

Published in final edited form as:

Patient Educ Couns. 2011 November ; 85(2): 194–200. doi:10.1016/j.pec.2010.10.005.

Examination of standardized patient performance: Accuracy and consistency of six standardized patients over time

Lori A.H. Erby^{a,*}, Debra L. Roter^a, and Barbara B. Biesecker^b

^aThe Johns Hopkins Bloomberg School of Public Health, Health, Behavior and Society, Baltimore, MD, USA

^bNational Human Genome Research Institute, Social and Behavioral Research Branch, Bethesda, MD, USA

Abstract

Objective—To explore the accuracy and consistency of standardized patient (SP) performance in the context of routine genetic counseling, focusing on elements beyond scripted case items including general communication style and affective demeanor.

Methods—One hundred seventy-seven genetic counselors were randomly assigned to counsel one of six SPs. Videotapes and transcripts of the sessions were analyzed to assess consistency of performance across four dimensions.

Results—Accuracy of script item presentation was high; 91% and 89% in the prenatal and cancer cases. However, there were statistically significant differences among SPs in the accuracy of presentation, general communication style, and some aspects of affective presentation. All SPs were rated as presenting with similarly high levels of realism. SP performance over time was generally consistent, with some small but statistically significant differences.

Conclusion and practice implications—These findings demonstrate that well-trained SPs can not only perform the factual elements of a case with high degrees of accuracy and realism; but they can also maintain sufficient levels of uniformity in general communication style and affective demeanor over time to support their use in even the demanding context of genetic counseling. Results indicate a need for an additional focus in training on consistency between different SPs.

Keywords

Standardized patient; Genetic counseling; Provider–patient communication; Accuracy; Consistency

1. Introduction

Since their introduction in the 1960s, the use of standardized patients (SPs) has become commonplace in the teaching and assessment of communication skills during health professional training programs, in objective structured clinical exams (OSCEs) for certification and licensing, and in research studies designed to examine some aspect of medical communication or to evaluate programs with medical-visit associated outcomes [1–3]. Despite the widespread use of SPs, performance studies are rare and limited in scope

© 2010 Published by Elsevier Ireland Ltd.

*Corresponding author at: The Johns Hopkins Bloomberg School of Public Health, Health, Behavior and Society, 624 N. Broadway, Room 755, Baltimore, MD 21205, USA. Tel.: +1 410 502 4414; fax: +1 410 955 7241. lerby@jhsph.edu, Lhamby@jhsph.edu (Lori A.H. Erby).

[4,5]. Most assessments focus on accurate portrayal of case specifics, usually a set of symptoms and medical history facts [5,6]. The more socio-emotional dimensions of a case, such as the patient's affective demeanor and general style of verbal and/or nonverbal communication are rarely addressed. Moreover, while many authors note that SP accuracy is monitored during training and sometimes throughout actual exercises, few report the results [7–10]. An exception is a series of studies by Tamblyn and colleagues in which an assessment of SP performance with medical students and family practitioners across multiple cases included history and physical exam items as well as elements related to the presentation of patient affect [5,11,12]. Average accuracy scores in regard to case specifics were greater than 90% in each study. The accuracy score for affective script items (89.5%) was only slightly lower than that for history items (93.5%) [11].

While performance variation in the context of training programs may only affect the quality of the individual learning exercise, the few studies designed to address SP performance variation among multiple SPs presenting the same case and across SPs over time suggest that potentially important sources of performance variation exist that could confound research study results or have more serious implications for conclusions drawn within certification or licensing exams [5,6,11,13].

The current study was designed to systematically and comprehensively assess the following research questions (1) what are the differences in performance of the same case portrayed by different SPs? and (2) how does SP performance on the same case differ over time? SP performance was assessed across four dimensions: (1) presentation accuracy of case specifics, including details of the family and medical history, and the portrayal of the psychosocial features of the case; (2) SPs' general style of verbal communication and verbal activity level; (3) SPs' affective demeanor; and (4) genetic counselors' perceptions of SP realism.

2. Methods

2.1. Overview

Data for this study come from the Genetic Counseling Video Project (GCVideo); a cross-sectional study of genetic counseling using SPs [14]. The study enrolled a national sample of 177 genetic counselors who conducted a simulated visit at one of two meetings of the National Society of Genetic Counselors (NSGC) (2003 and 2004). The counselors were free to choose either a routine prenatal or cancer case. Details regarding recruitment are published elsewhere [14].

A total of nine SPs participated in the study; six women and three men, equally representing Caucasian, African American, and Hispanic ethnicities. Each counselor was assigned to an SP such that the ethnicity of the patient and whether or not the patient was accompanied by her spouse was randomly determined. One hundred sixty-seven (94%) of the sessions were of sufficient quality to be transcribed and analyzed.

The Johns Hopkins Bloomberg School of Public Health Committee on Human Research approved the study.

2.2. SPs

The SPs were graduate students or their acquaintances. None were trained in genetic counseling or other clinical health care fields or had prior acting or SP experience. All were English-speaking. The Hispanic SPs were fluent in English but spoke with a recognizable accent.

2.3. Prenatal and cancer cases

The two study cases included (1) a woman seeking pre-amniocentesis counseling based on an indication of advanced maternal age and (2) a woman with a family history of breast and ovarian cancer seeking information about BRCA1/2 genetic testing. The patient in both cases was 38 years old, had a working class background, and a deep faith in God. The spouse was supportive of his wife. Neither the patient nor spouse was prepared to make a decision regarding genetic testing during the visit. The cases included items from the patient's medical history, family history, prior knowledge and beliefs, social and lifestyle information, and emotional reactions.

2.4. Training of SPs

All actors were cross-trained on both cases using a slightly abbreviated method based on common SP training practices [15]. Training consisted of four, two-hour group role-playing sessions. The focus of training was on the mastery of script items, general communication style, and appropriate affect. The SPs were instructed to follow the lead of the genetic counselor by providing information only when prompted. When closed-ended questions were asked, SPs provided simple direct responses without elaboration. In response to open-ended questions, however, a more detailed response was provided. In both cases, the patients were instructed to appear friendly and moderately anxious about testing.

Because of our interest in the communication of patients with limited literacy skills, instruction of the SPs emphasized a communication style thought to be consistent with that of a high school graduate. Not only was it stressed that the patient would have no prior exposure to genetic counseling and little specific knowledge of genetics, but it was also specified that she would be unlikely to initiate discussion of topics, ask questions, or disclose worries and concerns without encouragement and prompting [16].

2.5. Measures

The performance of the SPs was assessed through an analysis of the session videotapes and transcripts. Although both male and female SP performance was examined, the current analysis focuses solely on the female SP, as she was the primary patient.

2.5.1. SP mention of script items and presentation accuracy—Following a similar procedure to that of Tamblyn and colleagues [5], scoring sheets were developed using the items specified in the case outlines and applied to written transcripts. The prenatal case included 53 distinct items: 25 clinical (biomedical and family history information such as “I am 16 weeks pregnant” and “One of my male cousin’s sons is ‘not right’”) and 28 psychosocial items (verbal expressions of emotions, attitudes, beliefs, and social situation such as “I am mostly worried about making sure my pregnancy is healthy” and “I get a lot of support through God and prayer”). Similarly, the cancer case included a total of 55 items: 22 clinical and 33 psychosocial items. Each item within a session was assigned a score indicating the presence of a genetic counselor’s prompt for the information, the item’s mention during the session, and a dichotomous indication of presentation accuracy. The percentage of case items mentioned was calculated for each session.

Although SPs were instructed to reveal items within the case whenever prompted by the genetic counselor, there were instances in which SPs failed to disclose information in response to such a prompt. Accuracy scores for each session were calculated by dividing the total number of items that were given correctly by the SP by the total number of opportunities the SP had to provide the information (the sum of each SP’s mentioned items plus any unanswered genetic counselor prompts for scripted items). The percentage of case

items mentioned and the accuracy scores were calculated separately for clinical and psychosocial items.

2.5.2. SPs' general communication style—The general verbal communication style of the SPs was assessed through the application of the Roter Interaction Analysis System (RIAS). As has been described previously, the RIAS was adapted for use in genetic counseling and was applied directly to videotaped sessions without transcription by two coders with a high degree of reliability [14]. Coders applied a code from a list of mutually exclusive and exhaustive categories to each RIAS-defined utterance or complete thought expressed by each speaker within the session. The following six composite communication scores were created by combining individual RIAS codes assigned to each SP-expressed utterance: clinical information-giving (personal and family medical history), psychosocial information-giving (psychological and lifestyle information), question-asking (open and closed-ended questions in either the clinical or psychosocial realm), social talk (social conversation, approvals, compliments, laughter), emotional talk and partnership-building (empathy, showing concern, expressing reassurance or optimism, partnership), and facilitative talk (paraphrasing, checking for understanding, asking for reassurance, bidding for repetition). To examine differences in SPs' verbal activity levels, the ratio of SP to genetic counselor utterances was calculated for each session.

2.5.3. Affective demeanor—RIAS coders rated the warmth and anxiety levels of the SPs after each session on a 6-point scale. Higher scores on these ratings indicated greater degrees of the affect in question.

2.5.4. Genetic counselors' ratings of SP realism—After completing the simulated genetic counseling session, each genetic counselor was asked to rate how “real” the SP appeared to be, on a 4-point scale, from “not at all real” to “completely real”.

2.5.5. Time variables—Three sets of time variables were created to characterize each visit. To allow for the exploration of differences between multiple sessions performed by an SP within a single day, we created two dichotomous variables: one indicating whether or not the visit was the first visit of the day for that particular SP (to examine warm-up effects) and one indicating whether or not the visit was the fifth or later visit of the day for that SP (to examine the effect of fatigue). To allow for the exploration of differences between performances over consecutive days of taping within a conference, we created two additional dichotomous variables: one indicating whether or not the visit occurred on the first day of the conference and one indicating whether or not the visit occurred on the final day of the conference. Finally, to statistically account for differences that may have occurred between the two different years of taping the cancer case, we created a dichotomous variable to indicate the year.

2.6. Analyses

All analyses were conducted using Intercooled Stata 10 [17]. To explore performance differences among SPs, analysis of variance was performed for each outcome, including case and presence of the standardized spouse as dichotomous covariates.

In addition, a multivariate regression analysis was carried out to simultaneously examine the relationships of case, presence of spouse and the three sets of time variables with each specific measure of SP performance as the outcome variable. Because each SP saw many genetic counselors, observations of each outcome cannot be considered to be independent. In order to account for this, all differences over time were examined using Generalized

Estimating Equations (GEE) assuming an exchangeable within-subjects correlation structure and using model-based estimates of the standard errors [18].

3. Results

3.1. Description of the study population

The socio-demographic characteristics of the 91 prenatal and 76 cancer genetic counselors who participated in the study have been reported elsewhere [14]. In brief, the counselors were broadly representative of the membership of the NSGC.

3.2. SP performances

Each SP performed between two and eight sessions a day on each of the five consecutive days of each conference. The total number of visits for each SP varied from 24 to 33. The genetic counseling sessions ranged in length from 23 to 92 min, with average lengths of 45 and 52 min respectively for prenatal and cancer sessions.

3.3. SPs' performance of script items

Based on the results of the multivariate model, SPs tended to mention more of the scripted items in the prenatal (71% of clinical items; 51% of psychosocial items) than in the cancer case (52% of clinical items; 45% of psychosocial items) ($z = -5.93, p < 0.001$; $z = -1.73, p = 0.084$). Each of the scripted items in both cases was elicited by at least one genetic counselor. The female SPs also mentioned significantly more of their own scripted items when the standardized spouse was not present (68% of clinical items; 52% of psychosocial items) than when he was present (57% of clinical items; 45% of psychosocial items) ($z = -4.03, p < 0.001$; $z = -3.22, p = 0.001$).

Overall, accuracy of script item presentation was high across cases (clinical item accuracy averaged 91% and 89% for the prenatal and cancer cases, respectively; psychosocial item accuracy likewise averaged 92% and 89%). There were no statistically significant differences between the two types of cases or between cases with or without a standardized spouse in clinical item accuracy ($z = 0.83, p = 0.408$; $z = 0.64, p = 0.515$). SPs were significantly more accurate in their presentation of the psychosocial items in the prenatal case ($z = -3.11, p = 0.002$) and when the standardized spouse was not present ($z = -2.51, p = 0.012$).

Table 1 shows that there were some differences across the six SPs in both clinical and psychosocial item accuracy. However, there were no systematic differences between SPs in the percentage of script items mentioned.

3.4. SPs' general communication style

Across all sessions, almost half (45%) of SPs' talk was comprised of clinical information-giving, and one-quarter was characterized by psychosocial information-giving. Questions comprised a small proportion of SPs' talk (3%). Other categories of patient talk included social talk (9%), emotional talk and partnership-building (13%), and attempts to facilitate engagement (3%). SPs were less verbally active than genetic counselors, with a mean ratio of SP to counselor talk of $.23 \pm .09$ (~1:4). SPs' talk in the prenatal case was characterized by significantly greater proportions of clinical information-giving ($z = -2.08, p = 0.037$), marginally significantly greater proportions of social talk ($z = -1.90, p = 0.058$) and facilitative talk ($z = 1.68, p = 0.098$), and lower proportions of psychosocial information-giving ($z = 2.35, p = 0.019$) and emotional talk ($z = 3.60, p < 0.001$) when compared to the cancer case. SPs tended to be more verbally active in the prenatal than in the cancer case (0.25 and 0.22 respectively; $z = -3.01, p = 0.003$).

SPs' talk when the spouse was present was characterized by significantly greater proportions of social talk ($z = 3.22, p = 0.001$), marginally significantly greater proportions of question-asking ($z = 1.82, p = 0.069$), and marginally significantly lower proportions of clinical information-giving ($z = -1.88, p = 0.060$). SPs also tended to be more verbally active when the standardized spouse was present (0.25 vs. 0.22; $z = 2.40, p = 0.017$).

Adjusting for case differences and differences related to the presence of the spouse, there were some dissimilarities in the general communication styles across the various SPs (see Table 2). There was a statistically significant difference among SPs in question asking, with one SP consistently asking more questions than the others. Five percent of this patient's total talk was devoted to questions (an average of 9.1 questions per session), compared to an overall average of three percent for the other SPs (an average of 4.1 questions per session). SPs differed significantly in their use of social talk, emotional talk and partnership-building, and facilitative talk. There was a marginally significant difference in the amount of clinical information-giving, with one SP tending to devote fewer of her utterances to providing clinical details (39% vs. 46% for all others). There were no statistically significant differences among SPs in overall verbal activity. Using Cohen's f as an indicator of effect size, the detected communication differences would be characterized as medium to large differences in communication indicators.

3.5. SPs' affective demeanor

Ratings of SP demeanor by coders reflected low levels of anxiety, with an average score of 1.7 on a 5-point scale, and moderate levels of warmth, with an average score of 3.2. Anxiety scores did not significantly differ by case ($z = -1.33, p = 0.182$), but scores were higher on warmth in the prenatal than in the cancer case (3.2 and 2.9 respectively; $z = -2.00, p = 0.045$).

SPs were not rated as having different levels of warmth when performing with vs. without a standardized spouse ($z = -0.27, p = 0.785$). However, SPs were rated as more anxious when the visit had a spouse present (1.9 vs. 1.5; $z = 3.90, p < 0.001$).

One SP was rated consistently higher on both anxiety and warmth than others (see Table 2), with a mean anxiety rating of 2.4 and a mean warmth rating of 3.4 compared to an average of 1.5 and 3.1, respectively, for the other SPs. It should be noted that an anxiety rating of 2.5 reflects moderate levels of anxiety on the RIAS global affect scale.

3.6. Reality ratings across different SPs

Overall, the genetic counselors rated 24.4% of SPs as "completely real", 47.5% as "moderately real", and 26.3% as "somewhat real". Less than two percent of genetic counselors rated the patient's performance as "not at all real". Considering realism as a continuous variable, ratings for the individual SPs did not differ from one another (see Table 2), nor did ratings differ by the type of case ($z = 0.46, p = 0.649$) or by the presence of the spouse ($z = 0.05, p = 0.959$).

3.7. Differences in SP performance over time

Performance over time was explored in the same multivariate analyses described previously. There were no statistically significant differences in performance between the two years of taping the cancer case (data not shown). As can be seen in Table 3, there were no significant differences between performances during the first session of each day in comparison to later sessions when accounting for other sources of variance.

Likewise, no statistically significant differences were observed for any of the performance variables when comparing the last sessions of the day with earlier sessions. However, there was a marginally significant trend toward SPs being rated as less real during the last sessions of the day when compared with earlier sessions (2.8 vs. 3.0 on a 4 point scale; Cohen's $f^2 = 0.02$).

Examining trends in SP communication over several different days of performance, the SPs gave less psychosocial information (20% vs. 26%; Cohen's $f^2 = 0.03$) and asked more questions (4% vs. 3%; Cohen's $f^2 = 0.03$) on the first day. When comparing performance on the last day of each conference with performance on previous days, a similar but only marginally significant trend emerged, with SPs tending to have higher levels of psychosocial information-giving (30% vs. 24%; Cohen's $f^2 = 0.03$) and tending to ask fewer questions (2.7% vs. 3.4%; Cohen's $f^2 = .02$) on the last day of each conference. They also mentioned a significantly smaller percentage of the psychosocial script items (45% vs. 49%; Cohen's $f^2 = .02$) on the first day, although this difference was only marginally statistically significant. When comparing ratings of performance on the last day of each conference to previous days, SPs tended to be seen as less real on the last day of taping (2.7 vs. 3.0; Cohen's $f^2 = .02$).

4. Discussion and conclusions

4.1. Discussion

The SPs performed their cases with high degrees of accuracy and consistency over time during lengthy sessions with few breaks. Performance was generally consistent from session to session, with no statistically significant evidence of either a warm-up effect or an impact of fatigue as had been observed in a previous study [11]. There were also only a few examples of performance drift over the course of several days. Based on the observed effect sizes (Cohen's f^2), the statistically significant time differences are relatively small differences, and we had between 91% and 96% power to detect at least a medium-sized effect.

The most significant differences in our study were observed in the various performance characteristics between the six different SPs. While differences were observed between the SPs in general communication patterns, it should be noted that all categories of communication were in the same relative proportion, with the vast majority of SP's talk related to clinical information-giving in all visits. In considering the potential impact of the observed differences among the SPs, the distinction between statistically significant and clinically significant differences is important. The observed effect sizes (Cohen's f) indicate that the statistically significant differences would be considered to be medium to large differences that may be clinically meaningful, particularly when these differences occur on dimensions of communication that are the targets of a specific assessment or study. As the genetic counseling sessions in this study were often over an hour long and were verbally dominated by the genetic counselors, it is possible that even these relatively large differences had little effect on each genetic counselor's communication. However, interpersonal communication is highly reciprocal [19], and variation in general communication patterns or perceived affect between SPs could have led to variation in genetic counselors' behaviors. There is some evidence that counselors do change some aspects of their communication to match their patients' needs [20].

While the role of third parties in medical communication has been explored in several studies of actual patient provider communication [21–23], the impact of the presence of a standardized spouse on the performance of SPs has not been previously reported. The degree of tailoring within genetic counseling communication must also be considered when interpreting these observed differences. We cannot conclude that variations in performance

were driven solely by the SPs in our study because the SPs were trained to be responsive to cues provided by the genetic counselors, who may have driven the observed differences.

While other studies have previously noted differences in the accuracy of performance of script items between SPs, these have not generally included assessments of broader communication characteristics [5,6,11,13]. When considering these broader elements, the performance of each SP was likely shaped to some degree by her own personality [15]. For instance, a more affectively expressive individual may tend to be seen as more expressive when performing as an SP. There may be a tradeoff between increasing consistency of emotional expressivity and level of reality of the portrayal. It is notable, therefore, that each of the SPs received similarly high ratings on the measure of realism in spite of their differences in affect.

The balance between consistency and “realism” may shift based upon the needs of the individual training experience, assessment, or research study. In some instances, case consistency may outweigh the need for the SP to be seen as completely real; in others, the opposite may be true. In the context of high stakes exams, comparability in multiple performance areas over time and between different SPs is essential in order to assure that individual test-takers are faced with identical tasks. In a research setting, some variation may be tolerable as long as procedures are in place to allow for appropriate statistical controls. In contrast, in a training scenario, variation in some aspects of SP performance is unlikely to detract from the overall pedagogical mission [14]. In the current research project, consistency of the passive elements of the scripted case was important to our ability to capture genetic counselor-driven variation in communication, even with the possible tradeoff of reducing the overall realism of the case. Given the complexity of the communication task, observed accuracy deficits and variations in communication patterns indicate a need to define the minimal level of accuracy or consistency required for specific components of every SP task.

The SPs in our study differed significantly in their levels of question-asking, in spite of our training emphasis on how to respond to genetic counselors’ questions with an appropriate level of information and how to avoid asking questions that were not scripted. It is possible that individual SP characteristics play an important role here as well. Individuals who naturally communicate with an inquisitive style may be more likely to give in to a tendency to ask questions in a standardized medical encounter, suggesting a need for further emphasis in training on those aspects of a case which may be most unnatural for a given SP.

Although our study provides an unusually comprehensive analysis of variation in SP performance, there are several limitations. The genetic counselors in the study took time away from a conference to participate. We cannot rule out the possibility that observed changes in SPs’ performance and differences in ratings of SPs’ reality over time may have been driven in part by differences in the genetic counselors’ mood or engagement related to conference activities. It is also possible that the genetic counselors talked to other participants about the simulated cases. Although 84% of the genetic counselors overall reported that they had not discussed any aspect of the study with other participants, 39% of those who were videotaped on the last day of each conference reported that they had discussed “some aspects” of the case with other counselors. The generalizability of our findings is limited to some degree by the characteristics of our cases and of our SPs. It is possible that accuracy and consistency of performance may differ when SPs are scripted to be more active participants in the communication process. Also, as our study only examined the performance of our female SPs, we are limited in our ability to draw conclusions about male SP performance [15]. Finally, although the SPs in this study were asked to provide an assessment of each genetic counselor, our study was not designed to assess the reliability of

these assessments nor did we focus our training time on enhancing the reliability of these assessments as would have occurred if these were forming the basis for an examination. We cannot comment on the extent to which heightened scrutiny of the genetic counseling visit on the part of the SP might affect performance accuracy or consistency.

4.2. Conclusions

SPs are now a routine tool for medical education and communication research. Although our findings demonstrate a need for further attention to differences in performance between multiple SPs trained on the same case, the current findings demonstrate that well-trained SPs can not only perform the factual elements of a case with generally high degrees of accuracy and realism; but they can also maintain acceptable levels of uniformity in general communication style and affective demeanor over time in the demanding context of genetic counseling [24,25]. Genetic counseling sessions are far longer than most medical encounters, typically lasting from 30 min to an hour and a half [14,26–30]. In contrast, most medical cases using SPs range from 5 to 20 min [10,31]. Future research is needed to examine the ways in which SP characteristics such as personality might overtly influence performance accuracy and consistency as well as the degree to which such differences might be ameliorated by training.

4.3. Practice implications

Given the observation of some inconsistencies in performance between different actors portraying the same case, between actors performing with and without a standardized spouse, and to a lesser extent in performance over time, an increased emphasis on reproducibility in the training of SPs would be necessary before widespread use in high-stakes assessment of genetic counseling communication or in research settings in which the outcomes necessitate distinguishing between SP-driven and genetic counselor-driven differences in communication. In some SP exercises, the expected outcomes or goals may be such that even small variations in the performance of a specific aspect of the case may be of critical importance. Our findings emphasize the need to determine for each case the minimum levels of accuracy and consistency required on each specific aspect of communication, to provide a particular focus on those aspects during training so that each actor demonstrates the desired level prior to implementation in the field, and to monitor and provide feedback throughout the performance period.

We would further recommend that researchers using SPs use analyses that nest observations within SPs in order to increase analytic power.

Acknowledgments

This research was supported by grant 1R01HG002688-01A1, Genetic Counseling Processes and Analogue Client Outcomes, funded by the National Human Genome Research Institute of the NIH. This study was performed in partial fulfillment of the requirements for Dr. Erby's doctoral dissertation at the Johns Hopkins Bloomberg School of Public Health. The authors thank Rita Johnson for her transcription services, Erin McDonald for assistance in transcript coding, Mary Catherine Beach, Peter Zandi, and Ada Hamosh for their early insights, and the anonymous reviewers whose suggestions have considerably improved our manuscript. We would also like to thank the Johns Hopkins Bloomberg School of Public Health biostatistics consulting service for their helpful insights on the statistical modeling in this most recent version of the manuscript.

References

1. Makoul G. Commentary: communication skills: how simulation training supplements experiential and humanist learning. *Acad Med.* 2006; 81:271–274. [PubMed: 16501275]

2. Roter DL. Observations on methodological and measurement challenges in the assessment of communication during medical exchanges. *Patient Educ Couns.* 2003; 50:17–21. [PubMed: 12767579]
3. Barrows HS. An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Acad Med.* 1993; 68:443–451. [PubMed: 8507309]
4. Beullens J, Rethans J, Goedhuys J, Buntix F. The use of standardized patients in research in general practice. *Fam Pract.* 1997; 14:58–62. [PubMed: 9061346]
5. Tamblyn R, Klass D, Schnabi G, Kopelow M. The accuracy of standardized patient presentation. *Med Educ.* 1991; 25:100–109. [PubMed: 2023551]
6. Vu N, Steward D, Marcy M. An assessment of the consistency and accuracy of standardized patients' simulations. *J Med Educ.* 1987 December.62:1000–1002. [PubMed: 3681930]
7. Barrows H, Norman G, Neufeld V, Feightner J. The clinical reasoning of randomly selected physicians in general medical practice. *Clin Invest Med.* 1982; 5:49–55. [PubMed: 7116714]
8. Ainsworth M, Rogers L, Markus J, Dorsey N, Blackwell T, Petrusa E. Standardized patient encounters: a method for teaching and evaluation. *J Am Med Assoc.* 1991; 266:1390–1396.
9. Carney P, Dietrich A, Freeman D, Mott L. The periodic health examination provided to asymptomatic older women: an assessment using standardized patients. *Ann Intern Med.* 1993; 119:129–135. [PubMed: 8512162]
10. Hodges B, Regehr G, Hanson M, McNaughton N. An objective structured clinical examination for evaluating psychiatric clinical clerks. *Acad Med.* 1997; 72:715–721. [PubMed: 9282149]
11. Tamblyn R, Klass D, Schanbl G, Kopelow M. Factors associated with the accuracy of standardized patient presentation. *Acad Med.* 1990; 65:S55–S56. [PubMed: 2400506]
12. Tamblyn R, Abrahamowicz M, Berkson L, Dauphinee W, Gayton D, Grad R, et al. Assessment of performance in the office setting with standardized patients: first-visit bias in the measurement of clinical competence with standardized patients. *Acad Med.* 1992; 67:S22–S24. [PubMed: 1388544]
13. Badger L, deGruy F, Hartman J, Plant M, Leeper J, Ficken R, et al. Stability of standardized patients' performance in a study of clinical decision making. *Fam Med.* 1995; 27:126–131. [PubMed: 7737446]
14. Roter D, Ellington L, Erby LH, Larson S, Dudley W. The genetic counseling video project (GCVP): models of practice. *Am J Med Genet C Semin Med Genet.* 2006 Nov 15.142:209–220. [PubMed: 16941666]
15. Wallace, P. Coaching standardized patients: for use in the assessment of clinical competence. New York: Springer Publishing Company; 2007.
16. Roter, DL. Health literacy and the patient–provider relationship. In: Schwartzberg, JG.; VanGeest, JB.; Wang, CC., editors. *Understanding health literacy: implications for medicine and public health.* Chicago: American Medical Association; 2005. p. 87-100.
17. StataCorp LP. *Stata Statistical Software: Release.* 2007; 10
18. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics.* 1986; 42:121–130. [PubMed: 3719049]
19. Roter DL, Hall JA. Health education theory: an application to the process of patient-provider communication. *Health Educ Res.* 1991; 6:185–193. [PubMed: 10148690]
20. Pieterse AH, van Dulmen AM, Ausems MG, Beemer FA, Bensing JM. Communication in cancer genetic counseling: does it reflect counselees' previsit needs and preferences? *Brit J Cancer.* 2005; 92:1671–1678. [PubMed: 15841073]
21. Tsai MH. Who gets to talk? An alternative framework evaluating companion effects in geriatric triads. *Commun Med.* 2007; 4:37–49. [PubMed: 17714042]
22. Ishikawa H, Roter DL, Yamazaki Y, Hashimoto H, Yano E. Patients' perceptions of visit companions' helpfulness during Japanese geriatric medical visits. *Patient Educ Couns.* 2005; 61:80–86. [PubMed: 16242292]
23. Clayman ML, Roter D, Wissow LS, Bandeen-Roche K. Autonomy-related behaviors of patient companions and their effect on decision-making activity in geriatric primary-care visits. *Soc Sci Med.* 2005; 60:1583–1591. [PubMed: 15652689]

24. Trepanier A, Greb A, Kavanaugh M. Monitoring genetic counseling students' progress in developing practice-based competencies through standardized patient encounters. *J Genet Counsel.* 2002; 11:490–491.
25. Kinnersley P, Pill R. Potential of using simulated patients to study the performance of general practitioners. *Brit J Gen Pract.* 1993; 43:297–300. [PubMed: 8398247]
26. Hamby, L. The Johns Hopkins School of Hygiene and Public Health. 2001. Discussions of personal meaning in pre-amniocentesis genetic counseling [dissertation].
27. Aalfs CM, Oort FJ, de Haes HC, Leschot NJ, Smets EM. Counselor–counselee interaction in reproductive genetic counseling: does a pregnancy in the counselee make a difference? *Patient Educ Couns.* 2006; 60:80–90. [PubMed: 16332473]
28. Lynch H, Lemon S, Durhan C, Tinley S, Connolly C, Lynch J, et al. A descriptive study of BRCA1 testing and reactions to disclosure of test results. *Cancer.* 1997; 79:2219–2228. [PubMed: 9179070]
29. Kemel, Y. The Johns Hopkins School of Hygiene and Public Health. 2000. What happens during the prenatal genetic counseling session: exploratory study of genetic counseling [dissertation].
30. Butow P, Lobb E. Analyzing the process and content of genetic counseling in familial breast cancer consultations. *J Genet Counsel.* 2004; 13:403–424.
31. Harden R, Gleeson F. Assessment of clinical competence using an objective structured clinical examination (OSCE). *Med Educ.* 1979; 13:41–54. [PubMed: 763183]

Table 1

Variation across six standardized patients (SP) on performance of script items^{***}.

	Lowest mean SP score ± SD	Highest mean SP score ± SD	F-statistic	p-Value	Degrees of freedom	Effect size (Cohen's <i>f</i>) ^a
Percentage of script items mentioned ^a	57 ± 12%	73 ± 12%	1.34	0.2519	5	0.12
	44 ± 5%	54 ± 5%	1.61	0.1591	5	0.12
Accuracy ^b	85 ± 1%	95 ± 1%	3.55	0.0045	5	0.28
	86 ± 3%	96 ± 3%	5.84	0.0001	5	0.37

^a (Number of script items mentioned/total possible script items) × 100%. Prenatal case totals: 25 clinical items, 28 psychosocial items. Cancer case totals: 22 clinical items, 33 psychosocial items.

^b {Total number of items that were correctly stated/(genetic counselor prompts for information plus standardized patient-initiated information)} × 100%.

*** Analyses based on ANOVA with case and presence of spouse as covariates with 7 degrees of freedom.

Table 2

Variation in standardized patients' (SP) general communication, affective demeanor, and reality ratings^{***}.

	Lowest mean SP score ± SD	Highest mean SP score ± SD	F-statistic	p-Value	Effect size (Cohen's f)
SP general communication (mean use of each category as a percentage of total talk)	39% ± 4%	50% ± 4%	2.03	0.0789	0.18
Clinical information-giving	22% ± 2%	28% ± 1%	1.37	0.2543	0.11
Psychosocial information-giving	2% ± 0.4%	5% ± 0.4%	5.64	0.0001	0.40
Question-asking	6% ± 2%	14% ± 2%	11.74	<0.0001	0.57
Social Talk	11% ± 1%	15% ± 1%	2.96	0.0144	0.25
Emotional talk and partnership building	0.8% ± 0.2%	4% ± 0.2%	20.63	<0.0001	0.83
Facilitative talk	0.21 ± 0.02	0.26 ± 0.02	0.95	0.4495	0.08
Verbal activity level (SP talk/GC talk)	1.4 ± 0.2	2.4 ± 0.2	11.40	<0.0001	0.56
SP affective demeanor	2.8 ± 0.1	3.4 ± 0.1	3.61	0.0043	0.29
Reality rating ^c	2.8 ± 0.3	3.3 ± 0.3	1.22	0.3025	0.08

^a Rating on a 6 point scale, with higher values indicating higher levels of anxiety.

^b Rating on a 6 point scale, with higher values indicating higher levels of warmth.

^c Rating on a 4 point scale, with '4' indicating "completely real".

*** Analyses based on ANOVA with case and presence of spouse as covariates with 7 degrees of freedom.

Table 3

Variation in standardized patient performance over time^{***}.

	Sessions over a day			Days over a conference series				
	First session compared to subsequent sessions		Last session compared to subsequent sessions	First day compared to subsequent days		Last day compared to preceding days		
	z-Score	p-Value	z-Score	p-Value	z-Score	p-Value		
Percentage of items mentioned	0.28	0.776	-0.70	0.486	0.23	0.815	0.60	0.549
Psychosocial	-1.23	0.219	-0.31	0.754	-1.75	0.079	1.03	0.304
Accuracy								
Clinical	0.44	0.663	-0.47	0.642	-0.49	0.624	0.31	0.753
Psychosocial	-1.29	0.197	1.18	0.240	-1.03	0.302	-0.56	0.578
SP general communication								
Clinical information-giving	0.25	0.804	-0.14	0.888	0.93	0.354	-0.88	0.381
Psychosocial information-giving	-0.87	0.386	-0.44	0.662	-2.07	0.038	1.85	0.061
Question-asking	1.64	0.101	-0.44	0.661	2.13	0.033	-1.81	0.071
Social talk	0.09	0.930	-1.04	0.298	1.31	0.190	-1.48	0.140
Emotional talk and partnership building	0.68	0.498	-0.47	0.637	-0.33	0.740	-0.15	0.879
Facilitative talk	-1.38	0.167	0.33	0.740	-0.67	0.504	0.62	0.536
Verbal activity level (SP talk/GC talk)	-1.60	0.109	-0.63	0.526	-0.69	0.488	-0.73	0.463
SP affective demeanor	-0.81	0.416	-0.02	0.985	0.50	0.617	-0.76	0.448
Warmth	-0.30	0.763	-1.09	0.275	0.83	0.405	0.82	0.415
Reality rating	0.79	0.427	-1.95	0.051	0.62	0.538	-1.81	0.070

*** All analyses based on main effects observed in multivariate GEE with exchangeable correlations to account for nesting within SPs with each of these outcomes as the dependent variable, along with the following independent variables: case, presence/absence of spouse, five variables to account for timing differences over sessions within a day, across multiple days, and across the two taping periods. Model significance based on a Wald chi square test with 7 degrees of freedom.