18 of the 24 tasks, with most accomplishing 22 or 23, which is not surprising since the SEGUE Framework is used for teaching in the first-year of the curriculum. However, 5.5% of the students were identified as having trouble with communication skills during the assessments, accomplishing between 46% and 71% of the tasks. It appears that the SEGUE Framework has an acceptable degree of sensitivity in terms of its ability to detect differences in student performance.

5.3. Concurrent validity

While it is clear that the SEGUE Framework can produce a range of scores, it is essential to test the degree to which those scores reflect other judgments of students' abilities in the assessment context. One important indicator is concurrent validity, a type of predictive validity that focuses on an instrument's degree of correspondence to a contemporaneous criterion measure [10]. The clinical skills assessments offered a venue for testing the concurrent validity of the SEGUE Framework by assessing the correlation with SPs' confidence in the medical students' competence.

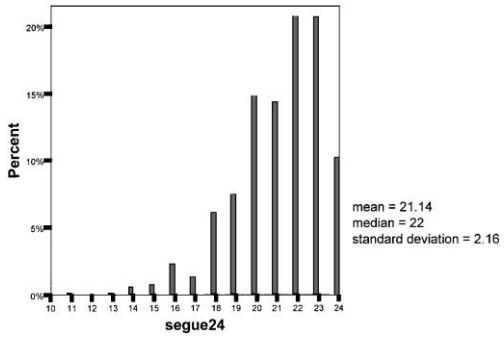

Fig. 1. Distribution of SEGUE scores in clinical skills assessments (1997–1999) (data from the 1997–1999 clinical skills assessments for students at the end of their second-year at Northwestern University Medical School ($n = 507$)).

5.3.1. Method

In the 1995 clinical skills assessment, SPs completed the SEGUE checklist, a brief physical examination checklist, and an item that read ''would you choose this student to be your doctor?'', accompanied by a five-point scale ranging from 1 = ''definitely not'' to 5 = ''yes, definitely''. Prior to administration of the 1995 assessment, two of the SEGUE items (*understand 20* and *understand 21*) were marked ''not applicable'' by the assessment team. Thus, SEGUE summary scores in this analysis are composites of 23 items (i.e. they do not include either *understand 20* or *understand 21*). Logic, not to mention a plethora of research on the relationship between communication and patient satisfaction [23,24], dictates that positive answers to these items should be accompanied by higher SEGUE scores if the framework is indeed measuring the quality of communication.

5.3.2. Results

The correlation between the ''would you choose this student to be your doctor?'' item and SEGUE summary scores from the 1995 clinical skills assessment was positive and relatively high ($r = 0.65$, $n = 134$, $P < 0.001$). It is important to note the possibility of an inflated correlation coefficient since the same SPs completed the SEGUE items and the ''choose'' item. That said, even a modest correlation may be considered fairly robust, given the complexities of testing predictive validity [10].

5.4. Construct validity

While construct validity is more of an issue for abstract variables (e.g. anxiety), it is both possible and worthwhile to test the extent to which results obtained using one measure of communication are similar to those derived from established measures in the same domain. The American Board of Internal Medicine (ABIM) developed a well-tested instrument called the ''patient satisfaction questionnaire'', which focuses explicitly on perceived quality of doctor–patient communication [25]. While somewhat unfortunately named, since it seems to address physician communication more than patient satisfaction, the ABIM questionnaire provides a useful tool against which to test the construct validity of the SEGUE Framework.

5.4.1. Method

The 1998 clinical skills assessment included 7 of the 10 items in the ABIM questionnaire, excluding two items about diagnosis and treatment, which were not applicable, as well

as one that was entirely redundant with items in the SEGUE Framework. The SPs responded to each of the ABIM items (e.g. no. 3: treating you like you are on the same level; never ''talking down'' to you or treating you like a child) by using a five-point scale labeled as 1 = ''poor'', 2 = ''fair'', 3 = ''good'', 4 = ''very good'', and 5 = ''excellent''. Composite scores were computed by summing the seven items and dividing the totals by seven to facilitate interpretation. The SEGUE summary scores in this analysis include 24 of the 25 items (i.e. they exclude *understand 21*, which was marked ''not applicable'' by the assessment team). Construct validity was evaluated through correlational analysis.

5.4.2. Results

While the ABIM composite scores ranged from 1.7 to 4.9, they clustered around the midpoint, with a mean of 3.15, median of 3.14, mode of 3 (S.D. = 0.42). There was a moderate correlation between ABIM and SEGUE scores ($r = 0.49$, $n = 169$, $P < 0.001$).

5.5. Internal consistency

Communication is multifaceted, and the SEGUE Framework was constructed to reflect the nature of doctor–patient communication. Given the nature of medical encounters, some tasks are accomplished in most, if not all, of the encounters (e.g. explore physical/physiological factors), while others provide better reflections of communication competence [26]. Since the framework was not generated by factor analysis and item-total correlational analysis, internal consistency is not an appropriate criterion for this instrument. In other words, it is both reasonable and logical to include a mix of items that tap somewhat different abilities [10,27].

6. The SEGUE Framework as a research tool

Recordings of simulated or actual medical encounters often provide raw data for research on communication in medicine. We have tested reliability in the research context using both videotapes and audiotapes of encounters. While both media can capture verbal communication and paralanguage (i.e. the nonverbal communication channel involving speech rate, pitch, tone, accent, etc.), video can pick up additional nonverbal channels such as proximity and body lean during the interaction.

6.1. Reliability in coding videotaped encounters

In initial testing of the SEGUE Framework for assessing physician–patient communication in videotaped encounters, interrater reliability was gauged via two exercises: (1) a *criterion run*, with the author and a research assistant as coders, to determine the extent to which coding rules required refinement; (2) a *test run*, with two naive coders who received a copy of the coding rules and participated in an intensive training session, to provide a more generalizable experience. Intrarater reliability was assessed during the *test run* as well.

6.1.1. Method

For the *criterion run*, interrater reliability was assessed by having the author and a research assistant each use the SEGUE Framework to evaluate communication in 16 videotaped physician–patient encounters. In an attempt to diversify the sample, four of the videotaped interactions were randomly selected from those collected in a large study of general practitioner–patient communication in Oxford [11] and 12 were drawn from the Rehabilitation Institute of Chicago training program [12]. Interactions were indeed diverse; they ranged from 4.25 to 47.92 min of time duration.

During the *test run*, two naive coders were provided with a copy of the coding rules. They participated in an intensive 2-h training session conducted by the author, which included viewing videotaped examples relevant to each task in the SEGUE Framework. These coders proceeded to code 52 videotaped encounters: 12 between residents and patients at the Rehabilitation Institute of Chicago, and 40 from the SP-based clinical skills assessment at Northwestern University Medical School. The length of these encounters ranged from 9.62 to 47.92 min for tapes from the Rehabilitation Institute of Chicago, and from 5.75 to 20.50 min for tapes from the clinical skills assessment. The two coders each reviewed an average of eight tapes per week, for a total of 7 weeks. At the end of each week, they met with the author to discuss process and progress. As the primary goal was to evaluate interrater reliability, the coders were not allowed to adjust past coding decisions. To examine intrarater reliability, on the last week of the test run, both coders received six tapes that they had reviewed earlier in the test run. Again, *Kn* was the appropriate reliability index.

6.1.2. Results

The combination of clearly worded items, detailed coding rules, and a simple, nominal response scale yielded high reliability coefficients. During the *criterion run*, *Kn* was 1.0 (i.e. perfect agreement) for 23 of the 25 tasks on the checklist (see Table 1). Coders disagreed once in their evaluation of *set 04* (make a personal connection during visit), $Kn = 0.88$, and once in their evaluation of *elicit 08* (explore psychosocial/emotional factors), $Kn = 0.91$. Given that the two coders in this criterion run were very familiar with the SEGUE Framework, the coding rules were clarified for *set 04* and *elicit 08*, and the test run was implemented to assess coding reliability in a more representative setting.

Table 1 shows that interrater reliability coefficients were quite acceptable during the *test run* (mean $Kn = 0.93$), though not as high as those achieved during the *criterion run*. The weekly meetings with these coders were useful in eliciting and extinguishing idiosyncratic coding standards,

and provided an important lesson; coding rules must be carefully reinforced during the coding process so coders will not incorporate individual biases into their coding decisions. During the last week of coding, the two raters were asked to view and evaluate six tapes that they had coded earlier in the test run. Their use of the checklist matched almost perfectly, yielding intrarater reliability coefficients of 1.0 for all but two items: *understand 20* (acknowledge patient's progress/accomplishments/challenges) $Kn = 0.83$; *understand 22* (express caring, concern, empathy) $Kn = 0.83$. Overall, intrarater reliability was very high (mean $Kn = 0.99$).

6.2. Reliability in coding audiotaped encounters

Research on physician–patient communication is often based on audiotaped encounters. We used the SEGUE Framework to code a sample of audiotapes to determine if interrater reliability was comparable to that obtained when using videotaped data.

6.2.1. Method

After a 1-day training session, two coders listened individually to the same 46 audiotapes. The tapes were drawn from a large sample of encounters between patients and a variety of specialist physicians in Northern California; interactions ranged in time duration from 3.9 to 17.8 min.

6.2.2. Results

As with the videotapes, interrater reliability for audiotaped medical encounters was high (mean $Kn = 0.87$). Due to the lack of visual cues, coders were unable to include two items in their work: *set 05* (maintain patient's privacy) or *elicit 14* (listen — give patient undivided attention).

6.3. An examination of communication in primary care encounters

The plan and procedure of the item-generation process, during which physicians listed and rated key communication tasks, was the main strategy for lending content validity to the SEGUE Framework. This final section on uses and characteristics of the framework brings the discussion full-circle, focusing on whether or not the tasks are actually relevant to, and accomplished in, an applied setting.

6.3.1. Method

Research assistants coded a total of 500 videotapes collected for a study of physician–patient communication in primary care. There are approximately 25 patients for each of the 20 academic general internists in this sample; about half of the encounters were taped in Chicago and the other half were recorded in Burlington, Vermont. Two of the items — *understand 20* (acknowledge patient's accomplishments, progress, challenges) and *understand 21* (acknowledge waiting time) — are not included in the analysis because they were not applicable in most cases.

Table 2
SEGUE tasks accomplished in primary care encounters[a]

SEGUE item	% Accomplished
Set 01	99.2
Set 02	90.6
Set 03	11.2
Set 04	70.8
Set 05	60.8
Elicit 06	99.6
Elicit 07	99.8
Elicit 08	64.8
Elicit 09	94.4
Elicit 10	57.8
Elicit 11	66.4
Elicit 12	60.0
Elicit 13	72.8
Elicit 14	88.0
Elicit 15	97.0
Give 16	87.7[b]
Give 17	99.0
Give 18	15.0
Give 19	93.8
Understand 20	n/a
Understand 21	n/a
Understand 22	99.2
Understand 23	98.0
End 24	14.0
End 25	92.4

[a] Data from a study of communication between general internists and patients ($n = 500$).

[b] The sample size is slightly reduced for this item since 13 visits did not include physical examinations ($n = 487$).

6.3.2. Results

As shown in Table 2, 20 of the 23 SEGUE tasks included in the analysis were accomplished in a majority of the 500 encounters; the number of tasks accomplished per encounter ranged from 11 ($n = 2$) to 22 ($n = 1$), with a mean of 17.31 (S.D. = 1.77). In this sample, practicing general internists accomplished *set 03* (outline agenda for visit) in only 11.2% of the encounters, *give 18* (encourage questions) in 15.0%, and *end 24* (ask if there is anything else the patient would like to discuss) in 14.0%. All three are patient-centered tasks [28] that focus on addressing patients' agendas.

7. Discussion

As more educational programs — whether medical schools, postgraduate tracks, or continuing medical education courses — focus attention on doctor–patient communication, the need for coherent approaches to teaching and assessment becomes self-evident. Work in different contexts over the past several years has demonstrated the SEGUE Framework's acceptability, educational impact, reliability, validity, and feasibility, five key indicators of a tool's utility [29,30]. The SEGUE Framework has gained considerable acceptance in teaching, assessment, and research venues ranging from pediatrics to geriatrics and from oncology to

psychology. In addition to use at different medical schools [7], the framework has demonstrated generalizability within the health professions; for example, it has been incorporated into work with osteopathic students at the Philadelphia College of Osteopathic Medicine in Pennsylvania and physical therapy students at Midwestern University in Illinois. In the research arena, it has been used to provide an indication of physician–patient communication and to evaluate the effect of educational programs. Consistent with the data reported herein, interrater reliability has been high in these studies. For instance, Bourke et al. [31] reported reliability above 0.90 when three independent researchers coded communication between oncologists and new patients.

In addition to providing a useful tool for teaching, assessment, and research, the SEGUE Framework offers an effective organizational structure for communication skills education. It outlines broad areas of the consultation, from setting the stage to ending the visit, and highlights specific communication tasks for students and physicians. In so doing, the framework reinforces the notion that physicians must adapt their communication to different situations and different patients. Further, the framework's task approach provides a point of departure for discussing relevant communication skills, strategies, and attitudes. The importance of an organizational structure cannot be understated. Medical schools and training programs for physicians employ a diverse set of methods for teaching communication skills, such as small-group discussions, lectures and presentations, interviews with real or simulated patients, and observation of others working with patients [7]. Without a framework to help structure and focus attention on communication, the teaching is likely to be inconsistent and ineffective. Similarly, the reliability and effectiveness of observation and feedback are likely to be compromised unless they are grounded in a coherent framework.

7.1. Considerations

Tamblyn et al [32] suggested that ''whenever possible, items should be constructed so that a single behavior is included in each item and the judgement required involves the simple documentation of the presence or absence of the behavior''. The SEGUE Framework is consistent with this approach to assessment, using discrete items, clear coding rules, and a nominal response scale. The combination of a nominal coding scheme and ample space for comments offers considerable fidelity and freedom, while allowing both SPs participating in clinical skills assessments and coders working on research studies to achieve high interrater reliability. Maintaining high reliability is not simply a function of the instrument; it requires intensive training, continuous spot-checking, and careful calibration.

Despite the space for — and importance placed on — narrative comments, the SEGUE Framework is essentially a checklist. While they may be relatively simple to use, all checklists have limitations, some more obvious than others. One very relevant line of research involves comparison of checklists and global ratings of clinical skills. While not focused on communication skills per se, results of these studies suggest that checklist scores are highly content specific [33] and may not be valid measures of increasing clinical competence [34]. Most of our work has been conducted in the context of a clinical skills assessment for second-year medical students, but our experience in using the SEGUE Framework to code encounters between general internists and a diverse sample of patients suggests that the task approach transcends the problem of content specificity. In other words, the key to generalizability may be the focus on tasks instead of specific questions and/or actions.

Interestingly, the results presented in this article suggest that medical students, who were taught communication skills via the SEGUE Framework, accomplished relatively more of the communication tasks than did practicing physicians, who had much more clinical experience but no such focused training. We are interested in pursuing research on predictive validity to learn how well the SEGUE Framework can predict performance in different content areas, at different phases of training, and at different levels of expertise. Moreover, we hope to study how the tasks themselves relate to outcomes of medical encounters. Just as research on physician–patient communication informed development of the SEGUE Framework, use of the framework will be enhanced by further research into these areas.

7.2. Practice implications

The SEGUE Framework provides a coherent structure for communication skills education. When used as a teaching tool, teachers and learners must be reminded that it is a flexible framework. When applied as an assessment instrument, it is important that evaluators utilize the space for narrative comments to provide the sort of detail that cannot be captured by a checklist. It appears that the framework is also valuable in the research context, particularly in terms of creating profiles of which tasks were accomplished by different doctors and/or in different settings.

7.3. Conclusion

There are, of course, other instruments (i.e. checklists, frameworks, guides, models) available in addition to the SEGUE Framework. These include the Arizona Clinical Interview Rating Scale [9], Bayer Institute for Health Care Communication E4 Model [35], Brown Interview Checklist [36], Calgary–Cambridge Observation Guide [16], Patient-Centered Clinical Method [28], Seven Task Model [1], and the Three Function Model [37]. Whichever instrument is used, it is essential to define patient-centered communication tasks [8,28,38] so they can be taught, assessed, and studied for the benefit of caregivers and the patients they serve.

Acknowledgements

This work was funded, in part, by grants from the Arthur Vining Davis Foundations (US) and the Ethox Foundation (UK). I am grateful to all of the patients, standardized patients, patient instructors, medical students, faculty members, physicians, and investigators who have participated in the teaching, assessment and/or research in which the SEGUE Framework has been used. Special thanks are also due to Drs. Raymond Curry and Theo Schofield for their helpful comments regarding this manuscript, and to Jason Thompson for his research assistance.

Appendix A

The SEGUE Framework *(long form)* Patient _____ Physician _____

Set the Stage	**Yes**	**No**
1. Greet pt appropriately		
2. Establish reason for visit: _____		
3. Outline agenda for visit (e.g., "anything else?", issues, sequence)		
4. Make a personal connection during visit (e.g., go beyond medical issues)		
→ 5. Maintain pt's privacy (e.g., knock, close door)		

Elicit Information	**n/a**	**Yes**	**No**
6. Elicit pt's view of health problem and/or progress (ideas, concerns)			
7. Explore physical/physiological factors (signs, symptoms)			
8. Explore psychosocial/emotional factors (e.g., living situation, family relations, stress)			
9. Discuss antecedent treatments (e.g., self-care, last visit, other care)			
10. Discuss how health problem affects pt's life (e.g., quality-of-life)			
11. Discuss lifestyle issues/prevention strategies (e.g., health risks)			
→ 12. Avoid directive/leading questions			
→ 13. Give pt opportunity/time to talk (e.g., don't interrupt)			
→ 14. Listen. Give pt undivided attention (e.g., face pt, verbal acknowledgement, nv feedback)			
→ 15. Check/clarify information (e.g., recap, ask "how much")			

Give Information	**n/a**	**Yes**	**No**
16. Explain rationale for diagnostic procedures (e.g., exam, tests)			
17. Teach pt about his/her own body & situation (e.g., provide feedback from exam/tests, explain anatomy/diagnosis)			
18. Encourage pt to ask questions			
→ 19. Adapt to pt's level of understanding (e.g., avoid/explain jargon)			

Understand the Patient's Perspective	**n/a**	**Yes**	**No**
20. Acknowledge pt's accomplishments/progress/challenges			
21. Acknowledge waiting time			
→ 22. Express caring, concern, empathy			
→ 23. Maintain a respectful tone			

End the Encounter	**Yes**	**No**
24. Ask if there is anything else pt would like to discuss		
25. Review next steps with pt		

If suggested a new or modified treatment/prevention plan:	**n/a**	**Yes**	**No**
26. Discuss pt's interest/expectation/goal for treatment/prevention			
27. Involve pt in deciding upon a plan (e.g., options, rationale)			
28. Explain likely benefits of the option(s) discussed			
29. Explain likely side-effects/risks of the option(s) discussed			
30. Provide complete instructions for plan			
31. Discuss pt's ability to follow plan (e.g., attitude, time, resources)			
32. Discuss importance of pt's role in treatment/prevention			

Comments:

Visit Date: / / Review Date: / / Reviewer _____

References

[1] Pendleton D, Schofield T, Tate P, Havelock P. The consultation: an approach to learning and teaching. Oxford: Oxford University Press, 1984.

[2] Muller S (chair). Physicians for the twenty-first century: report of the project panel on the general professional education of the physician and college preparation for medicine. J Med Edu 1984;59 (part 2).

[3] Smith DH, Pettegrew LS. Mutual persuasion as a model for doctor–patient communication. Theor Med 1986;7:127–46.

[4] Lancaster H. Where patients are partners. Wall Street J 1989;24:5D.

[5] Nelson A. Humanism and the art of medicine. JAMA 1989;262: 1228–30.

[6] Novack DH, Volk G, Drossman DA, Lipkin M. Medical interviewing and interpersonal skills teaching in US medical schools: progress, problems and promise. JAMA 1993;269:2101–5.

[7] Association of American Medical Colleges. Contemporary issues in medicine: communication in medicine (MSOP Report III). Washington, DC: Association of American Medical Colleges, 1999.

[8] Makoul G, Schofield T. Communication teaching and assessment in medical education: an international consensus statement. Pat Edu Counsel 1999;137:191–5.

[9] Stillman PL, Brown DR, Redfield DL, Sabers DL. Construct validation of the Arizona Clinical Interview Rating scale. Edu Psych Meas 1977;37:1031–8.

[10] Nunnally J, Psychometric theory. New York: McGraw-Hill, 1978.

[11] Makoul G, Arntson P, Schofield T. Health promotion in primary care: physician–patient communication and decision making about prescription medications. Soc Sci Med 1995;41:1241–54.
[12] Makoul G, Sliwa J. A context-based communication training program for resident physicians. Presented at the Teaching about Communication in Medicine Conference. Oxford, 1996.
[13] Schofield T, Arntson P. A model for teaching doctor–patient communication during residency. In: Stewart M, Roter D, editors. Communicating with medical patients. Newbury Park, CA: Sage, 1989, p. 138–52.
[14] Billings JA, Stoeckle JD. The clinical encounter: a guide to the medical interview and case presentation. Chicago: Year Book Medical Publishers, 1989.
[15] Folger JP, Hewes DE, Poole MS. Coding social interaction. In: Dervin B, Voight MJ, editors. Progress in communication sciences, vol. IV. Norwood, NJ: Ablex, 1984, p. 115–61.
[16] Kurtz SM, Silverman JD, Draper JD. Teaching and learning communication skills in medicine. Abingdon, UK: Radcliffe Medical Press, 1998.
[17] Curry RH, Makoul G. An active-learning approach to basic clinical skills. Acad Med 1996;71:41–4.
[18] Makoul G, Curry RH. Patient, Physician & Society: Northwestern University Medical School. Acad Med 1998;73:14–24.
[19] Brennan RL, Prediger DJ. Coefficient Kappa: some uses, misuses, and alternatives. Edu Psych Meas 1981;41:687–99.
[20] Cohen J. A coefficient of agreement for nominal scales. Edu Psych Meas 1960;20:37–46.
[21] Fleiss JL. Statistical methods for rates and proportions. New York: Wiley, 1981.
[22] Bakeman R, Gottman JM. Observing interaction: an introduction to sequential analysis. New York: Cambridge University Press, 1986.
[23] Ong LML, de Haes JCJM, Hoos AM, Lammes FB. Doctor–patient communication: a review of the literature. Soc Sci Med 1995;40: 903–18.
[24] Stewart MA. Effective physician–patient communication and health outcomes: a review. Can Med Assoc J 1996;152:1423–33.
[25] PSQ Project Co-investigators. Final report on the patient satisfaction questionnaire project. Philadelphia: American Board of Internal Medicine, 1989.
[26] Cegala DJ, McGee DS, McNeilis KS. Components of patients' and doctors' perceptions of communication competence during a primary care medical interview. Health Commun 1996;8:1–27.
[27] Bollen K, Lennox R. Conventional wisdom on measurement: a structural equation perspective. Psych Bull 1991;110:305–14.
[28] Stewart M, Brown JB, Weston WW, McWhinney IR, McWilliam CL, Freeman TR. Patient-centered medicine: transforming the clinical method. Thousand Oaks, CA: Sage, 1995.
[29] Tate P, Foulkes J, Neighbour R, Campion P, Field S. Assessing physicians' interpersonal skills via videotaped encounters: a new approach for the Royal College of General Practitioners membership examination. J Health Commun 1999;4:143–52.
[30] Ram P, Grol R, Rethans JJ, Schouten B, van der Vleuten C, Kester A. Assessment of general practitioners by video observation of communicative and medical performance in daily practice: issues of validity, reliability and feasibility. Med Edu 1999;33:447–54.
[31] Bourke T, Power K, McDermot E, Crown J, Fennelly D, Carr A. Improved doctor/patient indices in oncology: results of a prospective, controlled evaluation and intervention based on the E4 model. Presented to the European Society for Medical Oncology (23rd ESMO Conference). Athens, 1998.
[32] Tamblyn RM, Klass DJ, Schnabl GK, Kopelow ML. Sources of unreliability and bias in standardized-patient rating. Teach Learn Med 1991;3:74–85.
[33] Regehr G, Freeman R, Hodges B, Russell L. Assessing the generalizability of OSCE measures across content domains. Acad Med 1999;74:1320–2.
[34] Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M. OSCE checklists do not capture increasing levels of expertise. Acad Med 1999;74:1129–34.
[35] Keller V, Carrol G. A new model for physician–patient communication. Pat Edu Counsel 1994;23:131–40.
[36] Novack DH, Dube C, Goldstein MG. Teaching medical interviewing: a basic course on interviewing and the physician–patient relationship. Arch Internal Med 1992;152:1814–20.
[37] Cohen-Cole SA. The medical interview: the three-function approach. St. Louis: Mosby Year Book, 1991.
[38] The Bayer-Fetzer Conference on Physician–Patient Communication in Medical Education. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med 2001;76:390–393.