

# Validating Measures of Third Year Medical Students' Use of Interpreters by Standardized Patients and Faculty Observers

Desiree Lie, MD, MEd<sup>1,3</sup>, John Boker, PhD<sup>1</sup>, Sylvia Bereknyei, MS<sup>2</sup>, Susan Ahearn, RN<sup>1</sup>, Charlotte Fesko, BA<sup>1</sup>, and Patricia Lenahan, LCSW, MFT<sup>1</sup>

<sup>1</sup>University of California, Irvine School of Medicine, Irvine, CA, USA; <sup>2</sup>Stanford University School of Medicine, Stanford, CA, USA; <sup>3</sup>Department of Family Medicine, Orange, CA 9286, USA.

**BACKGROUND:** Increasing prevalence of limited English proficiency patient encounters demands effective use of interpreters. Validated measures for this skill are needed.

**OBJECTIVE:** We describe the process of creating and validating two new measures for rating student skills for interpreter use.

**SETTING:** Encounters using standardized patients (SPs) and interpreters within a clinical practice examination (CPX) at one medical school.

**MEASUREMENTS:** Students were assessed by SPs using the interpreter impact rating scale (IIRS) and the physician patient interaction (PPI) scale. A subset of 23 encounters was assessed by 4 faculty raters using the faculty observer rating scale (FORS). Internal consistency reliability was assessed by Cronbach's coefficient alpha ( $\alpha$ ). Interrater reliability of the FORS was examined by the intraclass correlation coefficient (ICC). The FORS and IIRS were compared and each was correlated with the PPI.

**RESULTS:** Cronbach's  $\alpha$  was 0.90 for the 7-item IIRS and 0.88 for the 11-item FORS. ICC among 4 faculty observers had a mean of 0.61 and median of 0.65 (0.20, 0.86). Skill measured by the IIRS did not significantly correlate with FORS but correlated with the PPI.

**CONCLUSIONS:** We developed two measures with good internal reliability for use by SPs and faculty observers. More research is needed to clarify the reasons for the lack of concordance between these measures and which may be more valid for use as a summative assessment measure.

**KEY WORDS:** interpreter use; evaluation of skills; assessment; validation.

J Gen Intern Med 22(Suppl 2):336–40

DOI: 10.1007/s11606-007-0349-3

© Society of General Internal Medicine 2007

Recent emphasis on cultural competence curricula in medical education has highlighted the need for training medical students and residents in the effective use of interpreters.<sup>1-3</sup> The knowledge and application of this particular skill set is distinctly described as a separate learning entity

(domain V) in the AAMC's tool for assessing cultural competency training (TACCT).<sup>3</sup> Several recent guidelines have reported areas of consistency and agreement, as well as controversy, about the best practices for interpreter use from both interpreter and clinician perspectives.<sup>4-8</sup>

Despite the need to train learners to use interpreters effectively and the availability of guidelines describing core skills, validated measures for evaluating these skills are not available. The challenges of designing such measures for clinical encounters are considerable. They range from the diverse and varying skill levels of interpreters available in practice settings to the difficulty of isolating communication skills necessary for using an interpreter to the shortage of faculty fluent in the patient's language to assess the verbal content of non-English language encounters. Partial provider/learner fluency in the patient's language may hinder or help the encounter when an interpreter is present. In addition, the manner in which clinicians utilize interpreters,<sup>9</sup> and clinician awareness of the health literacy of their limited English proficiency (LEP) patients are also important variables affecting physician-patient communication,<sup>10</sup> length of visits,<sup>11,12</sup> and the quality of health care and patient satisfaction.<sup>13,14</sup> Few attempts to design assessment tools of interpreter use have been reported in the literature, and those that have are case-specific and not generalizable across other encounters using interpreters.<sup>15</sup>

In this study, we used an empirical expert-based consensus framework to create two interpreter use skill assessment measures for clinical encounters where an interpreter is involved. We designed a study to test the internal reliability of the two measures and their performance compared to a widely used communication assessment measure in the setting of a standardized clinical examination.

## METHODS

We used an iterative process to create two measures: the interpreter impact rating scale (IIRS), for use by standardized patients (SPs), and the faculty observer rating scale (FORS), for use by faculty. The rationale for creating a new SP measure to assess interpreter-mediated communication was that additional or different observable skills apply to an encounter when an interpreter is present as a third party, which may impact patient/SP perception of the provider and perceived effectiveness of the information transfer.

A panel of 7 faculty, which included a residency program director, a medical education researcher, 2 course directors, an expert on cultural diversity education, a director of a training center, and a SP trainer performed the expert review of literature on existing practice guidelines for using interpreters.<sup>16-19</sup> This review was done in consultation with the professional interpreter services of the local university hospital. Five panelists were clinicians familiar with interpreter use. Two were bilingual in English and Spanish and 2 were bilingual in 1 language other than English. This panel adopted the following principles for the IIRS, the measure designed for use by the SP: (a) should be patient-centered; (b) can be used in any language encounter; (c) is capable of translation into another language; (d) can be used for other health professions; and (e) emphasizes observable behaviors. Care was taken to avoid 'ethnocentric' items. For example, item 5 (see Table 1, IIRS) defined "sat at a comfortable distance from me" according to the accepted norm of the culture or custom of the SP, and item 6 ("the trainee's nonverbal body communication was reassuring") can be defined for each case by the SP's particular background. The following criteria were used for the FORS: (a) should be applicable to any language encounter; (b) can be used by faculty not proficient in the language of the encounter; and (c) should be independent of the clinical case and be generalizable to different clinical situations. Thus, each measure was deliberately constructed to measure interpreter usage skill from two different viewpoints. The SP's perspective reflected participation in the encounter as a patient with emotional engagement within the encounter and used less observable behavior items, whereas the faculty rater perspective reflected more objective observation of communication skill behaviors as an uninvolved outsider and had more verbal items.

Prototype items were constructed, pilot tested, and refined from the analysis of videotapes of other CPX encounters using interpreters. As a result of this process, consensus among the panel produced 7 items for the IIRS (see Table 1, IIRS) and 11 items for the FORS (see Table 1, FORS). The last item of each measure was a global rating item. Each item on both scales used a 5-point Likert-type rating scale where a value of 1 represented "marginal/low" and 5 represented "outstanding" performance. Scale scores for the IIRS and FORS were computed by summing across obtained item ratings on each measure; higher scores were viewed as indicative of higher skill level. Percentage of maximum scale scores were computed to place both measures on the same metric, i.e., the denominator was a maximum score of 35 on IIRS and 55 on the FORS.

## VALIDATION

To perform the validation, we chose to compare the performance of the IIRS and the FORS with a commonly used scale for communication skills, the physician patient interaction (PPI) scale.<sup>20</sup> The PPI is a 7-item rating form and uses a 6-point Likert scale (ranging from 1="unacceptable" to 6="outstanding") to measure the competency of communication in the context of professional behavior. The 7 items are "appeared professionally competent", "effectively gathered information", "listened actively", "established personal rapport", "appropriately explored my perspective", "addressed my feelings", and "met my needs". These items were not constructed to take into account the presence of a third party interpreter or a language barrier as do the IIRS and FORS.

Table 1. IIRS and FORS Means and Variances

	Items	Mean±SD	Corrected item-scale correlation
IIRS	Combined total IIRS scale (Cronbach's $\alpha=0.90$ )	24.65±4.16	
1	Trainee showed direct eye contact with me during the encounter instead of at the interpreter most of the time.	3.74±0.62	0.77
2	Trainee directly addressed the issues translated that were of concern to me.	3.52±0.67	0.72
3	The trainee acknowledges and responds to my beliefs, concerns, and expectations about my problems.	3.35±0.83	0.83
4	The trainee asked me questions in the first person (example: "Do you feel..." rather than "interpreter, can you ask him if he...")	3.70±0.56	0.62
5	The trainee sat at a comfortable distance from me (not too close and not too far away).	3.61±0.66	0.64
6	The trainee's nonverbal body communication was reassuring (i.e.,—mannerisms, facial expressions, and body language).	3.48±0.51	0.58
7	Rate your overall satisfaction with the encounter.	3.26±1.25	0.89
FORS	Combined total FORS scale (Cronbach's $\alpha=0.88$ ; mean intraclass correlation coefficient=0.61 for 4 raters)	41.57±4.55	
1	The trainee adequately explained the purpose of the interview to the interpreter.	3.50±0.73	0.22
2	The trainee explained the interpreter's role to the patient at the beginning.	3.08±0.87	0.33
3	The trainee asked the patient one question at a time.	4.09±0.43	0.60
4	The trainee listened to the patient without unnecessary interruption.	4.18±0.40	0.68
5	The trainee asked questions to clarify his/her own understanding of the patient's answers.	3.54±0.53	0.65
6	The trainee presented information at a pace that was easy to follow for both patient and interpreter; that is, information was given in "digestible chunks".	4.01±0.59	0.52
7	The trainee maintained direct eye contact with the patient (instead of the interpreter).	3.75±0.65	0.44
8	The trainee addressed the patient in the first person and not as "he/she".	4.07±0.99	0.63
9	The trainee appropriately closed the encounter: at a minimum, asked the patient if he/she had any questions.	3.48±0.66	0.56
10	To what extent did the trainee keep the interpreter on track within his/her assigned role?	3.92±0.55	0.83
11	Global rating of trainee's effectiveness in using the interpreter for the patient encounter.	3.95±0.56	0.94

Each item on both scales used a 5-point Likert-type rating scale where a value of 1 represented "marginal/low" performance and 5 represented "outstanding" performance. Scale scores for IIRS and FORS were computed by summing across obtained item rating on each measure

## SETTING AND PROCEDURES

To control for variables such as patient performance, interpreter skill, and case details, we chose to validate our 2 assessment tools in the setting of a clinical practice examination (CPX) at a California medical school whose Institutional Review Board approved the study. The school was located near a city with the highest proportion of monolingual Latinos in the US.<sup>21</sup> Study participants were 92 third year medical students (MS3s) who participated in the required CPX after completing their required clinical clerkships. One of the cases in the CPX was a 15-minute station involving a monolingual Spanish-speaking patient. Spanish was chosen as the language because of local community demographics. Both the SP and the standardized interpreter (SI) (actors who were bilingual in Spanish and English) received 4, 2-hour training sessions as a pair to prepare for their roles. Two SP-SI pairs were used for the entire examination to standardize case performance. We standardized the rating performance of the SP-SI pairs by having SPs and SIs rate videotaped encounters before the examination. All encounters were recorded on DVD. The medical students had previously received instruction on the use of interpreters from a lecture, had practice interviewing with SP-SI pairs in their first year of medical school, and had experience using interpreters in clinical encounters during clerkships.

### Clinical Case Scenario and Student Task

A smoking cessation case involving a monolingual Spanish-speaking Latino patient wishing to quit smoking was designed and pilot tested with 39 MS3 in the preceding class cohort in May 2005. SPs and professional Spanish-speaking interpreters were used at the 2005 administration. Based on the findings, it was determined that the clinical tasks required of the students by the case (history taking and advising the patient) were at an appropriate level of challenge and fit within the 15 minutes provided for the encounter. This case was then administered in the 8-station CPX in May 2006 to the 92 MS3s. To control for performance in students with some degree of Spanish language proficiency, students were instructed not to use Spanish and that their skill using an interpreter, not Spanish proficiency, was being assessed.

### Raters and Data Collection

SPs rated all 92 students using the IIRS and the PPI during the encounter and submitted their scores online to a central databank. Subsequently, one-fourth (23 of 92) of the encounters were selected for further assessment by faculty raters using the FORS. The 23 encounters were selected by an independent trainer to generate a range of performances based on SP-generated IIRS student scores across the entire cohort. Students who self-reported any degree of Spanish fluency were excluded from the analysis.

Four faculty participated in the independent rating of the 23 encounters using DVD records. The students who were rated were not familiar or known to the faculty but had been taught by the faculty in lectures. One rater was a course director (DL), one was the content expert for Diversity (PL), one was a researcher (SB), and one was a SP trainer (CF). Only 1 rater was bilingual in Spanish and English. All 4 raters had some expertise in teaching students about the effective use of interpreters. Each rater reviewed 2 DVDs containing all 23

encounters in the same sequence with students identified by number only. After rating the 23 student encounters using the FORS, each rater submitted completed rating forms to a data entry administrator blinded to the student and faculty identities.

### Data Analysis

The process of designing and constructing the 2 measures was qualitatively analyzed to document the content and face validity of the measures. Item analysis was performed using conventional methods. Internal consistency reliability was assessed by Cronbach's coefficient alpha ( $\alpha$ ). Interrater reliability of the FORS was examined by the intraclass correlation coefficient (ICC) for 4 independent raters. Overlap of the IIRS, FORS, and PPI items and scale scores was analyzed by Pearson product-moment correlation coefficients. The nominal criterion for statistical significance of scale and item correlation coefficients was  $p < 0.05$ . No adjustment was made to this criterion for computing multiple correlation coefficients. All statistical analyses were computed with SPSS version 14.0 for Windows (SPSS, Chicago, IL, USA).

## RESULTS

### Reliability and Validity

Item means for the IIRS ranged from 3.26 ("rate your overall satisfaction with the encounter") to 3.74 ("trainee showed direct eye contact with me during the encounter instead of the interpreter most of the time") with SDs ranging from 0.58 to 0.89. Item means for the FORS ranged from 3.08 ("the trainee explained the interpreter's role to the patient at the beginning") to 4.18 ("the trainee listened to the patient without unnecessary interruption") with SDs from 0.40 to 0.87 (Table 1). Corrected item score-total scale score correlations (i.e., point-biserial correlation coefficients computed after removing the contribution of each item to the total score) ranged from 0.58 to 0.89 for the IIRS and 0.22 to 0.94 for the FORS. Values  $> 0.20$  are generally taken as evidence that the items comprising a measure are, in fact, measuring a similar construct.

The Cronbach coefficient  $\alpha$  reliability for the IIRS and FORS was 0.90 and 0.88, respectively. Each measure had internal consistency reliability that exceeded the conventional minimum standard of 0.75. The interrater agreement using the FORS was good with an observed mean ICC = 0.61 and median = 0.65 (range 0.20–0.86) across the 4 faculty raters (see Table 1).

Analysis at the individual item level showed no statistically significant correlations among separate IIRS and FORS item scores, including global rating items. In fact, no coefficients were significant, although approximately 4 significant coefficients would be expected by chance alone using the nominal  $\alpha = 0.05$  criterion. We found that the distribution of IIRS and FORS scores, when expressed as a percentage of the maximum, were similar; faculty were not more conservative and SPs were not more liberal in their respective ratings of student performance.

### Correlation of IIRS and FORS with PPI

The means and variances of the IIRS and PPI scores in the 23 randomly selected cases for the FORS substudy were not statistically different from the nonselected cases.

The FORS scale summary score did not correlate significantly with either the IIRS scale score ( $r=-0.20$ ;  $p=0.40$ ) or the 7-item communication skill-based measure of PPI ( $r=-0.22$ ;  $p=0.32$ ). However, the IIRS scale score correlated significantly ( $r=0.88$ ;  $p<0.0005$ ) with the PPI. Thus, the PPI score predicted 77% of the variance in the total IIRS scores (Table 2).

None of the PPI items correlated with the FORS scale score (range  $-0.42$  to  $-0.01$ ). In contrast, every PPI item was positively and significantly correlated ( $p<0.001$ ) with the IIRS scale score. The range of correlation coefficients with the total IIRS score was 0.65 (PPI 7: “met my needs”) to 0.85 (PPI 3: “listened actively”) (Table 2).

### CONCLUSIONS

We created and tested two new empirically designed measures for assessing student skills using highly skilled interpreters, using a Spanish-speaking case scenario during a high stakes examination. Using a subset of 23 clinical encounters, we examined SP and faculty ratings of student performance utilizing interpreters. The results suggest that the IIRS and the FORS each have acceptable levels of tangible content and face validity. It is reassuring that the PPI and the IIRS correlate well, suggesting that the IIRS may be a valuable measure for assessing interpreter use skills. The 2 measures also show high internal consistency reliability, and the FORS demonstrates an acceptable level of concordance among faculty observers.

The finding of the absence of correlation between items on the IIRS and the FORS is a surprise and of some concern, given that the verbal and nonverbal behaviors we sought to assess were identified *a priori* to represent desirable qualities of communication when an interpreter is present in a clinical encounter. We considered the reasons for this discordance by examining the individual items within the IIRS, PPI, and FORS. Items in the IIRS primarily address the SP’s assessment of the student’s verbal (items 2, 3, and 4; see IIRS Table 1) and nonverbal (items 1, 5, and 6; see IIRS Table 1) communication skills when a trained interpreter is present. Although very different in wording and content from the PPI items, the behaviors measured by the IIRS and the PPI both depend on SP assessment of the professional conduct and demeanor of the student. The high degree of correlation is not surprising considering that both scales were completed by the same SP on the same student for each encounter. Conversely, the FORS focuses on verbal behavior (Table 1: FORS items 1, 2, 3, 5, 6, 8, 9, and 10) demonstrated by the student, and, to a lesser

extent, nonverbal observable behaviors (Table 1: FORS items 4 and 7). Examining the individual behaviors assessed for the IIRS and the FORS, it is notable that only 2 items reflect exactly the same behaviors: “eye contact” and “addressing the patient in the first person”. The remaining 4 (observable behavior) items for the IIRS focus on positioning, nonverbal communication, and responding to the patient’s needs. The remaining 7 (verbal) items on the FORS focus on behaviors that the SPs were not trained to assess: explanation of the interview to the interpreter and patient, listening without interruption, closing the encounter appropriately, and keeping the interpreter on track. Furthermore, faculty rated the encounters as outside observers viewing a DVD, whereas SPs were participants in the encounter and subject to emotional, interpersonal, and other nuances. Although the psychometric properties of SP ratings are established,<sup>22</sup> other data suggest that the lens for examining these skills should be widened beyond SP rating scales.<sup>23</sup> A recent study on real clinical encounters suggested that the positive subjective experience of partnership in an encounter does not fully reflect communication skills as measured by the PPI, and articulates the need for the inclusion of affective relationship dynamics (such as trust and power) in the assessment of communication skills.<sup>24</sup> Furthermore, other factors such as primary childhood language and patient-provider ethnicity may also impact SP assessments of student communication skills.<sup>25</sup> In addition we speculate that faculty (compared to trained SPs) may rate according to global impression rather than individual behaviors.

Despite its surprising results, our study has several strengths. We used an iterative process based on the review of the literature and clinical expertise to design, pilot test, and refine the SP case and the IIRS and FORS measures. This process mimicked the plan-do-study-act (PDSA) model used in quality improvement.<sup>26</sup> Second, we used a highly structured and standardized setting of a high stakes assessment to test the performance of the new measures. This allowed us to control for variability in patient and interpreter responses and clinical tasks. Third, we used a variety of faculty raters with 1 quality in common: expertise in teaching and assessing medical students’ skill in using interpreters in clinical encounters.

Our study has some limitations. It was conducted at 1 institution for 1 class of medical students using a relatively small number of encounters with 1 case scenario. We examined only a Spanish language encounter. However, by standardizing the language used and the training of the interpreter in this field study, we have now provided a foundation to test the FORS with patient encounters in different languages and settings.

Table 2. IIRS and FORS Scale Scores: Correlations with PPI Scale Scores

n=23	IIRS		FORS	
	Pearson correlation coefficient	$\rho$ (2-tail)	Pearson correlation coefficient	$\rho$ (2-tail)
PPI 1: appeared professionally competent	0.85	<0.0005	-0.30	0.16
PPI 2: effectively gathered information	0.74	<0.0005	-0.41	0.07
PPI 3: listened actively	0.85	<0.0005	-0.13	0.55
PPI 4: established personal rapport	0.85	<0.0005	-0.18	0.42
PPI 5: appropriately explored my perspective	0.78	<0.0005	-0.01	0.98
PPI 6: addressed my feelings	0.83	<0.0005	-0.07	0.74
PPI 7: met my needs	0.65	0.001	-0.29	0.19
Total PPI score	0.88	<0.0005	-0.22	0.32

Clinical application of these measures presents challenges. The IIRS is limited by a need for intensive SP training and may not be easily adapted to real patient encounters where patients are often monolingual and *ad hoc* interpreters are common. Effective use of the FORS also depends on the skills of the interpreter. Furthermore, it is unclear from this study whether SP or faculty ratings should be considered the 'gold standard' for assessing student communication skills for using an interpreter. These limitations need to be addressed before either scale can be used in actual practice situations. Further research will involve demonstrating the face and construct validity of the FORS as used by faculty.

In summary, this study provides 2 reliable measures that can be used in standardized settings and have the potential to be adapted for real patient encounters. At our institution, the FORS is used in a formative way to provide feedback to clinician trainees after an encounter, and also as a faculty development tool. However, the validity of the FORS in actual clinical settings has not yet been demonstrated as this study was conducted in a standardized examination setting. Future plans include the examination of the IIRS and FORS in different language encounters, use in different health professions learners (such as nursing), and validation of the FORS against other faculty rating scales.

The US is becoming increasingly linguistically and culturally diverse.<sup>21,27</sup> Training future physicians to use interpreters according to accepted standards<sup>21,28,29</sup> is essential to the development of good clinicians.<sup>5,19</sup> As student learning is closely linked to assessment, this study represents a step toward the goal of setting objective educational standards in the use of interpreters.

---

**Acknowledgments:** This project was supported in part by grant awards from the National Institutes of Health (NIH), the National Heart, Lung and Blood Institute, K07 HL-04-012 (2004-9), and from the AAMC grant initiative "Enhancing Cultural Competence in Medical Schools" (2005-2008) supported by the California Endowment. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH or AAMC. The authors gratefully acknowledge the assistance of Sarah Pardee and Jennifer Encinas in the manuscript preparation and data presentation.

**Conflict of Interest:** None disclosed.

**Corresponding Author:** Desiree Lie, MD, MSED; Department of Family Medicine, 101 The City Drive South, Building 200, Rt 81, Room 512, Orange, CA 92868, USA (e-mail: dalie@uci.edu).

## REFERENCES

1. **Betancourt JR.** Cross-cultural medical education: conceptual approaches and frameworks for evaluation. *Acad Med.* 2003;78:560-9.
2. **Flores G, Gee D, Kastner B.** The teaching of cultural issues in U.S. and Canadian medical schools. *Acad Med.* 2001;75(5):451-5.
3. American Association of Medical Colleges. Tool for Assessing Cultural Competence Training (TACCT). <http://www.aamc.org/meded/tacct/start.htm>. Accessed on August 6, 2007.
4. **Romero CM.** Using medical interpreters. *Am Fam Phys.* 2004;69:2720-2.
5. **Karliner LS, Perez-Stable EJ, Gildengorin G.** The language divide. The importance of training in the use of interpreters for outpatient practice. *J Gen Intern Med.* 2004;19:175-83.
6. **Perez-Stable EJ, Napoles-Springer A.** Interpreters and communication in the clinical encounter. *Am J Med.* 2000;108:509-10.
7. **McLeod RP.** Your next patient does not speak English: translation and interpretation in today's busy office. *Adv Pract Nurs Q.* 1996;2:10-4.
8. **Jacobs EA, Shepard DS, Suaya JA, Stone EL.** Overcoming language barriers in health care: costs and benefits of interpreter services. *Am J Public Health.* 2004;94:866-9.
9. **Rivadeneira R, Elderkin-Thompson V, Silver RC, Waitzkin H.** Patient centeredness in medical encounters requiring an interpreter. *Am J Med.* 2000;108:470-4.
10. **Aranguri C, Davidson B, Ramirez R.** Patterns of communication through interpreters: a detailed sociolinguistic analysis. *J Gen Intern Med.* 2006;21:623-9.
11. **Fagan MJ, Diaz JA, Reinert SE, Sciamanna CN, Fagan DM.** Impact of interpretation method on clinic visit length. *J Gen Intern Med.* 2003;18:634-8.
12. **Tocher TM, Larson EB.** Do physicians spend more time with non-English-speaking patients? *J Gen Intern Med.* 1999;14:303-9.
13. **Kuo D, Fagan MJ.** Satisfaction with methods of Spanish interpretation in an ambulatory care clinic. *J Gen Intern Med.* 1999;14:547-50.
14. **Lee LJ, Batal HA, Maselli JH, Kutner JS.** Effect of Spanish interpretation method on patient satisfaction in an urban walk-in clinic. *J Gen Intern Med.* 2002;17:641-5.
15. **Zabar S, Hanley K, Kachur E, et al.** "Oh! She doesn't speak English!" Assessing resident competence in managing linguistic and cultural barriers. *J Gen Intern Med.* 2006;21:510-3.
16. U.S. Department of Health and Human Services, Office of Minority Health. National standards of cultural and linguistically appropriate services in health care. 2000. <http://www.omhrc.gov/templates/content.aspx?ID=87>. Accessed August 6, 2008.
17. **Hudelson P.** Improving patient-provider communication: insights from interpreters. *Fam Pract.* 2005;22:311-6.
18. **Kaufert JM, Putsch RW.** Communication through interpreters in healthcare: ethical dilemmas arising from differences in class, culture, language, and power. *J Clin Ethics.* 1997;8:71-87.
19. **Flores G.** The impact of medical interpreter services on the quality of health care: a systematic review. *Med Care Res Rev.* 2005;62:255-99.
20. **Makoul G.** Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med.* 2001;76:390-3.
21. U.S. Census Bureau. R1603. Percent of people 5 years and over who speak English less than "very well": 2004. Places within United States. Ranking table. Data set: 2004 American Community Survey. Available at: [http://factfinder.census.gov/home/saff/main.html?\\_lang=en&\\_ts](http://factfinder.census.gov/home/saff/main.html?_lang=en&_ts). Accessed August 6, 2007.
22. **Williams RG.** Have standardized patient examinations stood the test of time and experience? *Teach Learn Med.* 2004;16(2):215-22.
23. **Rose M, Wilkerson L.** Widening the lens on standardized patient assessment: what the encounter can reveal about the development of clinical competence. *Acad Med.* 2001;76(8):856-9.
24. **Saba GW, Wong ST, Schillinger D, et al.** Shared decision making and the experience of partnership in primary care. *Ann Fam Med.* 2006;4(1):54-62.
25. **Fernandez A, Wang F, Braveman M, Finkas LK, Hauer KE.** Impact of student ethnicity and primary childhood language on communication skill assessment in a clinical performance examination. *J Gen Intern Med.* 2007;22(8):1155-60.
26. **Speroff T, O'Connor GT.** Study designs for PDSA quality improvement research. *Qual Manag Health Care.* 2004;13:17-32.
27. **Shin HB, Bruno R.** Language use and English-speaking ability. United States Census, 2000: 1-3. <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>. Accessed August 6, 2007.
28. Civil Rights Act of 1964. <http://usinfo.state.gov/usa/infousa/laws/majorlaw/civilr19.htm>, July 2, 1964, Document Number PL 88-352, 88th Congress, H. R. 7152. Accessed August 6, 2007.
29. **Grubbs V, Chen AH, Bindman AB, Vittinghoff E, Fernandez A.** Effect of awareness of language law on language access in the health care setting. *J Gen Intern Med.* 2006;21:683-8.