2016

Standardized Patient Encounters Periodic Versus Postencounter Evaluation of Nontechnical Clinical Performance

T. Robert Turner

Mark W. Scerbo
Old Dominion University, mscerbo@odu.edu

Gayle A. Gliva-McConvey

Amelia M. Wallace

Original Publication Citation

Turner, T. R., Scerbo, M. W., Gliva-McConvey, G. A., & Wallace, A. M. (2016). Standardized patient encounters periodic versus postencounter evaluation of nontechnical clinical performance. *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare, 11*(3), 164-172. doi:10.1097/sih.0000000000000146

Standardized Patient Encounters

Periodic Versus Postencounter Evaluation of Nontechnical Clinical Performance

T. Robert Turner, PhD;
Mark W. Scerbo, PhD;
Gayle A. Gliva-McConvey;
Amelia M. Wallace

Introduction: Standardized patients are a beneficial component of modern healthcare education and training, but few studies have explored cognitive factors potentially impacting clinical skills assessment during standardized patient encounters. This study examined the impact of a periodic (vs. traditional postencounter) evaluation approach and the appearance of critical verbal and nonverbal behaviors throughout a standardized patient encounter on scoring accuracy in a video-based scenario.
Methods: Forty-nine standardized patients scored either periodically or at only 1 point in time (postencounter) a healthcare provider's verbal and nonverbal clinical performance during a videotaped standardized patient encounter. The healthcare provider portrayed in this study was actually a standardized patient delivering carefully scripted verbal and nonverbal behaviors in their portrayal of an actual physician. The encounter itself was subdivided into 3 distinct segments for the purpose of supporting periodic evaluation, with the expectation that both verbal and nonverbal cues occurring in the middle segment would be more challenging to accurately report for participants in the postscenario evaluation group as a result of working memory decay.
Results: Periodic evaluators correctly identified a significantly greater number of critical verbal cues midscenario than postencounter evaluators ($P < 0.01$) and correctly identified a significantly greater number of critical nonverbal cues than their postscenario counterparts across all 3 scenario segments ($P < 0.001$). Further, postscenario evaluations exhibited a performance decrement in terms of midscenario correct identifications that periodic evaluators did not ($P < 0.01$). Also, periodic evaluators exhibited fewer verbal cue false-positives during the first segment of the scenario than postscenario evaluators ($P < 0.001$), but this effect did not extend to other segments regardless of the cue type (ie, verbal or nonverbal).
Discussion: Pausing lengthier standardized patient encounters periodically to allow for more frequent scoring may result in better reporting accuracy for certain clinical behavioral cues. This could enable educators to provide more specific formative feedback to individual learners at the session's conclusion. The most effective encounter design will ultimately depend on the specific goals and training objectives of the exercise itself.
(*Sim Healthcare* 11:164–172, 2016)

Key Words: Standardized patients, Simulated patient encounter, Clinical performance evaluation, Human factors, Working memory, Attention, Medical evaluation, Medical simulation, Medical assessment, Educational testing and measurement.

In recent years, many healthcare-oriented training programs have been developed to enhance the cognitive (eg, decision making, situation awareness, problem solving) and interpersonal skills that collectively define one's nontechnical clinical competency. Studies in this area have suggested that didactic training and subsequent task exposure are not sufficient for developing interpersonal skills.[1] Rather, an effective training program for nontechnical skills must link meaningful, interactive learning opportunities with timely feedback to facilitate positive interpersonal development. A form of simulation-based training that has historically proven effective for interpersonal skills within the healthcare domain is that of standardized patients.[2] A standardized patient is defined as any individual who is trained to portray a patient with a specific condition in a realistic, standardized, and repeatable way (where portrayal/presentation varies on the basis of learner performance).[3]

Standardized patients can be leveraged for education and training objectives, including history taking/consultation, physical examination, and other clinical skills in a simulated clinical environment. They are also an invaluable source of performance feedback.[3] Standardized patients are optimal for simulation-based nontechnical skills training in some contexts, representing a higher degree of authenticity for learners

From the Department of Psychology (T.R.T.), Old Dominion University, Norfolk, VA; Department of Psychology (M.W.S.), Old Dominion University, Norfolk, VA; and Professionals Skills Teaching and Assessment (G.A.G.-M.), Sentara Center for Simulation and Immersive Learning Eastern Virginia Medical School (A.M.W.), Norfolk, VA.

Reprints: T. Robert Turner, PhD, James & Sylvia Earl SAIL Center, Anne Arundel Medical Center, 2001 Medical Parkway, Annapolis, MD 21401 (e-mail: rturner@facs.org).

The authors declare no conflict of interest.

DOI: 10.1097/SIH.0000000000000146

because they hone the requisite communication skills for common clinical activities such as patient interviews and communicating difficult information.[2] Further, the feedback provided by a trained standardized patient is critical for nontechnical clinical skill development and growth.

Standardized patients typically portray an ailing patient or, in some cases, another healthcare professional to create a standardized patient encounter in which trainees can gain experience practicing a variety of nontechnical skills through interactions with a live, responsive human being. While maintaining the guise of this scripted persona, the standardized patient will often simultaneously evaluate clinical performance and provide detailed feedback relating to specified performance criteria. However, rating clinical performance in real time can be challenging when coupled with the requirements of a realistic and dynamic patient portrayal. This may be especially true when rating nonverbal behaviors.[4]

Previous studies have addressed the quality of standardized patient evaluations and found it to be quite high, [5–7] yet evidence also suggests that some of the scoring variance is directly attributable to the individual standardized patients[8] and that standardized patients may be susceptible to cognitive challenges associated with the nature of observation and the conditions under which clinical behaviors are presented.[4] To realize the full potential of a standardized patient-based approach for nontechnical clinical skills training, it is important to first understand the capabilities and limitations of behavioral recognition during a standardized patient encounter. In this case, behavioral recognition denotes the successful observation, recognition, and subsequent reporting of specific verbal (ie, spoken words) or nonverbal (ie, tone of voice, body language, gestures, eye contact, facial expressions, or physical proximity) cues.

When evaluating verbal and nonverbal behaviors, there are limits to one's attention. Observing elements in the environment requires a balance of attention (bottom-up processing) and long-term memory (top-down processing). The perception and subsequent recall or recognition of cues can be impeded by omission (ie, not looking at a piece of information) or a narrowing of attention due to heavy task load or distractions. If attention becomes overloaded during a highly complex scenario, an observer might "miss" key information in the environment or subsequently fail to store that information in memory.[9]

Further, evaluations based on delayed (eg, postencounter) performance appraisal also present several challenges. One must accurately perceive, encode, and maintain task-relevant information while simultaneously attending to the scenario as it continues to unfold. Previous studies have documented difficulties in recalling information in both clinical and nonclinical contexts.[4,10] It is therefore important to understand what individuals can accurately report from a standardized patient encounter and be aware of cognitive limitations when considering the structure of standardized patient encounters for nontechnical skills training.

The ability to maintain and manipulate information is a function of working memory, which has been defined as a set of cognitive components involved in processing, briefly storing, and manipulating information to support task-relevant objectives.[11] Baddeley's working memory model[12,13] is composed of a central executive system with limited capacity that interacts with 3 other subsystems: the visuospatial sketchpad, phonological loop, and episodic buffer.[14] These working memory subsystems facilitate the processing of visual/spatial, verbal, and temporal information, respectively.

When multiple tasks are performed simultaneously and draw on the same working memory subsystem (eg, recalling different auditory messages presented to each ear), the resulting attentional overload will likely hinder performance.[15] On the other hand, multiple tasks drawing from separate working memory components can be performed together without exhibiting a significant impact on working memory integrity. It therefore stands to reason that standardized patients constructing speech-based dialogue while monitoring the same from a healthcare trainee will place a great deal of attentional load on the verbal component of working memory and may have difficulty encoding and maintaining some of the verbal information in working memory. Further, as the standardized patient encounter unfolds over time, information will continue to accumulate in the episodic buffer until it can be offloaded or data loss occurs.[16,17]

Previous studies have demonstrated that memory also depends in part on where an item occurs in a sequence of events. Items that occur in the middle of a sequence are less likely to be recalled than items occurring early or late in the sequence. Items occurring later in the sequence are recalled with the highest frequency.[18] These serial position effects have been demonstrated in verbal tasks[19] as well as in visual-spatial recall tasks.[20,21]

It is believed that serial position effects are caused by the different memory operations involved in the encoding of early- and late-sequence items.[22,23] Items occurring early in the sequence are likely to be encoded in long-term memory as a result of increased rehearsal and little interference from preceding items, thus resulting in a primacy effect.[22,24,25] Items occurring later in the sequence are likely still active in working memory, thus producing a recency effect.[22,24,26]

By extension, the relevant performance details most salient to standardized patients during postencounter evaluations may be influenced by their location within the standardized patient encounter. The first (or last) few minutes in the encounter may receive disproportionate weight during the postencounter evaluation due to a greater salience or memorability of information from those time segments, whereas most performance, which lies between these end points, may remain underrepresented. It is therefore a possibility that standardized patients categorize trainees schematically, on the basis of initial (or final) impressions rather than continuously updating observation data.

This present study extends previous work designed to explore factors that may influence recognition of verbal and nonverbal behavioral cues during clinical evaluation.[4,27] Turner et al[27] used a sample of university undergraduate students to evaluate video-recorded scenarios, which served as a pilot study for the current study. Results from that pilot study suggested that periodic recall may result in better verbal and nonverbal accuracy for clinical skills assessment than the traditional postencounter framework.

In the present study, it was therefore expected that periodic evaluation would enable participants to work from a smaller subset of information in working memory at any given time throughout the scenario as a result of offloading this information more frequently. Thus, the burden on working memory would be reduced,[28–30] resulting in more accurate cue recognition and improved scoring accuracy. The periodic evaluation benefit was expected to be most apparent during the middle segment of the encounter as a result of the challenges associated with recall of midsequence information.[22] We therefore hypothesized that participants in the periodic evaluation group would demonstrate better recognition accuracy for both verbal and nonverbal clinical behavioral cues than those in the postencounter evaluation group. Further, we hypothesized that recognition of verbal and nonverbal clinical behaviors would be more accurate during the first and third encounter segments than the middle segment, especially for postencounter evaluation participants who were unable to periodically offload information from working memory.

METHOD

Participants

Forty-nine professional standardized patients (23 males, 26 females) from the Sentara Center for Simulation and Immersive Learning at Eastern Virginia Medical School were recruited to participate in this institutional review board–approved study (Table 1). Recruitment was limited to individuals who were at least 18 years of age with self-reported normal or corrected vision and hearing. Each participant was financially compensated for their time.

Study Design

A 3 encounter segment (segment 1, patient history; segment 2, substance abuse; and segment 3, future goals) × 2 evaluation format (periodic evaluation vs. postencounter evaluation) mixed design was implemented in this study. Encounter segment was presented as a within-subjects factor and evaluation format was treated as a between-subjects factor. Participants were randomly assigned to 1 of the 2 evaluation conditions.

Procedure

Upon arrival, all participants were consented and completed a short demographic questionnaire. Participants were briefed on the scenario format and the instruments to be used in evaluating the verbal and nonverbal clinical behaviors of a videotaped "physician" (portrayed by a trained standardized patient) conducting a patient interview. The briefing lasted approximately 30 minutes and included a review of basic case details, an overview of the case presentation format, and an item-by-item review of each evaluation instrument. The evaluation instruments' content was familiar to participants because it was based on an existing set of evaluation criteria currently in use by standardized patients trained at this institution. As a result, most briefing focused on the modified evaluation format and procedural details.

Next, participants watched a 5-minute videotaped practice encounter and completed an initial performance evaluation to familiarize themselves with both the observation-evaluation task format and the behavioral evaluation instruments. Consistent with the subsequent experimental trials, the practice trial contained 6 verbal and 7 nonverbal target cues embedded in the video clip. Upon completion of the practice trial, participants were allowed to ask questions to ensure that they understood each item on the instruments and were comfortable with the scoring objectives. An experimenter reviewed participants' practice ratings with them immediately after the practice trial to ensure instrument and procedural comprehension. No time limit was placed on the question/answer phase of this study.

Once participants indicated that they were comfortable with implementing both the evaluation instruments and evaluation protocol, they were asked to observe each of the experimental encounter video clips and rate the standardized patient's clinical performance. Approximately half of the participants were randomly assigned to the periodic rating format (ie, at the end of each individual scenario segment), whereas the remaining participants rated the standardized patient's performance at only 1 point in time—the conclusion of the scenario.

In the periodic evaluation condition, the full-length scenario was briefly paused at 3 predetermined points to gather data from the participants. Pauses coincided with transitions between segments of the encounter (eg, between the patient history interview and discussion of substance abuse), and participants were given up to 6 minutes at each video freeze point to complete their verbal and nonverbal clinical performance evaluations before proceeding to the next segment. Rewinding and replaying video content during pauses were not permitted. Participants in the postencounter evaluation condition were asked to watch the full-length scenario without periodic freezes. Rather, the postencounter evaluation participants completed the performance evaluations once after the encounter was complete. These participants were given up to 18 minutes to complete the verbal and nonverbal evaluations for all 3 video segments to match the total amount of evaluation time provided to their periodic evaluation counterparts.

TABLE 1. Study Participant Demographics (N = 49)

Demographic Factor	Affirmative, %	Mean (SD)	Minimum	Maximum
Age, y	—	50.77 (16.83)	23.00	87.00
Experience, y	—	04.87 (04.75)	00.17	17.00
Sex, female	53.06	—	—	—
Formal acting experience	42.86	—	—	—
Completed instructor-level standardized patient training	22.45	—	—	—

TABLE 2. Behavioral Cues Comprising the Verbal and Nonverbal Checklist Inventories, Respectively

Verbal
Introduced self*
Addressed the patient by his/her surname (last name)†
Asked patient to state their own concerns, if they had any*
Set an agenda or sequence of discussion topics‡
Asked the patient for their narrative concerning key events†
Requested information to help establish a timeline of the chief complaint*
Used technical or medical jargon*
Verified information that the patient provided by stating it back to them†
Attempted to learn the patient's perspective and/or beliefs about the injury*
Inquired about the patient's feelings about the injury and if/how it has changed the patient's life
Addressed the impact of the injury on the patient's family†
Attempted to determine what financial and/or emotional support systems the patient could depend on during treatment‡
Used supportive comments to demonstrate empathy and acknowledge the patient's situation†
Encouraged the patient to ask questions‡
Admitted lack of knowledge or experience†
Attempted to determine whether the patient fully understood the information provided about injury, prognosis, and/or treatment options‡
Assessed the patient's motivation to change behavior, mindset, or personal habits‡
Explained any relevant investigations, tests, or interventions to the patient‡
Provided closure to the patient by discussing next steps, future goals, and/or when next meeting will occur
Invited the patient to contribute thoughts, ideas, suggestions, and/or preferences in determining the plan of care
Used a multiple or double-barreled question touching on more than 1 issue*
Nonverbal
Eye contact
Looked at watch
Looked at a pager or cell phone
Body language
Leaned far forward (toward patient), bracing torso with elbow on top of knee*
Leaned far backward (away from patient) into a slouching position*
Leaned to the side, bracing arm or elbow on counter top
Crossed arms†
Crossed legs‡
Crossed ankles
Head orientation
Tossed head backward, as if comprehending a key point of the conversation‡
Nodded head to affirm patient's statements*
Shook head, as if telling the patient "no"
Cocked head to one side
Facial expressions
Smiled at the patient, demonstrating acceptance‡
Smirked at patient, demonstrating sarcasm or derision†
Pressed lips together, demonstrating empathy or concern†
Frowned at patient, demonstrating condescension or judgment
Yawned‡
Gestures of the hand
Used slow, fluid, small (calm) hand gestures†
Used quick, erratic, large (aggressive) hand gestures†
Rubbed ear with hands
Scratched self on face ‡
Ran fingers through hair
Pointed index finger at patient
Touched mouth with finger/pen†
Clenched fist*
Direct touch
Shook patient's hand‡
Touched patient on the arm for encouragement or empathy
Touched patient on the shoulder for encouragement or empathy
Touched patient on the leg for encouragement or empathy
Touched patient on the back for encouragement or support
Tone/voice
Tone of voice was judgmental or condescending at times*
Tone of voice was empathetic at times‡
Interrupted the patient*
Cleared throat†
Tapping
Tapped pen, indicating impatience toward patient*
Tapped hands, indicating impatience toward the patient
Tapped feet, indicating impatience toward the patient

*Cue embedded in segment 1.
†Cue embedded in segment 2.
‡Cue embedded in segment 3.

Standardized Patient Encounter

The case adapted for use in this study was designed to address substance abuse[31] and was modified with embedded verbal and nonverbal behavioral cues using a carefully scripted dialogue and event sequences. The modified scenario was video recorded to produce a standardized stimulus for all participants. This case was selected primarily because it could be naturally subdivided into 3 qualitatively similar segments in accordance with the core skill areas defined by Wallace.[2] This resulted in the following 3 consecutive scenario segments of 6 to 7 minutes each: an interview eliciting the patient's history, discussion of emerging substance abuse concerns, and assessing the patient's motivation for behavior or lifestyle change. The full scenario was sufficiently complex to represent a typical standardized patient encounter and ran for approximately 18 to 20 minutes in length.

The videotaped encounter depicted a physician interviewing a patient about her medical history, substance dependency issues, and goals/motivation for behavior change successively. The encounter was filmed using 2 standardized patients, 1 portraying the physician and the other portraying the patient. Before recording the scenario, both standardized patients received a detailed script including both verbal dialogue and nonverbal actions to be captured in the video recording (including an event sequence detailing when these actions were to occur). This was to ensure standardization and thus experimental control of the video segments for all participants. The standardized patient "actors" spent several days mastering the scripts before conducting a small number of live rehearsals and ultimately a full day of video recording to produce the final set of video clips for this study.

Dependent Measures

Participants were asked to score clinical performance by completing 2 behavioral checklists. The first checklist was a modified version of the Master Interview Rating Scale (MIRS)[32] developed specifically for this study, focusing on 21 verbal clinical behaviors (Table 2) with a dichotomous yes/no response option for each item. Participants were responsible for

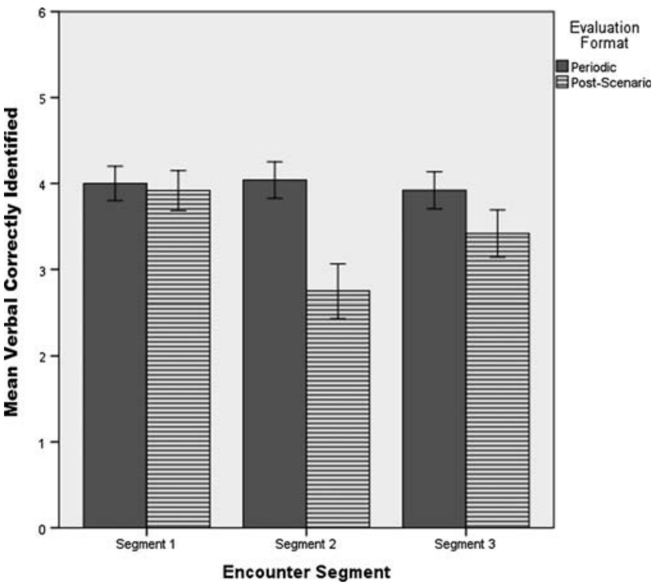

FIGURE 1. Participants' verbal correct identification scores by segment and evaluation (error bars depict ± 1 SE).

tracking all 21 verbal checklist items throughout the entire scenario. A total of 6 verbal cues from this checklist were embedded in each segment, for a total of 18 of a possible 21 verbal checklist cues actually appearing in the videotaped scenario.

A similar checklist of relevant nonverbal behaviors was also developed for this study on the basis of a sampling of published standardized patient practices and research.[4,33–35] The resulting checklist contained a total of 37 nonverbal behavioral items of interest for standardized patient evaluators (Table 2), of which 7 were embedded into each of the scenario's 3 segments. Thus, a total of 21 of a possible 37 nonverbal behaviors were embedded in the experimental scenario. Similar to the modified MIRS verbal checklist, nonverbal behaviors were scored via a dichotomous yes/no response indicating whether a given behavior was exhibited.

The verbal and nonverbal behavioral checklists resulted in 2 dependent measures each—a subset of checklist items that were correctly indicated as having occurred (ie, correct identifications) and a subset of checklist items that were incorrectly indicated as having occurred when in fact they did not (ie, false-positive reports). A naive third-party reviewer, who was not subject to the single-viewing constraint imposed on the study's participants, scored each video segment using both the verbal and nonverbal behavioral checklists to ensure that the appropriate embedded target cues were both present and unambiguous. The reviewer's checklist results were in complete agreement with the master list of embedded experimental cues. Thus, verbal and nonverbal clinical performance cues were considered to be of sufficient quality and distinctiveness across all segments for the purposes of this study.

RESULTS

Verbal Correct Identifications

An analysis of variance (ANOVA) conducted on verbal correct identification scores revealed a significant segment × evaluation interaction ($F_{2,94}$ = 3.73; P < 0.05, partial η^2 = 0.07; power = 0.62). The interaction was such that segment 2 verbal correct identification scores for periodic evaluation participants were significantly greater than those of postencounter evaluators (P < 0.01; Fig. 1). However, verbal correct identification performance did not significantly differ by group for segments 1 or 3 (Table 3).

Verbal False-Positive Reports

All participants demonstrated a significant verbal cue reporting deficit during segments 2 and 3 as evidenced by an increasingly greater proportion of false-positive reports in these segments. Analysis of variance results indicated a significant main effect for segment on participants' verbal false-positive reports ($F_{2,94}$ = 11.14; P < 0.001; partial η^2 = 0.19; power = 0.99; Fig. 2). Participants generally exhibited a greater number of false-positive reports in segments 2 [mean (SD), 4.86 (2.77); P = 0.001] and 3 [mean (SD), 5.45 (2.68); P < 0.001], than in segment 1 [mean (SD), 3.49 (2.21)]. Periodic evaluators exhibited significantly fewer false-positive reports for verbal cues during segment 1 than their postencounter comparators (P = 0.034), suggesting at least some periodic evaluation benefit in terms of mitigating verbal false-positive reports.

Nonverbal Correct Identifications

An ANOVA indicated significant main effects on the nonverbal correct identification scores for encounter segment ($F_{2,94}$ = 4.17; P = 0.01; partial η^2 = 0.09; power = 0.78) and evaluation ($F_{1,47}$ = 17.06; P < 0.001; partial η^2 = 0.27; power = 0.98; Fig. 3). With regard to encounter segment, the nonverbal correct identification scores were significantly higher in segments 1 [mean (SD), 3.31 (1.46); P < 0.05] and 3 [mean (SD), 3.43 (1.29); P < 0.01] than in segment 2

TABLE 3. Between-Group Comparisons of Verbal and Nonverbal Cue Recognition Performance Across Scenario Segments Reported as Means (SD)

		Correct Identifications			False-Positive Identifications		
	Cue Type	Periodic (n = 25)	Postscenario (n = 24)	P	Periodic (n = 25)	Postscenario (n = 24)	P
Segment 1	Verbal	4.00 (1.00)	3.92 (1.14)	0.786	2.84 (1.65)	4.17 (2.53)	0.034*
	Nonverbal	3.92 (1.53)	2.67 (1.09)	0.002†	2.60 (2.12)	2.79 (2.23)	0.759
Segment 2	Verbal	4.04 (1.06)	2.75 (1.54)	0.001†	4.68 (2.53)	5.04 (3.04)	0.652
	Nonverbal	3.56 (1.64)	1.96 (1.43)	0.001†	4.44 (3.18)	4.54 (2.50)	0.902
Segment 3	Verbal	3.92 (1.08)	3.42 (1.35)	0.155	4.96 (2.65)	5.96 (2.66)	0.195
	Nonverbal	3.80 (1.15)	3.04 (1.33)	0.038*	1.84 (1.68)	2.33 (1.24)	0.249

*Statistically significant at P < 0.05.
†Statistically significant at P < 0.01.
Data are presented as mean (SD) of the indicated behaviors.

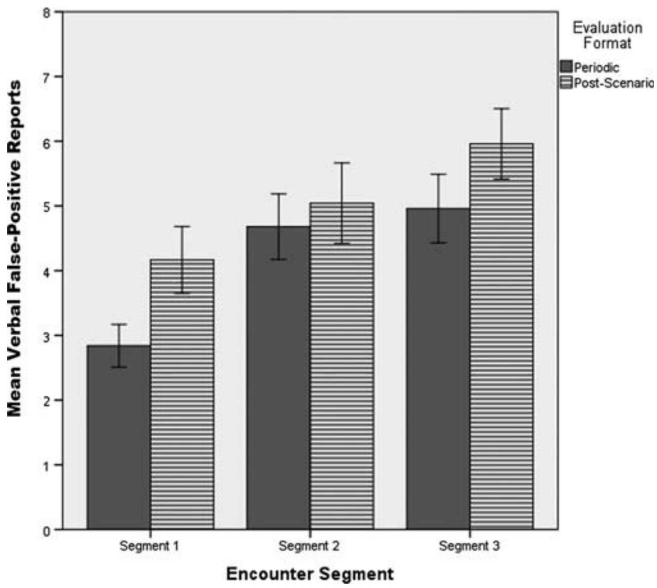

FIGURE 2. Participants' verbal false-positive reports by segment and evaluation (error bars depict ± 1 SE).

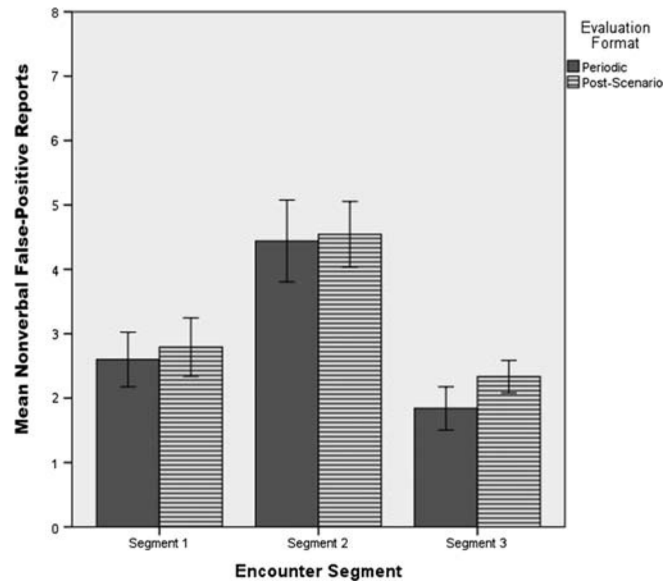

FIGURE 4. Participants' nonverbal false-positive reports by segment and evaluation (error bars depict ± 1 SE).

[mean (SD), 2.78 (1.72)]. With regard to evaluation format, periodic evaluators exhibited significantly higher nonverbal correct identification scores than postencounter evaluators across all 3 encounter segments (Table 3). Postencounter evaluators had significantly lower nonverbal correct identification scores in segment 2 than in either segment 1 (P = 0.01) or 3 ($P < 0.01$), constituting a decrement in segment 2 nonverbal cue recognition that periodic evaluation participants did not exhibit.

Nonverbal False-Positive Reports

An ANOVA indicated a significant main effect for segment on participants' verbal false-positive reports [$F_{2,94}$ = 23.24; $P < 0.001$; partial $\eta^2 = 0.33$; power = 1.0; Fig. 4]. Participants in both groups exhibited a significantly greater number of false-positive nonverbal cues in segment 2 [mean (SD), 4.49 (2.84)] than in segments 1 [mean (SD), 2.69 (2.15); $P < 0.001$] and 3 [mean (SD), 2.08 (1.48); $P < 0.001$].

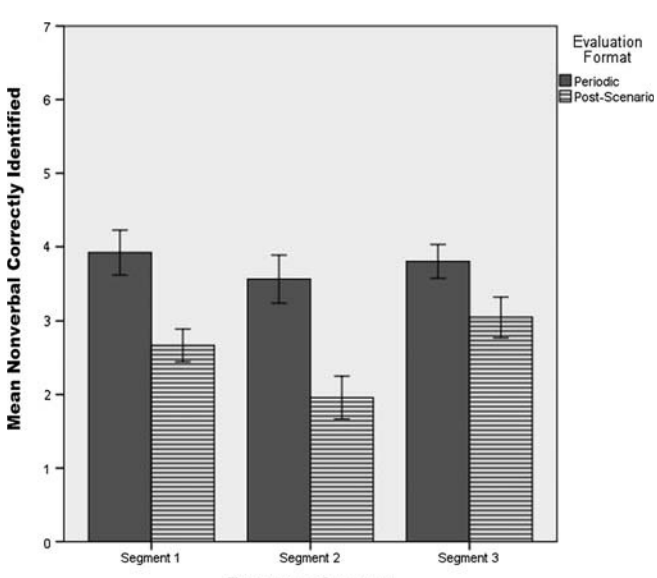

FIGURE 3. Participants' nonverbal correct identification scores by segment and evaluation (error bars depict ± 1 SE).

DISCUSSION

The aim of this study was to expand the results of a previous study,[27] suggesting that periodic evaluation of a standardized patient encounter may provide benefits more than that of the traditional postencounter evaluation format for clinical performance evaluation. The current study was conducted in an operational training environment drawing from a participant pool of trained standardized patients to determine whether the effects observed by Turner et al[27] are generalizable to a more ecologically valid setting.

We predicted that periodic evaluation participants would demonstrate superior verbal and nonverbal clinical cue recognition than their postencounter evaluation comparators as a result of being able to offload task-relevant information from working memory more frequently.[28–30] Further, it was predicted that periodic evaluation would be particularly beneficial for reporting information presented during the middle encounter segment, which theoretically represents a greater challenge for participants in terms of memory encoding and subsequent recall.[22,23] The results of this study generally support the previous hypotheses.

Periodic evaluation participants correctly identified more segment 2 verbal cues than did the postencounter evaluation participants. However, participants did not differ significantly in terms of verbal correct identifications for segments 1 or 3. We had anticipated a general performance decrement for all participants during the middle segment as a result of established serial position and memory effects,[22,23] but mitigation of this verbal cue performance decrement for participants in the periodic evaluation group is an encouraging result. It suggests that periodic offloading of standardized patient working memory to support more accurate midscenario verbal scoring might be a viable option for educators. In this case, periodic evaluation effectively reduced the number of critical verbal cues maintained in working memory from 18 to 6. As a result of this reduced load, periodic evaluation participants exhibited better reporting accuracy during an otherwise more perceptually challenging middle segment.

All participants reported an increasingly greater number of false-positive cues during segments 2 and 3, although periodic evaluation participants reported significantly fewer segment 1 verbal false-positives than their postencounter counterparts. This suggests at least some benefit of periodic evaluation in terms of mitigating verbal false-positives reports.

The strongest support for the benefit of periodic versus postscenario evaluation came in the form of correct identification of nonverbal cues. Periodic evaluators exhibited significantly better nonverbal correct identification scores than postencounter evaluators across all 3 encounter segments. Further, postencounter evaluators exhibited a performance decrement for nonverbal cue recognition in segment 2 that the periodic evaluation participants did not exhibit. Superior nonverbal correct cue identification performance across all segments combined with the absence of an anticipated middle-segment decrement for periodic evaluators suggests that a periodic scoring approach may result in enhanced nonverbal clinical skills reporting accuracy as well.

Despite the apparent benefits of periodic reporting to correct cue identification, both evaluation groups exhibited a significant increase in nonverbal false-positive reports during the middle segment. Periodic evaluation is therefore unlikely to mitigate the number of false-positive cues reported by evaluators beyond the segment 1 verbal cues discussed previously. Evidence for the benefit of a periodic versus postencounter evaluation is therefore limited primarily to enhanced (correct) identification of nonverbal cues throughout the entire scenario as well as enhanced correct identification of verbal cues during the otherwise challenging middle segment.

Study Limitations

There were several limitations in this study. The more passive nature of the video-based standardized patient encounter, as opposed to a live interaction-based scenario, is 1 potential study limitation. Rather than interacting with a live physician, participants observed a videotaped scenario and scored clinical performance at a predesignated time or times instead. This was done to control the quality and temporal location of embedded verbal and nonverbal cues across participants. As a result, participants were not exposed to the cognitive load associated with live standardized patient encounters.[4,10]

A second factor is that the presentation order of the encounter segments themselves was not counterbalanced. Randomizing the presentation order of encounter segments would have broken the comprehensible "flow" of scripted events. Thus, segments were presented sequentially (ie, patient history in segment 1, followed by a substance dependency discussion in segment 2 and then goals for behavioral change in segment 3) for logical continuity. Lack of randomization in the segment presentation sequence could have unintentionally introduced order effects into the response patterns of participants, but in this case, a predetermined presentation order was deemed necessary to preserve the integrity of a realistic patient encounter. In addition, all 3 encounter segments focused on a single aspect of a typical patient encounter—the patient interview. This was done intentionally to ensure that all 3 segments were qualitatively similar for investigative purposes, although other equally important patient encounter areas such as a physical examination are also worth investigating.

Another study limitation involves the complexity of detecting individual behavioral cues embedded in each segment of the scenario. Although the cues used in this study were not calibrated and balanced against an index of perceptual complexity, steps were taken to ensure that detection complexity for each cue was minimized. Whenever possible, verbal cues were articulated with the same language specified on the verbal behavior checklist to reduce ambiguity. For example, when the standardized patient was directed to portray empathy toward the patient, the keyword "empathize" served as a scripted marker of the corresponding verbal behavior empathetic tone (eg, "I can empathize"). It is possible that some cues were more vividly perceived than others as a result of the type of behavior exhibited. For instance, behaviors such as interrupting the patient might have proven more memorable than behaviors such as shaking the patient's hand as a result of the former's generally negative connotation.

Another feature and potential limitation of this study is that some participants were exposed repeatedly to the list of critical verbal and nonverbal clinical behaviors by means of repeated evaluation throughout the scenario. All participants were provided a brief (10 minute) overview and item-by-item explanation of the relevant performance checklists before beginning the study, so all participants were provided the same basis—albeit somewhat rudimentary—for the evaluation criteria. Despite receiving the same basic overview of the evaluation criteria, periodic evaluation participants were then asked to evaluate the standardized patient twice during the encounter, providing a working memory "refresher" in the form of multiple exposures to the evaluation instruments (and clinical behaviors of interest) that post-encounter evaluation participants did not receive.

Regarding the evaluation instruments used in this study, lists of verbal and nonverbal clinical performance items were presented to participants alongside dichotomous "yes/no" response options. Thus, performance was more reflective of participants' subsequent cue recognition processes rather than memory search and retrieval processes alone; this response format is consistent with standardized patient encounter assessments in which evaluators are asked to conduct assessments using a list of relevant behaviors as a memory cue (eg, Surgery Resident Objective Structured Clinical Examination evaluation instruments).[36] Standardized patient encounters often incorporate more complex assessment instruments designed to elicit not only whether a behavior was observed but also the perceived quality of the behavior (eg, using a Likert-type scale). However, the present study was concerned with participants' detection and recognition of cues rather than with perceived gradations of cue quality.

Another instrumentation issue relates to the number of independent verbal and nonverbal cues incorporated into the behavioral checklists. As previously noted, the verbal checklist used in this study was derived from the standard MIRS[32] assessment tool commonly used by practicing standardized patients to evaluate clinical performance in live scenarios. Although this study's verbal checklist did not contain the complete set of MIRS items, participants were responsible for a subset of 21 verbal behaviors, which could potentially occur at any given time throughout the scenario. The nonverbal checklist constituted an additional 37 behavioral items for which participants were responsible. It has been established that shorter evaluations result in greater standardized patient assessment accuracy, item discrimination, and reliability,[37] and this must be taken into account when devising an instrument for clinical performance assessment. However, the present study was designed to investigate a set of scenario configurations under which scoring performance may be impacted rather than on the design of an optimal evaluation instrument per se. Further, it is not atypical for standardized patients to rate learners for a large set of performance cues, particularly in more complex scenarios designed for advanced learners.

Another potential study limitation involves the necessarily brief orientation to the clinical evaluation instruments and case parameters. An actual standardized patient encounter should and often does involve more advanced preparation on behalf of the standardized patient evaluators. This includes sufficient time to memorize all relevant case materials and evaluation instruments and to discuss in-depth the case, goals, objectives, and any symptoms/conditions to be portrayed with responsible training personnel. They might also be given the opportunity to conduct dry-run rehearsals. Because of time limitations, experimental logistics, and resource availability, this level of advanced preparation was not practical for the present study. Despite receiving a necessarily brief case overview, all participants indicated that they fully understood the task requirements and evaluation items before observing the videotaped scenario. Regardless, the importance of dedicated standardized patient training time, checklist mastery, and mental fragmentation of cases to the overall standardized patient experience and assessment quality certainly warrant further research.

Finally, this study's participants were responsible for both identifying specific clinically relevant behaviors as well as the segment in which they occurred. Depending on case-by-case training or evaluation objectives, it may not be critical for evaluators to maintain temporal aspects of item occurrences in memory.

CONCLUSIONS

Certain aspects of a standardized patient's multiple roles within a simulated encounter might make them susceptible to a variety of psychological effects influencing the accuracy of clinical performance reporting. The current study suggests that pausing lengthier scenarios every few minutes to allow for more frequent scoring may improve reporting accuracy for verbal and nonverbal clinical behaviors at certain points in the scenario. Thus, a periodic evaluation approach might actually enhance the overall value of standardized patient encounters for formative training by allowing educators to provide more accurate, specific feedback than would otherwise be possible with the traditional postencounter approach. On the other hand, it is feasible that pausing scenarios periodically to allow for more frequent reporting introduces unanticipated challenges if the objective is to facilitate summative rather than formative assessment. Therefore, the most effective approach will ultimately be the one, which takes into consideration the established goals and objectives of the individual training exercise.

ACKNOWLEDGMENTS

The authors wish to acknowledge Eastern Virginia Medical School and the Sentara Center for Simulation and Immersive Learning for providing staff and materiel support for this study.

REFERENCES

1. Salas E, Diaz Granados D, Weaver SJ, King H. Does team training work? Principles for health care. *Acad Emerg Med* 2008;15:1002–1009.
2. Wallace P. *Coaching Standardized Patients: For Use in the Assessment of Clinical Competence.* 1st ed. New York: Springer Publishing; 2007.
3. Association of Standardized Patient Educators: "Standardized Patient" Definition. Association of Standardized Patient Educators Website. Available at: http://www.aspeducators.org. Accessed July 13, 2015.
4. Newlin-Canzone ET, Scerbo MW, Gliva-McConvey G, Wallace A. The cognitive demands of standardized patients: Understanding limitations in attention and working memory with the decoding of nonverbal behavior during improvisations. *Simul Healthc* 2013;8(4): 207–214.
5. De Champlain AF, Margolis MJ, King A, Klass DJ. Standardized patients' accuracy in recording examinees' behaviors using checklists. *Acad Med* 1997;72:S85–S87.
6. Furman GE. The role of standardized patient and trainer training in quality assurance for a high-stakes clinical skills examination. *Kaohsiung J Med Sci* 2008;24:651–655.
7. Williams RG. Have standardized patient examinations stood the test of time and experience? *Teach Learn Med* 2004;16:215–222.
8. van Zanten M, Boulet JR, McKinley D. Using standardized patients to assess the interpersonal skills of physicians: six years' experience with a high-stakes certification examination. *Health Commun* 2007;22: 195–205.
9. Jones DG, Endsley MR. Sources of situation awareness errors in aviation. *Aviat Space Environ Med* 1996;67:507–512.
10. Newlin-Canzone ET, Scerbo MW, Gliva-McConvey G, Wallace A. Attentional and mental workload demands in nonverbal communication. *Proceedings of the 55th Human Factors & Ergonomics Society Annual Meeting* 2011;1190–1195.
11. Baddeley AD, Logie RH. Working memory: The multiple-component model. In: Miyake A, Shah P, eds. *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control.* Cambridge: Cambridge University Press; 1999:28–61.
12. Baddeley AD, Hitch GJ. Working memory. In: Bower GH, ed. *The Psychology of Learning and Motivation: Advances in Research and Theory.* 8th ed. New York: Academic Press; 1974:47–89.
13. Baddeley AD, Wilson BA. Prose recall and amnesia: implications for the structure of working memory. *Neuropsychologia* 2002;40:1737–1743.
14. Baddeley AD, Hitch GJ, Allen RJ. Working memory and binding in sentence recall. *J Mem Lang* 2009;61:438–456.
15. Baddeley AD. The episodic buffer: a new component of working memory? *Trends Cogn Sci* 2000;4:417–423.
16. Moray N. Monitoring behavior and supervisory control. In: Boff KR, Kaufman L, Thomas JP, eds. *Handbook of Perception and Human*

Performance: Cognitive Processes and Performance. 2nd ed. New York: Wiley-Interscience; 1986:40.1–40.51.

17. Peterson L, Peterson MJ. Short-term retention of individual verbal items. *J Exp Psychol* 1959;58:193–198.
18. Deese J, Kaufman RA. Serial effects in recall of unorganized and sequentially organized verbal material. *J Exp Psychol* 1957;54:180–187.
19. Nairne JS, Whiteman HL, Woessner L. Symmetrical cuing effects for item and position information. *Am J Psychol* 1995;108:345–358.
20. Hay DC, Smyth MM, Hitch GJ, Horton NJ. Serial position effects in short-term visual memory: a SIMPLE explanation? *Mem Cognit* 2007;35:176–190.
21. Smyth MM, Hay DC, Hitch GJ, Horton NJ. Serial position memory in the visual-spatial domain: reconstructing sequences of unfamiliar faces. *Q J Exp Psychol A* 2005;58:909–930.
22. Atkinson RC, Shiffrin RM. Human memory: a proposed system and its control processes. *Psychol Learn Motiv* 1968;2:89–195.
23. Azizian A, Polich J. Evidence for attentional gradient in the serial position memory curve from event-related potentials. *J Cogn Neurosci* 2007;19:2071–2081.
24. Glanzer M, Cunitz AR. Two storage mechanisms in free recall. *J Verbal Learn Verbal Behav* 1966;5:351–360.
25. Rundus D. Analysis of rehearsal processes in free recall. *J Exp Psychol* 1971;89:63–77.
26. Postman L, Phillips LW. Short term temporal changes in free recall. *Q J Exp Psychol* 1965;17:132–138.
27. Turner TR, Scerbo MW, Gliva-McConvey G, Wallace A. Evaluating physician-patient communication: the effects of concurrent and retrospective assessment on the recall of verbal and nonverbal behaviors. In: *Proceedings of the 2013 International Annual Meeting of the Human Factors and Ergonomics Society.* San Diego, CA; 2013.
28. Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. *J Verbal Learn Verbal Behav* 1975; 14:575–589.
29. Hart SG, Loomis LL. Evaluation of the potential format and content of a cockpit display of traffic information. *Hum Factors* 1980;22:591–604.
30. Jacko JA. An empirical assessment of task complexity for computerized menu systems. *Int J Cogn Ergon* 1997;1:137–147.
31. Eastern Virginia Medical School: Substance Abuse Painter. *Training Scenario and Case Details*. 2012.
32. Master Interview Rating Scale: The clinical conversation and interview. *The Training Manual for Standardized Patients*. Norfolk, VA: Theresa Thomas Professional Skills Teaching and Assessment Center; 2005.
33. Collins LG, Schrimmer A, Diamond J, Burke J. Evaluating verbal and non-verbal communication skills, in an ethnogeriatric OSCE. *Patient Educ Couns* 2011;83:158–162.
34. Deladisma AM, Cohen M, Stevens A, et al. Do medical students respond empathetically to a virtual patient? *Am J Surg* 2007;193:756–760.
35. Griffith CH, Wilson JF, Langer S, Haist SA. House staff nonverbal communication skills and standardized patient satisfaction. *J Gen Intern Med* 2003;18:170–174.
36. American College of Surgeons: Surgery Resident Objective Structured Clinical Examination Manual with Examination Materials 2008. *DVD and CD-ROM* Available at: https://web4.facs.org/ebusiness/ProductCatalog/product.aspx?ID=124. Accessed July 13, 2015.
37. Yudkowsky R, Park YS, Riddle J, Palladino C, Bordage G. Clinically discriminating checklists versus thoroughness checklists: improving the validity of performance test scores. *Acad Med* 2014;89:1057–1062.