Self-Reported and Observer-Rated Empathy and Suicide-Risk Assessment Accuracy in Virtual Patient Interviews

Lilian Guo [1,2,3], Jimmy Pengyu Chen [1,2], Amna Maryam [1,2], Jason Tortora [1,2], Sarah Bloch-Elkouby [1,2]

1. Icahn School of Medicine at Mount Sinai, New York, NY, USA 2. Ferkauf Graduate School of Psychology, Yeshiva University, New York, NY, USA 3. Teachers College, Columbia University, New York, NY, USA

Icahn School of Medicine at Mount Sinai

Bloch-Elkouby Lab for Suicide Prevention and Psychotherapy Research

Ferkauf Graduate School of Psychology

Teachers College Columbia University

BACKGROUND

Therapist empathy is associated with better psychotherapy outcomes (Elliott et al., 2018). During an intake interview, clinicians must gather sensitive information while establishing a relationship with the patient. Empathic communication may help patients feel understood and facilitate discussion of suicidal thoughts and distress. However, an association with psychotherapy outcomes does not establish whether empathy is also associated with more accurate suicide-risk assessment.

Self-reported and observer-rated empathy capture different perspectives. Self-report measures assess individuals' perceptions of their empathic tendencies, whereas observer ratings assess empathy expressed during a particular interaction. The Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009) and Therapist Empathy Scale (TES; Decker et al., 2014) allow us to examine whether these perspectives correspond and whether either is associated with assessment accuracy.

Virtual patient (VP) intake interviews provide a controlled setting for examining interviewing and assessment skills (Fink et al., 2021). In this study, clinicians interviewed a VP with a predefined clinical presentation, allowing their assessments to be compared with common benchmark ratings. This approach reduces variation in the underlying patient presentation and makes it possible to examine whether differences in clinicians' empathy relate to differences in assessment accuracy.

Examining these relationships can help clarify what self-reported and observed empathy contribute to performance in simulated suicide-risk assessments and inform future research on how VP training can address both empathic communication and accurate identification of suicide risk and Suicide Crisis Syndrome features.

STUDY AIMS

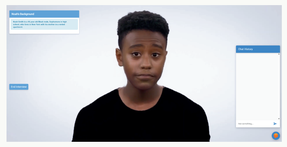
1. Examine associations of self-reported and observer-rated empathy with C-SSRS and SCS-C assessment inaccuracy.
2. Examine the correspondence between self-reported empathy and observer-rated empathy during a VP interview.

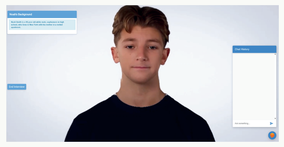
METHOD

Procedure and Sample

This secondary analysis used data from a larger study in which clinicians conducted an intake interview with a standardized virtual patient, "Noah." Participants encountered one of two patient race conditions: Black Noah or White Noah. Noah responded through pre-scripted text and video responses.

After excluding 41 participants flagged by the research team for data-quality concerns, 150 participants remained in the merged dataset. Sample sizes varied across analyses based on available scores (n = 140–148).





Empathy Measures

- Self-reported empathy: Total score on the 16-item Toronto Empathy Questionnaire (TEQ; Spreng et al., 2009), with reverse scoring of negatively worded items.
- Observer-rated empathy: Sum of eight consensus-rated items from the Therapist Empathy Scale (TES; Decker et al., 2014) used in this study to code VP interviews. All eight ratings were required to calculate the total.

Higher scores indicated greater empathy on both measures.

Assessment Inaccuracy

Clinicians' ratings were compared with benchmark ratings for Noah's scripted presentation.

- C-SSRS inaccuracy: A study-specific score reflecting the distance between the highest endorsed item level and the nearest acceptable benchmark level, 3 or 4.
- SCS-C inaccuracy: The proportion of 15 binary symptom ratings that disagreed with the benchmark. Rapid spikes and hypervigilance were benchmarked as absent; the remaining 13 symptoms were benchmarked as present. All 15 ratings were required.

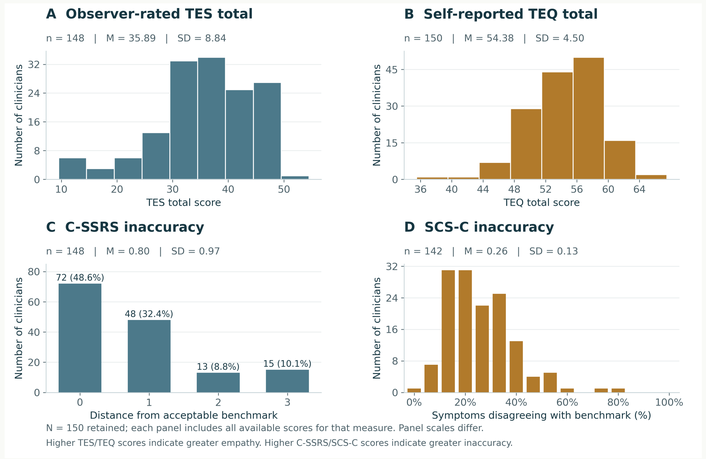
Higher scores indicated greater inaccuracy for both outcomes.

Statistical Approach

Two-sided Pearson correlations examined associations between each empathy total and each inaccuracy score, and between the two empathy totals. Analyses used complete observations for each variable pair. Results are reported with 95% confidence intervals and unadjusted p values.

RESULTS

1 Descriptive Statistics



N = 150 retained; each panel includes all available scores for that measure. Panel scales differ.
Higher TES/TEQ scores indicate greater empathy. Higher C-SSRS/SCS-C scores indicate greater inaccuracy.

2. Empathy and Assessment Inaccuracy

Variable pair	*n*	Pearson's r	95% CI	*p*
Observer TES × C-SSRS inaccuracy	146	0.006	[−.156, .169]	0.94
Self-report TEQ × C-SSRS inaccuracy	148	0.113	[−.049, .269]	0.172
Observer TES × SCS-C inaccuracy	140	−.052	[−.216, .115]	0.545
Self-report TEQ × SCS-C inaccuracy	142	−.100	[−.260, .066]	0.237
Self-report TEQ × Observer TES	148	≈ .000	[−.162, .161]	0.998

Note. Negative correlations indicate that greater empathy was associated with lower inaccuracy. All confidence intervals included zero.

Neither self-reported nor observer-rated empathy was significantly associated with C-SSRS or SCS-C assessment inaccuracy.

2. Self-Reported and Observer-Rated Empathy

Self-reported TEQ and observer-rated TES totals showed essentially no linear association, r ≈ .000, 95% CI [−.162, .161], p = .998, n = 148.

Self-reported empathy did not correspond to observer-rated empathy during this VP interview.

DISCUSSION

Neither self-reported nor observer-rated empathy was significantly associated with assessment inaccuracy in this standardized VP interview. One possible explanation is that recognizing and classifying suicide-risk indicators depends on skills such as assessment knowledge, systematic questioning, and interpretation of clinical information, which may be partly distinct from empathic communication. These skills were not directly tested as explanations in the present analysis. Although the observed correlations were small, the confidence intervals remain compatible with some modest associations.

The scripted VP format may also help explain these findings. In human interactions, empathy could influence trust, willingness to disclose suicidal thoughts, and the depth of information available to clinicians. A VP with predefined responses may offer fewer opportunities for these interpersonal processes to influence assessment accuracy. Empathy might therefore relate to engagement or disclosure in ways that are not captured by agreement with benchmark ratings. Whether these relationships differ between virtual and human encounters requires direct comparison.

The near-zero correlation between self-reported and observer-rated empathy suggests that the two scores were not interchangeable in this sample. The TEQ assesses general self-perceived empathic tendencies, whereas TES ratings reflect behavior during one specific interview. Differences in measurement perspective, interview context, or measurement error could contribute to the lack of correspondence. This result alone does not establish that either measure is unreliable.

Future research should examine internal consistency and agreement between observers before consensus, and evaluate the study's eight-item TES adaptation. Studies using multiple clinical scenarios and repeated interviews could help distinguish stable clinician characteristics from encounter-specific behavior. Comparisons with human standardized patients and real clinical encounters should assess disclosure, alliance, and engagement alongside assessment accuracy. Including assessment knowledge, training, and clinical experience could help clarify the contribution of empathy. Together, these approaches would test whether the present findings reflect the measures, the simulation context, or a weak relationship between empathy and benchmark-based accuracy.

References

Decker, S. E., Nich, C., Carroll, K. M., & Martino, S. (2014). Development of the Therapist Empathy Scale. Behavioural and Cognitive Psychotherapy, 42(3), 339–354. https://doi.org/10.1017/S1352465813000039

Elliott, R., Bohart, A. C., Watson, J. C., & Murphy, D. (2018). Therapist empathy and client outcome: An updated meta-analysis. Psychotherapy, 55(4), 399–410. https://doi.org/10.1037/pst0000175

Fink, M. C., Reitmeier, V., Stadler, M., Siebeck, M., Fischer, F., & Fischer, M. R. (2021). Assessment of diagnostic competences with standardized patients versus virtual patients: Experimental study in the context of history taking. Journal of Medical Internet Research, 23(3), e21196. https://doi.org/10.2196/21196

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. Journal of Personality Assessment, 91(1), 62–71. https://doi.org/10.1080/00223890802484381