Assessment

The Test of Accurate Perception of Patients’ Affect (TAPPA): An ecologically valid tool for assessing interpersonal perception accuracy in clinicians

Judith A. Hall a,*, Amy N. Ship b, Mollie A. Ruben a, Elizabeth M. Curtin a, Debra L. Roter c, Sarah L. Clever d, C. Christopher Smith b, Karen Pounds e

a Department of Psychology, Northeastern University, Boston, USA
b Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, USA
c Department of Health, Behavior, and Society, Johns Hopkins University Bloomberg School of Public Health, Baltimore, USA
d Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, USA
e School of Nursing, Bouvé College of Health Sciences, Northeastern University, Boston, USA

ARTICLE INFO

Article history:
Received 29 June 2013
Received in revised form 27 September 2013
Accepted 5 October 2013

Keywords:
Communication skill
Emotion recognition
Empathic accuracy
Provider-patient communication
Accuracy of interpersonal perception

ABSTRACT

Objective: A clinician’s ability to infer patients’ thoughts and feelings is a critical component of high quality care. The goal of this article is to present a new test to measure this ability in clinicians, called the Test of Accurate Perception of Patients’ Affect (TAPPA).

Methods: Audiovisual clips were taken from patients’ actual medical visits. The patients reviewed the videotape after the visit to identify their thoughts and feelings during the visit. This information was used to extract short audiovisual clips for which the correct answer was the patient’s report of the thought or feeling associated with that clip. The TAPPA contains 48 audiovisual clips, each responded to in a multiple choice format.

Results: The TAPPA showed good psychometric properties (optimal mean and good variance, adequate internal consistency, and strong re-test reliability) and convergent validity with other tests of emotion recognition. In addition, the test showed predicted better performance by female than male participants.

Conclusion: The TAPPA promises to be a valuable tool for research and education on provider–patient relationships and quality of care.

Practice implications: A tool for testing clinicians’ understanding of patients’ thoughts and feelings may contribute to better quality of care and to improved selection and training.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Effective clinicians need to be aware of their patients’ emotions, personality, cultural and personal background, values, and expectations. An ability to “read” the patient plays a role in accurate diagnosis, predicting and recognizing nonadherence, relationship quality, and recognizing how the patient is reacting to the clinician. Accurate perception of patients is routinely acknowledged in discussions of patient-centered practice [1–5]. Yet, compared to the huge volume of research on what clinicians say and do, there has been much less research on clinicians’ accuracy in perceiving patients, on clinicians’ accuracy of interpersonal perception in general, and on the correlates and consequences of their accuracy [6]. There is, however, a great deal of research that provides a strong rationale for pursuing this topic in the clinical domain [7,8]. The present article presents a new test to assess clinicians’ accuracy in understanding the thoughts and feelings of patients, called the Test of Accurate Perception of Patients’ Affect (TAPPA).

1.1. Methods for studying clinicians’ accuracy in interpersonal perception

Some observation-based coding systems for evaluating clinicians’ behavior include items describing the clinician’s ability to perceive the patient accurately, thus relying on observers’ impressions as an indicator of the clinician’s interpersonal judgment accuracy. For example, an observer’s impression of physician sensitivity to patients is included in the Medical Interaction Process System for evaluating oncologists’ communication skill [9]. However, observers’ impressions of skill in
interpersonal perception cannot substitute for empirically tested skill because observers’ impressions are only weakly (though positively) correlated with empirical measurements of such accuracy [7]. Among other problems, observers are likely to have difficulty separating the clinician’s perception accuracy from what the clinician says and does. It is an empirical question whether accurate perception and appropriate behavior are positively correlated; they are certainly not synonymous and they can be separately measured as components of the care process.

Another possible way to capture clinicians’ accuracy in interpersonal perception is to ask them to rate their own skill. However, research has found that self-reports of accuracy in interpersonal perception are positively but only weakly correlated with tested skill, meaning these reports cannot substitute for empirically measured skill [7].

Yet another method is based on an in vivo assessment wherein clinicians’ ratings of their patients’ states are correlated with the patients’ corresponding self-ratings. This method involves data collection from both clinicians and patients, preferably multiple patients per clinician, and preferably with a videotape record of the clinician’s behavior during the encounter that is used to disambiguate the clinician’s perceptive ability from the patient’s clarity of expression [6].

A more efficient way to measure skill in this domain is to test it objectively in a standardized testing paradigm in which a test-taker makes judgments about what messages or emotions are being conveyed in cues, and those judgments are then scored right or wrong according to independent criteria. A number of tests of accurate interpersonal perception exist that have established validity [10,11], yet with one exception they were not developed with clinicians in mind. The exception is the Patient Emotion Cue Test (PECT) [12], which was designed for clinician test-takers and which measures emotion recognition in combined verbal and nonverbal modalities as portrayed by an actress playing the part of a patient making utterances to an imagined physician. The PECT has good psychometric properties and promising convergent validity.

1.2. Test of Accurate Perception of Patients’ Affect (TAPPA)

The Test of Accurate Perception of Patients’ Affect (TAPPA) adds ecological validity by using video clips of actual patients in their routine medical visits as the stimuli, and by using scoring criteria that are based on the patients’ verbatim reports of their thoughts and feelings moment to moment as determined through post-visit videotape review. The full TAPPA contains 48 audiovisual clips showing patients interacting with their physicians. The items are grouped into four modules of 12 items each that were constructed to be approximately interchangeable, allowing for flexibility (e.g., test shortening, training protocols where some items are used for training and other items are used for outcome testing, group discussion in a training setting, etc.).

The test’s design was based on the empathic accuracy paradigm [13,14], whereby the person in the video watches his/her video right after the interaction and describes all of his/her thoughts and feelings as they occurred. These descriptions are the criteria for scoring the test-taker’s accuracy. In the TAPPA, each test item consists of approximately 30 s leading up to the moment when that thought or feeling occurred. Thus, a test-taker has some context to draw on in deciding what the patient’s thought or feeling was. The TAPPA uses a multiple-choice answer format consisting of the correct answer (what the patient actually said the thought or feeling was) and three plausible but incorrect answers.

The present article presents psychometric analysis, normative data, and several investigations of the test’s construct validity in seven groups of participants. Specifically, the TAPPA was examined for convergent validity with other tests of accuracy in recognizing emotions from verbal and/or nonverbal cues. Validity was further addressed by the prediction that women would score higher on the test than men, because this is typically the case on tests of emotion recognition [15,16].

2. Method

2.1. Video archive for construction of the TAPPA

2.1.1. Recruiting and videotaping patients

The protocol was approved by the Committee on Clinical Investigations of the Beth Israel Deaconess Medical Center, Boston, MA, and the Committee on Human Subject Research of Northeastern University. One of the investigators (ANS) contacted physicians in the Healthcare Associates to obtain help in identifying internal medicine patients who might be willing to be videotaped. Patients were selected to fit the following criteria: (1) they had upcoming medical appointments, (2) the visit was likely to contain substantive discussion, and (3) the physician thought the patient might be interested in participating. The co-investigator contacted potential patients via letters and phone calls. Patients who agreed to take part signed an informed consent prior to their scheduled medical appointments. The medical appointment was videotaped with the camera facing the patient. If an exam was performed, the lens cap was placed on the camera but audio remained on during this phase. The archive contained 24 patients of six physicians. Patients were paid $50.00 for participating.

2.1.2. Post-visit videotape review

Immediately after each visit, the videotape was uploaded to a computer. Patients watched the videotape and stopped it every time they remembered having a thought or feeling. They were instructed to dictate exactly what they had been thinking and feeling at that moment, during which a camera caught both the computer screen that the patient was watching (to determine the exact moment of the thought or feeling) and the patient’s dictated thought or feeling.

2.2. Developing the pool of potential items

A research assistant transcribed the patients’ dictated thoughts and feelings. Video clips were edited to a target length of about 30 s so that each clip provided some meaningful context and ended just after the moment when the patient identified his/her thought or feeling. This process yielded 503 potential test items.

2.3. Pilot studies

The 503 clips were randomized and compiled into 10 videotapes. Between 10 and 20 undergraduate pilot participants at Northeastern University watched each tape. Participants were asked to infer what the patient was thinking or feeling during each clip in an open-ended response format [13]. One research assistant scored each open-ended response on a scale from 0 to 2 (where 0 meant a complete mismatch with the patient’s commentary and 2 meant a nearly synonymous match). A second research assistant independently scored one participant from each of 10 tapes (259 thoughts/feelings). Reliability between the two coders was very good, r = .81, p < .001. Based on this scoring, each clip was assigned its difficulty level (i.e., how well the pilot participants were able to correctly infer the thought or feeling). A smaller pool of 75 clips was then selected for moderate difficulty level (not too hard or too easy) and the goal of maximizing
diversity in ethnicity, gender, and the affective states being expressed; these decisions were made by consensus among three of the authors (JAH, MAR, EMC). The 75-clip test was administered to a new group of 50 pilot participants, and further eliminations were made using the same criteria with the result being the final 48-item TAPPA.

2.4. Description of the final TAPPA

The TAPPA contains audiovisual clips of 14 patients (10 female, 4 male; 12 non-Hispanic white, 1 black, 1 multiracial). Each patient is shown from 1 to 10 times (mean = 3 times); when the same patient appeared in more than one clip, it was a different excerpt with different answer options. More than two clips from the same patient were never shown successively. Test items (audiovisual clips) range from 18 to 33 s in duration (M = 26 s). For each item there are four answer options, one of which is the verbatim text of the patient’s thought or feeling (in its entirety or shortened), which was ordered randomly among the three incorrect options. Incorrect answer choices are either pilot participants’ incorrect open-ended answers or were composed by a research assistant and reviewed by members of the research team. Incorrect answer choices are equivalent in length, complexity, and tone to the correct answers. After every item, the following question appears on the video screen: “What was the patient thinking or feeling the moment the tape stopped?” The following are the response options from one test item: “A. I wish I could express myself better. B. My doctor isn’t being very helpful, C. I’m curious about what this pain is, D. My body scares me with all its problems.” Another item read: “A. I felt sad remembering. B. I’m uncomfortable talking about this with the doctor. C. I feel guilty talking about my son’s medical issues, D. I’m embarrassed that my son has these problems.”

The 48 items fell into the following general categories: happy, relieved, appreciative, 11; upset, tense, worried, fearful, 10; confused, 6; annoyed, impatient, in disagreement, 5; embarrassed, 2; guilty, 2; and 16 that were unique (e.g., sad, distracted, disgusted with self, contemptuous, disappointed). No effort was made in the scoring to distinguish “thoughts” from “feelings,” as this was not a theoretically important distinction and such a distinction would be difficult to make. Each test item is scored as either correct or incorrect, meaning that total score can range from 0 to 48.

Administration can be done either via an electronic link whereby test-takers view the clips and enter their responses directly into the computer, or by showing a DVD accompanied by a paper answer sheet.

As mentioned earlier, the test consists of four 12-item modules. Researchers are free to use some or all of the modules depending on their needs. Analysis of the groups described below revealed that the first module had slightly weaker internal consistency than the remaining three modules. Therefore, it is recommended to drop that module if shortening is desired. Below, results are shown for both the full TAPPA and for the three-module version, for comparison purposes.

2.5. Recruitment of test groups and group characteristics

Table 1 shows the groups whose data are presented in the present article; none of the groups included the pilot participants alluded to earlier. Four of the groups were undergraduate students at Northeastern University who were either introductory psychology students majoring in psychology and interested in how people communicate. The other three groups were undergraduate nursing students at Northeastern University (mostly fourth and fifth year students), medical students at The Johns Hopkins University (mostly fourth year students), and internal medicine residents at the Beth Israel Deaconess Medical Center (second and third year residents). These three groups were recruited via email solicitation for paid volunteers sent by a faculty member in their respective programs. All participants signed an informed consent approved by the human subjects committees at their respective institutions and Northeastern University. Table 1 gives descriptive data for these seven groups.

2.6. Other measurements

Group 1 took the TAPPA on two occasions with an interval of 14 to 35 days (M = 21.17 days) for the purpose of assessing re-test reliability. Group 2 took only the TAPPA. Group 3 took one or two additional tests of interpreting affective cues, randomly determined (they did two tests if time allowed), and Group 4 took four such tests. The tests used with Groups 3 and 4 were: (1) Diagnostic Analysis of Nonverbal Accuracy Adult Faces (DANVA Faces) [10], consisting of 24 color slides of men and women expressing anger, fear, happiness, or sadness on their faces; (2) Diagnostic Analysis of Nonverbal Accuracy Adult Voices (DANVA Voices) [17], consisting of 24 vocal recordings of adult men and women reciting a standard sentence to express the same four emotions; (3) Profile of Nonverbal Sensitivity (MiniPONS version) [18], consisting of 64 2-s clips of an adult woman portraying 20 different affective states in various combinations of face, body, and content-free voice (sample item content: asking forgiveness, criticizing someone for being late, expressing motherly love), and (4) Patient Emotion Cue Test (PECT) [12], a 47-item test showing an adult female portraying a patient expressing five different affective states (confused, angry, happy, sad, anxious, and neutral) to an imagined physician, while varying the affective intensity of the verbal and nonverbal cues. Group 5 took the two DANVA tests. Higher scores on all of these tests indicated more accuracy.

Groups 5–7 took self-report scales to measure emotional self-awareness and empathy. The self-awareness scales belonged to the Trait Meta Mood Scales battery [19] and consisted of: (1) attention to one’s emotions, (2) clarity of one’s emotions, and (3) ability to repair one’s emotions. The empathy scale was the Empathic Concern scale from the Interpersonal Reactivity Index [20]. On these scales higher scores meant higher attention, clarity, repair, and empathy.

2.7. Procedure for administration of TAPPA and other instruments

Participants in Groups 1–5 came to a laboratory in the Psychology Department at Northeastern University. The TAPPA was administered on a laptop computer using the online survey.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Group characteristics.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>N</td>
</tr>
<tr>
<td>1 Undergraduates</td>
<td>24</td>
</tr>
<tr>
<td>2 Undergraduates</td>
<td>192</td>
</tr>
<tr>
<td>3 Undergraduates</td>
<td>79</td>
</tr>
<tr>
<td>4 Undergraduates</td>
<td>66</td>
</tr>
<tr>
<td>5 Nursing students</td>
<td>77</td>
</tr>
<tr>
<td>6 Medical students</td>
<td>44</td>
</tr>
<tr>
<td>7 Internal medicine residents</td>
<td>42</td>
</tr>
</tbody>
</table>

* Omit 2 males and 1 student with unreported gender; these students’ scores on the TAPPA were indistinguishable from the remaining students’ scores.
program Qualtrics. Groups 6 and 7 took the test on the same platform via an internet link. Other tests and self-report instruments, including demographic and other background questions, came after the TAPPA.

2.8. Statistical analysis

Standard descriptive and inferential statistics for within-group analyses were conducted using the IBM SPSS Statistics 20 program. As further described later, meta-analytic combination of effect sizes (correlations) across studies was accomplished using the Comprehensive Meta-Analysis program [21].

3. Results

3.1. Descriptive statistics and reliability of the TAPPA

Table 2 shows basic statistics for TAPPA total scores, expressed as the proportion of items correctly judged, as well as internal consistency reliability (Cronbach’s alpha) for the full TAPPA and for the three-module version. In all groups the scores had a symmetrical distribution. As the table shows, all groups scored very significantly higher than the guessing (chance) level, though not with a ceiling effect. Internal consistency indicated a moderately good degree of correlation among the test items, consistent with reliability figures for many tests of judging affect from nonverbal cues [22].

The re-test reliability in Group 1 over an average 3-week interval for the full TAPPA was strong, r(22) = .72, p < .001. For the three-module version, it was r(22) = .56, p < .01.

3.2. Gender differences on the TAPPA

The bottom row of Table 3 shows a meta-analytic summary of the findings across the six groups for which a gender difference could be examined. The last line shows the weighted (by group size) mean correlation between gender and accuracy, which was .16, p < .001, for the full TAPPA indicating better performance by women (point-biserial correlation, coded male = 0, female = 1). The average effect was also significant for the three-module version (p < .01). A point-biserial correlation of .16 corresponds to a male–female difference of .32 standard deviation units.

A meta-analytic contrast for the full TAPPA comparing the four non-clinician groups to the two clinician groups was not significant, p < .20, though the trend was for the clinician groups to show a bigger gender difference favoring women (mean weighted correlation = .28) than the non-clinician groups (mean weighted correlation = .12). Both of these correlations were significantly greater than zero (p < .05). The corresponding contrast between the clinician and non-clinician groups for the three-module TAPPA was also not significant (p < .23).

3.3. Correlations of the TAPPA with emotion recognition tests

Table 3 shows correlations of the TAPPA with the four tests of emotion recognition. As the meta-analytic statistics in the bottom row indicate, the DANVA Faces test had a significant overall correlation with both the full TAPPA and the three-module TAPPA. The DANVA Voices test had a marginally significant relation for the full TAPPA. The PECT had a particularly strong relation with the TAPPA (equally so for both the full and three-module TAPPA), which is especially notable because the TAPPA and the PECT have elements in common that set them apart from the other tests. First, both tests include both verbal and nonverbal cues, whereas the other tests include only nonverbal cues, conveyed via face, body, or content-free speech. Second, both tests were designed for clinician test-takers, with an actress-patient in the PECT and real patients in the TAPPA.

3.4. Correlations of the TAPPA with emotional awareness and empathy

There were no significant correlations of the TAPPA with the three emotional awareness scales nor the empathy scale, either for individual groups or in the meta-analysis across groups.

### Table 2

Descriptive statistics for the Test of Accurate Perception of Patients’ Affect (TAPPA).

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Undergraduates</td>
<td>.49</td>
<td>.10</td>
<td>.25–.65</td>
<td>.58 (.59)</td>
</tr>
<tr>
<td>2 Undergraduates</td>
<td>.48</td>
<td>.12</td>
<td>.21–.75</td>
<td>.56 (.55)</td>
</tr>
<tr>
<td>3 Undergraduates</td>
<td>.50</td>
<td>.09</td>
<td>.29–.67</td>
<td>.44 (.41)</td>
</tr>
<tr>
<td>4 Undergraduates</td>
<td>.51</td>
<td>.11</td>
<td>.29–.71</td>
<td>.52 (.55)</td>
</tr>
<tr>
<td>5 Nursing students</td>
<td>.55</td>
<td>.09</td>
<td>.33–.73</td>
<td>.61 (.48)</td>
</tr>
<tr>
<td>6 Medical students</td>
<td>.56</td>
<td>.10</td>
<td>.33–.75</td>
<td>.63 (.65)</td>
</tr>
<tr>
<td>7 Internal medicine residents</td>
<td>.49</td>
<td>.11</td>
<td>.23–.71</td>
<td>.63 (.65)</td>
</tr>
</tbody>
</table>

Note: Mean accuracy is significantly above the chance level of .25 (p < .001) in every group. Alpha for the three-module version is given in parentheses.

### Table 3

Correlations of the Test of Accurate Perception of Patients’ Affect (TAPPA) with gender and other emotion recognition tests.

<table>
<thead>
<tr>
<th>Group</th>
<th>Gender*</th>
<th>DANVA Faces</th>
<th>DANVA Voices</th>
<th>MiniPONS</th>
<th>PECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−.15 (−.15)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2</td>
<td>.16*</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>3</td>
<td>.15 (.00)</td>
<td>.24 (.16)</td>
<td>.21 (.19)</td>
<td>−.19 (−.17)</td>
<td>.56** (.56**)</td>
</tr>
<tr>
<td>4</td>
<td>.08</td>
<td>.15</td>
<td>.33**</td>
<td>.25*</td>
<td>.35*</td>
</tr>
<tr>
<td>5</td>
<td>.35* (.31*)</td>
<td>.14 (.15)</td>
<td>−.05 (−.13)</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>6</td>
<td>.23 (.16)</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>M_r</td>
<td>.16*** (.12**)</td>
<td>.16* (.15*)</td>
<td>.14* (.10)</td>
<td>.15 (.16)</td>
<td>.42*** (.42**)</td>
</tr>
</tbody>
</table>

Note: DANVA = Diagnostic Analysis of Nonverbal Accuracy. MiniPONS = “Mini” version of Profile of Nonverbal Sensitivity. PECT = Patient Emotion Cue Test. Correlation is missing if that test was not given. Correlations with other tests control for participant gender (partial correlation). M_r = mean correlation weighted by group size. Results for the three-module version are given in parentheses; that information is missing if the group took only two modules.

* p < .10; * * * p < .01; ** p < .05; *** p < .001.

* Coded 0 = male, 1 = female.

b All female.
4. Discussion and conclusion

4.1. Discussion

This article presented a new test of interpersonal perception accuracy that was designed with clinicians in mind. The TAPPA was designed to have strong ecological validity, being based on actual physician–patient interactions and employing the patients’ own reports as criteria for scoring accurate assessment of thoughts and feelings. Of course, relying on patients’ post-visit reports of their thoughts and feelings is not a perfect window into the thoughts and feelings they were having at those moments in the visit. However, it is a far more ecologically valid method than using posed expressions, which is the approach taken in the great majority of interpersonal perception tests [10–12]. Furthermore, patients were instructed not to guess or invent but to report only thoughts and feelings they distinctly remembered having, and research using this method has amassed an extensive track record [13,23].

Generally speaking, the full 48-item TAPPA and a shorter three-module (36 item) version performed similarly though as one might expect the shorter version sometimes had slightly weaker predictive validity. In seven groups of participants, the TAPPA had good psychometric properties in terms of mean accuracy that was above chance but without a ceiling effect, good range, adequate internal consistency, and very good re-test reliability. The test showed evidence of construct validity as well.

Consistent with a large published literature using other tests of interpersonal accuracy, women scored higher than men. The magnitude of this difference was in the same general range as found with other conceptually similar tests [15,16]. Because of the limited number of patients shown in the test, we did not separately analyze male and female patients. However, numerous studies on tests of interpersonal perception accuracy, including a meta-analysis, have failed to uncover an interaction of perceiver gender and target gender on accurate perception of cues [24,25].

The gender difference was not significantly different between the non-clinician and clinician groups (medical students and medical residents), but the clinician groups tended to show a bigger gender difference. This indicates that selection and/or the medical training environment did not erase, but perhaps even increased, the gender difference. This difference may have implications for the process of medical care insofar as accurate perception of patients’ thoughts and feelings is relevant for clinical performance. The TAPPA had significant positive correlations with several other tests of emotion recognition, in particular the Patient Emotion Cue Test (PECT), a test of emotion recognition that was also designed for clinician test-takers. That the TAPPA correlated more highly with the PECT than with more generic emotion recognition tests is evidence for the discriminant validity of the TAPPA and suggests that tests designed specifically for clinicians may prove to have superior ability to predict clinical performance compared to more generic tests.

The TAPPA was not correlated with self-report measures of emotional self-awareness and empathy. A previous meta-analysis on non-clinician groups found an overall positive relation between self-reported empathy and accurate interpersonal perception [7], but the relation in that meta-analysis was very modest in magnitude indicating that self-reported empathy is not inter-changeable with accurate perception of others. In a study of medical students using different tests of judging affective cues, accuracy was positively correlated with self-reported ability to name one’s own emotions (the same Clarity scale used in the present research) [26], but this was not found with the TAPPA. That study, like the present one, did not find significant relations for the other two self-awareness scales. Though further research is needed, it appears that self-reported insight into one’s own emotions and self-reported emotional regulation are not related to objectively measured accuracy in understanding other people’s thoughts and feelings.

4.2. Conclusion

Based on psychometric and construct validity evidence, as well as ecological validity, the Test of Accurate Perception of Patients’ Affect (TAPPA) is a promising tool for research on predictors of clinical performance and patient outcomes, and potentially for selection, assessment, and training of clinicians. Recent research demonstrates that interpersonal perception accuracy is a trainable skill [27,28].

4.3. Practice implications

Clinical performance and patient outcomes may be improved if clinicians’ accuracy of perceiving patients is assessed and remediated.

Role of funding

Funded by grants from the Agency for Healthcare Research and Quality and the Edward J. Stemmler Medical Education Research Fund of the National Board of Medical Examiners.

Conflict of interest

None.

Acknowledgments

The authors thank William Ickes for methodological advice and Bryanna Baptiste, Amanda Bertram, Jamie Chachere, Jennifer Glynn, David Kramer, Samuel Nevin, Michael O’Brien, Curtis Pitegoff, and Drew Thabault, for their assistance in conducting the research.

References


